

This work is protected by copyright and other intellectual property rights and duplication or sale of all or part is not permitted, except that material may be duplicated by you for research, private study, criticism/review or educational purposes. Electronic or print copies are for your own personal, non-commercial use and shall not be passed to any other individual. No quotation may be published without proper acknowledgement. For any other use, or to quote extensively from the work, permission must be obtained from the copyright holder/s.

Structural studies of the interaction of bacterial lipopolysaccharides with C-reactive protein and lung surfactant protein D

ABDUL S. AHMED

Ph.D.

2012

IMAGING SERVICES NORTH

Boston Spa, Wetherby
West Yorkshire, LS23 7BQ
www.bl.uk

**APPENDIX 1 HAS BEEN OMITTED FROM THIS DIGITAL
COPY AT THE REQUEST OF THE AWARDING UNIVERSITY.**

Abstract

In this thesis, recognition of the complex natural ligand lipopolysaccharide (LPS) by the innate immune proteins C-reactive protein (CRP) from *Limulus polyphemus* (Atlantic horseshoe crab), and human surfactant protein D, in the form of a recombinant fragment comprising the α -helical coiled-coil neck plus three lectin domains (CRD), has been studied using X-ray crystallography. The intact LPS from three bacteria, *E. coli* (0111.B4), *P. aeruginosa* (serotype 10) and *H. influenzae* type b strain RM153 Eagan (wild type and 4A mutant), were subjected to mild acid hydrolysis to separate the lipid A from the extended oligosaccharide fraction (OS) to improve solubility in crystallisation studies.

X-ray diffraction data collection using synchrotron radiation on crystals of *Limulus* CRP grown in the presence of intact *E. coli* LPS provided diffraction to 2.0Å resolution showing a large unit cell with multiple copies of the molecular aggregate of unknown size in the asymmetric unit. Structure solution was not attempted but has since been achieved by other members of the research group although significant difficulties remain with the refinement.

Crystallographic analysis at 1.7Å resolution of the recombinant fragment of human SP-D crystals with *H. influenzae* Eagan wild type and Eagan 4A OS shows no binding of the Eagan wild type OS, however, the Eagan 4A OS ligand is found complexed to the protein CRD in chains B and C in a calcium dependent manner. No ligand binding could be observed in chain A. In both chains the binding is found in one orientation involving the core L,D-heptose C₆- and C₇- hydroxyl groups of the glycerol side chain, despite the availability of a terminal core glucose bearing vicinal equatorial C₃ and C₄ hydroxyl groups. Interaction with Kdo, seen in a putative anhydro closed 5-membered ring form, is also evident suggesting that efficient recognition requires multiple binding interactions.

Contents

Abstract	i
Contents	ii
List of figures	x
List of tables	xx
Table of amino acids	xxiii
List of abbreviations and symbols	xxiv
Acknowledgements	xxvi
Chapter 1 Introduction to innate immune proteins and lipopolysaccharide	1
1.0 Introduction	2
1.1 Human C-reactive protein	3
1.1.1 Structure of CRP	5
1.1.2 Mechanism of phosphocholine and C1q binding	6
1.1.3 Synthesis of CRP	8
1.1.4 Defence against microbial pathogen	9
1.1.5 CRP binding to altered host cells	9
1.1.6 CRP and complement activation	10
1.1.7 <i>Limulus</i> CRP	13
1.1.7.1 Structure of <i>Limulus</i> CRP	13
1.1.7.2 Ligand binding and functions of <i>Limulus</i> CRP	14
1.2 Human lung surfactant protein D	16
1.2.1 Biosynthesis and secretion of SP-D	17
1.2.2 Structure of SP-D	18

1.2.2.1 N-terminal region	20
1.2.2.2 Collagen region	20
1.2.2.3 α -helical coiled coil	21
1.2.2.4 Carbohydrate recognition domain	23
1.2.3 Calcium ions in the CRD	24
1.2.4 Functions of SP-D	25
1.2.4.1 Microbial and organic antigen interactions	26
1.2.4.2 Immune modulation and control of inflammation	29
1.2.5 Mechanism of ligand binding	31
1.2.6 SP-D receptors	36
1.3 Lipopolysaccharide	37
1.3.1 Immune response and receptors to LPS	40
1.3.2 Structure of LPS	42
1.3.2.1 Lipid A region	44
1.3.2.2 Core oligosaccharide region	45
1.3.2.3 O-polysaccharide region (O-antigen)	46
1.3.3 Biosynthesis: Lipid A and core region	47
1.3.4 Biosynthesis: polysaccharide and O-polysaccharide	50
1.3.5 <i>Escherichia coli</i> LPS	52
1.3.5.1 <i>E. coli</i> LPS structure	52
1.3.5.2 <i>E. coli</i> O-polysaccharide	54
1.3.6 <i>Pseudomonas aeruginosa</i> LPS	55
1.3.6.1 <i>P. aeruginosa</i> LPS core and O-polysaccharide	55
1.3.7 <i>Haemophilus influenzae</i> LPS	57

1.3.7.1 Pathogen and vaccine	58
1.3.7.2 Structural diversity	59
1.3.7.3 Inner and outer core structure	61
1.3.7.4 Biosynthesis of the inner core	63
1.3.7.5 Biosynthesis of the outer core	64
1.4 Aims and objective of the research project	66
1.4.1 Structural studies using <i>Limulus</i> CRP with <i>E. coli</i> , <i>P. aeruginosa</i> and <i>H. influenzae</i> LPS and polysaccharide ligand	66
1.4.2 Structural studies using rfhSP-D with <i>H. influenzae</i> Eagan wild type LPS and mutant 4A	68
Chapter 2 Experimental Techniques	69
2.1 Principals of the technique used	70
2.1.1 PC affinity chromatography	70
2.1.2 Hydrolysis	72
2.1.3 Mass spectroscopy	74
2.1.3.1 ESI-MS	75
2.1.3.2 MALDI-MS	75
2.1.3.3 Mass spectrum interpretation	77
2.1.3.4 Structural analysis of LPS using MS	77
2.1.4 Protein crystallography	80
2.1.4.1 Crystallisation	81
2.1.4.2 X-ray diffraction setup	82
2.1.4.3 X-ray diffraction analysis	83

2.1.4.4 X-ray generators	83
2.1.4.5 Synchrotron radiation source	84
2.1.4.6 Production of X-ray diffraction	85
2.1.4.7 Braggs Law	86
2.1.4.8 X-ray diffraction data collection	87
2.1.4.9 Diffraction data processing	89
2.1.4.9.1 Auto indexing	89
2.1.4.9.2 Truncate	90
2.1.4.9.3 Structure factor	91
2.1.4.9.4 R factors	93
2.1.4.9.5 Calculation of the electron density map	94
Chapter 3 Experimental: Materials and Methods	95
3.1 Isolation of <i>Limulus</i> CRP from <i>Limulus</i> plasma using affinity chromatography	96
3.2 Lipopolysaccharide	98
3.2.1 Hydrolysis of <i>E. coli</i> LPS	99
3.2.2 Hydrolysis of <i>P. aeruginosa</i> LPS	100
3.2.3 Hydrolysis of <i>H. Influenzae</i> Eagan LPS	100
3.2.4 Hydrolysis of <i>H. influenzae</i> Eagan 4A LPS	100
3.3 MS analysis of <i>E. coli</i> polysaccharide by ESI	101
3.4 MS analysis of <i>E. coli</i> lipid A using MALDI	101
3.5 MS analysis of <i>H. influenzae</i> Eagan 4A oligosaccharide using MALDI	101
3.6 Preparation of rfhSP-D	102

3.7 Preparation of <i>Limulus</i> CRP	103
3.8 Crystal tray set up	103
3.8.1 Crystallisation trials using <i>Limulus</i> CRP	104
3.8.2 Crystallisation trials using <i>Limulus</i> CRP with <i>E. coli</i> LPS and polysaccharide	105
3.8.3 Crystallisation trials using <i>Limulus</i> CRP with <i>P. aeruginosa</i> LPS and polysaccharide	105
3.8.4 Crystallisation trials using <i>Limulus</i> CRP with <i>H. influenzae</i> Eagan oligosaccharide	106
3.9 Structural studies of rfhSP-D with <i>H. influenzae</i> Eagan oligosaccharide	107
3.10 Structural studies of rfhSP-D with <i>H. influenzae</i> Eagan 4A oligosaccharide	107
3.11 Cryoprotection	108
3.12 X-ray diffraction data collection and processing	108
3.13 Structure solution and refinement (rfhSP-D / Eagan 4A oligosaccharide)	109
Chapter 4 Hydrolysis & mass spectroscopy of bacterial LPS:	
Results and discussion	110
4.1 <i>E. coli</i> LPS	111
4.2 <i>P. aeruginosa</i> LPS	111
4.3 <i>H. influenzae</i> Eagan and Eagan 4A LPS	111
4.4 ESI-MS analysis of <i>E. coli</i> polysaccharide	111
4.5 MALDI-MS analysis of <i>E. coli</i> lipid A	111
4.6 MALDI-MS analysis of <i>H. influenzae</i> Eagan 4A oligosaccharide	112

4.7 Discussion	114
4.7.1 Hydrolysis of LPS	114
4.7.2 <i>E. coli</i> LPS hydrolysis and structural characterisation	117
4.7.3 <i>P. aeruginosa</i> LPS hydrolysis and structural characterisation	124
4.7.4 <i>H. influenzae</i> Eagan LPS hydrolysis and structural characterisation	125
4.7.5 <i>H. influenzae</i> Eagan 4A hydrolysis and structural characterisation	128
4.8 Summary	133
 Chapter 5 Crystallisation of bacterial LPS and polysaccharide with <i>Limulus</i> CRP: Result and discussion	 134
5.1 Isolation of <i>Limulus</i> CRP from plasma using affinity chromatography	135
5.2 Crystallisation of <i>Limulus</i> CRP	136
5.3 Crystallisation of <i>Limulus</i> CRP with <i>E. coli</i> LPS and polysaccharide	137
5.4 Crystallisation of <i>Limulus</i> CRP with <i>P. aeruginosa</i> LPS and polysaccharide	143
5.5 Crystallisation of <i>Limulus</i> CRP with <i>H. influenzae</i> Eagan oligosaccharide	145
5.6 Discussion	152
5.6.1 Crystallisation of <i>Limulus</i> CRP with <i>E. coli</i> LPS and polysaccharide	152
5.6.2 Crystallisation of <i>Limulus</i> CRP with <i>P. aeruginosa</i> LPS and polysaccharide	158
5.6.3 Crystallisation of <i>Limulus</i> CRP with <i>H. influenzae</i> Eagan oligosaccharide	159
5.7 Summary	160
5.7.1 <i>Limulus</i> CRP with <i>E. coli</i> LPS and polysaccharide	160
5.7.2 <i>Limulus</i> CRP with <i>P. aeruginosa</i> LPS and polysaccharide	160

5.7.3 <i>Limulus</i> CRP with <i>H. influenzae</i> Eagan polysaccharide	161
Chapter 6 Structural studies of rfhSP-D with <i>H. influenzae</i> Eagan and Eagan 4A oligosaccharide: Result and Discussion	162
6.1 Crystallisation and X-ray diffraction of rfhSP-D with <i>H. influenzae</i> Eagan oligosaccharide	163
6.2 Crystallisation and X-ray diffraction of rfhSP-D with <i>H. influenzae</i> Eagan 4A oligosaccharide	163
6.3 Data collection and processing of rfhSP-D with <i>H. influenzae</i> Eagan 4A oligosaccharide	165
6.3.1 Rigid body	165
6.3.2 Map calculation, model building and electron density map	167
6.3.3 Electron density map overall structure	167
6.3.4 Electron density of Eagan 4A oligosaccharide in the binding site	170
6.4 Discussion	173
6.5 Summary	181
Chapter 7 Conclusions and future work	182
7.1 Conclusions	183
7.1.1 Crystallisation studies using <i>Limulus</i> CRP with LPS.	183
7.1.2 Crystallisation studies using rfhSP-D and <i>H. influenzae</i> Eagan and 4A.	184
7.2 Future work	186
References	188

Appendix	233
Appendix 1 Publication: <i>Haemophilus influenzae</i> evades innate immune Defence mediated by surfactant protein D by blocking multiple Interactions with LPS core	
Appendix 2 MALDI-MS of <i>Haemophilus influenzae</i> Eagan 4A oligosaccharide	
Appendix 3 MALDI-MS of <i>E. coli</i> lipid A	

List of figures

Chapter 1

- Figure 1.1** Structure of phosphocholine. 4
- Figure 1.2** Structure of the teichoic acid that makes up the CPS of *Streptococcus pneumoniae*. The repeat unit (n) contains two PCho ligands allowing recognition by CRP. Structure modified from Fischer *et al.*, 1997. 4
- Figure 1.3** Structure of CRP subunits arranged in a pentameric fashion. Gray spheres show the two calcium ions in each subunit on the 'B face' [Thompson *et al.*, 1999]. 5
- Figure 1.4** a) Mechanism of PCho binding and b) the 'B' face on which PCho binding takes place [Thompson *et al.*, 1999]. 7
- Figure 1.5** The complement pathway showing the three pathways that converge to the common complement protein C3. 12
- Figure 1.6** Structure of Phosphoethanolamine. 15
- Figure 1.7** Synthesis of SP-D as well as SP-A located on the long arm of chromosome 10. Image taken from Sorensen *et al.*, 2007. 18
- Figure 1.8** Schematic description of the formation of the SP-D tetramer. Each polypeptide subunit is trimerised which is then further oligomerised by being tethered together at the N-terminus by disulphide bonds. Image adapted from Clark and Reid, 2002. 19
- Figure 1.9** A) Structure of recombinant fragment of SP-D with each chain A, B and C shown in different colours. B) electrostatic surface potential as viewed directly on top of the carbohydrate recognition domain. Blue regions show the positive charge distribution from Lys246 and Lys252 and the red regions indicate negative charge distributions from Glu232. Green dots show Ca^{2+} ions [Hakansson *et al.*, 1999]. 22
- Figure 1.10** Schematic representation taken from Hakansson and Reid, 2000 showing the human SP-D CRD as a double loop showing the series of β sheets and α -helices. Calcium binding and ligand

- recognition is shown in the second loop (circled red).
Encircled numbers denote amino acids that ligate to the principal calcium ion and are also involved in ligand binding. 23
- Figure 1.11** Illustration of high mannose oligosaccharide linked to protein (glycoprotein) on the surface of IAV and HIV virus (dark sphere). The high mannose oligosaccharide unit is N-linked to amino acid (green box).
Structure drawn using CS Chem. Draw Pro. Version 4.5. 29
- Figure 1.12** A) Glucosylceramide and B) phosphatidylinositol. 30
- Figure 1.13** Ligand binding region on the second loop of rfhSP-D via coordination with the principal Ca^{2+} ion (Ca1) [Shrive *et al.*, 2003; 2009]. 32
- Figure 1.14** C-terminal sequence of selected collectins (hSP-D numbering) Known neck/CRD structures are underlined, reported ligand bound structures are highlighted. Residues that bind to both calcium and ligand are underlined. Residues that flank the ligand binding site (SP-D325 and 343) are highlighted.
Taken from Shrive *et al.*, 2009. 33
- Figure 1.15** Ca^{2+} dependent monosaccharide binding by rat MBP-A.
A) Schematic representation of the ternary complex with hydrogen and coordination bond stabilisation. B) MBP-A complex with mannose (pdb code 2msb). The blue sphere represents the principal calcium ion.
Image taken from Zelensky and Gready (2005). 35
- Figure 1.16** Structures of monosaccharides with their respective ring numbering (red) A) α -D-glucopyranose (pyranoside in the chair conformation), B) β -fructose (furanoside), C) α -D-galactopyranoside and D) the amine sugar α -N-Acetyl-D-glucosamine. The anomeric carbon involved in the glycosidic bond is at position 1 (red numbering) except in B which is at the C₂ anomeric position.
Drawn using CS Chem. Draw Pro. Version 4.5. 38
- Figure 1.17** The formation of maltose, 4-O-(α -D-glucopyranosyl)- β -D-glucopyranose from the glycosidic bond of two glucose units.

Drawn using CS Chem. Draw Pro. Version 4.5.	39
Figure 1.18 The formation of a cyclic carbohydrate ring structure from open chain forms [Solomons, 1992]. The two anomers at the C ₁ anomeric carbon position are known as β when OH is on the same side of the ring as the CH ₂ -OH on C ₅ and α where OH is on the opposite side of the ring. (-OH in red).	40
Figure 1.19 The three regions of LPS. The inner core is linked to the lipid A via the ketosidic linkage. The molecule projects out from the cell membrane, anchored via the fatty acid residue. Drawn using CS Chem. Draw Pro. Version 4.5.	42
Figure 1.20 Chemotypes of LPS. Further variations in the Rd chemotype are possible arising from attachment of Hep and Phosphate in various forms [Nigam, 1975].	43
Figure 1.21 a) β-(1-6)-linked glucosamine disaccharide with (R)-3-hydroxyacyl group. Further ester linked fatty acid at R ₁ and R ₂ leads to variations. b) <i>Escherichia coli</i> lipid A. The hydroxyl group at C ₆ (blue circle) indicates the position of glycosidic attachment of the remaining part of the LPS. Structure modified from Odegaard <i>et al.</i> , 1997.	45
Figure 1.22 Schematic illustration of the O-polysaccharide repeat unit attached to the outer core. The inner core is composed of Kdo and heptose.	47
Figure 1.23a Biosynthesis of preformed lipid A from UDP-GlcNAc [Price <i>et al.</i> , 1994; Jackman <i>et al.</i> , 2000].	48
Figure 1.23b Fully assembled lipid A of <i>Escherichia coli</i> by condensation of UDP-2,3-di-acylglucosamine with -2,3-di-acylglucosamine [Price <i>et al.</i> , 1994; Jackman <i>et al.</i> , 2000].	49
Figure 1.23c Attachment of KDO unit to the lipid A region carried out by WaaA/KdtA [Price <i>et al.</i> , 1994, Jackman <i>et al.</i> , 2000].	50
Figure 1.24 Glycosidic bond formation leading to oligo and polysaccharide. The red semi circle shows the binding pocket in the GT-A enzyme [Lairson <i>et al.</i> , 2004].	51

- Figure 1.25** *E. coli* LPS showing the lipid A, inner core and the O-antigen polysaccharide. The inner core contains two Kdo units and three heptose unit. Modified from Amor *et al.*, 2000. 53
- Figure 1.26** Structure of the five outer core variations of *Escherichia coli* LPS. Attachment of O-antigen (red) has only been determined for R1 and R2 core types. Structure adapted from Amor *et al.*, 2000. 53
- Figure 1.27** Structures of O-polysaccharide from *Escherichia coli* a) strain 086:K62:B7 b) strain 62D₁ and c) K-12. [Yi *et al.*, 2006; Staff *et al.*, 1999; Stevenson *et al.*, 1994]. 54
- Figure 1.28** Structures of the *Pseudomonas aeruginosa* inner core and outer core but lacking the O-polysaccharide A) serotype 06 mutant strain A28 and B) LPS of cystic fibrosis isolate [Masoud *et al.*, 1995; Knirel *et al.*, 2001]. 56
- Figure 1.29** Structure of smooth LPS of *Pseudomonas aeruginosa* showing repeat unit of O-polysaccharide attached to the outer core [Sadovskya *et al.*, 2000]. A) serotype 06 mutant strain A28 and B) LPS of cystic fibrosis isolate [Masoud *et al.*, 1995; Knirel *et al.*, 2001]. 56
- Figure 1.30** *Haemophilus influenzae* serotype b glycoconjugate protein vaccine contains 7 repeat units of the capsular polysaccharide unit. The capsular polysaccharide b repeat unit of polyribose-ribitol phosphate (PRP) is shown in brackets [Verez-Bencomo *et al.*, 2004]. 59
- Figure 1.31** The common structural architecture of *H. influenzae* LPS, showing the three regions, lipid A, inner core, and the outer core comprised of various oligosaccharide extension (R¹ to R³). The numbers indicate the glycosidic bond on the pyranose ring. 60
- Figure 1.32** Structural heterogeneity of a single strain of NTHi serotype f as a result of various oligosaccharide chain attachments from the inner core [Yildirim *et al.*, 2003]. 61
- Figure 1.33** Structure of the NTHi strain 486 showing the attachment of sialic acid to the digalactoside extension from the terminal heptose unit, along with substitution of PCho and PEtn on the inner

heptose unit [Mansson <i>et al.</i> , 2001].	63
Figure 1.34 Schematic illustration of the oligosaccharide and non carbohydrate units coded by phase variable genes. Genes involved in the biosynthetic pathway of the inner core Hep are shown in bold.	65
Chapter 2	
Figure 2.1 Structure of agarose beads, a polysaccharide consisting of 1,3-linked-D-galactopyranose and 1,4-linked 3,6-anhydro- α -L-galactopyranose.	71
Figure 2.2 Schematic illustration of affinity chromatography. A) Protein of interest is reversibly bound to matrix containing ligand whilst the impurities are washed down with the wash buffer, B) the elution buffer contains a competitive ligand that has higher binding affinity towards the protein of interest and forms a reversible bond with the protein. Other techniques of detachment/removal of protein from matrix ligand may include changes in pH strength or ionic strength.	72
Figure 2.3 Mechanism of acid hydrolysis of β -(1-4)-Glucosidic bond of maltose. Drawn using CS Chem. Drawpro. Version 4.5	73
Figure 2.4 Structure of MALDI matrix A) CHCA, B) DHB, C) Sinapinic acid, D) DHAP, E) 3-hydroxypicolinic acid, F) DAN, G) MSA.	76
Figure 2.5 Fragmentation of methylated glycoside bond cleavage in MS producing Y_i and B_i ions with sodium adduct.	79
Figure 2.6 The stages involved in X-ray crystallography to determine the overall 3D structure of a protein. Image taken from www.doly.biochem.arizona.edu	81
Figure 2.7 Schematic illustration of synchrotron radiation caused by electrons travelling at relativistic velocity in an orbit. Image taken from http://all-physics.net	85
Figure 2.8 Illustration of X-ray waves forming constructive and destructive interference as a consequence of their phase difference	

[Glusker and Trueblood, 1985].	86
Figure 2.9 Bragg's construction of "reflection" of incident waves from planes hkl. If the waves are in phase then constructive interference will occur and if waves are out of phase then destructive interference will occur.	87
Figure 2.10 Unit cell (red) that makes up the crystal lattice. The lattice points are shown as dots.	88
Chapter 3	
Figure 3.1 Schematic illustration of crystal tray set up.	104
Chapter 4	
Figure 4.1 MALDI-MS of <i>E. coli</i> lipid A in the negative ion mode using 2,5-DHB as the matrix.	112
Figure 4.2 MALDI-MS of <i>H. influenzae</i> Eagan 4A oligosaccharide in the negative ion mode using 2,5-DHB matrix.	113
Figure 4.3 The mechanism of acid catalysed hydrolysis of LPS using 2% aqueous acetic acid at 100 °C for 2.5 hours, producing oligo/polysaccharide and lipid A. α -Kdo (red) and lipid A (green) structure are attached via C ₁ and C ₆ respectively. Extended oligosaccharide (R) in the α anomeric position attachment is on α -Kdo at the C ₅ position.	115
Figure 4.4 Further reaction of the acid labile α -Kdo in the hydrolysis process leading to olefinic acid (if R=H or its derivatives) through β -elimination of the phosphate group [Danan <i>et al.</i> , 1982; Auzanneau <i>et al.</i> , 1991].	116
Figure 4.5 Olefinic acid (A and B) undergoes intramolecular addition of the alcohol at position A) 8-OH to the olefinic bond to form a 6 membered cyclic 4,8-anhydro Kdo derivative. B) 7-OH to the olefinic bond to form a 5 membered cyclic 4,7-anhydro Kdo derivative. Both forms of intramolecular cyclisation lead to β and α anomeric carbon at position 1 and further keto/enol tautomerism formation (brackets) [Danan <i>et al.</i> , 1982; Auzanneau <i>et al.</i> , 1991].	116

- Figure 4.6** Structure of *E. coli* 0111.B4 [Morrison *et al.*, 1975].
LPS I contains 11 repeat units (n) and LPS II form contains 3 repeat units of the antigenic side chain. Where R₁, R₂ or R₃ could be Glc-GlcNAc or Hep. 119
- Figure 4.7** Structure of *E. coli* 0111.B4 used for ESI mass analysis.
Considering R=O then the molecular weight of LPS I (n=11 subunits) is 12324.7 and LPS II (n=3 subunits) is 6022.3 Da. 120
- Figure 4.8** Structure of lipid A with molecular mass of 1798.8 containing two amide linked and four ester linked fatty acid residues [Qureshi *et al.*, 1983; Que *et al.*, 2000]. 124
- Figure 4.9** Structure of the fully extended wild type *H. influenzae* Eagan wild type LPS deduced from NMR, NOESY connectivities [Masoud *et al.*, 1997]. The lipid A shown is attached to the core oligosaccharide via a single Kdo (red) with 1-5 glycosidic linkage. The oligosaccharide contains a PPEtn substituted from the inner core HepII and the subpopulation with phosphate substituted from KDO-4. The molecular weight calculated is 2779.40 Da.
Drawn using CS Chem. Draw Pro. Version 4.5. 126
- Figure 4.10** Structure of the fully extended native *H. influenzae* Eagan LPS containing the pentameric *anhydro* Kdo ring (red) with molecular weight 1746.4 Da. Drawn using CS Chem. Draw Pro. Version 4.5. 128
- Figure 4.11** Schematic illustration of the glycosidic linkages in *H. influenzae* Eagan wild type. The Eagan 4A is a gene construct by gene rfaF (red arrow) [Masoud *et al.*, 1997]. 129
- Figure 4.12** Extended glycoform of Eagan 4A (red structure), red dashed line shows extension of glycoform coded by lex2 gene. Mary Deadman and co workers (unpublished results, IMM, Oxford). 130
- Figure 4.13** Hydrolysis of Eagan 4A using 2% aqueous acetic acid at 100°C to give oligosaccharide with the *anhydro* Kdo ring form (574.48 Da). 130

Figure 4.14 Structure of Eagan 4A as deduced from the MALDI-MS after hydrolysis (992.4 Da).	132
Chapter 5	
Figure 5.1 Elution profile of <i>Limulus</i> pentraxins. The first mark (M) indicates elution of CRP from approximately 83 min to 180 min from addition of 10mM PC in calcium wash buffer. Second mark indicates elution of SAP from approximately 350 min to 369 min using 30mM PE. The flow rate was at 0.5 ml/min.	135
Figure 5.2 Toblerone shaped crystals from trial 1 using <i>Limulus</i> CRP with <i>E. coli</i> LPS obtained using 50mM MES pH 7.0, 9% PEG 6K and 10mM CaCl ₂ . The largest crystal measures 0.2 x 0.18mm.	137
Figure 5.3 Crystals produced in trial 3 using the same conditions and reagents as trial 1 (50mM MES pH 7.0, 9% PEG 6K and 10mM CaCl ₂). A) needle shaped crystals in well B1 B) flat rectangular crystals in well B2 and C) flat and toberlone shaped crystal in well B3.	139
Figure 5.4 Crystals produced in trial 5 using <i>Limulus</i> CRP with <i>E. coli</i> polysaccharide. A) triangular bi-pyramidal shaped crystals seen in wells A1, A2, A3 and B) multi sided rod shaped crystal seen in well B1.	142
Figure 5.5 Torpedo shaped crystals seen in well B3 in trial 2 using <i>Limulus</i> CRP with <i>P. aeruginosa</i> LPS.	144
Figure 5.6 triangular bi-pyramidal crystals observed in wells B1 and B3 in trial 4 using <i>Limulus</i> CRP with <i>P. aeruginosa</i> polysaccharide.	145
Figure 5.7 Central portion of a diffraction image of <i>Limulus</i> CRP with <i>E. coli</i> LPS taken on station SRS station 14.1. $\Delta\phi = 1^\circ$, crystal to detector distance 95mm (maximum resolution 2.0Å), exposure time 45 seconds. Maximum resolution shown here is 3Å.	155
Figure 5.8 Schematic illustration demonstrating selective binding of	

various CRP's (here CRP-1 and 2) with semi smooth LPS (black) and unrecognised smooth LPS (blue). CRP-3 (red) is the forms that has no recognition ability with either semi or smooth LPS and hence crystallised out as shown in the inset photo 1. Drawn using CS Chem. Draw Pro. Version 4.5.

156

Chapter 6

Figure 6.1 X-ray diffraction pattern of rfhSP-D with *H. influenzae* Eagan 4A soak at 30mM. Data was collected on station 14.1 on ADSC Quantum 4R CCD, Daresbury SRS.

164

Figure 6.2 Backbone of the rfhSP-D trimer (chain A blue; chain B magenta and chain C yellow) with three calcium ions in the CRD and the fourth calcium ion located in the central axis of the trimeric neck. Ligand is seen bound in chains B and C only. Image displayed in *O* and taken using Snapshot.

168

Figure 6.3 View down the trimeric axis of rfhSP-D. Chains A, B and C are shown in blue, magenta and yellow respectively. Calcium ions are shown as green spheres with ligand binding seen in chains B and C. Image displayed in *O* and taken using Snapshot.

169

Figure 6.4 Electron density of the bound ligand in chain B with its associated linkages. Electron density of the Hep1 unit can be seen clearly coordinated to the Ca^{2+} ion (green sphere) by C_6 and $\text{C}_7\text{-OH}$. Substitution of $\beta\text{-Glc}$ at Hep $\text{C}_4\text{-OH}$ is not present in the density. The pyranose ring structure representing Kdo is distorted at $\alpha\text{-Hep-(1-5)-}\alpha\text{-Kdo}$. Map visualised in the *O* graphics package and contoured at 1rms.

171

Figure 6.5 Bound Eagan 4A in subunit B of rfhSP-D. The ligand is bound via coordination of Hep1 to the calcium ion Ca1 and protein side chains. Figure generated using MOLSCRIPT (Kraulis, 1991)

172

Figure 6.6 The influence of the variable crystal contacts on the orientation and interactions of bound mannobiose and Arg343. A, B and C: $\text{man}\alpha 1\text{-2man}$ chains A, B and C respectively. D, E and F: $\text{man}\alpha 1\text{-4man}$ chains A, B. In all figures the symmetry contact protein molecule and the bound terminal mannose are represented in grey, with selected residues of the hSP-D monomer

represented by yellow main chain bonds and white side chain bonds. The crystal contacts with chains B and C, in both structures, are very similar but not identical. Of the six subunit structures, the only pairs with similar ligand orientation and similar ligand non bonded contacts are chains B and C in man α 1-4man and chain A in both man α 1-2man and man α 1-4man, although in the latter case the two orientations of the Arg343 side chain are significantly different. Image taken from Shrive *et al.*, 2009.

176

Figure 6.7 Electron density for the bound Glc-Hep and the putative *anhydro* Kdo ligand. Original Kdo numbering retained. Density labelled A corresponds to the putative position of O2' (original Kdo numbering) in either or both of the enantiomeric substituent off KdoC4. The glucose is not visible in the electron density. Figure generated using the PyMOL Molecular Graphics System Version 1.4 (Schrödinger, LLC, 2011).

179

Figure 6.8. The Hep-Kdo (*anhydro*) disaccharide in the ligand binding site. Kdo O6' is positioned to interact with Asp325, while the enantiomeric C4 substituent is not identifiable in the electron density except for an indication of O2' (original Kdo numbering), directed towards Arg343, which may be similarly positioned in both enantiomers. Original Kdo numbering retained. Figure generated using the PyMOL Molecular Graphics System Version 1.4 (Schrödinger, LLC, 2011).

180

List of Tables

Chapter 1

Table 1.1 The binding affinity of each collectin towards monosaccharide ligands which is in part determined by the residues flanking the binding region and in part by the residues which directly bind to the ligand.	34
---	----

Chapter 2

Table 2.1 Types of crystal lattice and relationships between their cell parameters.	89
--	----

Chapter 4

Table 4.1 Proposed structures of <i>E. coli</i> oligosaccharide possible after hydrolysis of native LPS and the possible m/z values scanned in the positive and negative ion mode. Molecular mass of Hex (Glc/Gal) 180.2, Hep 210.2, P 98.0, Kdo 238.2, PEtn 126.1, Col (colitose) 148.1, GalNAc/GlcNAc 221.2 and lipid A 1798.0 Da. R could be Glc-GlcNAc or Hep. * (loss of water for glycosidic bond formation 18.0 Da).	121
Table 4.2 ESI-MS in the positive ion mode. Observed m/z values of <i>E. coli</i> polysaccharide and the percentage abundance and the possible molecular weights calculated.	122
Table 4.3 ESI negative ion mode. Observed m/z values of <i>E. coli</i> polysaccharide and the possible molecular weight calculated. * theoretical values assigned to deduce possible ionisation that would lead to proposed structure.	123
Table. 4.4 Negative ion MALDI-MS of Eagan 4A. Molecular mass calculated using hydrated mass units, Hex (Glc) 180.2, Hep 210.2, Kdo 238.2, <i>anhydro</i> Kdo 220.2, P 98.0, PEtn 126.08, lipid A 953.0 (molecular mass determined using Chem. Draw Ultra Pro 4.5).	131

Chapter 5

Table 5.1 Conditions used in the crystallisation of <i>Limulus</i> CRP at a concentration of 7.5mg/ml. small irregular crystals could be seen in the highlighted wells.	136
Table 5.2 Precipitant well buffer investigated in trial 5 using <i>Limulus</i> CRP with <i>E. coli</i> polysaccharide. Highlighted wells indicate growth of crystals.	141
Table 5.3 Crystallisation conditions for <i>Limulus</i> CRP with <i>P. aeruginosa</i> LPS in trial 2. Highlighted conditions indicate formation of small irregular crystals.	143
Table 5.4 Crystallisation conditions in trial 4 using <i>Limulus</i> CRP with <i>P. aeruginosa</i> polysaccharide. Highlighted wells indicate growth of triangular bi-pyramidal crystals.	144
Table 5.5 Crystallisation conditions for <i>H. influenzae</i> oligosaccharide with <i>Limulus</i> CRP. Highlighted wells indicate growth of small triangular bi-pyramidal crystals.	146
Table 5.6 Crystallisation conditions for <i>H. influenzae</i> Eagan wild type oligosaccharide with <i>Limulus</i> CRP. Highlighted wells indicate growth of small crystals.	146
Table 5.7 Crystallisation conditions used in trial 3 for <i>H. influenzae</i> oligosaccharide with <i>Limulus</i> CRP.	147
Table 5.8 Crystallisation conditions used for trial 4 for <i>H. influenzae</i> Eagan wild type oligosaccharide with <i>Limulus</i> CRP.	148
Table 5.9 Crystallisation conditions in trial 5 for <i>H. influenzae</i> Eagan wild type oligosaccharide with <i>Limulus</i> CRP. Grainy crystals seen in highlighted well A1.	149
Table 5.10 Crystallisation conditions in trial 6 for <i>H. influenzae</i> Eagan wild type oligosaccharide with <i>Limulus</i> CRP. Blue highlight well indicates flat glass shaped crystals.	150

Table 5.11 Crystallisation conditions in trial 7 for <i>H. influenzae</i> Eagan wild type oligosaccharide with <i>Limulus</i> CRP. Small rectangular glass shaped crystals seen in highlighted wells B2 and B3.	151
--	-----

Chapter 6

Table 6.1 Crystallographic data collection and refinement statistics for <i>H. influenzae</i> Eagan wild type with rfhSP-D compared with statistics obtained previously for native rfhSP-D* by [Shrive <i>et al.</i> , 2003].	166
--	-----

Table of amino acids

Amino acid	Three letter code	Single letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

List of abbreviations and symbols

CRP	C-reactive protein
LBP	Lipid binding protein
α	alpha
β	beta
CPS	C-polysaccharide
PCho	Phosphocholine
PEtn	Phosphoethanolamine
kDa	kilo Dalton
TNF- α	Tumour necrosis factor alpha
IL-1	Interleukin 1
LPS	Lipopolysaccharide
LOS	Lipooligosaccharide
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser enhancement spectroscopy
cDNA	cloned deoxyribonucleic acid
SP-D	Surfactant protein D
SP-A	surfactant protein A
rfhSP-D	Recombinant fragment of human surfactant protein D
CRD	Carbohydrate recognition domain
ManNAc	N-acetylmannosamine
Fuc	Fucose
Mal	Maltose
Gal	Galactose

GlcNAc	N-acetylglucosamine
Lac	Lactose
GalNAc	N-acetylgalactosamine
Glc	Glucose
KDO	3-deoxy-D-manno-oct-2-ulosonic acid
UDP-GlcNAc	Uridine-di-phosphate Glucosamine
PPEtn	pyrophosphate
OAc	O-acetyl group
ESI-MS	Electrospray ionisation mass spectroscopy
MALDI-MS	Matrix assisted laser desorption ionisation mass spectroscopy
MPD	Methyl pentane diol
sCD14	Soluble CD14
mCD14	Membrane bound CD14
TLR-4	Toll-like receptor 4
EDTA	Ethylenediaminetetraacetic acid

Acknowledgements

The work was carried out entirely by me except where mentioned in the text and as detailed below. It constitutes a significant extension of previous knowledge within the structural biology research group at Keele University.

To my supervisor Trevor Greenhough for providing me with the funding and the opportunity to work in this exciting and great project, I have appreciated that you have always found the time for helping me and guiding me throughout the project, especially with obtaining the X-ray crystallography data at Daresbury and helping me with the interpretation and processing of the data using the silicon graphics suite. I would also like to thank you for all the help during the writing process of the manuscript.

To Annette Shrive for all the input and advice with data and structure analysis.

To Ian Burns for technical assistance throughout, including demonstrating the techniques for using the affinity chromatography instrument, providing me with the initial *Limulus* CRP sample and also showing me how to set up the crystallisation trials. Thank you for all the practical tips regarding the lab work.

To Heidi Fuller at RJAH Oswestry for doing the MALDI MS on the *H. influenzae* Eagan 4A polysaccharide and the *E. coli* lipid A sample.

To Tony Curtis, my advisor, for his scientific guidance, answering my carbohydrate chemistry questions and encouragement when my morale was diminishing with the project, especially during the first year. Thank you!

Also my thanks go to my research group Jenny Patterson and Amy Shaw for some great scientific discussions in the office and in the lab and also for a lot of entertaining moments during the years.

To my parents, my brother and sister and my wife for their patience and support whilst doing my PhD project.

Last but not least, my two sons, Samin and Ayan, for keeping me very busy, especially during the writing-up of this thesis.

Chapter 1

Introduction:

Innate immune proteins and lipopolysaccharides

1.0 Introduction

The human body is constantly under attack by pathogenic microbes such as bacteria, yeast, fungi and viruses. The initial defence against these pathogens is mounted by the host first line of defence known as the innate immune system, discovered by Russian immunologist Ilya Mechnikov in the early 1900s who observed that “phagocytic” starfish cells such as macrophages could engulf completely and destroy the foreign pathogens [see review by Tauber, 2003]. The innate immune system functions very rapidly and uses a variety of mechanism for detecting such pathogens, primarily to ward off the invader and eliminate the infection in the host before its spread. The adaptive immune system also mounts a host defense mechanism, which provides specific recognition of foreign antigens and immunological memory of infection by production of B and T cells. In subsequent attacks pathogens encountered before can promptly be eliminated [see review by Janeway 2001, 2002].

The two classes of innate immune proteins discussed in this thesis are C-reactive protein which is released in the circulatory system [Tillet and Francis, 1930] and human lung surfactant protein D (SP-D) which is released in the lungs [Pattle, 1955]. These proteins are pattern recognition molecules that recognise invariant molecular constituent such as lipopolysaccharide (LPS) that is conserved on the outer membrane surface of all gram-negative bacteria and is involved in the pathogenesis and manifestation of bacterial infection and is the major cause of sepsis. Understanding the molecular mechanism of how C-reactive protein and SP-D recognise LPS [Weiser *et al.*, 1998; Kuan *et al.*, 1992; Ng *et al.*, 2004] may lay the foundation for development of potential therapeutic agents [Crouch, 2010; Clark and Reid, 2003; Hirschfield and Pepys, 2003].

1.1 Human C-reactive protein

Human C-reactive protein (CRP) is an acute phase protein whose serum concentration increases within 24 hours of first stimulus by up to 1000 fold over normal serum concentration (6.8-8.2 µg/ml) [Claus *et al.*, 1976]. In a normal healthy individual, the base line of CRP can vary depending on sex, obesity and genetic polymorphism in the CRP gene [Hirschfield and Pepys, 2003]. The increase in serum concentration may be a result of tissue damage (physiological trauma) and inflammation. The liver is the main site of production under the influence of certain cytokines and hormones produced during an acute phase reaction [Steel and Whitehead, 1994].

At the time of its discovery in 1930 CRP was known as C-reactive substance, because it caused precipitation with “fraction C” of the serum of patients infected with *Streptococcus pneumonia* [Tillet and Francis, 1930]. This was later characterised by calcium dependent precipitation of the pneumococcal C-polysaccharide (CPS) [Macleod and Avery, 1941]. The cause of the CRP-CPS interaction was later found to be the phosphocholine (PCho) component on the teichoic acid that makes up the CPS (Figure 1.1 and 1.2) which is regarded as the principal ligand for CRP [Kaplan and Volanakis, 1974]. In the present era of medicine CRP is an exquisitely sensitive systemic marker of inflammation and tissue damage and is widely used to assess cardiovascular disease [Ridker, 2003; Hirschfield and Pepys, 2003].

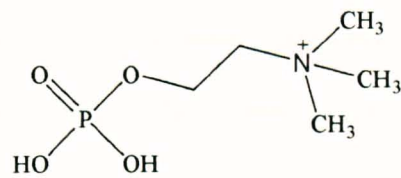


Figure 1.1 Structure of phosphocholine.

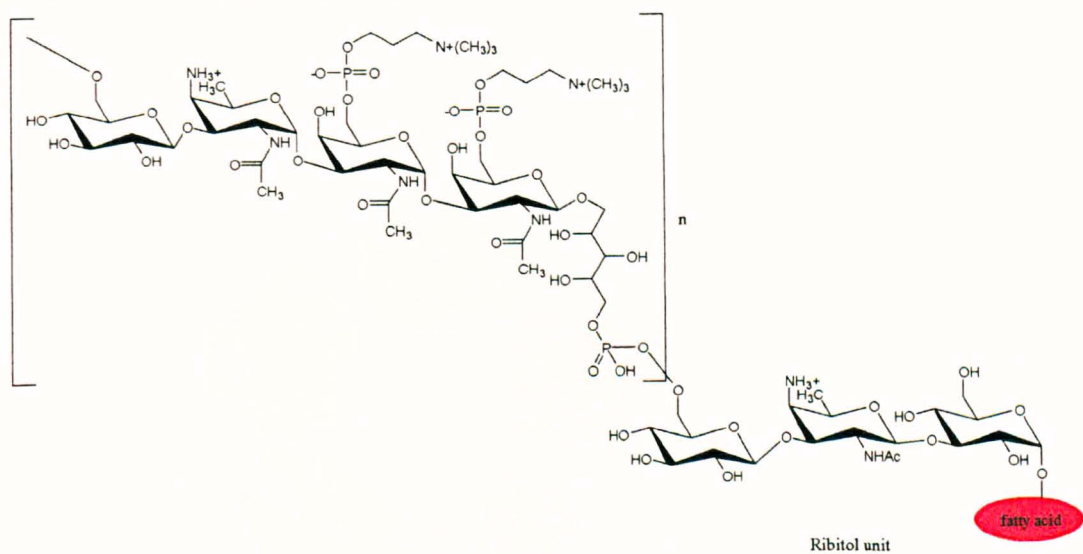


Figure 1.2 Structure of the teichoic acid that makes up the CPS of *Streptococcus pneumoniae*. The repeat unit (n) contains two PCho ligands allowing recognition by CRP. Structure modified from Fischer *et al.*, 1997.

1.1.1 Structure of CRP

Human CRP belongs to a group of proteins classed as the pentraxins because it is composed of five subunits each having a molecular weight of 23kDa [Gotschlich and Edelman, 1965; Osmand *et al.*, 1977]. The initial amino acid sequence indicated 206 amino acids in a single polypeptide chain [Oliveira *et al.*, 1979], later confirmed by cDNA sequencing [Woo *et al.*, 1985; Lei *et al.*, 1985]. The three dimensional crystal structure of the CRP in its native state has been obtained by Shrive and co workers (Shrive *et al.*, 1996) and PCho ligand bound by Thompson and co workers (Thompson *et al.*, 1999) revealing that each subunit is arranged in a pentameric planar arrangement with all subunits tilted and orientated in the same direction (Figure 1.3).

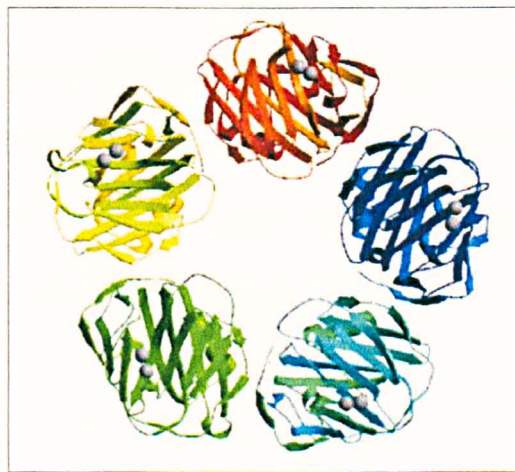


Figure 1.3 Structure of CRP subunits arranged in a pentameric fashion. Gray spheres show the two calcium ions in each subunit on the ‘B face’ [Thompson *et al.*, 1999].

Each pentamer is described as having two faces (designated the “A” face and “B” face here), each subunit having two β sheets with a single α -helix of residues 168-176 on the “A” face folded up against one of the β sheets. This A face is regarded as providing the binding site for a single C1q molecule and/or Fc gamma receptors [Thompson *et al.*, 1999; Agrawal *et al.*, 2001]. The “B” face contains two calcium ions that are bound 4Å apart; the first calcium ion is coordinated by Asp60, Asn61, Glu138, Asp140 and the main chain carbonyl oxygen of residue Gln39. The second calcium ion is coordinated by loop residues Gln138, Asp140 and Gln150 [Shrive *et al.*, 1996; Thompson *et al.*, 1999].

1.1.2 Mechanism of PCho and C1q binding

The crystal structure of CRP with PCho (ligand) demonstrates binding on the ‘B face’ and involves coordination bonds between the phosphate and the calcium ions adjacent to the hydrophobic pocket made by Phe66, Leu64 and Thr76 [Thompson *et al.*, 1999] Figure 1.4. Two key residues mediating the PCho binding are Phe66 and Glu81 located at the opposite end of the hydrophobic binding pocket to the calcium ions. Phe66 provides hydrophobic interactions with the methyl groups of PCho whereas Glu81 interacts with the positively charged choline nitrogen [Thompson *et al.*, 1999].

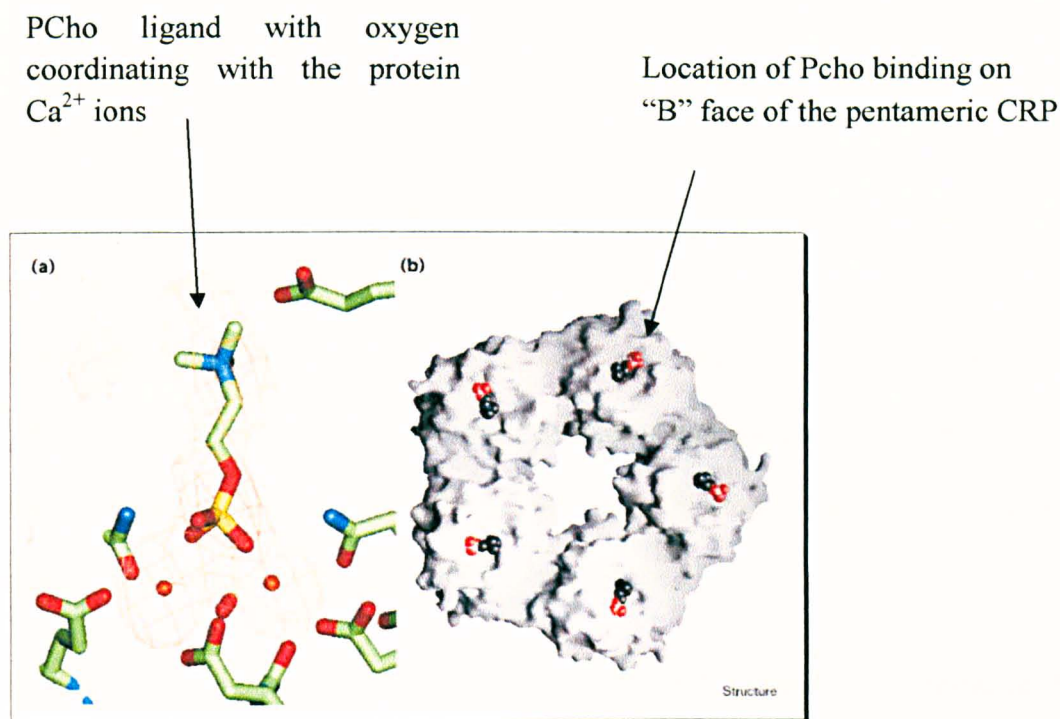


Figure 1.4 a) Mechanism of PCho binding and b) the ‘B’ face on which PCho binding takes place [Thompson *et al.*, 1999].

The orientation of the PCho binding against the two residues also forms a secondary hydrophobic cavity on the CRP surface lined by atoms Glu81, Gly79, Asn61 and Thr76 and this cavity allows the study of structure activity relationship when looking at potential drugs to block possible harmful effects of CRP [Thompson *et al.*, 1999] a theory also supported by earlier work [Srinivasan *et al.*, 1994]. The ‘A face’ is regarded as the effector face allowing binding of C1q and Fc gamma receptors. This binding region has a furrow with a negative charge distribution, with residues Asp112 and Tyr175 being the key mediators for the recognition of C1q [Agrawal *et al.*, 2001] along with other residues along the furrow comprising of Ser5, Arg6, Gln203, Pro206, Trp187, Arg188, Asn160, Gly177, Leu176 and His95 [Thompson *et al.*, 1999]. The importance of these residues has been

demonstrated using mutagenesis studies [Agrawal and Volanakis, 1994; Agrawal *et al.*, 2001].

1.1.3 Synthesis of CRP

CRP synthesis is rapidly up regulated during the acute phase, principally in the hepatocytes, under the influence of a cascade of cytokines involving interleukin 6 (IL-6) which is enhanced by IL-1 β [Hurlimann *et al.*, 1966; Mackiewicz *et al.*, 1991]. Expression of CRP during a systemic response to inflammation is mainly conducted by the liver although extra-hepatic expression of CRP has been found in alveolar macrophages [Gabay and Kushner, 1999; Dong and Wright, 1996].

Although the principal inducer of hepatic CRP is interleukin 6, enhanced by interleukin 1 β , it is also found that IL-1, glucocorticoids and complement activating factor C5a also act synergistically with IL-6 [Toniatti *et al.*, 1990; Ganapathi *et al.*, 1991; Szalai *et al.*, 2000]. Using transgenic mice, Szalai and co workers have shown that IL-6 is not sufficient alone for the induction of the CRP gene [Szalai *et al.*, 1998]. The gene for CRP is located on the long arm of chromosome 1 between 1q21 and 1q23, containing 2263 nucleotides consisting of two exons separated by an intron of 278 base pairs [Woo *et al.*, 1985].

The first exon encodes for a leader peptide and the first amino acid of the mature protein and the second exon encodes the remaining 204 amino acids. The coding region 1q21 and 1q23 is also found to be flanked by other genes with roles specifically coding for immunity and defence, for example serum amyloid protein (SAP) which is another pentraxin sharing sequence and structure homology with CRP [Shrive *et al.*, 1999; Emsley *et al.*, 1994; Hind *et al.*, 1985]. It is thought that these two proteins, both of which are almost always found in other species of mammals and in fish [Shrive *et al.*, 1999; Pepys *et al.*, 1978; Robey *et al.*,

1983; Maudsley *et al.*, 1987] may have evolved from a common ancestor. Both pentraxins are present in the invertebrate *Limulus polyphemus* (horseshoe crab), a living fossil which has existed on earth for over 500 million years [Robey and Liu, 1981; Marchalonis and Edelman, 1968; Shrive *et al.*, 1999].

1.1.4 Defence against microbial pathogens

CRP conjugates with pathogens such as the C polysaccharide (CPS) of *Streptococcus pneumonia* and the LPS of *Haemophilus influenzae* via the PCho ligand [Mold *et al.*, 1981; Volanakis *et al.*, 1971; Weiser *et al.*, 1998]. An *in vivo* study showed that transgenic mice expressing human CRP, when infected with *Streptococcus pneumonia*, showed reduced bacteraemia, prolonged median survival and increased survival rate [Szalai *et al.*, 1995; 1996]. Once the pathogen is recognised via the PCho ligand on the surface of such microbes, a secondary binding can occur with the C1q complement protein [Kishore and Reid, 2000] and the complement system is activated via the classical pathway [Kaplan and Volanakis, 1974]. CRP also promotes phagocytosis by neutrophils and blood derived macrophages [Kindmark, 1971].

1.1.5 CRP binding to host altered host cells

Kushner and co workers (1961) have initially shown that CRP has the ability to bind to damaged or necrotic cell membranes and not to intact cell membranes. This occurs because in damaged cells [Li *et al.*, 1994; Narkates *et al.*, 1982] the polar head group exposes the PCho to CRP. Compromised cells such as apoptotic and necrotic cells are recognised by CRP and subsequently removed by phagocytosis preventing autoimmune disease [Kim *et al.*, 2003; Hanayama *et al.*, 2004].

Recognition and binding of CRP to nuclear constituents such as histones, nucleosomes and chromatin, and polycations occurs at a much higher affinity compared to PCho, enabling clearance of cellular debris and hence a decreased hazard of autoimmunity [Robey *et al.*, 1985].

1.1.6 CRP and complement activation

The complement system is regarded as the host defence mechanism against infection and also plays a crucial part in humoral system removing endogenous ligands such as dying cells and DNA molecules, complement consists of approximately more than 30 plasma and cell-surface proteins (that circulate in the blood) and sets in motion the biological activities as the defence's final outcome: opsonisation of pathogens; chemotaxis and activation of leucocytes; and direct killing of pathogens [Fujita, 2002]. The complement system is triggered by various stimuli and is actually a series of chain reactions consisting of three activation pathways: the classical, lectin and alternative pathway, which cleave the third complement component, termed C3. Cleavage of this component is the main step in the complement chain reaction which brings about the inflammation response and removal of the pathogen [Fujita, 2002] Figure 1.5.

The classical pathway: this pathway is activated by antibody-antigen complexes such as the interaction between C1q and CRP molecule. The mode of activation in this pathway is that the head of a single C1q molecule [Kishore and Reid, 2000] recognises and binds to the CRP pentamer causing the activation in the C1r and C1s serine proteases for the proteolytic cleavage of C4 and C2 to form a C3 convertase C4b and C2a [Agrawal *et al.*, 2001; Kaplan and Volanakis, 1974; Fujita, 2002]. Activation of the classical pathway by CRP is restricted to the formation of C3 convertase followed by the formation of opsonic

fragments C3 and C4 which coat the surface of CRP bound antigen for macrophages and neutrophils [Berman *et al.*, 1986]. This process promotes an anti-inflammatory response, because the latter part of the classical pathway (formation of C5 convertase and C5) which promotes release of pro inflammatory mediators such as TNF- α , IL-1 and membrane attack complex C5-C9 (associated with inflammation) is inhibited by the large numbers of CRP molecules, which recruit and bind factor H (a complementary regulatory protein) that accelerates the decay of C3 and C5 convertase [Mold *et al.*, 1984; 1999].

The lectin pathway: carbohydrates such as mannose and N-acetylglucosamine on the surface of microbes are recognised and bound by mannose binding lectin (MBL) or ficolins [Mizuno *et al.*, 1981; Drickamer *et al.*, 1986], leading to the subsequent activation of the MBL-associated serine proteases (MASP1 and MASP2) [Takahashi *et al.*, 2008; Matsushita and Fujita, 1992; Sato *et al.*, 1994; Thiel *et al.*, 1997; 2000]. Similar to the function of C1s in the classical pathway, MASP2 causes cleavage of C4 and C2 to form C3 convertase C4b and C2a, whereas MASP1 is able to cleave C3 directly resulting in the activation of the alternative pathway.

The alternative pathway: this pathway does not rely on antigen-antibody interactions for its activation but depends on the spontaneous conversion of circulating C3 in the blood to activated C3b forming C3bBb in the presence of factor D and factor B. This pathway provides a positive feedback loop that boosts production of C3 [Brouwer *et al.*, 2006].

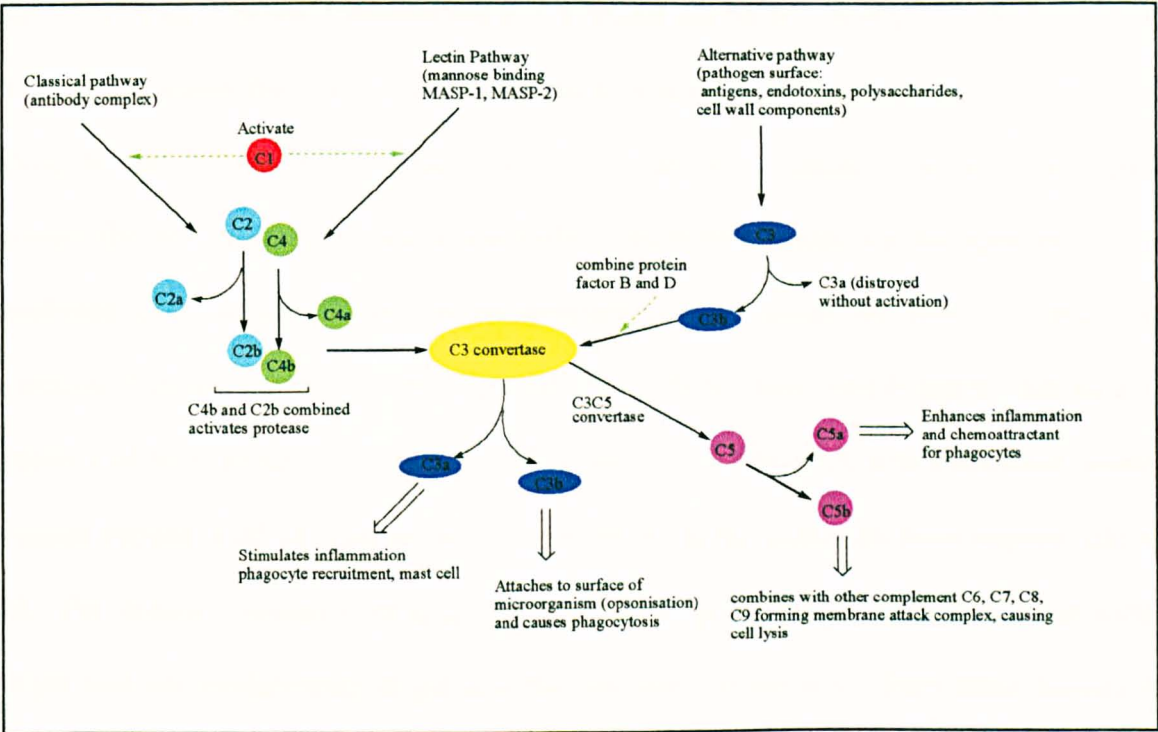


Figure 1.5 The complement pathway showing the three pathways that converge to the common complement protein C3. Figure adapted from Fujita, 2002.

1.1.7 *Limulus* CRP

Unlike human CRP, which is an acute phase protein, *Limulus* CRP is constitutively expressed in the creature's haemolymph at a concentration of 5-6 mg/ml [Robey and Liu, 1981] and, because the creature is found to lack immunoglobulins, CRP can be considered acting as the creature's immunoglobulin in its immune system. *Limulus* is a marine creature (horseshoe crab), a species hundreds of millions of years old that appears to have developed a remarkable ability in its immune system, hence its survival throughout evolution. Nguyen and co workers [Nguyen *et al.*, 1986a] have established the amino acid sequence of three subunits of *Limulus* CRP termed 1.1, 1.4 and 3.3, with a common amino-terminal sequence of 44 residues and a carboxyl-terminal sequence from residue 206 to 218. The primary structure obtained by Nguyen [Nguyen *et al.*, 1986a] also provides insight into the evolutionary origin and the deviation in sequence from other species in which it is found, such as humans and rabbits. Nguyen and co workers hypothesise that human and *Limulus* CRPs diverged 500 million years ago [Nguyen *et al.*, 1986a; Shrive *et al.*, 1999; 2009]. Nguyen and co workers [Nguyen *et al.*, 1986b] found three separate CRP genes suggesting that there is more than one form of *Limulus* CRP present in a single pool (human CRP is coded by a single gene). As yet, no crystal structure has been reported for the *Limulus* CRP.

1.1.7.1 Structure of *Limulus* CRP

Early research suggested that *Limulus* CRP has a molecular weight of 500 kDa, consisting of two non identical subunits with a molecular weight of 18 kDa to 24 kDa [Robey and Liu, 1981]. Electron microscopy suggested that the subunits are aggregated in a doubly stacked cyclic hexamer (human CRP is a single pentamer) [Fernandez-Moran *et al.*, 1968]. Quigley and co workers in 1994 identified a trace CRP, limulin, similar to *Limulus* CRP

but having sialic acid binding activity [Quigley *et al.*, 1994]. Since there is more than one form of protein present coded by multiple genes, each form of CRP could be specifically tuned for specific ligand binding properties [Nguyen *et al.*, 1986b; Shrive *et al.*, 1999] hence providing optimum immunity to the creature in parallel to human immunoglobulins. Tennent and co workers have found that there is the tendency of one form of CRP amongst the isoprotein mixture to predominate in the plasma [Tennent *et al.*, 1993]. Indeed this has been found by Iwaka and co workers who identified and isolated three groups of CRP from the Japanese horseshoe crab *Tachypleus tridentatus*, termed tCRP-1, tCRP-2 and tCRP-3. With the exception of tCRP-3, the other two forms could constitute isoprotein mixtures [Iwaka *et al.*, 1999]. tCRP-1 is found to be the counterpart of *Limulus* CRP, tCRP-2, a counterpart of limulin and tCRP-3 shows similarity to the *Limulus* amino acid sequence but the structure as revealed by electron microscopy has a slightly bigger diameter (14.5nm for tCRP-3 and 10nm in *Limulus* CRP) [Iwaka *et al.*, 1999].

1.1.7.2 Ligand binding and function of *Limulus* CRP

Similarly to human CRP, *Limulus* CRP also recognises and calcium-dependently binds to CPS containing PCho (Figure 1.1) and to phosphoethanolamine (PEtn) (Figure 1.6), but with higher affinity for PEtn as compared to human CRP, which has higher affinity for PCho [Shrive *et al.*, 1999; Robey and Liu, 1981; Oliviera *et al.*, 1980].

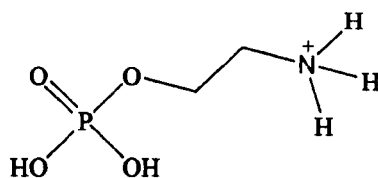


Figure 1.6 Structure of Phosphoethanolamine.

tCRP-1 and tCRP-2 showed binding activity to PCho-agarose but tCRP3 showed no binding [Iwaka *et al.*, 1999]. tCRP-2 and tCRP-3 showed haemolytic and haemagglutinating activity [Iwaka *et al.*, 1999]. Tan and co workers [Tan *et al.*, 2005] show that a single isoform of CRP-2 from the horseshoe crab *Carcinoscorpius rotundicauda* has binding activity against the gram negative bacteria *Escherichia coli* and *Pseudomonas aeruginosa* in a calcium independent manner as well as agglutinating activity against the polysialic capsule (poly- α (2 \rightarrow 8)-Neu5Ac) of *Escherichia coli* K1. In a similar work Ng and co workers [Ng *et al.*, 2004] have found that *Limulus* CRP was able to bind to the LPS from *Pseudomonas aeruginosa*. The ability of *Limulus* CRP to recognise and bind to PCho or PEtn (Figure 1.1 and 1.6) present on microbial surfaces indicates its pivotal role in the creature's immune function, as well as other biological functions that are still not determined because there are several forms of CRP with the various forms showing differing binding activities.

1.2 Human lung surfactant protein D

Lung surfactant protein D (SP-D) constitutes approximately 10% of the surfactant protein present in the lungs, including SP-A, SP-B, SP-C and SP-D [Clark and Reid 2003]. These lung surfactant mixtures reduce alveolar surface tensions and airspace potency as well as providing innate immune protection in the lung environment. Surfactant proteins are recognised as central to respiratory distress syndrome in infants born prematurely, due to deficiency in SP-B and SP-C.

SP-D is a glycoprotein belonging to the collectin family of innate immune molecules, which also includes mannan-binding lectin (MBL), SP-A, collectin of 43kDa (CL-43), collectin of 46kDa (CL-46), conglutinin, CL-L1 and CL-P1 [Ohtani *et al.*, 1999; 2001; Koenraad van de Wetering *et al.*, 2004; Holmskov, 1994]. SP-D is involved in the first line of defence in the lung against a broad spectrum of inhaled pathogens by binding to pathogens and promoting their elimination from the lung by various mechanisms that shall be discussed in later sections. In principal the biological function of SP-D can be categorised in terms of four biological areas; pathogen and organic antigen recognition, immune modulation, control of inflammation and antioxidant properties within the lungs [Clark and Reid, 2003; Clark, 2010].

The biological secretion of SP-D in structural form is higher order multimers, which increases its binding affinity towards microbial pathogens. Each part of the whole length of the protein is considered to be involved in the preservation of certain functions.

There is increasing interest in the construction of recombinant fragments of human SP-D that are much smaller than intact SP-D, consisting of the neck region and the carbohydrate recognition domain (CRD), but which maintain important biological functions.

Preservation of activity is the key to drug development towards affecting the pathogenesis of human lung disease. According to Clark and co-workers the areas of interest and potential use for the recombinant form of SP-D lie in the management of neonatal chronic lung disease, cystic fibrosis, emphysema, microbial infection and asthma [Clark and Reid, 2003].

1.2.1 Biosynthesis and secretion of SP-D

The major site of production of SP-D appears to be the lungs, where it is secreted by alveolar type II cells and by a subset of bronchiolar epithelial cells, the non-ciliated Clara cells where it is stored and regulated [Crouch *et al.*, 1991; 1992; Voorhaut *et al.*, 1992]. Extra pulmonary sites of expression have been determined such as mucosal surfaces in human tissues [Madsen *et al.*, 2000; Oberley *et al.*, 2004] and amniotic epithelial cells [Miyamura *et al.*, 1994]. Hormones and transcription factors control the expression of SP-D within and outside of the lungs [Dulkerian *et al.*, 1996; Lin and Floros, 2002], levels increasing with gestation from 16 weeks gestation in human foetus [Dahl *et al.*, 2005]. SP-D is coded by a single gene termed SFTPD [Crouch *et al.*, 1993; Kolble *et al.*, 1993] which is localized to region 10q22.2-23.1 on the long arm of chromosome 10 (Figure 1.7) [Floros and Hoover, 1998; Sorensen *et al.*, 2007]. The 10q22.2-23.1 region is also responsible for the coding of another collectin member SP-A, which consist of two genes SFTPA1 and SFTPA2 [White *et al.*, 1985; Katyal *et al.*, 1992; Sorensen *et al.*, 2007].

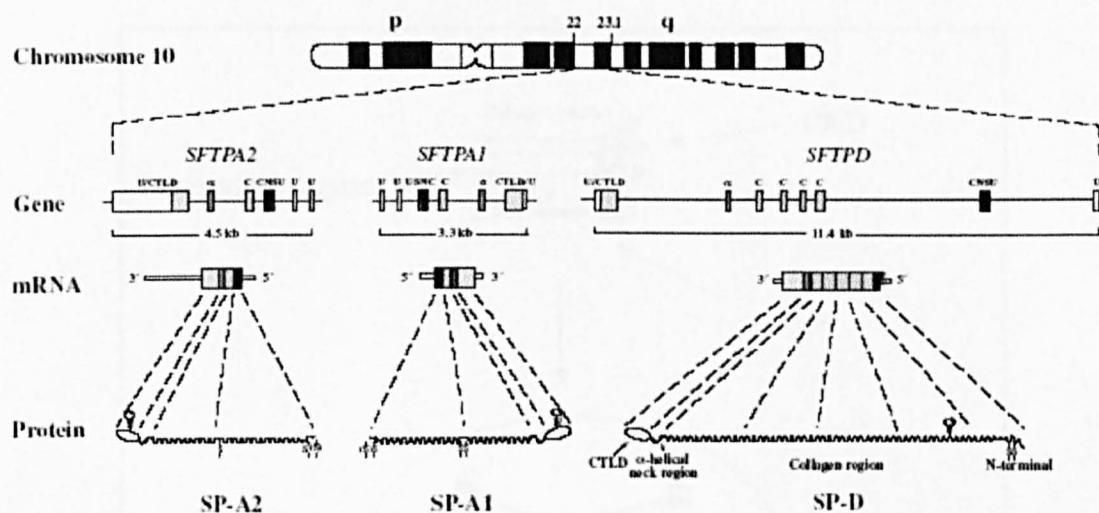


Figure 1.7 Synthesis of SP-D as well as SP-A located on the long arm of chromosome 10. Image taken from Sorensen *et al.*, 2007.

1.2.2 Structure of SP-D

Actively secreted SP-D in the lung is assembled as multimers [Crouch *et al.*, 1994]. The general architecture is seen in all members of the collectins except that SP-D is the longest of the series. The basic unit of the SP-D molecule (as well as all collectin members) is a trimer, having a molecular mass of 43 kDa, each chain consisting of four discrete structural domains; a short cysteine rich N-terminal domain, a collagenous domain, a coiled coil neck domain and a C-terminal known as the carbohydrate recognition domain (CRD). The trimer is formed from this single polypeptide chain by association at the N terminal domain [Koenraad van de Wetering *et al.*, 2004]. The trimer then has the ability to form tetramers by being associated at the N-terminus region (Figure 1.8).

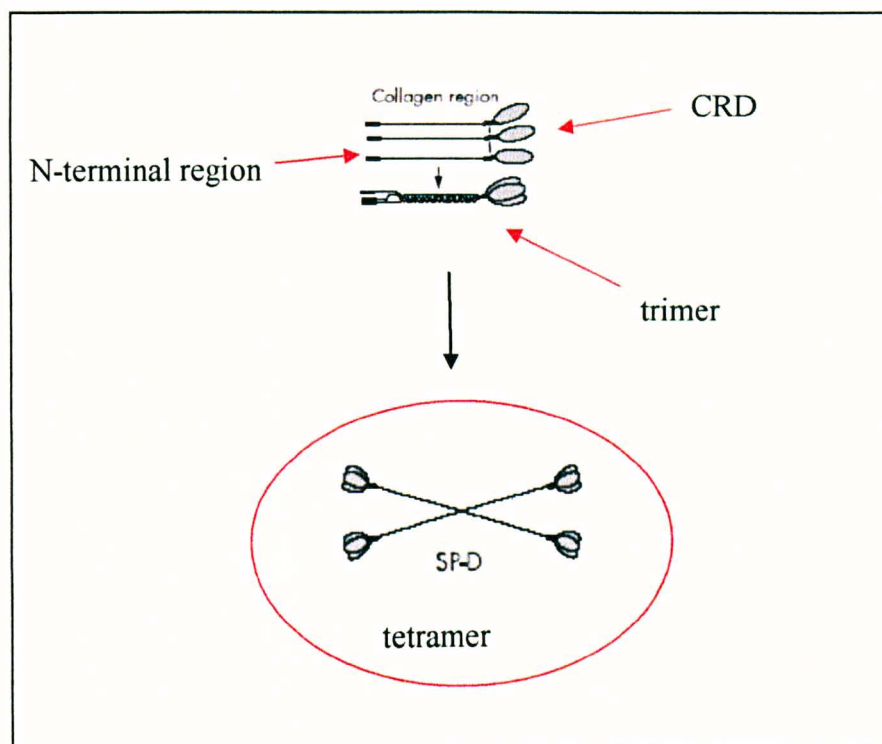


Figure 1.8 Schematic description of the formation of the SP-D tetramer. Each polypeptide subunit is trimerised which is then further oligomerised by being tethered together at the N-terminus by disulphide bonds. Image adapted from Clark and Reid, 2002.

The first crystal structure of SP-D [Hakansson *et al.*, 1999] was that of the recombinant trimeric fragment rfhSP-D, expressed in *Pichia pastoris*, consisting of the neck region and the globular carbohydrate recognition domain. Subsequently the structure of the recombinant fragment expressed in *Escherichia coli* was shown in the native state and with various ligands bound [Shrive *et al.*, 2003; 2009].

The crystal structure reveals the neck and the three carbohydrate binding domains and each polypeptide chain is labelled A, B and C. The structure reveals similarities with the crystal

structure of other collectin family members (Figure 1.9a) [Weis *et al.*, 1991; 1992; Sheriff *et al.*, 1994; Weis and Drickamer, 1994; Head *et al.*, 2003].

1.2.2.1 N-terminal region

The SP-D N-terminal domain contains 25-28 amino acids containing two conserved cysteine residues. Cysteine residues are involved in interchain disulphide crosslinks stabilising the trimer as well as linking them to form tetramers and higher order oligomers [Crouch *et al.*, 1994a]. At least two cysteine residues are required for multimerisation in the N-terminal region [Brown-Ausberger *et al.*, 1996; McCormack *et al.*, 1997; 1999; Wallis and Drickamer, 1999]. It has also been demonstrated that deletion of cysteines in this region leads to the formation of trimers [Brown-Ausberger *et al.*, 1996; McCormack *et al.*, 1997; 1999]. CL-L1 contains a single cysteine residue and is secreted as a trimer [Ohtani *et al.*, 1999]. Conflicting evidence suggest that other factors could be responsible for multimerisation since CL-43, containing two cysteine residue, is found only as a trimer, whilst in SP-D, cysteine occupies the same position as CL-43 yet displays highly multimerised forms [Rothman *et al.*, 1997; Lim *et al.*, 1994; Van Eijk *et al.*, 2000].

1.2.2.2 Collagen region

The collagen region is characterised by the collagen Gly-X-Y repetitive pattern, where X and Y can be any amino acid but usually prolines and hydroxyprolines forming a left-handed helix. The length of this region differs between collectins with SP-D having the longest length with 177 amino acid residues [Hakansson and Reid, 2000]. In the trimer the chains are coiled around each other in a right handed fashion orientating the glycine residues in the interior of the helix. Further stabilisation of trimeric structure is achieved by hydrogen bonds between the N-H of glycine and the C=O group of the amino acid in

the X position. The collagen region can also be N-glycosylated or O-glycosylated [Crouch *et al.*, 1994a; b; Colley and Baenziger, 1987]. The polypeptide chain in MBL and SP-A has a bend not seen in SP-D or other collectin members because the repeat motif Gly-X-Y is interrupted thereby producing a kink into the protein [Koenraad van de Wetering *et al.*, 2004]. Triple helix formation in MBL mediates binding of other macromolecules such as C1q and serine proteases (MASP-1 and MASP-2), thereby triggering the complement cascade [Matsushita and Fujita 1992; Thiel *et al.*, 1997].

1.2.2.3 α -helical coiled coil

The α -helical neck region, also important in forming and stabilising the trimer, is characterised by the heptad amino acid repeat pattern a-b-c-d-e-f-g, where residues a and d are hydrophobic amino acid (Val204, Leu207, Val211, Leu214, Val218, Leu221, Phe225, Tyr228 and Glu232) [Hakansson *et al.*, 1999; Hakansson and Reid, 2000]. In SP-D, 28 amino acids make right handed α -helices that twist around in a left handed manner, the coil being broken by Pro235 that is unable to form hydrogen bonds due to a *cis* peptide bond. The Tyr228 residue of chains A and B are exposed forming hydrogen bonds with water molecules whereas this residue in chain C is completely buried in the coiled coil. Three glutamic acid residues from each chain are in close contact with each other as well as with Lys246 residues from the carbohydrate recognition domain. The constellation found in this region has a positive electrostatic surface potential due to the presence of Lys246 and Lys252 and the negative electrostatic potential due to Glu232 residues (Figure 1.9b) [Hakansson *et al.*, 1999].

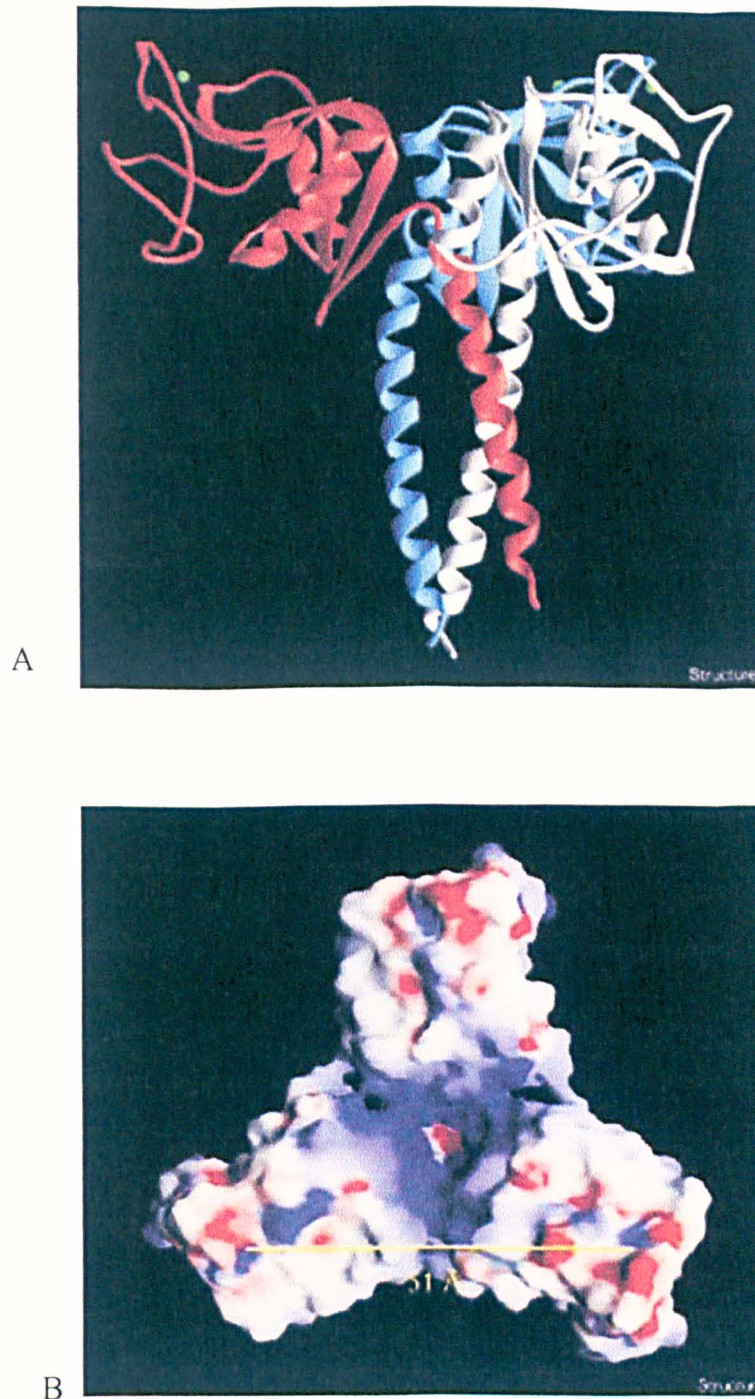


Figure 1.9 A) Structure of recombinant fragment of SP-D with each chain A, B and C shown in different colours. B) Electrostatic surface potential as viewed directly on top of the carbohydrate recognition domain. Blue regions show the positive charge distribution from Lys246 and Lys252 and the red regions indicate negative charge distributions from Glu232. Green dots show Ca^{2+} ions [Taken from Hakansson *et al.*, 1999].

1.2.2.4 Carbohydrate recognition domain

A great amount of structural information on the CRD has been obtained from many collectins using X-ray crystallography of not only SP-D but also rat and human mannan-binding protein [Weis *et al.*, 1991; 1992; Sheriff *et al.*, 1994; Weis and Drickamer, 1994], SP-A and SP-D [Head *et al.*, 2003; Hakansson *et al.*, 1999; Shrive *et al.*, 2003; 2009]. These studies show that the CRD is described as having a double loop made up of 115-130 amino acid residues folded into two anti parallel β sheets. The CRD is stabilised by two disulphide bridges between the β sheets and hydrophobic and polar interactions with the α -helical coiled coil. The second loop includes the calcium and ligand recognition sites described as having a shallow groove [Weis *et al.*, 1991; 1992; Ng *et al.*, 1996; Hakansson *et al.*, 1999; Hakansson and Reid, 2000; Shrive *et al.*, 2003; 2009] Figure 1.10.

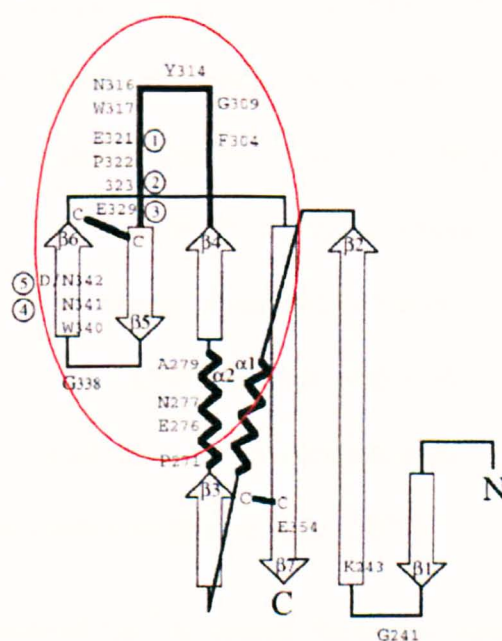


Figure 1.10 Schematic representation showing the human SP-D CRD as a double loop showing the series of β sheets and α -helices. Calcium binding and ligand recognition is shown in the second loop (circled red). Encircled numbers denote amino acids that ligate to the principal calcium ion and are also involved in ligand binding. Taken from Hakansson and Reid, 2000.

1.2.3 Calcium ions in the CRD

Each CRD (subunits A, B and C) contains three Ca^{2+} ions located on the surface of the second loop [Hakansson *et al.*, 1999; Shrive *et al.*, 2003; 2009] and an additional calcium ion has been found recently by Shrive and co workers [Shrive *et al.*, 2003; 2009] located in the central axis of the trimeric pore of rfhSP-D. The first Ca^{2+} (known as the principal ligand binding Ca^{2+} ion) is involved in carbohydrate ligand binding which is coordinated through five amino acid residues Glu321, Asn323, Glu329, Asn341 and Asp342 (Figure 1.10).

The second and third Ca^{2+} are in close vicinity to the ligand binding Ca^{2+} ion, separated by 3.9-4.0Å and coordinated to the lectin by residues Asp297, Glu301, Asp324, Asp330 and also Lys329. The third Ca^{2+} is coordinated by three acidic side-chains of Glu301, Asp330 and water molecules [Shrive *et al.*, 2003; 2009]. The fourth Ca^{2+} ion is coordinated by three Glu232 side chains which also interact with Lys246 of the same chain. The central pore of the trimeric axis is therefore positively charged due to the exposed lysine (Lys). The fourth Ca^{2+} ion detected by Shrive and co workers (2003), using rfhSP-D in the native state, was found to be absent when rfhSP-D was crystallised with maltose. However in subsequent studies using simple ligands with rfhSP-D (discussed in later chapters), this Ca^{2+} was always present when a concentration equal to or greater than 2mM was present in the crystallisation conditions [Shrive *et al.*, 2009].

Although the principal Ca^{2+} is crucial for ligand binding as observed by Shrive and co workers in the rfhSP-D bound to maltose carbohydrate ligand [Shrive *et al.*, 2003, 2009] and in similar studies using MBL, [Weis *et al.*, 1991; Sheriff *et al.*, 1994; Weis and Drickamer, 1994] no definite function or participation has been recorded for the remaining

Ca^{2+} ions. On the contrary, according to Shrive and co workers [Shrive *et al.*, 2003] the role of the two Ca^{2+} ions in the vicinity of the principal Ca^{2+} ions may be stabilisation of the ligand binding loop. The fourth Ca^{2+} ion may function as a molecular switch, keeping the symmetrical Glu232 and Lys246 conformation in all the three chains, possibly by keeping a tight coordination bond with Glu232. It is found that in the absence of this fourth Ca^{2+} there appears to be a conformational change affecting residues Glu232 and Tyr228 [Shrive *et al.*, 2009]. An alternative function could be to neutralise the charge density on the surface of the lectin [Hakansson *et al.*, 1999].

1.2.4 Functions of SP-D

SP-D has a diverse role of functions such as clearance of bacteria and fungi, viral neutralization, clearance of host altered cells and dying cells such as apoptotic and necrotic cells from lung circulation, and down regulation of allergic reaction and resolution of inflammation as well as maintenance of surfactant homeostasis in the lung [Hartshorn *et al.*, 2007; 1998; Korfhagen *et al.*, 1998; Clark and Reid, 2002; Madan *et al.*, 2001; Kishore *et al.*, 2006].

Outcome of microbial recognition is to remove the pathogen from the lung which may involve microbial agglutination, enhanced binding of microorganisms, phagocyte activation and opsonisation [Lawson and Reid, 2000]. These effector mechanisms rely on the nature of the secreted SP-D in the lungs, for example, oligomerisation into multimers greatly enhances the avidity of protein-ligand interactions causing massive agglutination by forming bridges with the ligands on the surface of different individual bacteria. Higher order multimers promote aggregation of pathogens more effectively compared to dodecamers or trimers [Lawson and Reid, 2000].

1.2.4.1 Microbial and organic antigen interactions

In vitro and *in vivo* experiments using recombinant SP-D (rfhSP-D) expressed in yeast and *Escherichia coli* have shown that this small fragment consisting of 177 C-terminal residues (Gly179-Phe355) with a molecular weight of 60kDa (entire SP-D molecule is 540 kDa) [Shrive *et al.*, 2003, 2009] shows significant beneficial biological activity. *In vivo* studies have used transgenic mice and mouse models of infection and inflammation in the lungs [Madan *et al.*, 2001a; 2001b]. These experiments used a model of pulmonary hypersensitivity induced by *Aspergillus fumigates*, an opportunistic fungal pathogen in the lungs of immune compromised or neutropenic patients [Madan *et al.*, 2001a; 2001b; 1997].

The beneficial effect of rfhSP-D is brought about by the ability of the SP-D molecule to interact with the glycosylated antigens and allergens of *Aspergillus fumigates* and hence inhibit specific IgE binding to these allergens [Madan *et al.*, 2001a; 2001b; 1997]. Recognition was found to occur by interactions with mannan and β -(1-6) linked glucan, a structural component of the cell wall. Mannan is a branched homopolymer of mannose-residues coupled to each other via varying glycosidic linkages. SP-D subsequently aggregates the particle [Allen *et al.*, 2001].

Inhibition of allergen caused by house dust mite (*Dermatophagoides pteronyssinus*) and extract has also been reported with binding to the glycoprotein allergen inhibiting the binding of allergen-specific IgE to mite allergens [Wang *et al.*, 1996]. SP-D can interact with gram negative bacteria by binding to membrane anchored rough forms of lipopolysaccharide which contains saccharide units displaying vicinal equatorial hydroxyl groups. Kuan and co workers have reported that *Escherichia coli* expressing O-deficient or short O-antigen LPS are recognised by binding to the inner core containing glucose and

heptose [Kuan *et al.*, 1992]. SP-D only recognises *Escherichia coli* strains that are O-antigen deficient or contain short O-antigen regions [Kuan *et al.*, 1992]. Once bound it causes massive aggregation leading to uptake of bacteria by neutrophils. In a similar work, another group of workers [Ofek *et al.*, 2001; Sahly *et al.*, 2002] have demonstrated the recognition and binding of SP-D to strain specific bacteria such as the respiratory pathogen *Klebsiella pneumonia*, and it is found that SP-D binds to the inner core of the LPS of strains that possess a polysaccharide capsule (envelope) on the outer membrane. This envelope restricts the accessibility of binding to the mannose unit, but efficiently binds to the O3 and O5 variant form lacking the capsule containing O-antigen repeat unit -3)- α D-Man-(1-3)- α D-Man-(1-2)- α D-Man-(1-2)- α D-Man-(1. Binding of the core oligosaccharide is also influenced by the number of repeating saccharide units associated with the O-antigen [Ofek *et al.*, 2001; Sahly *et al.*, 2002].

Mycobacterium tuberculosis and *Mycobacterium avium*, another lung pathogen that resides in the macrophage and evades detection, is bound by SP-D through recognition of the terminal mannosyl-oligosaccharide of lipoarabinomannan, reducing the uptake of bacteria by alveolar macrophages [Ferguson *et al.*, 1999; 2002]. Koenraad van de Wetering and co workers (2001) have shown that SP-D recognises the gram positive bacteria *Bacillus subtilis* and *Staphylococcus aureus* via the CRD by binding to the cell wall components of lipoteichoic acid and peptidoglycan, which are the two major cell wall components. SP-D also recognises and interacts with viruses by recognising and binding to the glycoprotein expressed on the viral envelope proteins, in calcium dependent and sugar sensitive interactions [Meschi *et al.*, 2005; Hickling *et al.*, 1999; Hartshorn *et al.*, 2000].

Human immunodeficiency virus (HIV) is recognised by SP-D by calcium-dependent binding to the envelope protein (gp120) which contains highly conserved mannosylated oligosaccharides, inhibiting the viral replication [Meschi *et al.*, 2005]. It is also found that native dodecameric SP-D binds more strongly than native trimeric SP-D [Meschi *et al.*, 2005].

Respiratory syncytial virus, commonly prevalent in neonates and infants, is bound by SP-D by recognition of the highly glycosylated protein on the surface of the virus (Figure 1.11) [Hickling *et al.*, 1999]. Influenza A virus (IAV) is bound by SP-D via the CRD recognising hi-mannose oligosaccharides in a calcium dependent manner on the two envelope glycoproteins, haemagglutinin and neuraminidase [Hartshorn *et al.*, 2000]. Binding to these molecules prevents the virus from binding to host respiratory cells. SP-D binding to neuraminidase inhibits neuraminidase enzyme activity and causes massive aggregation as well as enhancing the binding and uptake of IAV by neutrophils [Hartshorn *et al.*, 1994; 1997; 2000; Crouch *et al.*, 2009].

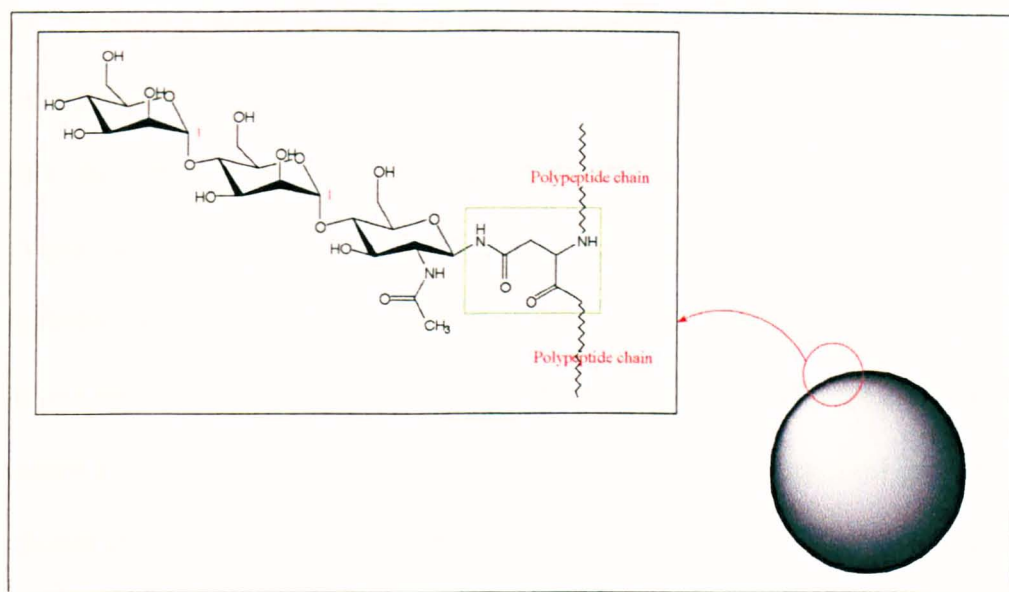


Figure 1.11 Illustration of high mannose oligosaccharide linked to protein (glycoprotein) on the surface of IAV and HIV virus (dark sphere). The high mannose oligosaccharide unit is N-linked to amino acid (green box) Structure drawn using CS Chem. Draw Pro. Version 4.5.

1.2.4.2 Immune modulation and control of inflammation

Secondary immune responses are achieved by modulating immune cell functions including macrophages, monocytes and neutrophils. Nadesalingham and co workers have shown that SP-D binds to several classes of antibody molecules and enhances the phagocytosis of IgG-coated molecules [Nadesalingham *et al.*, 2005], bridging the innate and adaptive responses and modulating the regulation of inflammation [Sano *et al.*, 2000; 2006; Vandivier *et al.*, 2002]. SP-D also binds to proteoglycans with long glycosaminoglycan chains such as decorin and this may be related to reducing tissue specific inflammation [Nadesalingham *et al.*, 2003] SP-D also recognises surfactant components such as glucosylceramide, which is the most abundant neutral glycolipid present on lymphoid cells, phosphatidylinositol and myo inositol (Figure 1.12) [Ogasawara *et al.*, 1995; Persson

et al., 1992]. SP-D is seen to bind to glucosylceramide through interactions of the carbohydrate binding sequence of the CRD with the glycosyl moiety. SP-D also binds to self derived ligands such as apoptotic cells and regulates inflammatory responses [Vandivier *et al.*, 2002]. Binding to apoptotic cells occurs because these contain carbohydrate polymers such as DNA on their surfaces. Accumulation of such cells as well as excess free DNA is known to cause inflammation and septic shock, and therefore removal of such particles maintains inflammation-free tissue and prevents autoimmune disease [Savill *et al.*, 2002]. Nucleic acid is bound via the CRD as well as the collagen-like regions and reduces the generation of anti-DNA autoantibody [Palaniyer *et al.*, 2005].

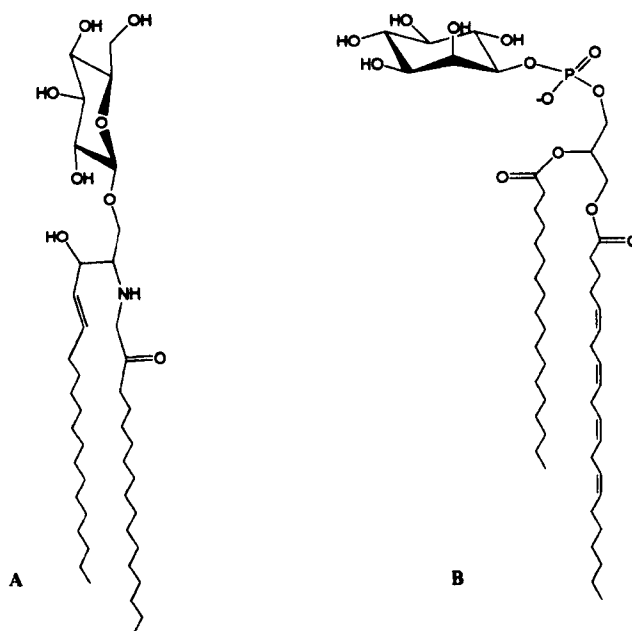


Figure 1.12 A) Glucosylceramide and B) phosphatidylinositol.

1.2.5 Mechanism of ligand binding

The mechanism of how SP-D binds to microbial pathogens or organic antigens (non microbial pathogen) targets has been studied in great depth using binding assays and from crystal structures of a recombinant fragment of the homotrimeric SP-D [Weis *et al.*, 1992, Ng *et al.*, 1996, Shrive *et al.*, 2003, 2009; Crouch *et al.*, 2009; 2005; Allen *et al.*, 2004; 2001]. The main component of the ligand recognised by SP-D as well as other collectin member is one or more simple sugar molecules such as glucose (Glc), rhamnose (Rham), galactose (Gal), mannose (Man), N-acetyl glucosamine (GlcNAc) and N-acetyl galactosamine (GalNAc) that form much larger complex structures such as lipopolysaccharide, teichoic acid, capsular polysaccharide and glycoprotein, which make up the microbial surface components.

The ability of SP-D to distinguish microbial cell surface from host cell surface lies in the geometry of the binding sites on the CRD that are spaced at a distance of 50Å [Hakansson *et al.*, 1999; Head *et al.*, 2003; Shrive *et al.*, 2003; 2009]. This is considered too far apart to bind to host cell surface carbohydrate structures but matches densely populated carbohydrate structures present on microbial cell surface. Binding takes place in a shallow pocket peripherally located on the surface of the CRD some 30Å from the threefold symmetry axis [Weis *et al.*, 1992; Ng *et al.*, 1996; Shrive *et al.*, 2003; 2009; Crouch *et al.*, 2009] and involves coordination with the principal Ca^{2+} ion (Figure 1.13).

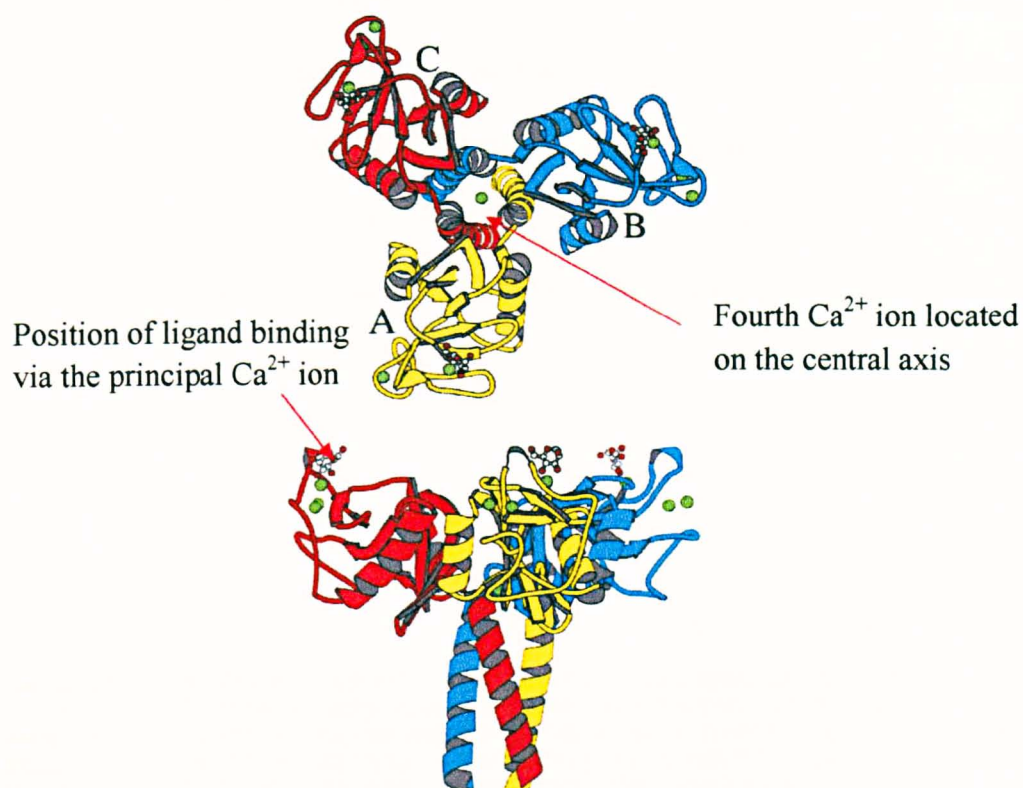


Figure 1.13 Ligand binding region on the second loop of rfhSP-D via coordination with the principal Ca^{2+} ion (Ca1) [Shrive *et al.*, 2003; 2009].

Sequence alignments within the CRD [Shrive *et al.*, 2009] show that the carbohydrate binding site is conserved among the collectins (Figure 1.14). The residues Asp325 and Arg343, which flank the ligand-binding pocket, modulate and fine tune the carbohydrate specificity through non-specific binding and by influencing the ability of the carbohydrate to enter the binding pocket by contributing to steric hindrance [Shrive *et al.*, 2009]. This region also provides varying binding affinity towards different monosaccharide ligands. (Table 1.1).

human SP-D	320	330	340	350	
<u>human SP-D</u>	SLVYSNWAPGEP <u>NDD</u> ---	GGSEDCVEIFTNGKWN <u>DRAC</u> GE-KRLVVCEF	355		
CL-46	SLVYSNWASGEPNNNA-GQ	PENCVQIYREGKWN <u>DVPC</u> SE-PLLVICEF	351		
conglutinin	ILVYSNWADGEPNNSDE-GQ	PENCVEIFPDGKWN <u>DVPC</u> SK-QLLVICEF	351		
<u>human MBL</u>	RLTYTNWNEGEPNNA---	GSD <u>EDCV</u> LLLLKNGQWN <u>DVPC</u> ST-SHLAVCEFPI	228		
<u>rat MBP-A</u>	RLTYSNWKKDEPN <u>DH</u> ---	GSGEDCVTIVDNGLWN <u>DISC</u> QA-SHTAVCEFPA	221		
<u>rat MBP-C</u>	RVRYTNWNEGEPNNV---	GSGENCVLLTNGKWN <u>DVPC</u> SD-SFLVVCEFS	226		
<u>rat SP-A</u>	SVNYTNWYPGEPR <u>GQ</u> ---	GK-EKCVEMYTDGTWN <u>DRGC</u> LQ-YRLAVCEF	228		
human SP-A1	PVNYTNWYRGEP <u>AGR</u> ---	GK-EQCVEMYTDGQWN <u>DRNCL</u> Y-SRLTICEF	228		

Figure 1.14 C-terminal sequence of selected collectins (hSP-D numbering) Known neck/CRD structures are underlined, reported ligand bound structures are highlighted. Residues that bind to both calcium and ligand are underlined. Residues that flank the ligand binding site (SP-D325 and 343) are highlighted. Taken from Shrive *et al.*, 2009.

Collectin	Saccharide selectivity
SP-A (human)	ManNAc > Fuc, Mal > Glc > Man >> Gal, GlcNAc
SP-D (human)	Mal > Man, Glc > Lac, Gal > GlcNAc > Fuc
MBL (human)	GlcNAc > Fuc, Man, ManNAc > Mal > Glc >> Gal, GalNAc
CL-P1 (human)	GalNAc > Fuc, Gal>> Glc, GlcNAc, Man
CL-L1 (human)	Gal, Man, Fuc > GlcNAc >> GalNAc
Conglutinin (bovine)	GlcNAc > Fuc, Glc > Man, ManNAc >> Mal, Gal, GalNAc, Lac
CL-43 (bovine)	Man > ManNAc > Fuc, GlcNAc > Glc, Mal > Gal, Lac >> GalNAc
CL-46 (bovine)	GlcNAc >> ManNAc, Man > GalNAc > Man, Glc, Fuc>Gal

Table 1.1 The binding affinity of each collectin towards monosaccharide ligands which is in part determined by the residues flanking the binding region and in part by the residues which directly bind the ligand. Taken from Hickling *et al.*, 2004.

The amino acid residues also specify whether the CRD is a “mannose” or “galactose” type monosaccharide recognition domain, by considering the hydrogen-bond donor and acceptor groups within the Ca^{2+} binding site. The two residues (E321 and P322 (SP-D numbering)) are crucial for determining “mannose-binding” type CRDs and in galactose-binding type CRDs the two residues are QP [Drickamer, 1992; Weis *et al.*, 1992; Ogasawara *et al.*, 1994; Kolatkar and Weis, 1996; Meir *et al.*, 2000; Shrive *et al.*, 2009].

Mutational studies using MBP-A and SP-A have demonstrated that changing the amino acid residues in the binding site that coordinate to the Ca^{2+} changes the specificity of the molecule from mannose type to Gal type [Drickamer, 1992; Weis *et al.*, 1992; McCormack *et al.*, 1994].

The mode of ligand binding was first demonstrated by Weis and co workers (1992) using rat MBP-A and N-glycan Man₆-GalNAc₂-Asn (oligosaccharide). A ternary complex is produced between the terminal mannose moiety of the oligosaccharide, the principal calcium ion (coordinated by residues Glu185, Asn187, Glu193, Asn205, and Asp206) and the protein. The ternary complex is stabilised by coordination and hydrogen bonds (Figure 1.15).

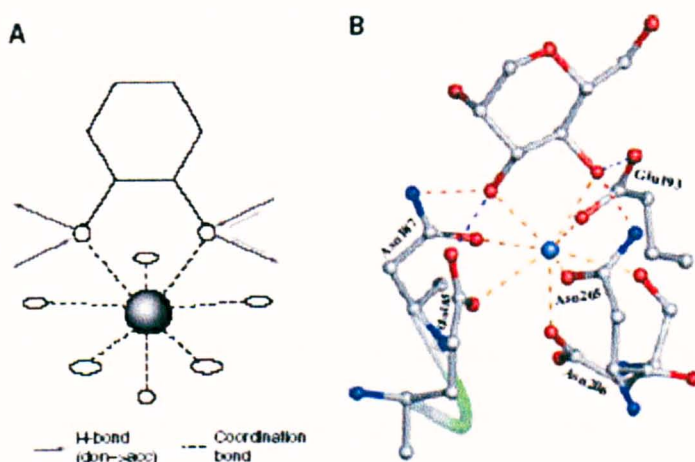


Figure 1.15 Ca²⁺ dependent monosaccharide binding by rat MBP-A. A) Schematic representation of the ternary complex with hydrogen and coordination bond stabilisation. B) MBP-A complex with mannose (pdb code 2msb). The blue sphere represents the principal calcium ion. Image taken from Zelensky and Gready (2005).

Using the simple carbohydrate ligands man α 1-2man, man α 1-4man, maltose, galactose and inositol monophosphate, Shrive and co-workers (2003; 2009) demonstrated structurally that SP-D binds to these ligands by recognising the pair of hydroxyl groups. In the case of mannan type ligands this utilises the C₃ and C₄ equatorial OH groups, for galactose binding the C₁ and C₂ OH groups are available (Figure 1.16). However in the event of galactose

being involved in glycosidic bond which uses the C₁ OH groups, then SP-D would not be able to recognise and bind to Gal type ligands since galactose does not provide any other vicinal equatorial OH groups. Inositol monophosphate is seen to bind via the equatorial C₄ and C₅ OH groups. The CRD also displays another binding mode influenced by residues Asp325 and Arg343 in the binding region in SP-D (variable equivalent in other collectins). The monosaccharide can be bound in either of two orientations that differ by 180° rotation without causing structural change to the binding site [Ng *et al.*, 1996; Shrive *et al.*, 2009]. These observations have been interpreted in terms of the two residues influencing interference with the α1-anomeric carbon directed towards His189 or Val194 in MBP-A and MBP-C respectively, or the freedom of rotation favouring the orientation with minimum steric hindrance [Ng *et al.*, 2002].

1.2.6 SP-D receptors

Most receptors for SP-D are believed to be located on the surface of macrophages, interaction causing agglutination and opsonisation. Glycoprotein with a molecular weight of 340 kDa (Gp340) obtained from lung lavage from a patient with alveolar proteinosis is reported to be a putative receptor of SP-D [Holmskov, 1999]. Gp340 is a member of the scavenger receptor cysteine-rich super family of proteins involved in innate immune defence and is associated with alveolar macrophages. It is highly glycosylated and allows binding of the SP-D in a calcium dependent manner involving protein-protein interactions [Holmskov, 1999]. SP-D also binds to alveolar macrophages in a calcium dependent manner involving the CRD; the binding affinity can be inhibited by saccharides [Crouch, 2001]. CD14 is another receptor recognised by SP-D. CD14 is located on the surface of macrophages and neutrophils, with SP-D binding via the CRD - sugar interaction, thereby altering the lipopolysaccharide signalling pathway [Crouch, 2001].

1.3 Lipopolysaccharide

Lipopolysaccharide (LPS) is an important constituent of the outer membrane of gram negative bacteria. The cell wall of gram negative bacteria is very thin consisting of a thin layer of peptide polysaccharide conjugate known as peptidoglycan that is situated between the cytoplasmic inner and outer-phospholipid bilayer membrane. In contrast, gram positive bacteria have lipoteichoic acid on the surface of a relatively thick cell wall peptidoglycan.

Lipopolysaccharides (and lipoteichoic acid) contain lipids (fatty acids) and carbohydrates, typically of the pyranose or furanose ring structure (Figure 1.16) glycosidically linked to form disaccharides and polysaccharides (Figure 1.17). The sugar rings are in the chair form as opposed to the boat conformation since the former is the most stable and preferred mode, with OH adopting axial or equatorial positions [Solomons, 1992]. The resulting saccharides exhibits a vast diversity in complexity as well as structural and functional properties, with varying linkages, either α - α' , α - β' , β - α' or β - β configurations at the C₁ anomeric carbon position.

As a result, LPS can form a diverse range of networks that may differ not only by the type and number of monosaccharides, but also in how the monosaccharides are joined together and the types of substitution, including amine groups at the non anomeric –OH group to form an amino sugar, for example D-glucosamine. D-glucosamine may also be acetylated forming N-acetyl-D-glucosamine (figure 1.16). The molecular mass of LPS can be in excess of 10kDa. Analogous to LPS, lipooligosaccharides (LOS) are much smaller, lacking the O-antigen part and tend to be more intricately branched compared to LPS [Masoud *et al.*, 1997].

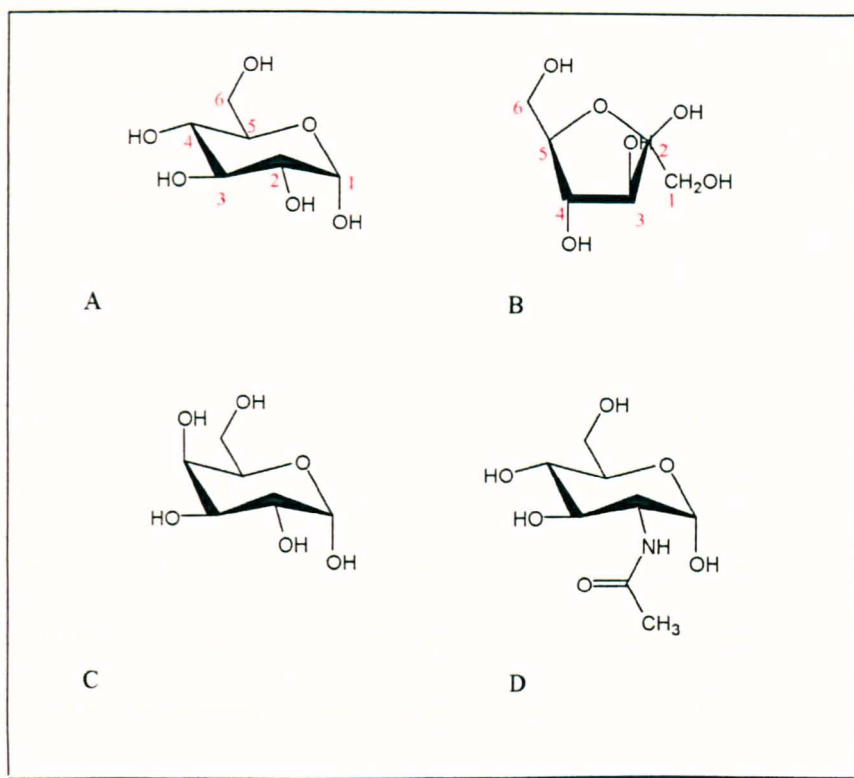


Figure 1.16 Structures of monosaccharides with their respective ring numbering (red) A) α -D-glucopyranose (pyranoside in the chair conformation), B) β -fructose (furanoside), C) α -D-galactopyranoside and D) the amine sugar α -N-Acetyl-D-glucosamine. The anomeric carbon involved in the glycosidic bond is at position 1 (red numbering) except in B which is at the C₂ anomeric position.

The carbohydrate structures of LPS can be grouped into two classifications termed aldoses or ketoses depending on the functionality of $-\text{CH}=\text{O}$ and $\text{C}=\text{O}$ respectively. In aqueous solution carbohydrates rarely exist in the chain form but rather they form cyclic structures of either six membered (pyranose) or five membered rings (furanose). This is achieved by intramolecular condensation between the C₅-OH group and the C=O group forming a cyclic hemiacetal group with a new asymmetric center at the C₁ position. This position is

known as the anomeric carbon position (C_1 anomeric position) because during hemiacetal formation the OH group at this position can adopt either one of the two spatial orientations with respect to the orientation of $-CH_2OH$ at the C-5 position.

If the OH group on the C_1 anomeric position is on the opposite side of the $-CH_2OH$ at C-5, then the anomeric isomer is said to be alpha (α), while if the OH group is on the same side as the $-CH_2OH$ at C-5 then the anomeric isomer is said to be the beta (β) isomer. The formation of glycoside occurs in the presence of acid. For example the formation of maltose, 4-O-(β -D-glucopyranosyl)- β -D-glycopyranose, which can be hydrolysed back to monosaccharide units in acidic solution (Figure 1.17). Although carbohydrates exist as the cyclic form in aqueous solution, the ring opens and recloses in mild acidic conditions, giving opposite anomeric configurations at equilibrium (Figure 1.18)

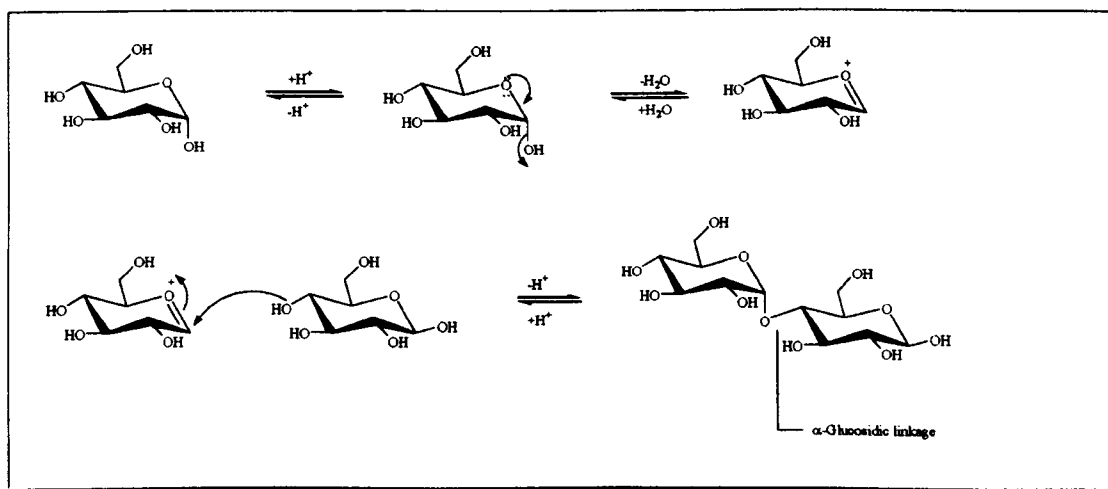


Figure 1.17 The formation of maltose, 4-O-(α -D-glucopyranosyl)- β -D-glycopyranose from the glycosidic bond of two glucose units.

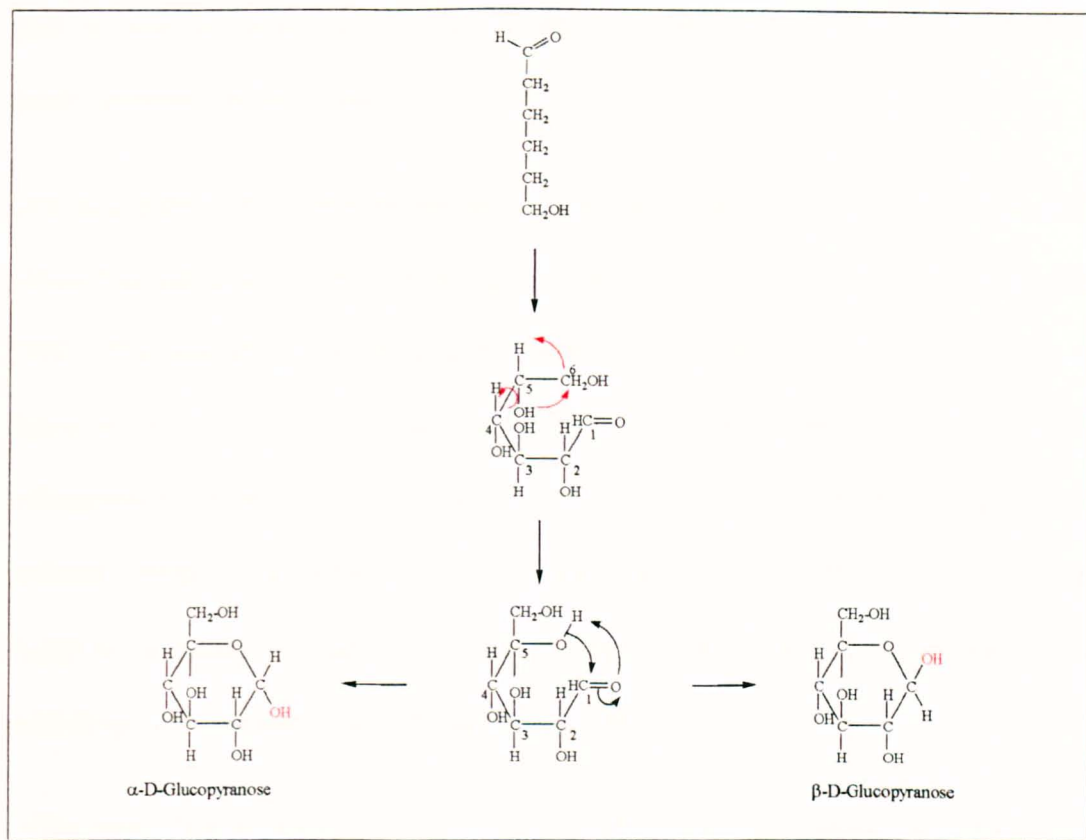


Figure 1.18 The formation of a cyclic carbohydrate ring structure from open chain forms [Solomons, 1992]. The two anomers at the C₁ anomeric carbon position are known as β when OH is on the same side of the ring as the CH₂-OH on C₅ and α where OH is on the opposite side of the ring. (-OH in red).

1.3.1 Immune response and receptors to LPS

The release of LPS (also referred to as endotoxin) from the surface of the bacterial cell, as a result of bacterial cell growth or the destruction of the bacterial cell by antimicrobial agent such as antibiotics or by host immune cells, elicits a wide variety of inflammatory responses even for extremely small amounts of endotoxin (ng to pg) [Raetz and Whitfield, 2002]. The main role of LPS on the bacterial outer membrane is to protect the organism from antibiotics and environmental stress, and to aid bacterial growth and colonisation [Raetz and Whitfield, 2002]. LPS is a pattern recognition target for innate immune proteins

such as lung surfactant proteins and C-reactive protein and is known to activate the host innate immune defence system.

LPS is a potent stimulator of the immune system, increased presence of the LPS in the serum leading to septic shock and organ failure [Bone, 1991; Parrilo, 1993; Martich *et al.*, 1993]. The biological toxicity causing septic shock is due to the lipid A region which is found to activate monocytes and macrophages to cause excessive production of the pro-inflammatory mediators TNF- α and IL-1. These factors cause a chain reaction mediating systemic toxicity [Calandra *et al.*, 1990; Klosterhalfen *et al.*, 1992]. Tetracyclins have been found to be good candidates for potential drugs aimed at preventing systemic toxicity by inhibiting cytokine secretion [Shapira *et al.*, 1996].

CD14 and TLR-4 have been identified as the prime receptors for LPS causing cell signalling once LPS is recognised [Wright *et al.*, 1989; Hailman *et al.*, 1994; Tobias *et al.*, 1995; Poltorak *et al.*, 1998]. CD14 is membrane bound on the surface of macrophages and B cells, anchored by a glycosylphosphoinositol tail. There are two forms of CD14, soluble sCD14 and mCD14. sCD14 occurs in plasma and helps LPS signalling in cells that lack membrane bound CD14 such as endothelial and epithelial cells, and mCD14 is found anchored to the membrane of myeloid cells via the glycosylphosphoinositol tail [Hazirot *et al.*, 1988].

Initial detection of LPS, as free “floating” molecules, fragments or still bound to the bacterial membrane is through complex formation by lipid binding protein (LBP), which transports and forms a ternary complex with the CD14 receptor on the surface of macrophages and B cells [Hailman *et al.*, 1994]. This complexation then induces the dimerisation of the second receptor TLR-4 which is constitutively associated with an

adaptor protein MD-2 [Palsson-Mcdermott and O'Neil, 2004]. This leads to the activation of multiple signalling molecules within the cellular matrix inducing the transcriptional activation of inflammatory mediators such as TNF- α and IL-1 [Palsson-Mcdermott and O'Neil, 2004].

1.3.2 Structure of LPS

The molecular architecture of LPS is common to all gram negative bacteria. It consists of the membrane anchoring lipid A moiety which is linked to the core region via a unique eight carbon sugar called Kdo (3-deoxy-D-*manno*-oct-2-ulosonic acid). The core region mostly contains non stoichiometric repeat units of L-glycero-D-manno-heptose (Hep), galactose (Gal) and glucose (Glc) and may also be attached to a repeat unit of polysaccharides making up the O-antigen moiety (O-polysaccharide) (Figure 1.19) [Nigam, 1975; Fridrich and Whitfield, 2005; Raetz and Whitfield, 2002].

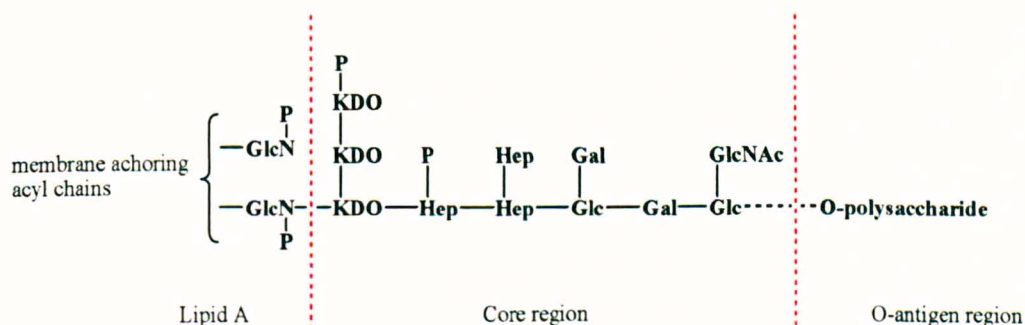


Figure 1.19 The three regions of LPS. The inner core region is linked to the lipid A via the ketosidic linkage. The molecule projects out from the cell membrane, anchored via the fatty acid residue. Drawn using CS Chem. Draw Pro. Version 4.5.

Mutant chemotypes can exist for the same species of bacterial strains (see below) which may lack the O-polysaccharide (termed rough chemotype) or may contain large numbers of repeat units making up the O-polysaccharide (termed smooth chemotype). These chemotypes are named according to the degree of polysaccharide chain length, with Re having the smallest polysaccharide chain length and Ra having the longest polysaccharide chain length (Figure 1.20) [Reys *et al.*, 2005; Nigam, 1975]

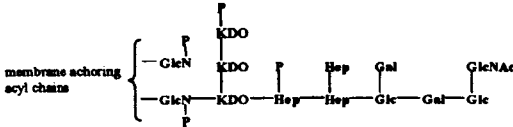
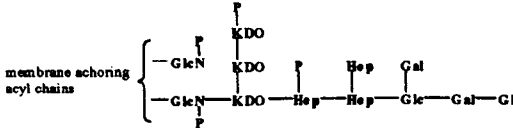
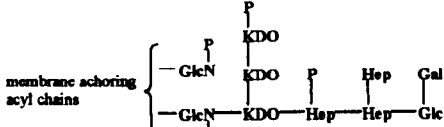
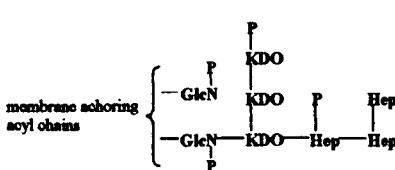
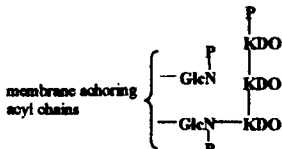
Chemotype	Probable structure
Ra	
Rb	
Rc	
Rd	
Re	

Figure 1.20 Chemotypes of LPS. Further variations in the Rd chemotype are possible arising from attachment of Hep and Phosphate in various forms [Nigam, 1975].

1.3.2.1 Lipid A region

The lipid A region is mostly found conserved throughout the spectrum of LPS and is responsible for most of the endotoxic activity. The conserved structural backbone is β -1'6-linked glucosamine disaccharide. The disaccharide contains α -glycosidic and non glycosidic phosphoryl groups at position 1 and 4' substituted with (R)-3-hydroxyacyl groups in amide and ester linkages (Figure 1.21).

The degree of variation of lipid A between species mostly arises in the chain length of hydroxyacyl groups C_{12} – C_{14} and the position of acyl groups [Odegaard *et al.*, 1997; Mikhail *et al.*, 2005]. For example in *Escherichia coli*, the ester and amide (R)-3-hydroxyacyl chain is further esterified with fatty acids. The degree of pathophysiological activity may be correlated to the level of acylation, attachment of phosphate groups, and in certain bacterial groups the presence of amino arabinose and ethanolamine groups that may also be substituted at the 1 position [Reatz and Whitfield, 2002].

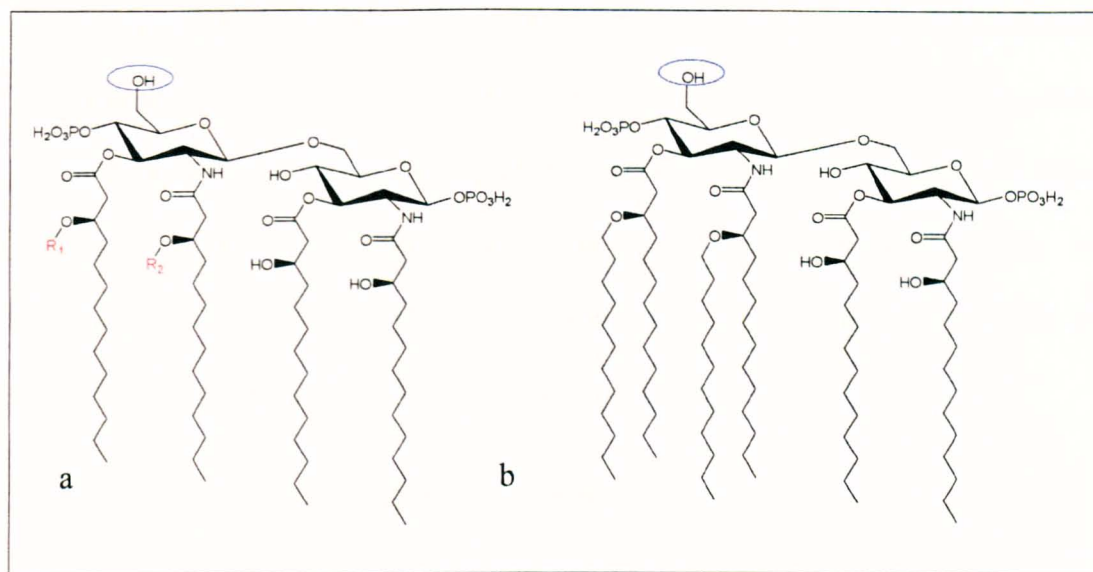


Figure 1.21 a) β -(1-6)-linked glucosamine disaccharide with (R)-3-hydroxyacyl group. Further ester linked fatty acid at R_1 and R_2 leads to variations. b) *Escherichia coli* lipid A. The hydroxyl group at C_6 (blue circle) indicates the position of glycosidic attachment of the remaining part of the LPS core. Structure modified from Odegaard *et al.*, 1997.

1.3.2.2 Core oligosaccharide region

This region contains a short series of sugars divided into two parts; the inner core proximal to the lipid A region is conserved in all gram negative bacteria and contains a common acidic sugar L-glycero-D-manno-heptose (Heptose) and 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo). The core is attached to the lipid A region at the reducing end via a ketosidic linkage providing stability in the outer membrane (Figure 1.22)

The Kdo and heptose can be further substituted by additional Kdo and heptose units providing various scaffolds of the inner core structures. The core may have other monosaccharides (discussed later for specific bacterial LPS) or charged species (non carbohydrate substituents) such as phosphates or ethanolamine attached [Knirel *et al.*, 2001; Cox *et al.*, 2002; Mansson *et al.*, 2003; Masoud *et al.*, 1997; Qureshi *et al.*, 1988;

Amor *et al.*, 2000; Schweda *et al.*, 2007]. The core provides LPS with the stability in the outer membrane [Heinrichs *et al.*, 1998]. The virulence of a strain is also dependent on the oligosaccharide chains and the degree of attachment of non carbohydrate substituents [Schweda *et al.*, 2007]. The oligosaccharides in the core also provide antigenic determinants to which specific immune responses are directed and also provide antigenic serotyping of bacterial strains within the same species.

1.3.2.3 O-polysaccharide region (O-antigen)

The O-antigen region consists of repeating units of sugars in the range of 15-25 repeating monomers of three to seven sugar subunits which may be identical or different [Al-Tahhan *et al.*, 2000; Stenutz *et al.*, 2006], each of which is interlinked by glycosidic linkages (Figure 1.22). The O-antigen region shows a large amount of variation in the structure of LPS between bacterial species or within the same strain. These variations may include sugar type, nature of linkage, and the degree of substitution such as acetyl, ketal and glycosyl residues. The variability in this region also provides the basis of serotyping bacterial species within the same species [Stenutz *et al.*, 2006].

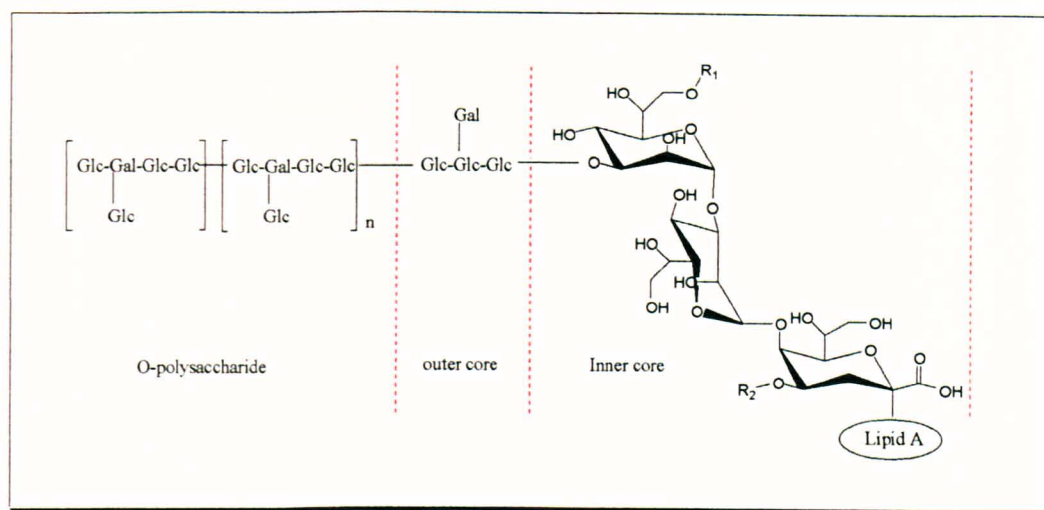


Figure 1.22 Schematic illustration of the O-polysaccharide repeat unit attached to the outer core. The inner core is composed of Kdo and heptose units. Drawn using CS Chem. Draw Pro. Version 4.5.

1.3.3 Biosynthesis: lipid A and core region

Biosynthesis takes place on the cytosolic side of the inner membrane, requiring sequential transfers of sugars and fatty acid by enzymatic pathways (Figure 1.23) [Price *et al.*, 1994; Jackman *et al.*, 2000]. The main precursor in the initial step of the synthesis requires UDP-GlcNAc [Bulawa *et al.*, 1984]. UDP-GlcNAc is initially acylated at the ester linkage by fatty acid (R-3-myristic acid, C₁₄-OH) from a carrier protein by LpxD enzyme. The second stage requires the deacylation at the amide position by LpxC deacetylase followed by amide acetylation by another fatty acid (R-3-hydroxymyristate, C₁₄-OH) catalysed by enzyme LpxD to produce precursor UDP-2,3-di-acylglucosamine. This is then hydrolysed and condenses with -2,3-di-acylglucosamine, catalysed by the LpxB enzyme, forming the intermediate lipid A.

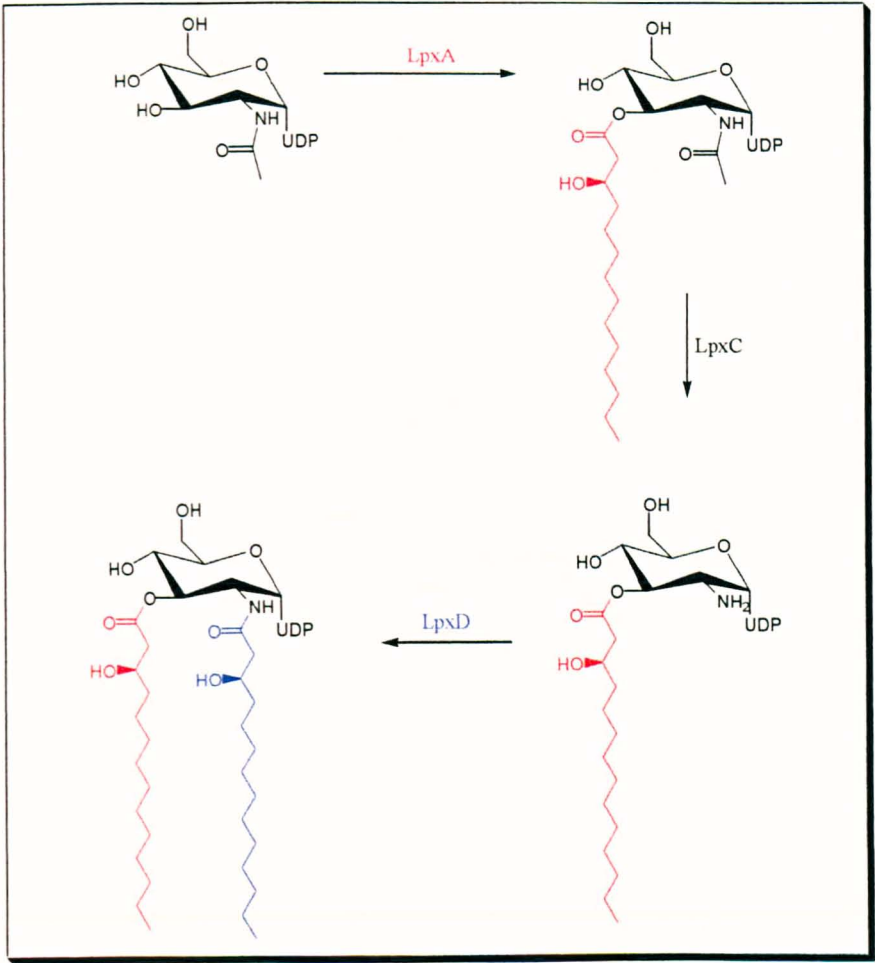


Figure 1.23a Biosynthesis of preformed lipid A from UDP-GlcNAc [Price *et al.*, 1994; Jackman *et al.*, 2000].

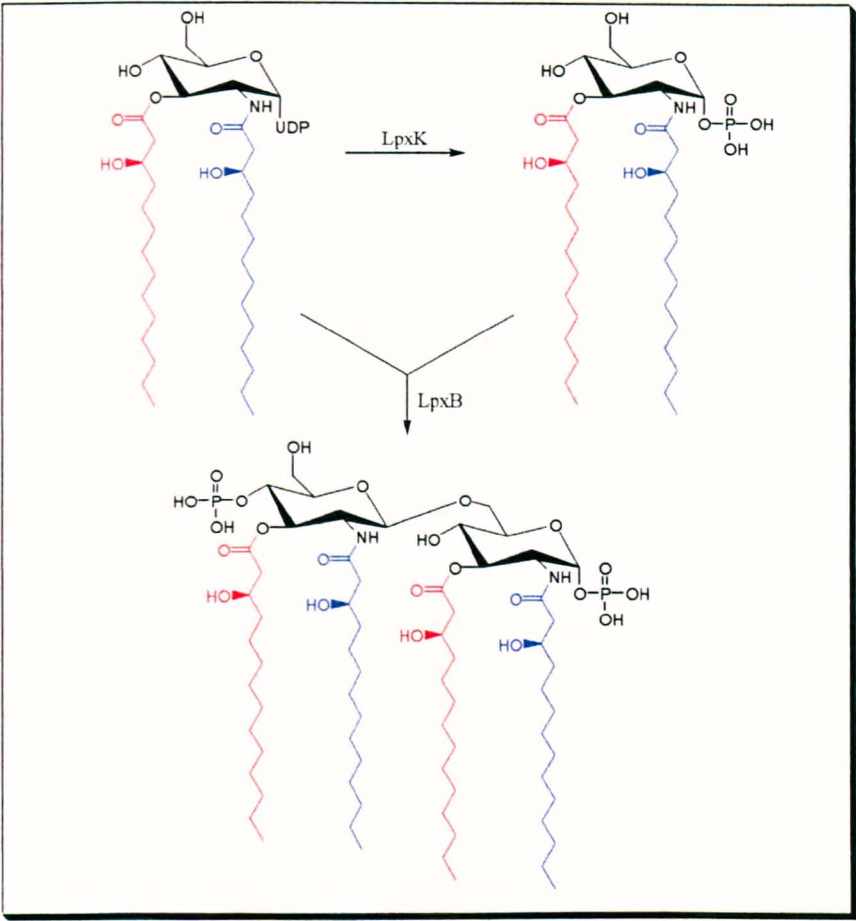


Figure 1.23b Fully assembled lipid A of *Escherichia coli* by condensation of UDP-2,3-diacylglucosamine with 2,3-diacylglucosamine [Price *et al.*, 1994; Jackman *et al.*, 2000].

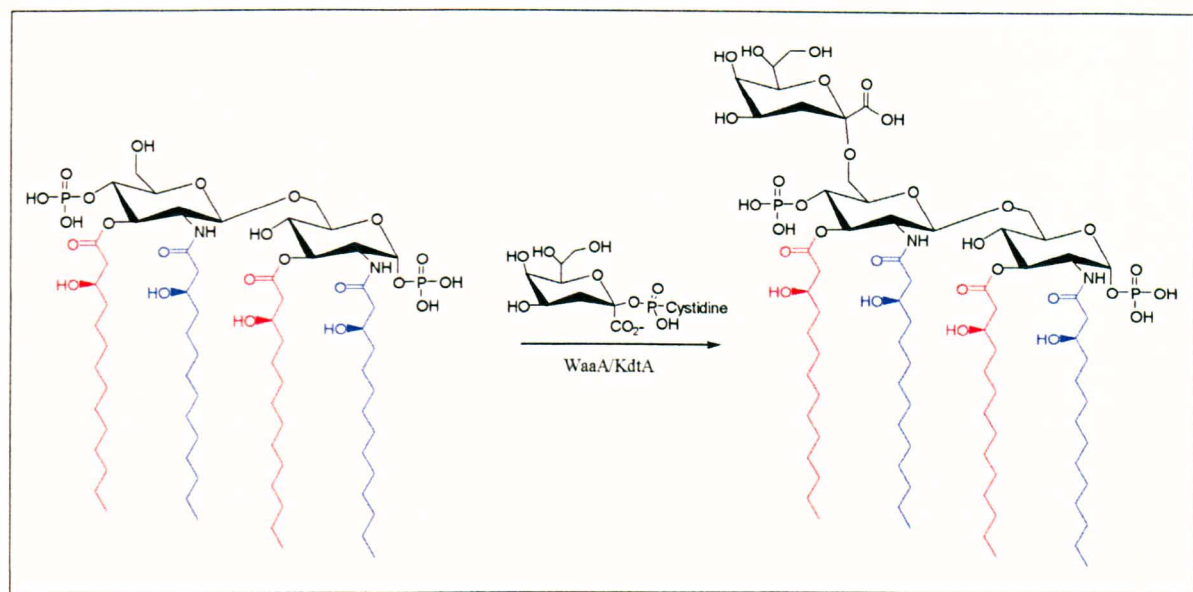


Figure 1.23c Attachment of Kdo unit to the lipid A region carried out by WaaA/KdtA [Price *et al.*, 1994, Jackman *et al.*, 2000].

1.3.4 Biosynthesis: Polysaccharide and O-polysaccharide

Depending on the bacterial species, synthesis of the oligo and polysaccharide is directed by various gene clusters. One such mechanism that operates in the biosynthesis utilises glycosyl transferase (GT) which catalyses glycosidic bond formation, using activated sugar donors containing nucleoside phosphate, to an “acceptor molecule”, usually a sugar (Figure 1.24). Two distinct GT structural folds have been identified termed the GT-A and GT-B folds [Lairson *et al.*, 2004].

The formation of a glycosidic linkage usually occurs by the S_N2 mechanism with either retention or inversion (α or β) of configuration at the anomeric carbon position. The synthesis occurs in the GT binding pocket containing a divalent metal ion, usually Mn^{2+} (in GT-A) and a carboxylic group acting as the nucleophile.

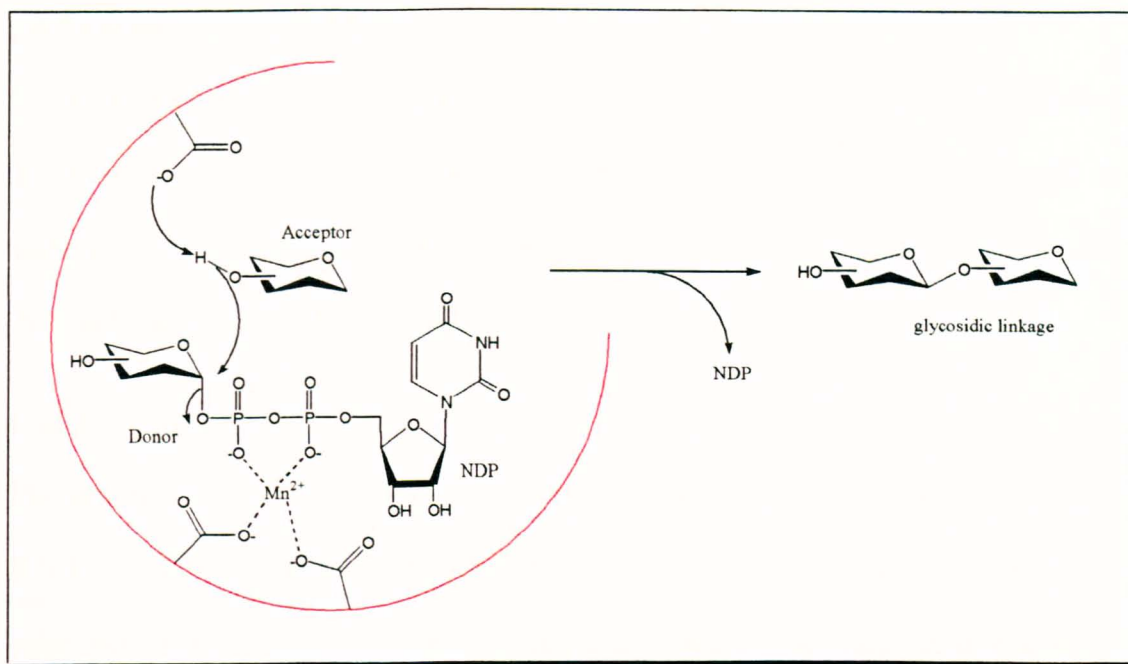


Figure 1.24 Glycosidic bond formation leading to oligo and polysaccharide. The red semi circle shows the binding pocket in the GT-A enzyme [Lairson *et al.*, 2004].

O-polysaccharide synthesis occurs on the cytosolic face of the cytoplasmic membrane and involves linkage of monosaccharides to undecaprenol phosphate (lipid carrier). Following the sequential addition of monosaccharides, the assembled oligosaccharides are then translocated by a flippase enzyme (*wzx*) to the periplasm face of the membrane and then polymerised by another enzyme (*wzy*) initiating from the reducing end. The fully assembled unit is then transported to the outer surface of the bacterial membrane. The linear or branched form of polysaccharide depends on the linkage α or β polymerisation [Yi *et al.*, 2006; Liu *et al.*, 1996; Burrows *et al.*, 1999; Marolda *et al.*, 2004].

1.3.5 *Escherichia coli* LPS

E. coli is an anaerobic bacteria that lives in the human colon and is non pathogenic unless the natural flora is imbalanced and they travel from the colon into the blood stream, leading to three major clinical syndromes (I) sepsis/meningitis, (II) urinary tract infection, (III) diarrhoea which is further classified in to five categories [Staff *et al.*, 1999].

1.3.5.1 *E. coli* LPS structure

The structure of *E. coli* LPS is only conserved in the inner core, composed of heptose and di-Kdo units, with variations arising in the outer core and the O-antigen repeat unit. The outer core variations arise from the linkages of oligosaccharide, the types of hexose and the side chains attached. Based on these variations, five outer core regions have been identified designated K-12 and R1 to R4 [Amor *et al.*, 2000]. Although the common structure is composed of carbohydrate backbone and two side chain residues, the variations arise from the order of the hexoses in the backbone, the linkages of the side chains and their nature and position (Figure 1.25 and 1.26).

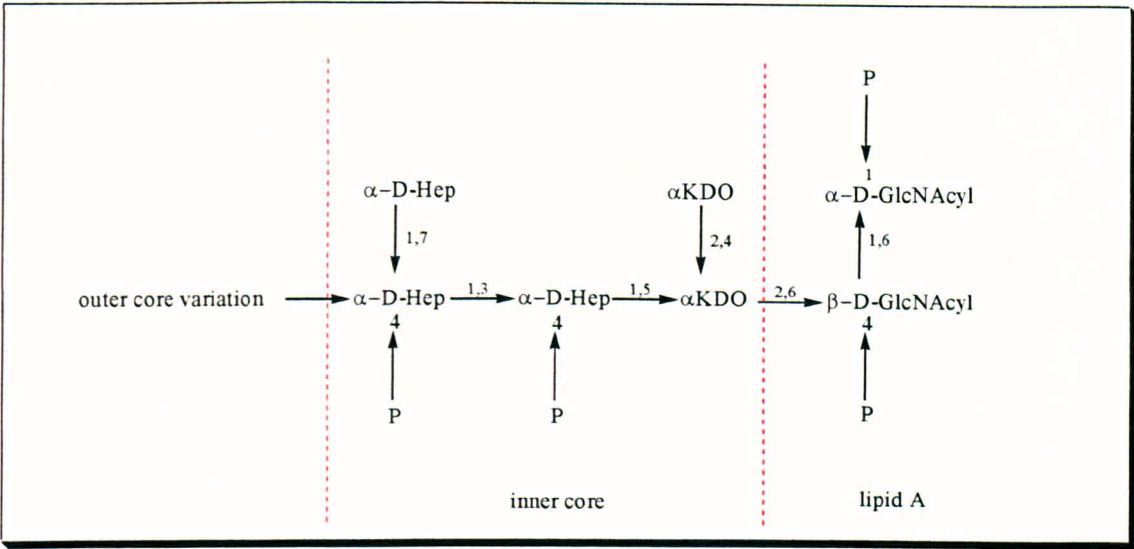


Figure 1.25 *E. coli* LPS showing the lipid A, inner core and the O-antigen polysaccharide. The inner core contains two Kdo units and three heptose unit. Modified from Amor *et al.*, 2000.

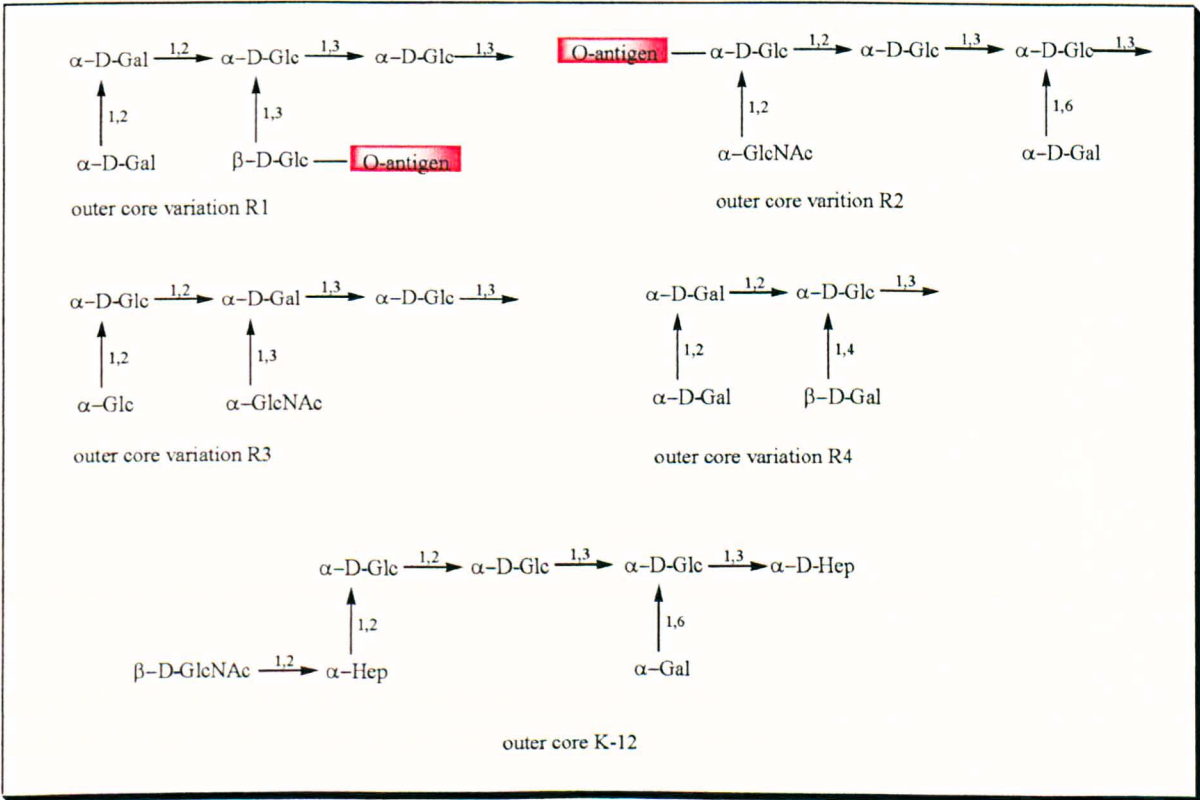


Figure 1.26 Structure of the five outer core variations of *Escherichia coli* LPS. Attachment of O-antigen (red) has only been determined for R1 and R2 core types. Structure adapted from Amor *et al.*, 2000.

1.3.5.2 *E. coli* O-polysaccharide

E. coli LPS exists in the smooth form with a variable repeat unit of O-polysaccharide and the rough form lacking the O-polysaccharide region (rough LPS). Serotyping *E. coli* is based on the O-polysaccharide and over 120 serotype groups have been identified [Amor *et al.*, 2000]. The repeat units are made up of five to six monosaccharide units and vary between strains and virulence (Figure 1.27).

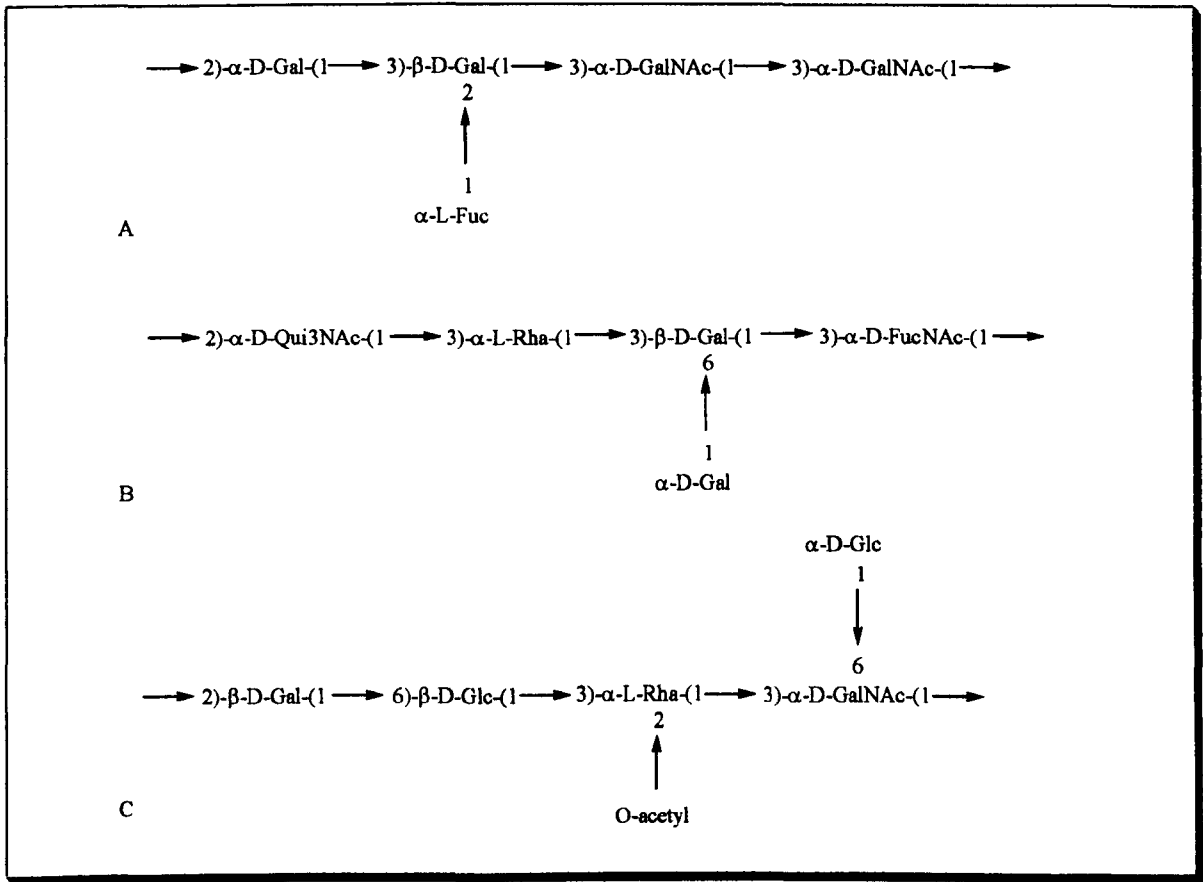


Figure 1.27 Structures of O-polysaccharide from *Escherichia coli* a) strain 086:K62:B7 b) strain 62D₁ and c) K-12. [Yi *et al.*, 2006; Staff *et al.*, 1999; Stevenson *et al.*, 1994].

1.3.6 *Pseudomonas aeruginosa* LPS

P. aeruginosa is an opportunistic gram negative bacterial pathogen that infects hospitalised patients including those with severe burn wounds, cystic fibrosis and cancer patients with a compromised immune system. It is generally found in soil and on the surface of plants and only infects humans under compromised tissue conditions (i.e. damage to the skin or mucosa membrane) when it can cause a range of clinical symptoms such as urinary tract infection, respiratory system infections and dermatitis.

Twenty major serotypes have been identified based on their O-antigens (Figure 1.28 and 1.29) [Liu *et al.*, 1983; 1990]. In 1994 de Kievit [De Kievit *et al.*, 1994] used monoclonal antibodies to demonstrate that the inner core was more conserved than the outer core. The LPS can be found as smooth, semi rough and rough depending on the degree of O-antigen polysaccharide [Bystrova *et al.*, 2004; Knirel *et al.*, 2001; Sadovskaya *et al.*, 2000]. The smooth strain that possesses the O-antigen polysaccharide has enhanced virulence and the ability to survive in the host compared to strains that lack O-polysaccharide, which are more susceptible to the bactericidal effects of serum complement [Hancock *et al.*, 1983; Schiller, 1988; Cryz *et al.*, 1984; Tang *et al.*, 1996].

1.3.6.1 *P. aeruginosa* LPS core and O-polysaccharide

The inner core region contains two Kdo units and two heptose units (Figure 1.28 and 1.29). The outer core is composed of D-Glc, D-Gal, L-Rham and L-Ala. The O-polysaccharide antigen shows a broad spectrum of monosaccharide or substituted monosaccharide units [Bystrova *et al.*, 2004; Sadovskaya *et al.*, 2000].

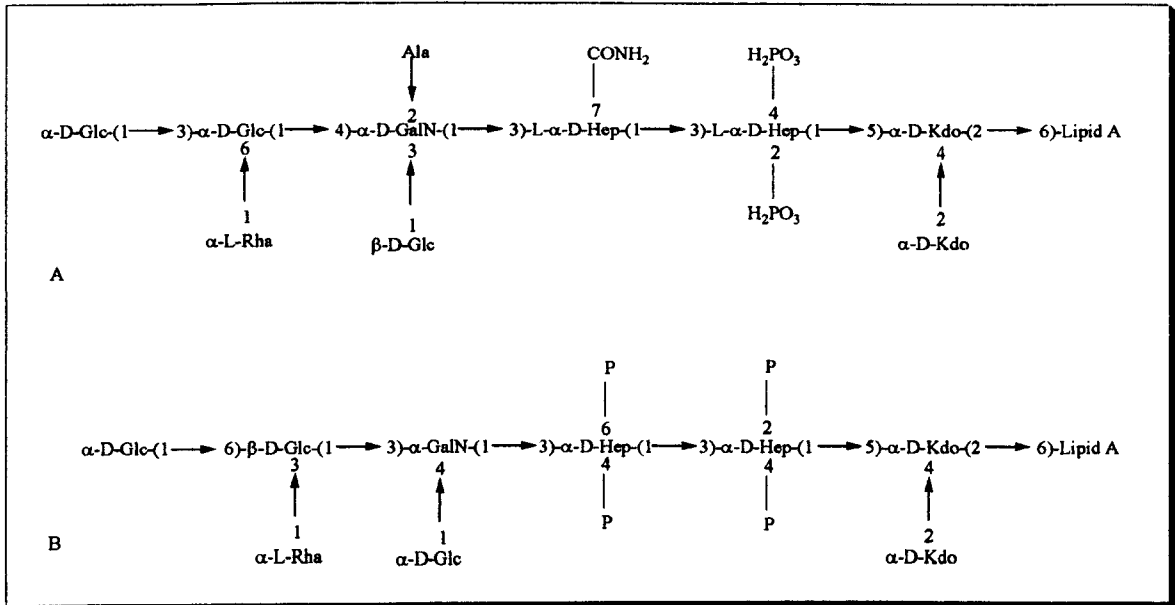


Figure 1.28 Structures of the *Pseudomonas aeruginosa* inner core and outer core but lacking the O-polysaccharide A) serotype 06 mutant strain A28 and B) LPS of cystic fibrosis isolate [Masoud *et al.*, 1995; Knirel *et al.*, 2001].

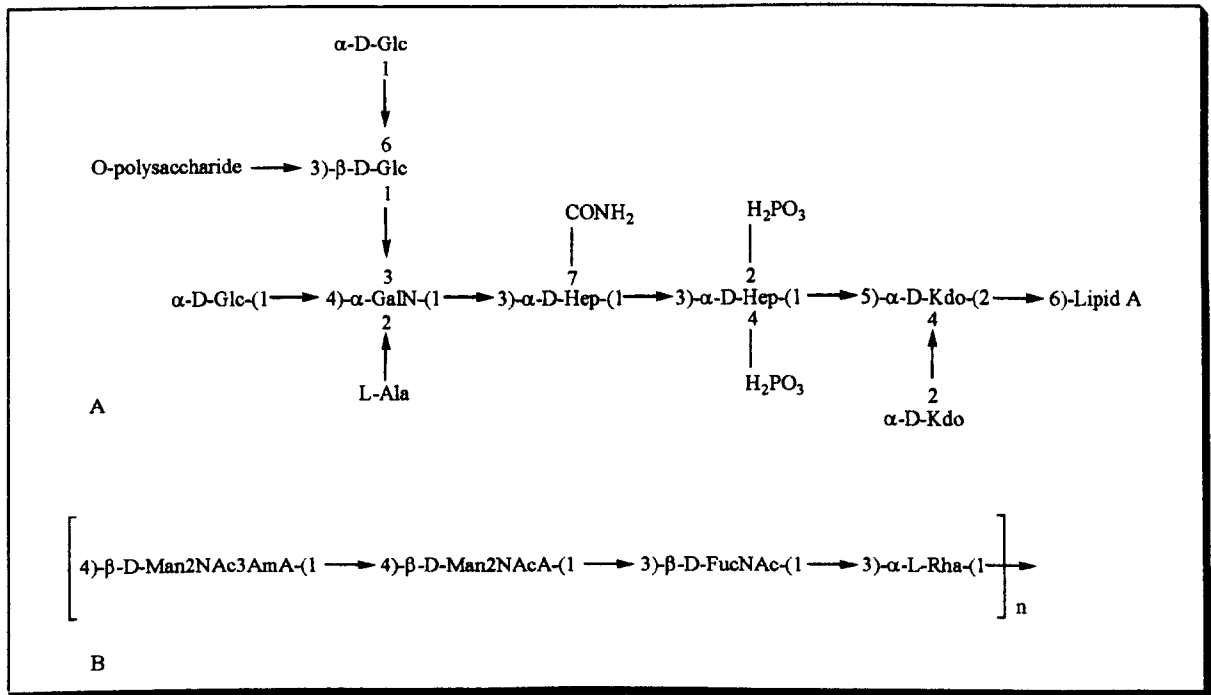


Figure 1.29 Structure of smooth LPS of *Pseudomonas aeruginosa* showing repeat unit of O-polysaccharide attached to the outer core [Sadovskya *et al.*, 2000]. A) serotype 06 mutant strain A28 and B) LPS of cystic fibrosis isolate [Masoud *et al.*, 1995; Knirel *et al.*, 2001].

1.3.7 *Haemophilus influenzae* LPS

H. influenzae is a small rod shaped gram negative bacteria that is not found in the environment and does not colonise, or infect, animal species except for humans. It is therefore exclusively a human pathogen [Gilsdorf, 1998]. It begins colonising from birth onwards and may persist for long periods of time [Faden *et al.*, 1995; Groeneveld *et al.*, 1990; Spinola *et al.*, 1986] and is responsible for the majority of childhood respiratory disease on a global scale. It mostly resides in the nasopharynx and translocates to the middle ear, and has been known to cause symptoms such as mild sore throat and middle ear infection (otitis media) particularly in young children. In 1892 *H. influenzae* was considered to be responsible for the epidemic influenza outbreak, a belief that was later proved to be incorrect. As the name suggests, it requires the haemin that is present in the blood.

There are two classes of *H. influenzae*, non capsulated (nontypeable - NTHi) and capsulated (typeable). The capsulated (typeable) form is further subdivided into six serotypes (a through to f) depending on the makeup of its capsular polysaccharide (an envelope that surrounds the bacterial cell). Capsule expression, such as that of serotype b composed of polyribose-ribitol phosphate (PRP), is relatively unstable and loss of expression occurs due to loss of genetic material [Kroll, 1992] mutating the capsule form into a non capsulated form. Bacterial classification, particularly of the capsulated forms, is determined by using the slide agglutination test [Falla *et al.*, 1994; LaClair *et al.*, 2003; Shively *et al.*, 1981]. Recent developments include a molecular typing system such as PCR [Billal *et al.*, 2007], which can accurately define bacterial strains compared to the conventional slide agglutination test. PCR methods identify and amplify the *bexA* gene,

which is necessary for polysaccharide export and expression. This gene is located on the cap locus, a 17kb DNA fragment which is common to all six serotypes [Luong *et al.*, 2004]. The strains that cannot be identified as one of the six serotypes are hence categorised as nontypeable strains (NTHi).

1.3.7.1 Pathogen and vaccine

H. influenzae show a broad spectrum of pathogenicity depending on its surface antigens and variants [Foxwell *et al.*, 1998]. Capsular serotypes, of which serotype b is the most virulent form on a global scale, is responsible for life-threatening, sometimes fatal infections such as meningitis and bacteraemia especially in young children and infants [Moxon 1985; Moxon *et al.*, 1974]. Non capsular (NTHi) strains rarely cause such life-threatening disease but exist as a commensal pathogen in the human host [Foxwell *et al.*, 1998], with the onset of otitis media (middle ear infection) and mild respiratory tract infection in patients with chronic bronchitis and cystic fibrosis [Foxwell *et al.*, 1998; Moxon *et al.*, 1990; Buttery and Moon, 2002].

In the present era, with modern advanced vaccines, the virulent strain b responsible for such life-threatening disease is completely eradicated in western countries [Muhlemann *et al.*, 1996] but there has been evidence of newly emerging strains with potentially lethal virulence [Adderson *et al.*, 2001]. Vaccination against the widespread serotype b strain is achieved by using glycoconjugate vaccines (conjugated to a carrier protein) based on the capsular polysaccharide of b (Figure 1.30) [Verez-Bencomo *et al.*, 2004; Kandil *et al.*, 1997], however this vaccine is only effective against serotype b and provides no protection against other serotypes, even though there have been no reports of invasive disease caused by the other serotypes (a, c through to f).

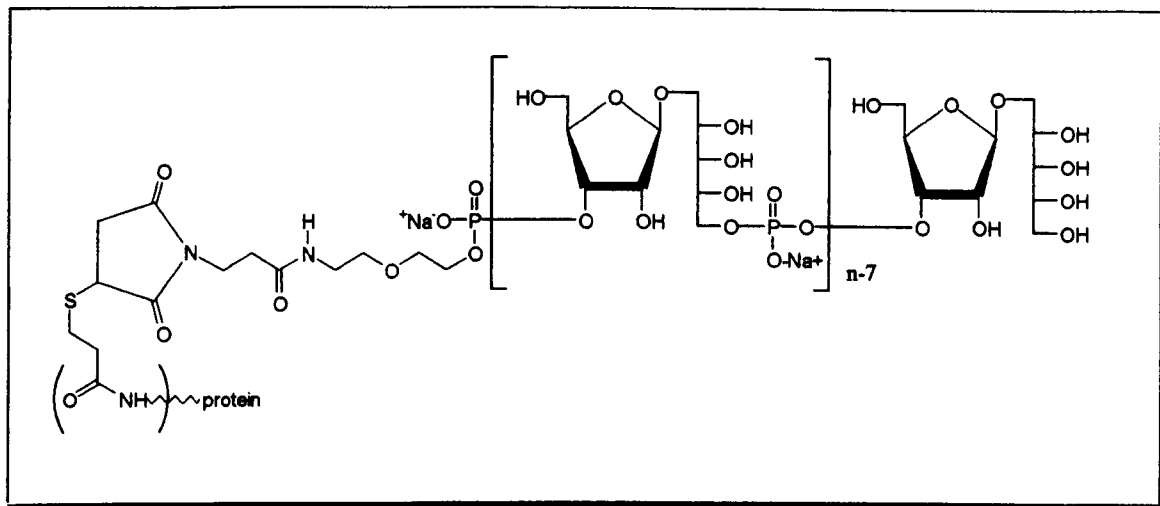


Figure 1.30 *Haemophilus influenzae* serotype b glycoconjugate protein vaccine contains 7 repeat units of the capsular polysaccharide unit. The capsular polysaccharide b repeat unit of polyribitol-phosphate (PRP) is shown in brackets [Verez-Bencomo *et al.*, 2004].

1.3.7.2 Structural diversity

While the virulence of *H. influenzae* is caused by the capsular polysaccharide, the surface of the outer membrane of these bacteria is also covered with LPS. This antigen, which also has virulence potential, is composed of much shorter oligosaccharides and is termed lipooligosaccharide, which also shows strain to strain heterogeneity [Inzana, 1983].

Over the last few decades, research has been carried out to determine the structure of LPS using various techniques, such as mass spectroscopy (MS) and high field nuclear magnetic resonance spectroscopy (NMR). These have demonstrated antigenic heterogeneity in the lipid A moiety as well as the core polysaccharide or short carbohydrate side chains [Apicella *et al.*, 1985; Masoud *et al.*, 1997; Mansson *et al.*, 2001; Mansson *et al.*, 2002;

Risberg *et al.*, 1997; Yildirim *et al.*, 2003; Landerholm *et al.*, 2004; Campagnari *et al.*, 1987; van Alpen *et al.*, 1990].

An extensive structural study from NMR and MS indicates that LPS contains a tri heptose inner core attached to a single Kdo unit (Figure 1.31). This inner core provides a framework for further attachment of oligosaccharide extension [Schweda *et al.*, 2007] and non carbohydrate substituents, giving an enhanced ability to colonise and to evade detection by the host immune system [Schweda *et al.*, 2007; Tinnert *et al.*, 2005; Schweda *et al.*, 2000; Lysenco *et al.*, 2000; Yildirim *et al.*, 2005].

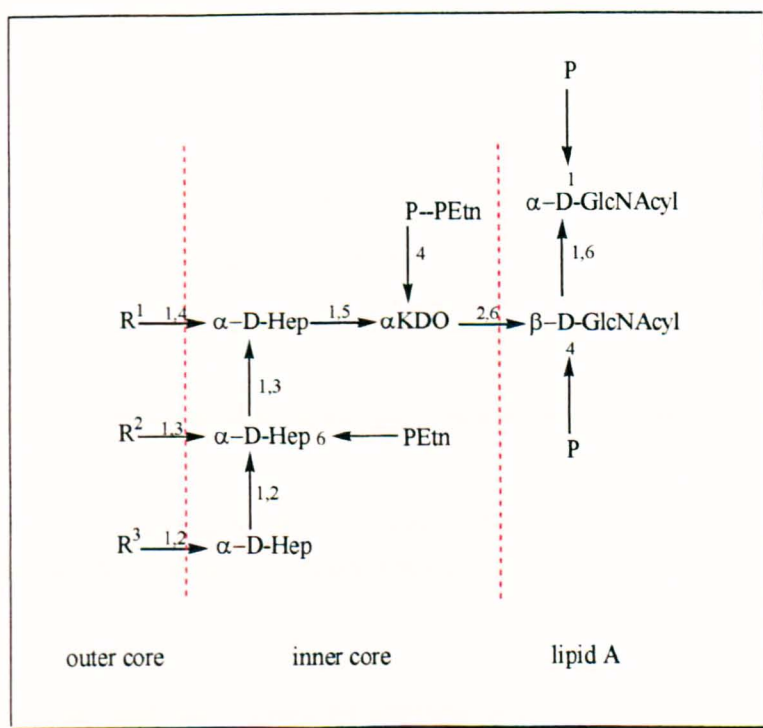


Figure 1.31 The common structural architecture of *H. influenzae* Eagan LPS, showing the three regions, lipid A, inner core, and the outer core comprised of various oligosaccharide extension (R^1 to R^3). The numbers indicate the glycosidic bond on the pyranose ring.

1.3.7.3 Inner and outer core structure

H. influenzae is classified as the rough form of LPS because it lacks the highly variable O-side-chain repeating unit, but has an extensive repertoire of non repeating oligosaccharide, non saccharide, and organic molecules attached to the inner core (Fig. 1.32) such as Glc, Gal, GlcNAc, phosphate, PPEtn, and N-acetyl-neuraminic acid, within a single strain.

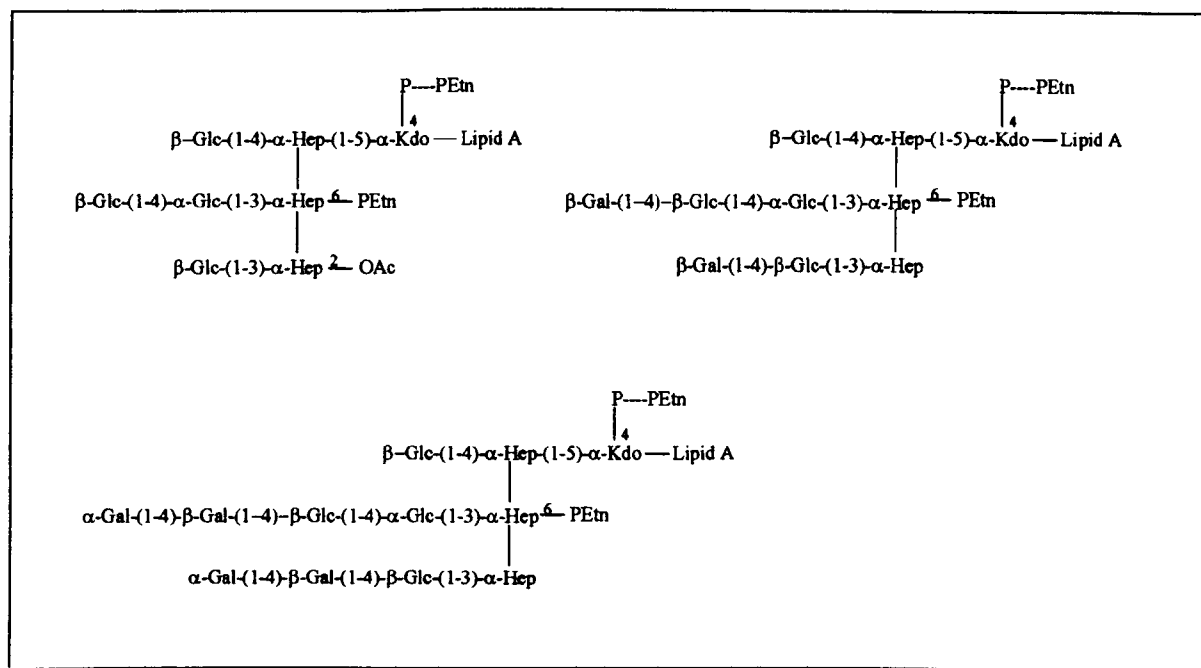


Figure 1.32 Structural heterogeneity of a single strain of NTHi serotype f as a result of various oligosaccharide chain attachments from the inner core [Yildirim *et al.*, 2003].

In comparison to the virulence caused by capsular polysaccharide, namely the serotype b, the virulence contribution from LPS depends on the type of oligosaccharide and non saccharide attachments. For example, presence or absence of sialic acid on the LPS modifies the antigenic specificity, and variants possessing phosphocholine residues alter the immunological reactivity therefore making the organism more susceptible to the

bacterial activity of the human serum component C-reactive protein (Figure 1.33) [Lysenco *et al.*, 2000].

The ability to invade and colonise the host within the epithelial linings is achieved by the structural motif digalactoside α -Gal-(1-4)- β -Gal, which mimics the host glycolipid, enabling the organism to camouflage and evade detection. The polysaccharide is also found substituted in certain strains by *O*-acetylation, again enabling it to evade killing by antibodies and human serum [Lysenco *et al.*, 2000; Weiser *et al.*, 1998; Swords *et al.*, 2000; Schweda *et al.*, 2007].

Phosphocholine is commonly found in LPS at various locations, and certain strains have been found to carry more than one PCho at different locations of the oligosaccharide, further increasing the virulence potential of the particular strain [Lysenco *et al.*, 2000; Schweda *et al.*, 2000]. In all strains investigated the Kdo unit is substituted at the 4 position by phosphate or pyrophosphate, providing further sub-populations. The exact physiological significance of this substitution is still unclear. The second heptose is also found substituted by phosphoethanolamine and again, the function is still not determined.

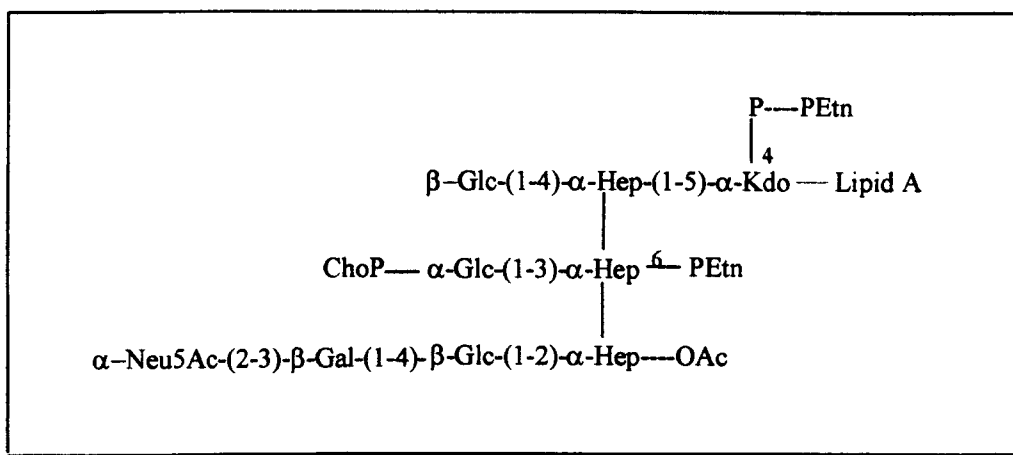


Figure 1.33 Structure of the NTHi strain 486 showing the attachment of sialic acid to the digalactoside extension from the terminal heptose unit, along with substitution of PCho and PEtn on the inner heptose unit [Mansson *et al.*, 2001].

1.3.7.4 Biosynthesis of the Inner core

The genetic mechanism that determines the biosynthesis of the triheptosyl inner core has been determined from a number of studies using *H. influenzae* strain Rd and the homologous strain RM 118 (Rd⁻), and also using type b strains Eagan (RM 113) and RM 7004 [Fleischmann *et al.*, 1995; Hood *et al.*, 1996a; 1996b].

The inner core triheptosyl framework is directed by Opsx, which adds the first Heptose (HepI) to the Kdo unit, rfaF adds the second Heptose (HepII) to HepI and orfH adds the third Heptose substituting from HepII. The genes responsible for directing Glc, which is commonly found in all *Haemophilus influenzae* LPS, substituted from each Hep (HepI-HepIII), are IgtF, lic2C and IpsA respectively.

1.3.7.5 Biosynthesis of the outer core

The variations in the outer core of a single strain are a result of phase variable expression of certain genes (Figure 1.34). Phase variation is a molecular switch process related to the expression of genes in chromosomal loci *lic1*, *lic2*, *lic3*, *igtC*, *lex2*, and *Ipt6* (Weiser *et al.*, 1989; 1990; Hood *et al.*, 2001; Griffin *et al.*, 2003; High *et al.*, 1993]. Phase variation occurs at a high frequency resulting in reversible loss or gain of sugars and non sugar units. This mechanism is thought to provide an adaptive mechanism that is advantageous for the survival of the bacterium and leads to virulence behaviour [Schweda *et al.*, 2007].

For example the disaccharide α Gal- β Gal which is a phase variable epitope of *lex2* substituted from Glc on HepI is found on the glycosphingolipids of some epithelial cells, and so allows molecular mimicry of surface structures of the host, allowing it to avoid detection by the host immune system [Weiser *et al.*, 1990].

- *Lic1*-this codes for the addition of PCho to the LPS at various locations between species.
- *Lic2* and *igtC*- involved in biosynthesis of the galabiose component of the globoside oligosaccharide
- *Lic3*- the locus *lic3A* and *lic3B* encodes an α -2,3-sialyltransferase, which is responsible for addition of sialic acid (N-acetylneuraminic acid or Neu5Ac) to a terminal lactose on the HepIII.
- *Lex2*- this is shown to encode a β 1(1-4)-glucosyltransferase that adds the second β -glucose to the first β -glucose from the first HepI.
- *Ipt6*- this codes for the addition of PEtn to HepII

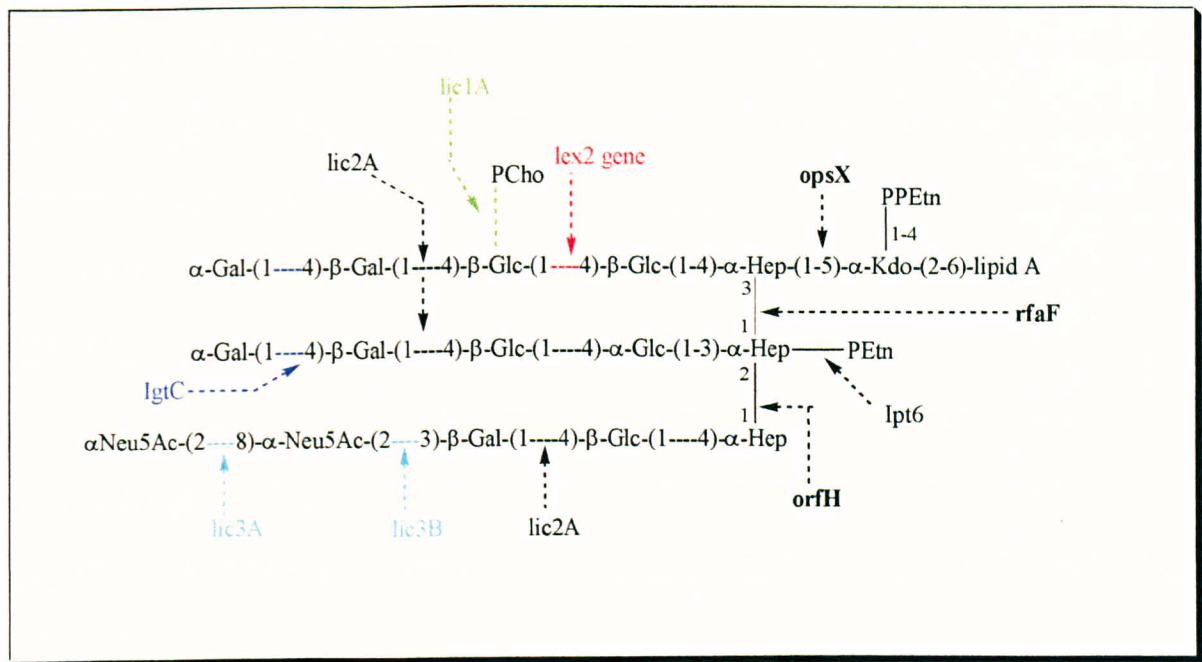


Figure 1.34 Schematic illustration of the oligosaccharide and non carbohydrate units coded by phase variable genes for *H. influenzae* Eagan LPS. Genes involved in the biosynthetic pathway of the inner core heptose are shown in bold.

1.4 Aims and objectives of the research project

The work described in this thesis is a continuation of the work carried out by Professor Greenhough and Dr. Shrive on structural studies using X-ray crystallography of the innate immune proteins SP-D and CRP. Reported structures of ligand-bound complexes using SP-D and CRP are restricted to simple but representative ligands (Crouch *et al.*, 2009; 2006; 2007; Wang *et al.*, 2008; Shrive *et al.*, 2003; 2009; Thompson *et al.*, 1999) but no study has been conducted using a much larger complex ligand such as lipopolysaccharide (LPS) with these proteins. The overall aim is to provide insight in to recognition and binding of LPS by SP-D and CRP through three-dimensional structural studies of protein-LPS complexes.

The barrier that has been hampering crystallisation studies using LPS is the difficulty in getting the LPS molecule to dissolve in crystallisation conditions, this is due to the hydrophobic lipid A region. Using LPS from *E. coli* strain 0111.B4, *Pseudomonas aeruginosa* and *H. influenzae* Eagan and the Eagan 4A LPS. one approach is to sever the lipid A-Kdo glycosidic bond under hydrolysis using 2% acetic acid to obtain the soluble polysaccharide region and discard the lipid A molecule. The soluble polysaccharide may be characterised using electrospray mass spectroscopy and matrix assisted laser desorption ionisation mass spectroscopy prior to crystallisation.

1.4.1 Structural studies using *Limulus* CRP with *E. coli*, *P. aeruginosa* and *H. influenzae* LPS and polysaccharide ligand.

Crystallisation of a single *Limulus* CRP obtained from the plasma of the Atlantic horse shoe crab (*Limulus Polyphemus*) has proved to be a formidable challenge in terms of isolation and crystallisation due to the multiple isoform of the CRP in the plasma (Nguyen

et al., 1986b; Bailey, PhD thesis 2008, Keele University). Numerous crystallisation conditions have been explored but none have provided suitable crystals for X-ray diffraction and as such no crystal structure has been reported (Bailey, PhD thesis 2008, Keele University). It is noted that since there is more than one form of the CRP present then each form could be specifically tuned for specific ligand binding properties [Nguyen *et al.*, 1986b; Iwaka *et al.*, 1999; Ng *et al.*, 2004]. Similar to human CRP, *Limulus* CRP is characterised by its high Ca-dependent binding affinity for phosphocholine (PC) and phosphoethanolamine (PE) [Schwalbe *et al.*, 1992]. Using the assumption of selective binding and the presence of PC and PE ligand present on LPS, using these LPS may lead to the possibility of “selectively” isolating pure single CRP complexed with LPS [Iwaka *et al.*, 1999; Ng *et al.*, 2004] which may provide insight into the structural recognition of PC and PE present on the outer or inner structures of the LPS.

In this work the CRP isolated by PC affinity chromatography, but still containing a heterogeneous mixture of CRP (Bailey, PhD thesis 2008, Keele University), is to be co-crystallised with the various LPS. No optimised crystal conditions have previously been determined so various conditions using different buffers will be investigated to seek suitable crystals for data collection using X-rays generated at synchrotron radiation sources (Daresbury SRS and the Diamond Light Source).

1.4.2 Structural studies of rfhSP-D with *H. influenzae* wild type LPS and mutant 4A.

Structural studies of the interaction between hSP-D and simple ligands have utilised a recombinant homotrimeric fragment (rfhSP-D) comprised of the neck region plus three CRDs, which has been shown to be therapeutically active in murine models of pulmonary hypersensitivity and infection induced by an opportunistic fungus, *Aspergillus fumigatus* [Madan *et al.*, 2001; 2010]. The crystal structure of rfhSP-D obtained at high resolution using synchrotron radiation [Shrive *et al.*, 2003; 2009] shows monosaccharide recognition achieved via the terminal stereochemically equivalent hydroxyl pair. The well characterised *H. influenzae* strain RM153 (Eagan) LPS has a rough phenotype, lacking the O—specific side chain, and the rfaF mutant Eagan 4A which contains a single heptose (HepI) provides an ideal opportunity to establish the structural basis of recognition by SP-D. Studies on *E. coli*, *Salmonella minnesota* and *Bordetella* species suggest that the lung collectins bind to the LPS core and strongly implicate heptose and glucose residues in this binding [Kuan *et al.*, 1992; Schaeffer *et al.*, 2004] Using the optimised conditions established by Shrive and co workers, hydrolysed *H. influenzae* Eagan LPS and the mutant 4A strain will be soaked into the rfhSP-D crystals for data collection at synchrotron radiation sources.

Chapter 2

Experimental techniques

2.1 Principals of the techniques used

This section briefly describes the fundamentals behind the methods employed in the experimental section and the protocols used in obtaining the data. The key areas that the experimental section is based on are:

- PC affinity chromatography
- Hydrolysis of the LPS and product isolation
- Structural analysis of LPS using electrospray ionisation mass spectroscopy (ESI-MS) and matrix assisted laser desorption ionisation mass spectroscopy (MALDI-MS)
- Protein crystallography and X-ray diffraction

2.1.1 PC affinity chromatography

Affinity chromatography was first developed in 1968 by Cuatrecasas and co workers [Cuatrecasas *et al.*, 1968; Cuatrecasas, 1970; Cuatrecasas and Anfinsen, 1971]. This technique is widely used for the purification and isolation of many biologically active proteins and relies on the reversible and specific interactions of the protein with the ligand, for example, interactions include the binding of an enzyme with an inhibitor, of antibody with antigen and of lectin with polysaccharide.

The protein to be purified is passed through a column that is packed with insoluble cross-linked polymer or gel (matrix) that has a specific competitive ligand covalently attached. Proteins that do not have the specific binding affinity for the ligand attached to the matrix will pass unretarded through the column, whereas proteins that have the specific binding affinity towards the ligand that is covalently attached to the matrices will be retained in the column (Figure 2.2). The bound protein is then eluted from the ligand by reverse inhibition

by changing the parameters such as the pH, salt concentration or by addition of a competitive ligand which has a much higher binding affinity than the matrix ligand towards the retarded protein.

The matrix used in affinity chromatography are agarose (derived from agar) or Sepharose (modified form of agar) that are made up of linearly linked sugar molecules (Figure 2.1), these matrices are widely used because the hydroxyl groups on the sugar residues can be easily altered to covalently attach the ligand [Aplin and Hall, 1980]. The matrix forms a loose porous lattice, permitting uniform and unimpaired flow of the molecules, without having any interaction with the proteins.

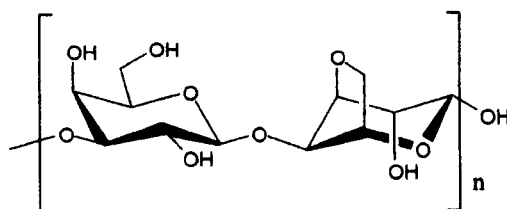


Figure 2.1 Structure of agarose beads, a polysaccharide consisting of 1,3-linked-D-galactopyranose and 1,4-linked 3,6-anhydro-α-L-galactopyranose.

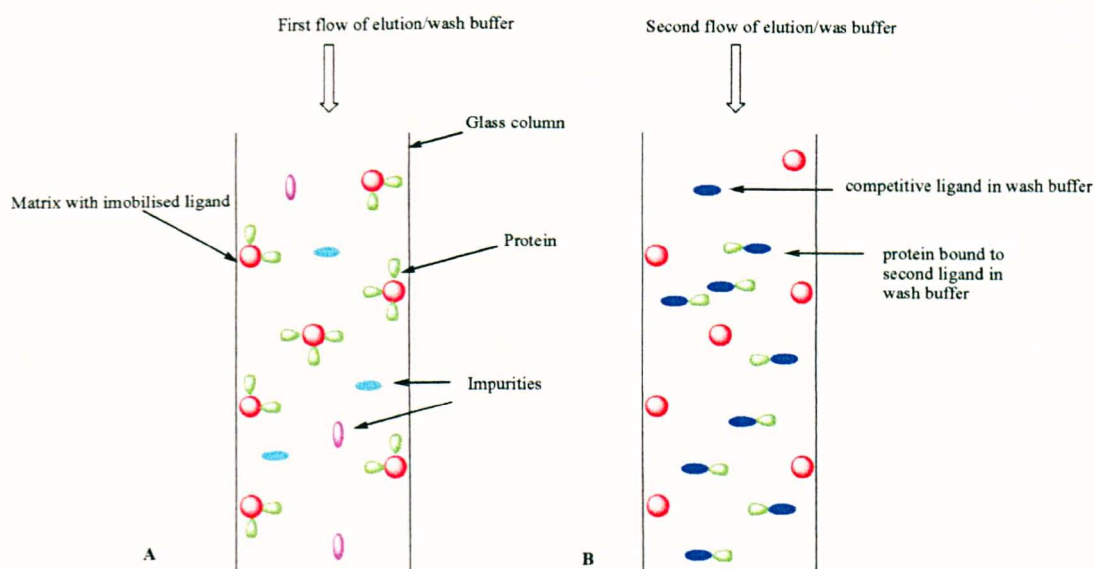


Figure 2.2 Schematic illustration of affinity chromatography. A) Protein of interest is reversibly bound to matrix containing ligand whilst the impurities are washed down with the wash buffer, B) the elution buffer contains a competitive ligand that has higher binding affinity towards the protein of interest and forms a reversible bond with the protein. Other techniques of detachment/removal of protein from matrix ligand may include changes in pH strength or ionic strength.

Monitoring of the product from affinity chromatography may involve the use of a UV detector which provides an elution profile of the protein from the column. The elution profile shows a plot of absorbance vs. time.

2.1.2 Hydrolysis

Since 1930 [Apicella *et al.*, 1994 and refs therein] there have been number of methods described for the separation of LPS from the bacterial membrane, such as extractions using trichloroacetic acid, ether, EDTA and pyridine. The most common method used is hot

phenol-water extraction. LPS are very large complex molecules and are therefore difficult to characterise using a single analytical method. One chemical method used to dissociate the lipid A from the polysaccharide is using acid hydrolysis, which introduces cleavage of the glucosidic linkages containing hemiacetal and hydroxyl terminal groups. This mechanism is dependent upon pH and the rate of hydrolysis is appreciable even below 100°C. The mechanism starts with proton from the acid interacting with the glucosidic oxygen linking the two sugar units, forming a conjugate acid. After the addition of water free sugar and proton are liberated (Figure 2.3).

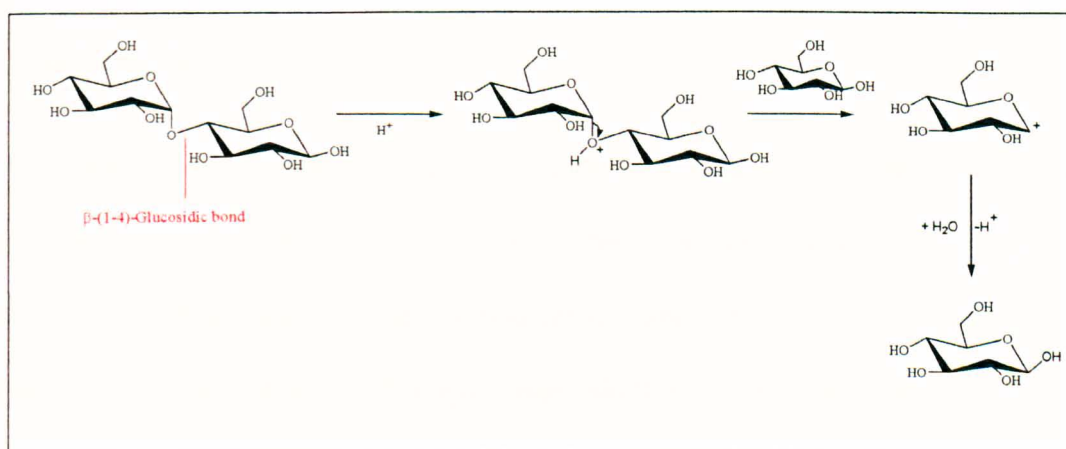


Figure 2.3. Mechanism of acid hydrolysis of the β -(1-4)-glucosidic bond of maltose. Drawn using CS Chem. Drawpro. Version 4.5

Hydrolysis carried out on LPS using 2% aqueous acetic acid always causes the cleavage of the glucosidic bond between the ketosidic and the lipid A bond [Masoud *et al.*, 1997; Apicella *et al.*, 1984; 1985]. This is found to be a very practical way of removing the lipid A whilst leaving the polysaccharide intact. Hydrolysis using this approach enables the polysaccharide to remain intact for the purpose of ligand binding with CRP and rfhSP-D.

2.1.3 Mass spectroscopy

Mass spectroscopy (MS), especially matrix assisted laser desorption ionisation (MALDI) and electro-spray ionisation (ESI), are widely used for structural characterisation and the determination of the structure of intact biomolecules such as proteins that have been separated using PAGE, oligo/polysaccharides, DNA and glycoconjugates [Kirpekar *et al.*, 1998; Zaia, 2004; Harvey, 2003]. Using these two techniques, both of which are employed in this work, provides details with great accuracy without causing structural destruction or multiple fragmentation that leads to complications in interpretation. The MS is based on three principals; formation of gaseous phase ions, mass analysis and detection [Hoffman and Stroobant, 2007].

ESI and MALDI are considered as soft ionisation techniques, which use minimal internal energy which is transmitted to the analyte for the ionisation process and hence does not cause too much fragmentation and is suited for low and extremely high molecular weight biomolecules. Ionisation occurs through proton addition or proton abstraction to form ions represented as $[M+nH]^+$ or $[M-nH]^-$ for positive or negative ions respectively, where M is the molecular mass, and n is the integer of the number of charges equivalent to protons H (1.00 proton charge).

In the presence of salts, ionisation can occur by the addition of cations or anions to analyte molecules, shown by the formulae $[M+Na]^+$ or $[M+Cl]^-$. The ions travel at a common velocity in vacuum and are analysed using quadrupole (Q), quadrupole ion trap (QiT), time of flight (ToF) or Fourier transform ion cyclotron resonance (FT-ICR) techniques [El-Aneed *et al.*, 2009]. The ions are manipulated by electric current or by magnetic field

based on their mass to charge ratio (m/z). Ions arriving at the detector are interpreted and displayed as a spectrum of abundance of ions against m/z values.

2.1.3.1 ESI-MS

Electro spray ionisation mass spectroscopy allows samples to be directly injected into the MS ionisation chamber from an online coupling device such as GC and HPLC systems. The samples, which can range from low to high mass, are dissolved in solvent such as ethanol, acetonitrile and water, in the presence of acid such as acetic acid or trifluoro acetic acid to aid protonation. The sample is then introduced into the MS chamber in the form of an aerosol, through a thin needle, which has a high electrical potential. The aerosols become highly charged droplets and as the solvent evaporates (desolvates) the size of the droplet is constantly reduced. There are two proposed methods by which gas phase ions, mainly multiply charged species $[M+/-nH]^{n+/-}$, are generated from charged droplets [Kearle, 2000]

- Ion evaporation: as the repulsive force of the ions on the droplet exceeds the surface tension of the solvent the ions desorb into gaseous phase
- Charge residual: the solvent is continuously evaporated followed by droplet fragmentation forming multiply charged ions.

2.1.3.2 MALDI-MS

Matrix assisted laser desorption ionisation MS (MALDI-MS) produces gaseous phase ions mainly as singly charged species $[M+/-H]^{+/-}$ from ionisation of the sample co-crystallised with matrix, by use of high intensity pulsed laser energy/radiation such as from a nitrogen laser at 337 nm, or from Nd:YAG lasers at 266 nm or 355 nm [Zhu *et al.*, 1995]. The sample is dissolved in a solvent such as acetonitrile, ethanol, methanol or water and in the

presence of acid such as acetic acid and trifluoro acetic acid, with an excess amount of matrix and air dried on MALDI target before irradiation with the laser. Gaseous phase ions are generated as the matrix absorbs the laser energy and is excited, transferring its energy to the analyte. Subsequently desorption and ionisation takes place. The matrix is usually a crystalline aromatic compound and there is no predetermined choice of matrix that will lead to desorption and ionisation of the analyte but rather the choice of matrix is experimental. Matrices commonly used in the structural analysis of proteins, oligonucleotides and oligosaccharides are 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid), α -cyano-4-hydroxycinnamic acid (alpha-cyano or alpha-matrix CHCA), 2,5-dihydroxybenzoic acid (DHB) and 1,5 diaminonaphthalene (DAN), 3-hydroxypicolinic acid and 5-methoxysalicylic acid (MSA), 2,6-dihydroxyacetophenone (DHAP), and 1,5-diaminenaphthalene (DAN) [Cohen et al., 1996; Zhao *et al.*, 1997; Distler *et al.*, 2001].

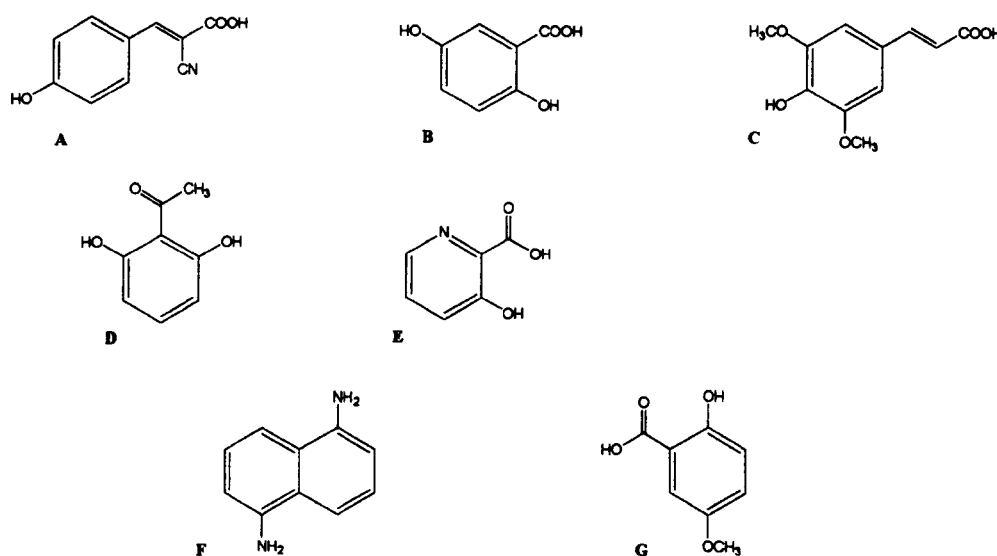


Figure 2.4 Structure of MALDI matrix A) CHCA, B) DHB, C) Sinapinic acid, D) DHAP, E) 3-hydroxypicolinic acid, F) DAN, G) MSA.

2.1.3.3 Mass spectrum interpretation

The spectrum is a representation of abundance against m/z values of charged ions, with a series of peaks related to the degree of protonation or deprotonation (charge). Several peaks differing by a single charge may arise. Multiple charge formation from high molecular weight sample tends to provide lower m/z values. For a single peak that is singly charged or multiply charged then the molecular weight can be calculated as follows for positive ionisation mode.

$$m/z = [M + nH]^{n+} / nH \text{ where } H \text{ is the charge of a proton (1.00)}$$

for negative ionisation the equation is as follows:

$$m/z = [M - nH]^{n-} / nH \text{ where } H \text{ is the charge of a proton (1.00)}$$

A spectrum with multiple peaks in which successive peaks at m/z_1 and m/z_2 differ by a single charge can be given by the series (shown for positive ionisation mode)

$$m/z_1 = [M + nH]^+ / nH$$

$$m/z_2 = [M + nH + 1]^+ / nH + 1$$

then, solving simultaneous equations gives the charge by which it is ionised

$$n = (m/z_2 - 1) / (m/z_1 - m/z_2)$$

2.1.3.4 Structural analysis of LPS using MS

ESI and MALDI are used to determine the composition of sugars and non carbohydrate substituents [Li and Richards, 2007; Masoud *et al.*, 1997; Mikhail *et al.*, 2005]. To obtain a molecular mass and to determine the composition of LPS in the native state is very complicated due to the solubility in solvents use for mass spectrometry. Therefore preliminary work on the LPS is carried out using deacylation, methylation and dephosphorylation [Masoud *et al.*, 1997 and Mikhail *et al.*, 2005].

Running solvents such as acetonitrile/water or methanol/water (with added NH_3 in negative ionisation mode and HOAc in positive ionisation mode) are used to enhance the ionisation process. The distribution of molecular ions can be observed as doubly, triply or quadruply charged species. The analysis can be complex and is beyond the scope of this thesis, in general due to the complex nature of LPS. Further structural information may be obtained by fragmenting the molecule using tandem ESI-MS, which uses triple quadrupole (TQ) ESI-MS/MS, or an ion trap (IT) ESI-MS and employing collision induced dissociation (CID). These techniques allow the elucidation of non carbohydrate substituents in oligosaccharide samples [Li and Richards, 2007].

Structural analysis of oligosaccharide is mainly deduced from ESI-MS on O-deacylated LPS, core oligosaccharide, and permethylated dephosphorylated oligosaccharide [Masoud *et al.*, 1997; Yildirim *et al.*, 2003] as well as the hydrolysed LPS. Glycosidic cleavage occurs at the reducing end if a terminal HexNAc residue is present. Methylation of the glycoses allows easy identification between terminal and substituted units. It also allows distinction between ions generated by fragmentation from cleavage of single glycosidic linkages and inner fragments from cleavage of two glycosidic linkages. ESI-QIT-MS

performed in the negative ion mode is used to determine the structure of lipid A [Mikhail *et al.*, 2005].

Dephosphorylation and permethylation with the addition of sodium acetate is seen to increase the MS response, with additional glycoforms being observed in the full MS scan compared to underivatised samples [Mansson *et al.*, 2003; Reinhold *et al.*, 1996]. Fragmentation of such oligosaccharides occurs by cleavage of the glycosidic bond with oxygen being retained at the reducing end leading to two products ions B_i or Y_i (Figure 2.5) [Reinhold *et al.*, 1996; Viseux *et al.*, 1998].

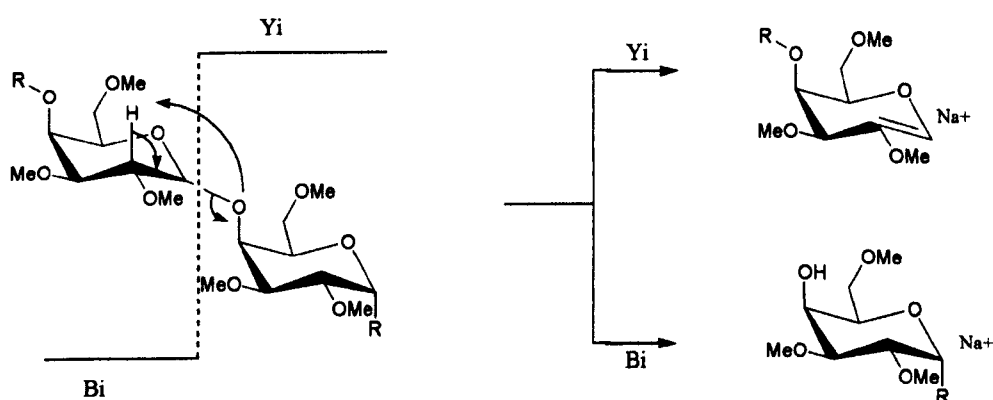


Figure 2.5 Fragmentation of methylated glycoside bond cleavage in MS producing Y_i and B_i ions with sodium adduct.

2.1.4 Protein crystallography

Protein crystallography is a technique which relies on crystals of pure protein in order to study the structure at the atomic level. Protein crystals are ordered three dimensional arrays of molecules that have the ability to diffract X-rays, thus enabling the study of the structure [Glusker and Trueblood, 1985; Rhodes, 1993; 2000; Blow, 2002].

Growing protein crystals of good quality for structural studies using X-rays may depend initially on the quality and concentration of protein. Most crystallisation trials employ the use of protein in stabilising buffers to prevent denaturation or compromising the stability in any form. Typically crystallisation uses protein concentrations of around 10mg/ml. Other factors that influence the quality of crystal growth are temperature, pH strength of buffers and choice of buffers/precipitants.

Three main types of precipitants that have been commonly used are salts such as ammonium sulphate, organic solvents such as 2-methyl-2,4-pentanediol (MPD) and organic polymers such as polyethylene glycol (PEG). There is an enormous choice of commercially “ready made” crystal trial “starting kits” that contain varying composition of precipitants and buffers at different pH that enable screening for optimal conditions that would favour crystal growth. These provide a starting point for optimising conditions in order to achieve better crystals [Jancarik and Kim, 1991].

The process of determining the 3D structure of the macromolecule requires several steps (Figure 2.6):

- Initial crystal growth
- X-ray diffraction data collection
- Computational analysis and

- Construction of the electron density map with subsequent refinement to obtain the 3D structure.

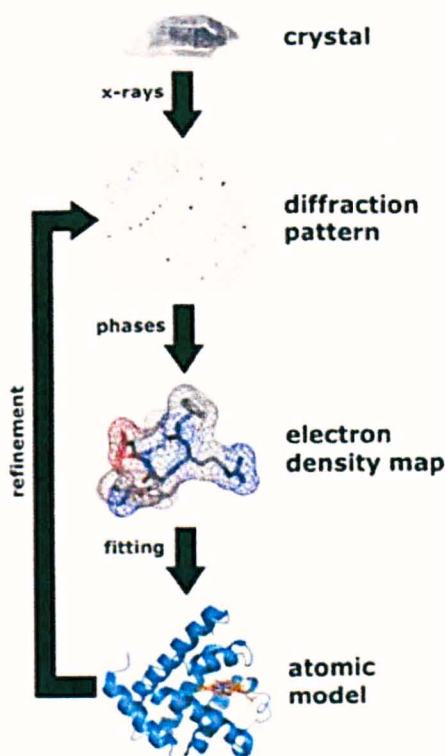


Figure 2.6 The stages involved in X-ray crystallography to determine the overall 3D structure of a protein. Image taken from www.doly.biochem.arizona.edu

2.1.4.1 Crystallisation

The initial stage of crystallisation of protein requires obtaining the protein either by expression or purification and quantifying the concentration of protein in the buffer. Protein concentrations are determined from the absorbance of UV at 280nm using a spectrophotometer. There are two types of crystallisation set up that are routinely used; the sitting drop vapour diffusion method and the hanging drop vapour diffusion methods. The former method is used in this project which involves using 24 well plates containing 1ml of precipitant buffer. Each well also contains a small plastic “micro-bridge” containing a

hollow groove on to which the protein sample and an aliquot of the well precipitant buffer (mother liquor) is mixed normally in an equal volume ratio, for example 2 μ l of each.

The well is covered using a glass cover slip sealed with vacuum grease and kept at a constant temperature. The process of crystallisation is that the concentration of the solubilised protein in the precipitant buffer gradually increases as the water content equilibrates between the mother liquor and the drop on the micro bridge groove. This increase in concentration of protein to a point of super saturation leads to the formation of ordered aggregates known as the nucleation phase followed by steady crystal growth in the metastable phase [Luft, 1994; Wilson, 1991; Glusker and Trueblood, 1985; Gernert, 1988].

2.1.4.2 X-ray diffraction setup

Crystals for X-ray diffraction trials are mounted in a nylon loop and also flash frozen in liquid N₂ and stored. Prior to X-ray diffraction it is necessary to “cryoprotect” the crystal using reagents such as glycerol, PEG 400, ethylene glycerol and methyl pentane diol. Cryoprotection is very critical since it prevents formation of ice crystals within the crystal which disrupt the crystal lattice.

The success of cryoprotection can be judged by the glassy appearance of the crystal in the nylon loop and the absence of ice rings in the diffraction pattern. There is no accurate method to determine the suitability of cryoprotectant and in many instances it can lead to the destruction of the crystal used for analysis, and is therefore a matter of trial and error. Making cryoprotectant is carried out by substituting a volume of water used in making the mother liquor for the above mentioned reagents acting as the cryoprotectant.

The crystal is then attached to the X-ray diffraction apparatus via the goniometer head and placed in a stream of liquid nitrogen [Hope, 1990; Glusker and Trueblood, 1985]. Once the crystal is accurately mounted, focused X-rays emerging from a collimator using adjustable slits to 0.1-0.3mm diameter [Smyth and Martin, 2000] strike the crystal to produce a diffraction pattern recorded on a detector (for example a charged couple device, CCD) [Rhodes, 2000]. The X-rays are focused and collimated to ensure that the beam is parallel and lacking significant “crossfire”.

2.1.4.3 X-ray diffraction analysis

X-rays are ideal to measure the inter-atomic bond length because they have a wavelength of 0.1 to 100Å, the shorter wavelengths being close to the 1.5Å distance of C-C bonds. The quality of X-ray diffraction depends primarily on the crystal, but also on the mode of X-ray generation, that will also determine the time of data collection and the quality of the data (i.e. recorded intensity). X-rays can be generated by X-ray generators and synchrotron sources. The latter is far more advanced and widely used in macromolecular structural determination and data collection is very rapid i.e. minutes as opposed to hours or days using X-ray generators [Helliwell, 1992; Ealick and Walters, 1993].

2.1.4.4 X-ray generators

In a simple laboratory set up “in house”, X-rays are produced in an X-ray tube. The chamber consists of tungsten filament (cathode) and a metal target usually copper at the opposite end of the anode which is rotating. Electrical current is applied through the tungsten filament with a large potential difference (kilovolts) placed between the cathode and anode.

Since the chamber is in a vacuum, the electrons (generated in the tungsten filament) move at a relatively high velocity from the cathode to the anode. The collision of the electrons with the metal atoms in the target (anode) causes the movement of electrons from the inner valency shell of the metal which is quickly occupied by electrons from the outer shell. The transition of electrons from the outer shell (lower energy) to inner shell (higher energy) generates X-rays which are then directed through the X-ray tube towards the crystal mounted on a goniometer placed at a distance (from the X ray tube). The crystal can be rotated to provide data collection at various angles. The wavelength of X-rays produced depends on the structure of the atom being bombarded with electrons, e.g. $\text{CuK}\alpha$ has a wavelength of 1.542\AA (Rhodes, 2000).

2.1.4.5 Synchrotron radiation source

X-rays generated at a synchrotron radiation source (eg European Synchrotron Radiation facility, Grenoble, Daresbury SRS, England and the new facility at Diamond England) are extremely intense and provide a high signal to noise ratio in the diffraction image. Data collection can be achieved rapidly (shorter exposure time before the crystal is destroyed by X-rays).

Synchrotron radiation is produced by highly charged particles such as electrons that circulate at relativistic velocities confined by magnetic fields in an evacuated toroidal cavity in a storage ring. Insertion devices such as undulators and wigglers cause the charged particles to change their circular path, emitting intense X-rays in a narrow cone in the forward direction at a tangent to the circular path (Figure 2.7). The electromagnetic radiation range emitted is from the ultra violet to the X-ray region. The selection of a particular wavelength is chosen using monochromators.

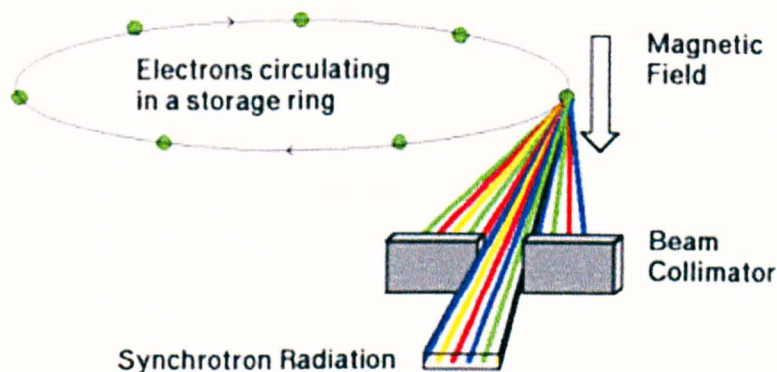


Figure 2.7 Schematic illustration of synchrotron radiation caused by electrons travelling at relativistic velocity in an orbit. Image taken from <http://all-physics.net>

2.1.4.6 Production of X-ray diffraction

X-ray waves have both phase and amplitudes and are deflected by the electron clouds of each and every atom of the molecule. The diffracted X-rays from the atoms of the crystal are detected on a CCD detector and interpreted as dots (diffraction maxima) [Gruner, 1994]. The diffraction maxima provide a measure of the intensity (I) of the diffracted X-rays. If the diffracted X-ray waves are in phase then constructive interference will occur and the amplitudes will add up producing a wave with higher amplitude; this will be seen as intense dark spots on a detector. If the waves are out of phase from the point of origin then destructive interference will occur cancelling the waves and the resulting wave will have no amplitude and therefore will not be detected (i.e. no spot will be detected) (Figure 2.8).

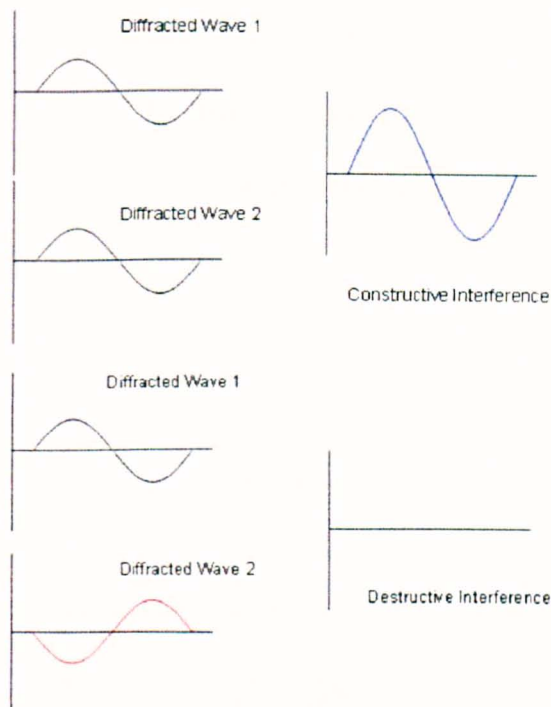


Figure 2.8 Illustration of X-ray waves forming constructive and destructive interference as a consequence of their phase difference [Glusker and Trueblood, 1985].

2.1.4.7 Braggs Law

This is a mathematical description of the diffraction of X-rays from atoms in a plane (considering that crystals consist of orderly arrangements of atoms in planes)

$$n\lambda = 2d_{hkl} \sin\theta$$

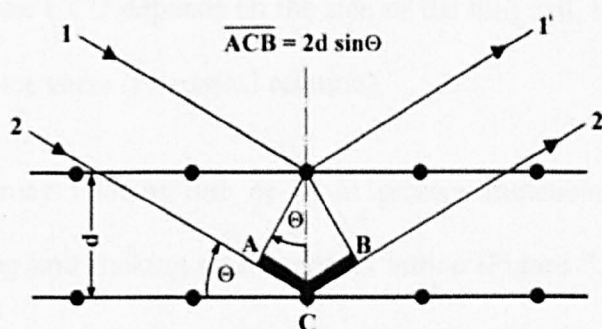
Where: n is the integer (1, 2, 3, etc) or order of reflection

λ is the wavelength of the X-rays (Å)

d is the interplanar distance hkl (miller indices of the lattice planes)

2θ is the diffraction angle

The above equation can be described in terms of a diffraction pattern by atom(s) in crystal planes. If the distance travelled by two X-rays differs by an integral number of wavelengths $n\lambda$, then X-rays 1 and 2 will be in phase leading to constructive interference (Figure 2.9).



Where d is the inter planer distance

Figure 2.9 Bragg's construction of "reflection" of incident waves from planes hkl . If the waves are in phase then constructive interference will occur and if waves are out of phase then destructive interference will occur.

2.1.4.8 X-ray diffraction data collection

The purpose of data collection is to measure the diffracted intensities (I_{hkl} which is measured experimentally on a CCD) and standard uncertainties (σI) for all hkl [Leslie, 2006]. Measuring the intensities is the first stage of data collection. These will eventually be used in calculation of "structure factors" from which the electron density map can be calculated and constructed. The crystal mounted on the goniometer is rotated through a succession of small angles ϕ , typically 0.2 to 2° [Leslie, 2006; Glusker and Trueblood

1985], and all reflections being recorded on the CCD. The rotation of the crystal is perpendicular to the beam, maximising the amount of data collected for each image, as the crystal lattice planes sweep through the X-ray beam [Smyth and Martin, 2000]. A resolution of 3\AA is regarded as sufficient to detect the amino acid side chains in the electron density map [Smyth and Martin, 2000]. The quality and completeness of the data collection will depend on the crystal quality and the length of time the crystal can stay in the X-ray beam before deterioration caused by radiation damage. The spacing of the diffraction maxima on the CCD depends on the size of the unit cell, the closer the spacing the large unit cell and vice versa (reciprocal relation).

The unit cell, which may contain one or more protein molecules related by crystal symmetry, is a repeating unit making up the crystal lattice (Figure 2.10). The dimensions of the unit cell are given as a , b , and c and three angles α , β and γ , these measurements also indicate which type of unit cell the crystal lattice has formed [Blow, 2002] (Table 2.1)..

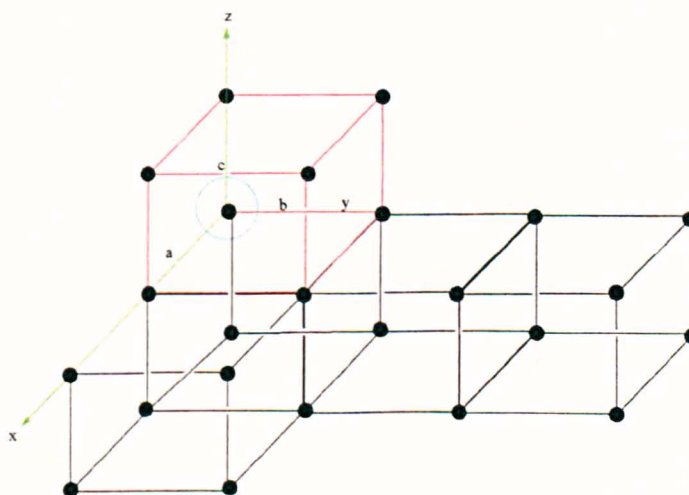


Figure 2.10 Unit cell (red) that makes up the crystal lattice. The lattice points are shown as dots.

Lattice	Unit Cell Conditions				
Monoclinic	a_1	\neq	a_2	\neq	a_3
	$\alpha = \gamma = 90^\circ, \beta$				
Orthorhombic	a_1	\neq	a_2	\neq	a_3
	$\alpha = \beta = \gamma = 90^\circ$				
Tetragonal	a_1	$=$	a_2	\neq	a_3
	$\alpha = \beta = \gamma = 90^\circ$				
Cubic	a_1	$=$	a_2	$=$	a_3
	$\alpha = \beta = \gamma = 90^\circ$				
Rhombohedral	a_1	$=$	a_2	$=$	a_3
	$\alpha = \beta = \gamma < 120^\circ \neq 90^\circ$				
Hexagonal	a_1	$=$	a_2	\neq	a_3
	α	$=$	β	$=$	90°
	$\gamma = 120^\circ$				
Triclinic	a_1	\neq	a_2	\neq	a_3
	$\alpha \neq \beta \neq \gamma$				

Table 2.1 Types of crystal lattice and relationships between their cell parameters. Taken from [Blow, 2002].

2.1.4.9 Diffraction data processing

Data processing involves generating a systematic listing of all the recorded reflections with their Miller indices, intensity (I_{hkl}) and standard uncertainties (σI). The intensities and the phases (α) contain the structural information, and it is this information that needs to be translated.

2.1.4.9.1 Auto indexing

The interactive computational programme MOSFLM [Leslie, 1993, 1992] is used to determine the unit cell and space group and to assign Miller indices (h, k, l) to the diffraction spots. A list of possible unit cells with penalty points is output from which one

is chosen for further refinement of the cell by taking blocks of images. Predicted diffraction spots based on the unit cell dimensions and orientation are then superimposed to fit the observed diffraction spots. If the observed and predicted spots do not match, then parameters such as crystal to detector distance, unit cell parameters and the direct beam coordinates will be fine tuned within the MOSFLM programme. Spot intensities are then measured. Scaling and merging of intensities is done using SCALA from CCP4 (1994). The output of scaling contains the index of each spot and its measured intensity I and σI .

2.1.4.9.2 Truncate

Truncate of the CCP4 suite is used to calculate the amplitudes $|F_{hkl}|$ from measurement of the intensities (I_{hkl}) of all reflections (hkl). The structure factor amplitude $|F_{hkl}|$ is described as the square root of the measured intensities. Equation 1

$$|F_{hkl}| = \sqrt{I_{hkl}} \quad \text{Equation 1}$$

Where $\underline{F}_{hkl} = |F_{hkl}| e^{i\alpha}$ Equation 2

The structure factor \underline{F}_{hkl} which allows calculation of the electron density equation cannot be calculated since the phase (α) information is lost.

2.1.4.9.3 Structure factor

For a given diffraction maximum hkl the structure factor \underline{F}_{hkl} provides details of the types and arrangements of the atoms in the unit cell (relationship between the structure and its diffraction pattern) and is described by the equation.

$$\underline{F}_{hkl} = \sum_{j=1}^N f_j \exp [2\pi i (h x_j + k y_j + l z_j)] \quad \text{Equation 3}$$

where the summation is over all N atoms ($j=1-N$), f_j is the scattering factor for j^{th} atom in the unit cell and xyz are the atomic coordinates of atom j in the unit cell.

To solve these equations and calculate \underline{F} , the phase needs to be computed (see equation 2). There are two main methods used to obtain calculated phase (α_{cal}) for protein diffraction data:

- The isomorphous replacement method [Blow, 2002] relies on incorporating a heavy atom, such as a gold, platinum or mercury, within the crystal without distorting the conformation of the protein or unit cell. The diffraction data from the protein plus heavy atom may be used to locate the heavy atom and then calculate its position x,y,z and hence the heavy atom structure factors (Equation 3) and phases (Equation 2). This information may then be used to estimate phases for the protein itself.
- The molecular replacement method [Rossman and Blow 1962], is widely used to estimate the phase, this method uses the availability of a closely related protein structure or a homologous section of a related protein (a search model) and tries to place it in the unit cell in the same position (translation) and orientation (rotation)

as the unknown structure. Structure factors calculated from this model may then be used as estimates of the structure factors of the protein itself.

The search model, which is a homologous structure closely related to the structure under investigation, perhaps the same protein from a different crystal form, is allowed to “fit” in to the new unit cell to determine the orientation, position and the coordinates of the unknown molecule in the new unit cell. In order to achieve this, there are two sets of parameters that need to be satisfied; three rotation angles followed by three translational vectors. Simply, structure factors amplitudes are calculated from the known molecule as it is rotated and translated within the unit cell and these are compared with the observed structure factors throughout the rotation and translation range. For highly homologous structures there will normally be a region of high correlation between the two sets of data which corresponds to the actual position and orientation of the unknown structure, ie the x, y and z (Equation 3).

The programme used for molecular replacement in this work was AMORE from the CCP4 suite [CCP4, 1994]. The model is then refined, initially using rigid body refinement, to determine the best fit of the model to the structure by minimising the difference between the structure factor amplitudes $|F_{hkl}|_{cal}$ calculated from the model and the experimental or observed structure factor amplitude $|F_{hkl}|_{obs}$. This is an iterative process within CNS [Brunger *et al.*, 1998] and is performed each time the model is changed with the agreement between observed and calculated structure factors expressed as the R factor or reliability index.

2.1.4.9.4 R factors

The “goodness of fit” between observed and calculated structure factors is shown by the R factor (reliability factor) expressed mathematically by the following formula:

$$R = \sum(|F_{hkl}|_{\text{obs}} - |F_{hkl}|_{\text{calc}}) / \sum(|F_{hkl}|_{\text{obs}}) \quad \text{Equation 4}$$

The R factor measures the discrepancy between the observed structure amplitudes $|F|_{\text{obs}}$ and the calculated structure factor amplitudes $|F|_{\text{calc}}$. The value of R reduces with model improvement suggesting good agreement between experimental and calculated data [Brunger, 1992] i.e a good model fit with observed data. For the best fit between model data set and observed data set the R value should be as low as possible (in the region of 0.2 (20%) is often achieved) and initial high values are improved with subsequent refinement. Although a low R is aimed for, it may sometimes be achieved with an incorrect structure [Kleywegt and Brunger, 1996]. Cross validation of the R value during refinement is carried out by calculating R_{free} (free R value), which is based on e.g 5% of the data (complementary test set) that is excluded from modelling and refinement process. The remaining 95% of the data (working set) is used in the refinement process.

The R_{free} value has been shown to correlate with the accuracy of atomic models. Models that show R_{free} values greater than 0.40 will probably contain errors and be based on unsatisfactory refinements even though the R-factor itself is low [Kleywegt and Brunger, 1996].

2.1.4.9.5 Calculation of the electron density map

Construction of the electron density map, which is a 3D contour map of the structure, within which the protein structure will be built, is calculated in the CCP4 suite [1994] by using SFALL, SigmaA and FFT, where SFALL calculates the Structure factor (F_{calc}) based on the atomic positions (Equation 3), SigmaA combines the calculated α with experimentally obtained data and FFT constructs the map. The electron density ($\rho(x,y,z)$) map is calculated by

$$\rho(x,y,z) = 1/V \sum_{hkl} \underline{F}_{hkl} \exp[-2\pi i(hx+ky+lz)] \quad \text{Equation 5}$$

where the summation is over all hkl . The structure factor \underline{F} is calculated from the measured amplitudes $|F_{hkl}|$ combined with the estimated (calculated) phase (Equation 2), or from known atomic positions (Equation 3) as determined via molecular replacement. The quality of the electron density map depends largely on the quality of the estimated/calculated phase which tends to bias the structure towards the model, known as Model bias. The map calculated is often a construct of $2F_{\text{obs}} - F_{\text{calc}}$ (observed and calculated structure factor) which is a combination of an $F_{\text{obs}} - F_{\text{calc}}$ map to identify errors in a model and an F_{obs} map of the structure itself. The map is used to guide the construction of a new model, including the addition of non-protein atoms such as ligands, using graphics programmes such as O [Jones *et al.*, 1991]. Cycles of structure factor calculation, refinement, map calculation and model building lead to the final structure.

Chapter 3
Experimental:
Materials & Methods

3.1 Isolation of *Limulus* CRP from *Limulus* plasma using affinity chromatography

The *Limulus* plasma was kindly provided by Professor Peter Armstrong, University of California San Diego, obtained from the Atlantic horse shoe crab *Limulus Polyphemus*. The *Limulus* plasma contains two components; CRP and SAP that need to be isolated using PC and PE affinity chromatography, a procedure that is routinely used in the Greenhough laboratory [Shrive *et al.*, 1999]. The instrument used is a BIORAD Biological LP system and a 25ml volume PE agarose column that is assembled in the cold room at a temperature of 4°C.

The PE agarose column is washed down (equilibrated) with calcium wash buffer consisting of 500mM Tris base, 100mM CaCl₂ and 1.5M NaCl at a pH 7.4. The calcium wash buffer is made by dissolving Tris base (60.57g) in deionised water (200ml) to which was added CaCl₂ dihydrate (14.7g) followed by adding another 100ml of deionised water and NaCl (87.66g), and then adding a further 500ml of deionised water. At this point the pH of the final solution was adjusted from 10.7 to a final pH of 7.4 using HCl and the solution was made up to 1L using a volumetric flask.

The procedure for affinity chromatography carried out in the cold room is as follows: 7.5ml of *Limulus* plasma was diluted with 15ml calcium wash buffer and filtered through 0.2µm cellulose filter to remove any particles. Using the BIORAD Biological LP system the plasma was loaded on to the PE column at a flow rate of 0.3ml/min and then eluted down the column using 75ml of calcium wash buffer containing 10mM PC to elute the CRP fraction. Elution of the CRP fraction was monitored using UV absorbance at 280nm linked to an LP data view programme on a Biorad Biological HR controller (situated outside the cold room) and according to the CRP elution profile approximately 45ml of the

CRP elution fraction was collected in a glass beaker until the UV absorbance at 280nm read zero.

After collection of the CRP fraction the column was again equilibrated using calcium wash buffer before eluting the column with 75ml calcium wash buffer containing 30mM PE to elute the SAP, the elution of SAP was also monitored at 280nm and according to the elution profile approximately 12ml of the SAP elution was collected in the glass beaker until the elution profile read zero.

After collecting the two fractions, the PE column was finally equilibrated with calcium wash buffer containing 10mM EDTA to elute any calcium dependently bound proteins and the column regenerated using regeneration buffer A (500mM NaCl, 200mM Tris, 10mM EDTA), regeneration buffer B (500mM NaCl, 50mM Tris, 10mM CaCl₂) and regeneration buffer C (150mM NaCl, 50mM Tris, 10mM CaCl₂, 0.02% NaN₃).

The CRP and SAP fractions were isolated from the PC and PE ligand using dialysis buffer (Tris 0.5M, NaCl 1.5M, EDTANa₂ 0.1M pH 7.4). This was carried out by placing the CRP (and SAP) fractions in dialysis tubing (Medicell) with a molecular cut off of 12-14kDa and left immersed in the dialysis buffer, under stirring for approximately 36 hours at 4°C with frequent change of dialysis solution. The solution was then dialysed back in the calcium wash buffer to restore the lost calcium ion from the previous dialysis using EDTA.

The volume of the CRP fraction after dialysis was concentrated to approximately 4ml (~4mg/ml) using centrifugal speed of 2424rpm (HERMLE Z400k) in a 15ml Amicon ultra filter tube that has a molecular cut off at 10kDa.

Protein concentration in the final solution was determined by measuring absorbance at UV 280 and 320nm. This was carried out by placing the sample in a Quivet Cell and analysing using Quivet spectrophotometer. This gave separate readings of absorbance at 280 and 320nm UV wavelength. The calculation of protein was done by taking the absorbance at 280nm minus the absorbance at 320nm and then multiplied by the dilution factor then divided by the extinction coefficient (Protein Concentration= $(A_{280} - A_{320}) \times \text{Dilution factor}$ then corrected by dividing the concentration using the *Limulus* CRP extinction coefficient of 1.5)

For protein crystallography 7.0-7.5mg/ml of CRP was used. This was obtained by concentrating 1ml of CRP solution at a concentration of 4.27mg/ml to 568 μ l using centrifugal filters (Amicon Ultra). An aliquot of 7 μ l of this solution was diluted (1 in 10 dilution) with 63 μ l of calcium wash buffer (500mM Tris base, 100mM CaCl₂, 1.5M NaCl at pH 7.4) and placed in the Quivet spectrophotometer and measured at UV 280 and 320nm.

3.2 Lipopolysaccharide

Lipopolysaccharide of *E. coli* 0111.B4 and *P. aeruginosa* serotype 10 strain ATCC27316 were purchased from Sigma.

H. influenzae Eagan LPS is an invaluable starting material that is not available commercially, but was kindly provided by Professor Moxon (The Institute of Molecular Medicine, Oxford). The type b strain Eagan used in these experiments was a disease isolate obtained from the Netherlands, the strains being grown at 37 °C in brain-heart infusion broth (BHI) supplemented with haemin (10 μ g/ml) and NAD (2 μ g/ml). For selection after

transformation from wild to mutant LPS, Kanamycin (10µg/ml) was added to the growth medium, so that the wild-type and mutant bacteria could be distinguished.

3.2.1 Hydrolysis of *E. coli* LPS

Preparation of the polysaccharide from LPS is achieved by hydrolysis in accordance with described methods (Masoud *et al.*, 1997) using 2% aqueous acetic acid and refluxing at 100°C for 2.5 hours using a sand bath and heating mantel.

E. coli LPS strain 0111.B4 was obtained by phenol-chloroform extraction. The LPS is in the form of a very light and fluffy white powder and so extreme care was taken when weighing out 25mg of the LPS into a 25ml round bottom flask (containing a small magnetic stirrer).

To this was added 2% aqueous acetic acid (5ml) and gently stirred using a magnetic stirrer device. The round bottom flask was then attached to a reflux condenser and the mixture heated using a sand bath to approximately 100 °C (the round bottom flask could also be heated in an oil bath but the former was chosen due to experimental practicality and safety). The hydrolysis timing of 2.5 hours was started when the temperature of the sand bath reached 100°C. After this time the heating (but not the water to the reflux condenser) was turned off and the round bottom flask was raised out of the sand bath and allowed to cool sufficiently (to hold in the palm of the hand). When the flask had cooled down sufficiently, the water supply to the reflux condenser was turned off and the flask removed from the reflux condenser and the contents transferred immediately using an automatic pipette to a 15ml plastic centrifuge tube and then further cooled for 30-60 minutes at 4°C in the fridge. The tube was then placed in a centrifuge (HERMLE Z400) at 4°C and centrifuged at a speed of 2424g for 30 minutes.

The top supernatant was carefully removed using an automatic pipette so as not to disturb the pellets that had settled at the bottom of the 15 ml tube and was transferred to another tube and stored at 4°C. The pellets in the tube included a small amount of the supernatant that could not be completely removed. Deionised water (1ml) was added followed by chloroform (1ml) which dissolved the pellet.

The solution was mixed thoroughly and allowed to settle out. Once the solution was settled, two liquid fractions could be seen, the top layer being the aqueous layer and the bottom layer the chloroform. The top aqueous fraction was removed using an automatic pipette and combined with the previous supernatant whilst the chloroform layer was transferred to another tube (1.5ml eppendorf tube).

In order to determine if the hydrolysis had completed, the aqueous supernatant was again hydrolysed under the same conditions using the procedure described above. The aqueous supernatant was lyophilised using savant speed vacuum.

3.2.2 Hydrolysis of *P. aeruginosa* LPS

Commercially purchased (Sigma) *P. aeruginosa* serotype 10 strain ATCC27316 LPS (50mg) was hydrolysed according to the methods described in section 3.2.1

3.2.3 Hydrolysis of *H. Influenzae* Eagan wild type LPS

H. influenzae Eagan wild type b (50mg) was hydrolysed according to the methods described in section 3.2.1

3.2.4 Hydrolysis of *H. influenzae* Eagan 4A LPS

Eagan 4A (53mg) was hydrolysed according to the methods described in section 3.2.1

3.3 MS analysis of *E. coli* polysaccharide by ESI

Samples were analysed on a Waters ZQ400 mass spectrometer (Swansea mass spec services, England) fitted with an electrospray ion source. Backbone polysaccharide (3mg) was dissolved in water (200µl) which was then mixed with 50% aqueous acetonitrile containing 1% acetic acid for mass spectral analysis in either the negative or positive-ion mode. Samples were injected by infusion with loop injection by an Agilent HP1100 LC in a stream of methanol/water (1:1) and the mass spectrometer was scanned from m/z 4000 to 14000.

3.4 MS analysis of *E. coli* lipid A using MALDI

Lipid A sample (2mg) was dissolved in a mixture of chloroform/methanol (4:1) before being mixed with matrix solution (1:1) which is a saturated solution of 10mg/ml 2, 5-dihydroxybenzoic acid in 50% acetonitrile. The sample mixtures were allowed to dry at room temperature prior to mass analysis. The mass spectra were obtained in the negative ion mode on a Applied Biosystems Voyager DE Pro Spectrometer equipped with a nitrogen laser (337) and delayed extraction optics (RJAH, Oswestry, UK).

3.5 MS analysis of *H. influenzae* Eagan 4A oligosaccharide using MALDI

The oligosaccharide preparation from *H. influenzae* Eagan 4A was analysed by matrix-assisted laser desorption ionisation. *H. influenzae* Eagan 4A polysaccharide (4mg) was dissolved in deionised water (100µl) mixed using a vortex. An aliquot (1µl) was mixed with matrix solution (10mg/ml 2, 5 dihydroxy-benzoic acid in methanol 100%) and air dried on the MALDI target. The mass spectra were obtained in the negative ion mode on a 4700 Reflector Spectrometer equipped with a nitrogen laser (337) and delayed extraction

optics (RJAH, Oswestry, UK). An equimolar mixture of external peptide standard was used for mass calibration.

3.6 Preparation of rfhSP-D

The rfhSP-D was kindly provided by Paul Townsend, MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford. Preparation by Paul Townsend was as follows:

The rfhSP-D was expressed in *E. coli*. The cDNA for the fragment, Gly 179-Phe 345 (for the neck/CRD including a short region of the collagen stalk (8 Gly-X-Y) and representing residues 179-355, was cloned from human lung library DNA. This was then inserted into pET-21d vector (Novagen, Nottingham, UK) and the plasmid was transformed into BL21 (λ DE3) pLysS and a single colony selected and re-plated. Once grown overnight at 37°C a single colony was inoculated into 25ml LB broth 25µg/ml ampicillin and grown overnight with shaking. 5mls was taken from the overnight culture and inoculated into 500mls LB broth 25µg/ml ampicillin in 2L baffle flasks the following day. When the OD₆₀₀ reached 0.61mg/ml IPTG was added for 3 hours to induce protein expression. The cells were then collected by centrifugation, lysed and sonicated (rfhSP-D is expressed in insoluble inclusion bodies). The pellet was solubilised, and then clarified by centrifugation and refolded by overnight dialysis of 6M-0M urea 20T, 150mM NaCl pH 7.4 for 2 nights. Correctly folded rfhSP-D was separated from denatured rfhSP-D by adsorption onto maltose-agarose (Sigma-Aldrich, Poole, UK) using 20mM Tris, 150mM NaCl, 10mM CaCl₂ pH 7.4 and eluted with 20mM Tris, 150mM NaCl, 5mM EDTA pH 7.4.

rfhSP-D was further purified by using a Superose 6/100 column (Amersham Pharmacia, UK), which had been equilibrated with 20mM Tris, 150mM NaCl, 5mM EDTA pH 7.4. A

single peak corresponding to a molecular weight of 60 kDa was observed. The recombinant preparation was judged to be pure by using SDS-PAGE, immunoblotting and amino terminal sequencing. A batch of rfhSP-D at a stock concentration of 1mg/ml was dialysed into 10mM Tris, 10mM CaCl₂, 140mM NaCl, 0.02% NaN₃ at pH 7.5.

3.7 Preparation of *Limulus* CRP

The CRP was obtained from PC and PE affinity chromatography as described in experimental section 3.1. The *Limulus* CRP used was at a concentration of 7.5mg/ml in buffer (Tris base 500mM, CaCl₂ 100mM, NaCl 1.5M, pH 7.4).

3.8 Crystal tray set up

The sitting drop vapour diffusion method (Figure 3.1) was used for crystal trials using 24 well Linbro plates. Precipitant buffer that had been previously filtered through acetate filter with a 0.2µm pore was placed in each well with a total well volume of 1ml. The amount of each component of the well mixture (precipitant buffer) was calculated by using the formulae;

$$(X / Y) \times 1000 = \text{volume } (\mu\text{l}) \text{ of reagent to be place in the linbro tray well.}$$

Where X is the required concentration or percentage of reagent required (see results and discussion), and Y is the concentration or percentage of the stock solution. This is then multiplied by the volume of well mixture (1000 µl) and the remaining volume is made up using deionised water.

A 2µl aliquot of the well precipitant buffer (see result, chapter 5) was placed on the micro bridge (Molecular Dimension) followed by adding protein with or without the LPS or the polysaccharide as described in chapter 5. Each well was covered with a glass cover slip

(Menzel-Glaser, Braunschweig, Germany) and sealed using high vacuum grease (Dow Corning, USA) and placed in a cupboard at room temperature (18-20°C).

The trays were viewed through an Olympus SZ-PT microscope. Photographs were taken using an Olympus Camedia C-3040ZOOM camera mounted on the microscope. On certain trays the glass cover slip becomes misty with condensation impairing visualisation so the glass cover slip was replaced with a new one prior to viewing under the microscope.

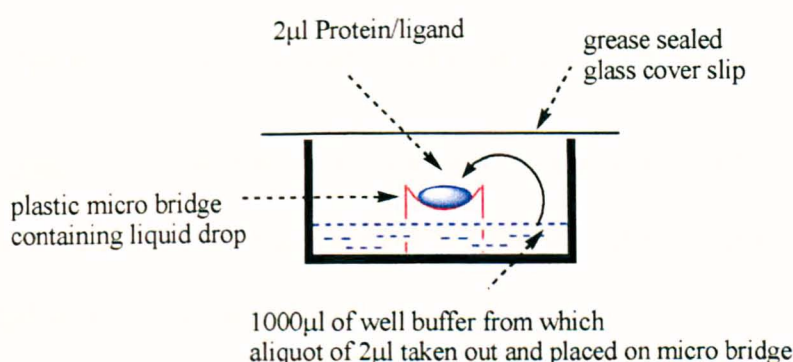


Figure 3.1 Schematic illustration of crystal tray set up.

3.8.1 Crystallisation trials using *Limulus* CRP

2µl of *Limulus* CRP at a concentration of 7.5 mg/ml in buffer (150mM NaCl, 50mM Tris, 10mM CaCl₂ pH 7.4) was placed on the microbridge, 2µl of well precipitant buffer were added. The tray was sealed and placed in a cupboard.

3.8.2 Crystallisation trials using *Limulus* CRP with *E. coli* LPS and polysaccharide

Trial 1. From stock concentration of *E. coli* LPS (9.3mM), 0.43µl was taken and added to the microbridge containing 2µl of *Limulus* CRP (7.5mg/ml) in buffer followed by adding 2µl of well precipitant buffer.

Trial 2. From stock concentration of *E. coli* LPS (9.3mM) 1.7µl was added to 8µl of *Limulus* CRP (7.5mg/ml) in buffer. The mixture was left at room temperature (18-20°C) overnight to incubate. 2µl of the incubated sample was then taken and placed on the microbridge to which 2µl of well precipitant buffer was added.

Trial 3. This trial is a repeat of trial 1 setting up 12 identical wells.

Trial 4. From stock solution of *E. coli* polysaccharide (11.7mM) 0.34µl was taken and added to the microbridge containing 2µl of *Limulus* CRP (7.5mg/ml) in buffer followed by adding 2µl of well precipitant buffer.

Trial 5. From stock solution of *E. coli* polysaccharide (11.7mM) 3.4µl was taken and was mixed with 20µl of *Limulus* CRP (7.5mg/ml) in buffer. The mixture was allowed to incubate at room temperature overnight. 2µl aliquot was then taken and placed on the microbridge to which 2µl of precipitant well buffer was added.

3.8.3 Crystallisation trials using *Limulus* CRP with *P. aeruginosa* LPS and polysaccharide

Trial 1. From stock concentration of *P. aeruginosa* LPS (5mM) 0.80µl was taken and added to the microbridge containing 2µl of *Limulus* CRP (7.5mg/ml) in buffer followed by adding 2µl of well precipitant buffer.

Trial 2. From stock concentration of *P. aeruginosa* LPS (5mM) 8.0µl was added to 20µl of *Limulus* CRP (7.5mg/ml) in buffer. The mixture was allowed to incubate at room temperature overnight. 2µl aliquot was then taken and placed on the microbridge to which 2µl of precipitant well buffer was added.

Trial 3. From stock concentration of *P. aeruginosa* polysaccharide (5.7mM) 7.0µl was added to 20µl of *Limulus* CRP (7.5mg/ml) in buffer. The mixture was allowed to incubate at room temperature overnight. 2µl aliquot was then taken and placed on the microbridge to which 2µl of precipitant well buffer was added.

Trial 4. 12.6µl of the polysaccharide was mixed with 36µl of *Limulus* CRP (7.5mg/ml) in buffer. The mixture was allowed to incubate at room temperature overnight. 2µl aliquot was then taken and placed on the microbridge to which 2µl of precipitant well buffer was added.

3.8.4 Crystallisation trials using *Limulus* CRP with *H. influenzae* Eagan wild type oligosaccharide

Trial 1. From stock concentration of the oligosaccharide (108mM) 0.4µl was mixed with 20µl of *Limulus* CRP (7.5mg/ml) in buffer. The mixture was allowed to incubate at room temperature overnight. 2µl aliquot was then taken and placed on the microbridge to which 2µl of precipitant well buffer was added.

Trial 2. 0.6µl of the oligosaccharide was mixed with 36µl of *Limulus* CRP (7.5mg/ml) in buffer. The mixture was allowed to incubate at room temperature overnight. 2µl aliquot was then taken and placed on the microbridge to which 2µl of precipitant well buffer was added.

Trial 3 and 4. 0.6 μ l of the oligosaccharide was mixed with 36 μ l of *Limulus* CRP (7.5mg/ml) in buffer. The mixture was allowed to incubate at room temperature overnight. 2 μ l aliquot was then taken and placed on the microbridge to which 2 μ l of precipitant well buffer was added.

Trial 5 and 6. 0.6 μ l of the oligosaccharide was mixed with 36 μ l of *Limulus* CRP (7.5mg/ml) in buffer. The mixture was allowed to incubate at room temperature overnight. 2 μ l aliquot was then taken and placed on the microbridge to which 2 μ l of precipitant well buffer was added.

Trial 7. 0.3 μ l of the oligosaccharide was mixed with 18 μ l of *Limulus* CRP (7.5mg/ml) in buffer. The mixture was allowed to incubate at room temperature overnight. 2 μ l aliquot was then taken and placed on the microbridge to which 2 μ l of precipitant well buffer was added.

3.9 Structural studies of rfhSP-D with *H. influenzae* Eagan oligosaccharide

Co-crystallisation trials with *H. influenzae* Eagan oligosaccharide at 15mM concentration was carried out with rfhSP-D protein solution (8mg/ml protein in 10mM Tris, 140mM NaCl, 10mM CaCl₂, pH 7.5) and precipitant buffer (16% PEG 4K in 0.1M Tris pH 7.5). Stock concentration of the oligosaccharide 85mM was made by dissolving 3.2mg of the oligosaccharide in 24.60 μ l of hSP-D buffer. For crystal trial 8.8 μ l from 85mM concentration was added to 42 μ l hSP-D protein in buffer.

3.10 Structural studies of rfhSP-D with *H. influenzae* Eagan 4A oligosaccharide

Co-crystallisation trials with *H. influenzae* Eagan 4A oligosaccharide at 30mM concentration was carried out with rfhSP-D protein solution (8mg/ml protein in 10mM

Tris, 140mM NaCl, 10mM CaCl₂, pH 7.5) and precipitant buffer (16% PEG 4K in 0.1M Tris pH 7.5). Crystals were soaked with ligand and prepared for cryocooling by addition of ligand (*H. influenzae* Eagan wild type oligosaccharide or Eagan 4A oligosaccharide) to MPD cryobuffers prepared using MPD in precipitant buffer. Successive 2 µl aliquots of increasing concentrations (5, 10, 15, 20%) of MPD/Ligand cryobuffer were added to the well at 10 minute intervals, followed by addition of a further 2µl aliquot of 20% MPD/Ligand cryobuffer, and an exchange of the resulting buffer with 20% MPD/Ligand cryobuffer.

3.11 Cryoprotection

Crystals that had been grown in the presence of ligand (co-crystals) and appeared suitable for X-ray diffraction data collection were prepared for cryocooling by adding 2-4µl of MPD cryoprotectant at increasing concentration of 5, 10, 15 and 20% in the respective well precipitant buffer. 2µl of each concentration (5-20%) was added in approximately 10 minutes intervals followed by adding a second 2µl aliquot of 20% MPD. A 10 µl volume from the drop on the micro bridge was then removed and replaced by 10 µl of 20% MPD (or in the case of rfhSP-D with *H. influenzae*, 20% MPD containing ligand 30mM). Cryoprotection was carried out on site immediately prior to X-ray data collection.

3.12 X-ray diffraction data collection and processing

X-ray diffraction experiments were carried out at Daresbury SRS, UK on station 14.1 using a wavelength of 1.488Å. Crystals were cryoprotected and mounted in a loop and flash frozen in a stream of nitrogen vapour at 100K. Crystal mounting on the goniometer was carried out approximately 30 minutes after the final exchange of cryobuffers.

The crystal to detector distance was set to capture the entire visible diffraction pattern, ranging from 95mm for 2.0Å data to a maximum of 170mm (3.0Å). For those crystals which produced a useful diffraction pattern (ordered, medium to high resolution), data was collected over 180° at 1°Δφ. Diffraction patterns were collected on an ADSC QR4CCD. Data processing was carried using MOSFLM version 6.2.3 and scaled using SCALA (CCP4). MOSFLM was used interactively to determine the orientation matrix, the crystal lattice and the unit cell from a single image prior to processing and scaling of the complete data set.

3.13 Structure solution and refinement (rfhSP-D / Eagan 4A oligosaccharide)

The previously reported native structure of rfhSP-D (PDB code 1pw9, Shrive *et al.*, 2003] was used as the model for rigid body refinement (CNS). The B factor was set to 12 and water molecules were omitted from the search model.

Electron density maps were calculated using CNS [Brunger *et al.*, 1998] and the CCP4 program suite [CCP4, 1994] including density modification (solvent flattening and histogram matching, but not NCS averaging). Model building of the structure was carried out using maximum likelihood refinement with CNS [Brunger *et al.*, 1998] alternated with rounds of manual model building with the program O [Jones *et al.*, 1991]. Topology and parameter files for ligand were obtained from the HIC-Up server [Kleywegt *et al.*, 2003]. The quality of the final structure was verified by MolProbity [Lovell *et al.*, 2003] and PROCHECK [Laskowski *et al.*, 1993].

Chapter 4
Hydrolysis & mass spectroscopy
of bacterial LPS:
Results & discussion

4.1 *E. coli* LPS

The aqueous supernatant fraction from the hydrolysis of *E. coli* LPS that was lyophilised using Savant Speed Vac produced 18.4 mg of the polysaccharide as a white powder whilst removal of the chloroform produced 4 mg of lipid A.

4.2 *P. aeruginosa* LPS

P. aeruginosa serotype 10 strain ATCC27316 LPS 50 mg was hydrolysed to produce 24.5 mg of white powder after lyophilisation of the aqueous supernatant. The amount of yield of *P. aeruginosa* was not determined due to the uncharacterised structure of this strain (see discussion).

4.3 *H. influenzae* Eagan and Eagan 4A LPS

Hydrolysis of *H. influenzae* Eagan (53mg) produced 33mg of the oligosaccharide as a white powder and 13.5mg of lipid A. Hydrolysis of Eagan 4A (50mg) produced 18mg of oligosaccharide as a white powder.

4.4 ESI-MS analysis of *E. coli* polysaccharide

ESI-MS analysis of *E. coli* polysaccharide failed to provide any meaningful data in either negative or positive ion mode.

4.5 MALDI-MS analysis of *E. coli* lipid A

The Lipid A fraction analysed using MALDI-MS produced multiple molecular ions (Figure 4.1). The molecular ion at m/z 1797.65 $[M-H^+]$ at 100% intensity is consistent with the molecular mass of lipid A 1798.00Da. The other peaks in the spectra could not be identified.

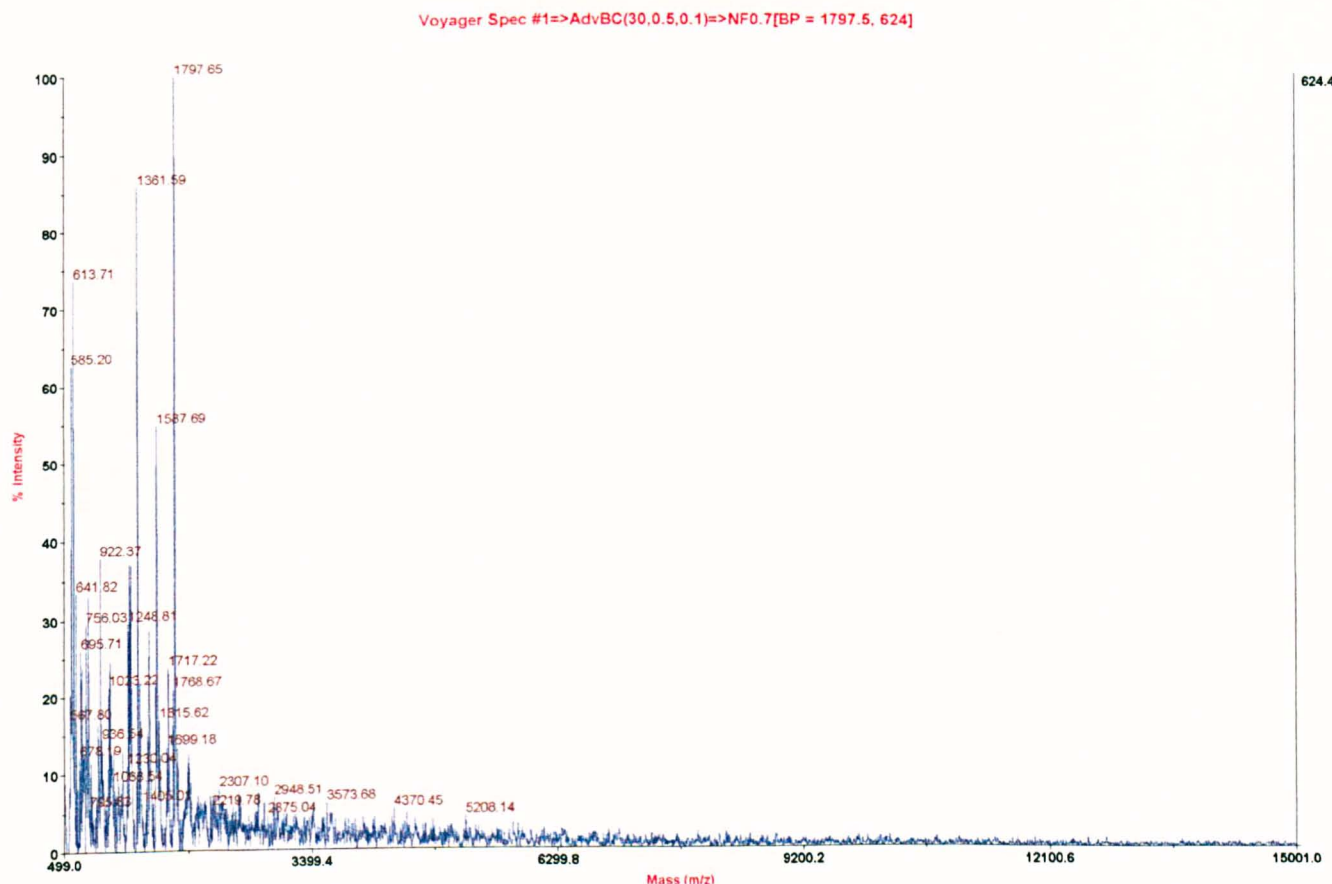


Figure 4.1 MALDI-MS of *E. coli* lipid A in the negative ion mode using 2,5-DHB as the matrix.

4.6 MALDI-MS analysis of *H. influenzae* Eagan 4A oligosaccharide

Analysis of the oligosaccharide by MALDI using the matrix 2, 5-dihydroxy benzoic acid in methanol in the negative ion mode identified a molecular ion at m/z 590.87 $[M-H]^-$ at 100% intensity (Figure 4.2). This is identified as the polysaccharide containing Kdo-Hep-Glc (Figure 4.14). The other molecular ions could not be identified.

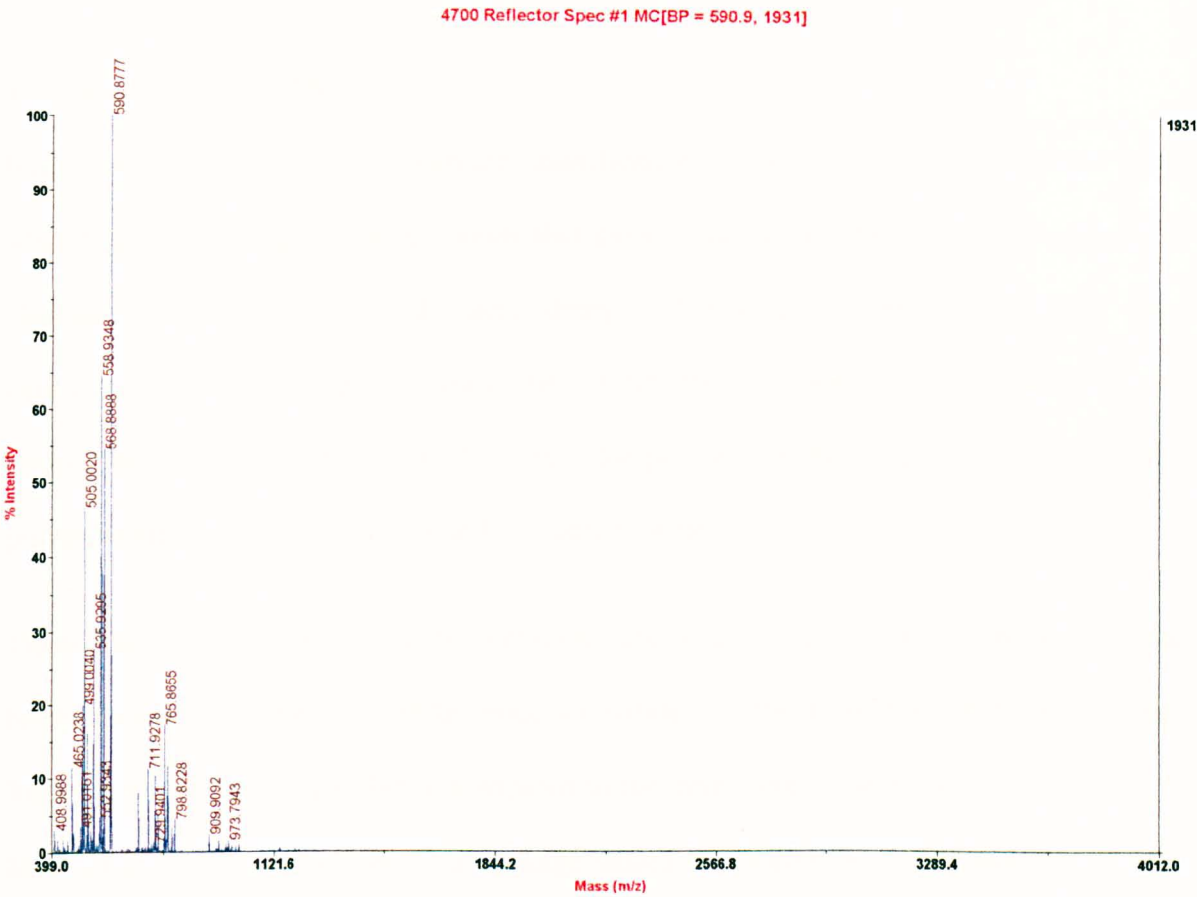


Figure 4.2 MALDI-MS of *H. influenzae* Eagan 4A oligosaccharide (Kdo-Hep-Glc) in the negative ion mode using 2,5-DHB matrix.

4.7 Discussion:

4.7.1 Hydrolysis of LPS

Hydrolysis is one method of chemical modification of the lipopolysaccharide and takes advantage of cleaving glycosidic bonds that are acid labile and the ease with which bond cleavage occurs depends on the acid strength. The method used in this work, acid hydrolysis is quite mild and only dissociates the glycosidic linkage between the lipopolysaccharide endotoxin (lipid A) from the polysaccharide molecule, resulting in the polysaccharide becoming more readily soluble in water.

There have been no reports of the remaining glycosidic bonds in the lipopolysaccharide being cleaved, however the PPEtn group substituted on the 4 position of the Kdo is also found to be extremely acid labile compared to the inner core substitution and according to Masoud and co workers (1997) this linkage would be hydrolysed during isolation of LPS from bacterial cultures using phenol-water extraction, followed by extensive dialysis of the aqueous phase. The subpopulation with this group in *H. influenzae* is found to account for more than 50% of the LPS when precipitated with ethanol from aqueous phase following from phenol-water extraction [Masoud *et al.*, 1997].

Acid hydrolysis using 1-2% aqueous acetic has been used to delipidate the ketosidic linkage joining lipid A to the polysaccharide via the Kdo-2 position. The mode of hydrolysis is assumed to proceed via S_N2 nucleophilic substitution producing exclusively the water soluble polysaccharide and the water insoluble lipid A (Figure 4.3). Under these conditions *anhydro* Kdo is formed from the rearrangement of Kdo with the loss of water after β -elimination of the phosphate or pyrophosphate group from Kdo O4 position [Danan *et al.*, 1982; Auzanneau *et al.*, 1991]. It is proposed that β -elimination of phosphate

is accompanied by pyranose ring opening. Subsequent ring closure leads to alternative cyclic structures, dominated by a 4,7 closure to form *anhydro* Kdo in the form of a substituted furanoid derivative with C4, which carries an extended substitution produced by the 4,7 closure, becoming a chiral centre (Figure 4.4 and 4.5)

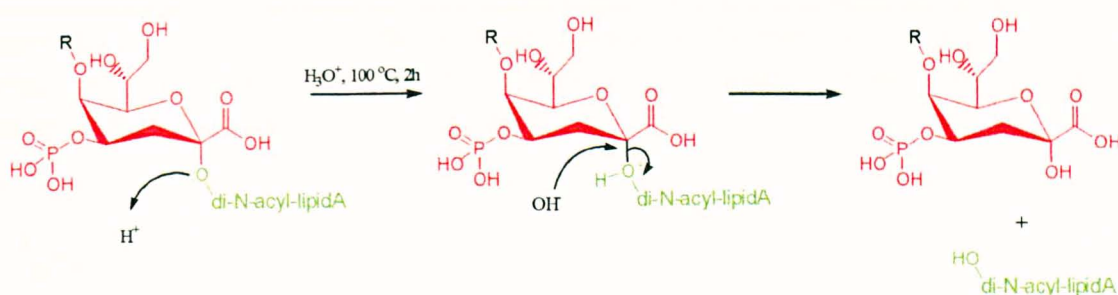


Figure 4.3 The mechanism of acid catalysed hydrolysis of LPS using 2% aqueous acetic acid at 100°C for 2.5 hours, producing oligo/polysaccharide and lipid A. α -Kdo (red) and lipid A (green) structure are attached via C₁ and C₆ respectively. Extended oligosaccharide (R) in the α anomeric position attachment is on α -Kdo at the C₅ position.

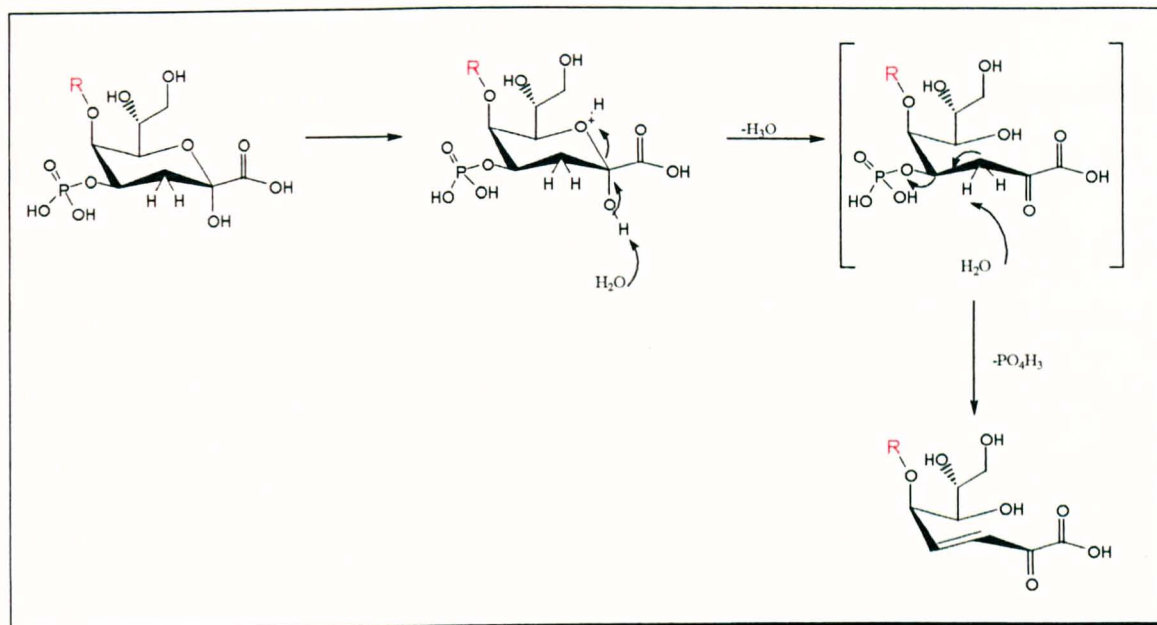


Figure 4.4 Further reaction of the acid labile α -Kdo in the hydrolysis process leading to olefinic acid (if $R=H$ or its derivatives) through β -elimination of the phosphate group [Danan *et al.*, 1982; Auzanneau *et al.*, 1991].

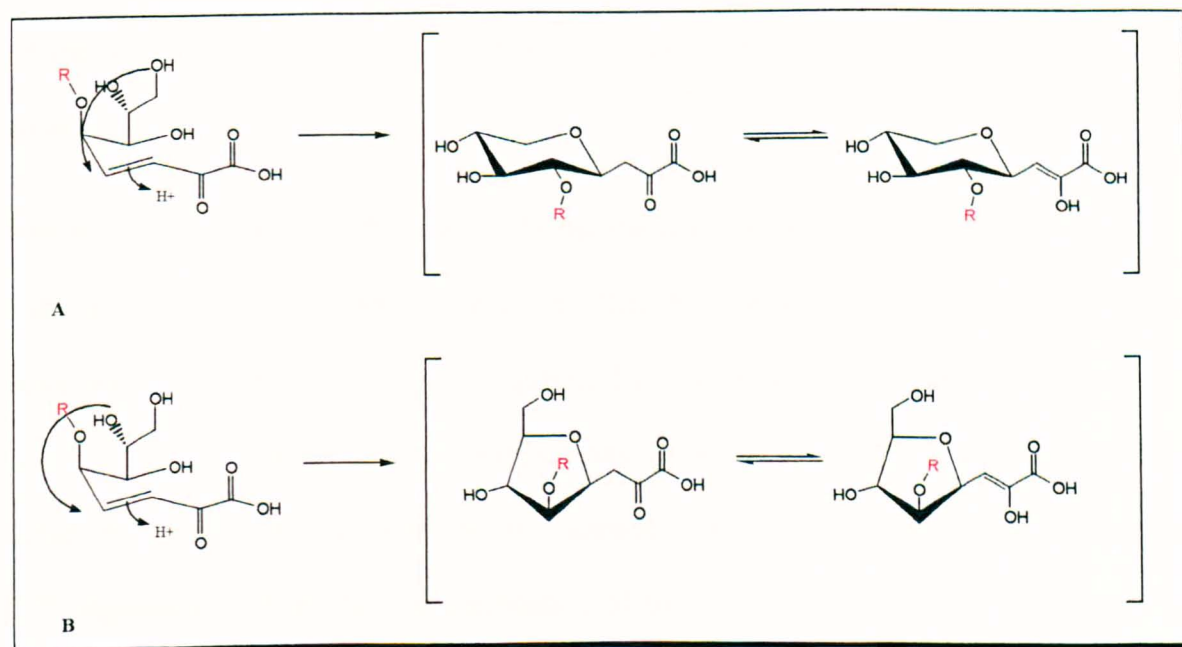


Figure 4.5 Olefinic acid (A and B) undergoes intramolecular addition of the alcohol at position A) 8-OH to the olefinic bond to form a 6 membered cyclic 4,8-*anhydro* Kdo derivative. B) 7-OH to the olefinic bond to form a 5 membered cyclic 4,7-*anhydro* Kdo derivative. Both forms of intramolecular cyclisation lead to β and α anomeric carbon at position 1 and further keto/enol tautomerism formation (brackets) [Danan *et al.*, 1982; Auzanneau *et al.*, 1991].

4.7.2 *E. coli* LPS hydrolysis and structural characterisation

The two fractions consisting of water soluble polysaccharide and water insoluble lipid A were separated out using two phase aqueous/organic separation method termed Bligh/Dyer mixture [Bligh and Dyer 1959; Zohou *et al.*, 1999]. Initially, after centrifugation at ultra high speeds the lipid A, which is insoluble in water and also denser than water, settled at the bottom of the centrifuge tube whilst the oligosaccharide remained in soluble form in the aqueous solution. This was easily pipetted out leaving the lipid A along with small amounts of aqueous solution remaining at the bottom of the tube. Adding a small amount of chloroform to dissolve the lipid A, and then adding an equal volume of water produced two phases, the top layer being aqueous and the bottom the chloroform fraction. The top layer was carefully removed and combined with the previous aqueous fraction for freeze drying, while the chloroform layer was removed under reduced pressure to give lipid A as an oil.

Freeze drying the aqueous fraction producing the crude oligosaccharide as a white powder and analysing the two fractions indicates that the hydrolysis went to completion as demonstrated by re-hydrolysing the aqueous fraction obtained from the hydrolysis of *E. coli* 0111.B4 LPS, which did not produce any further lipid A after ultra centrifugation. Since this project is not based on any accurate analytical or quantitative analysis of LPS/oligosaccharide, no further purification of the crude oligosaccharide as reported by other groups of workers was carried out, for example gel chromatography to separate the polysaccharide fractions [Masoud *et al.*, 1997].

The C₄ of the *E. coli* Kdo is substituted by a phosphate or PPEtn (ester bond) which is also acid labile and would be prone to hydrolysis under these conditions [Masoud *et al.*, 1997].

This would mean that it would be soluble in the aqueous fraction along with the polysaccharide and would constitute part of the white powder after hydrolysis and freeze drying.

Using chemical analysis, the structure of the *E. coli* 0111.B4 has been characterised but no MS structural analysis has been carried out. This particular strain, which exists as a heterogeneous population (Figure 4.6) [Morrison *et al.*, 1975], is characterised as smooth form containing O-polysaccharide antigen. This strain contains two dominant forms of LPS, termed LPS I containing 11 repeat units and LPS II containing 3 repeat units of the O-polysaccharide antigen (figure 4.6).

The repeat units are composed of five glycosidically bound saccharides units ($-\alpha\text{Col}-1\rightarrow6-\beta\text{GlcNAc}-1\rightarrow2\alpha\text{Glc}[1\rightarrow4\text{Col}]1\rightarrow2\alpha\text{Gal}$). The inner core contains 3 Kdo and 3 heptose unit with the outer core containing GlcNAc-Glc-Gal -Glc. Heterogeneous mixtures have been identified which arise from substitution from the heptose unit by either $\text{Glc}1\rightarrow4\text{GlcNH}_4$ or another heptose unit. The attachment of PEtn may occur on either of the heptoses of the inner core.

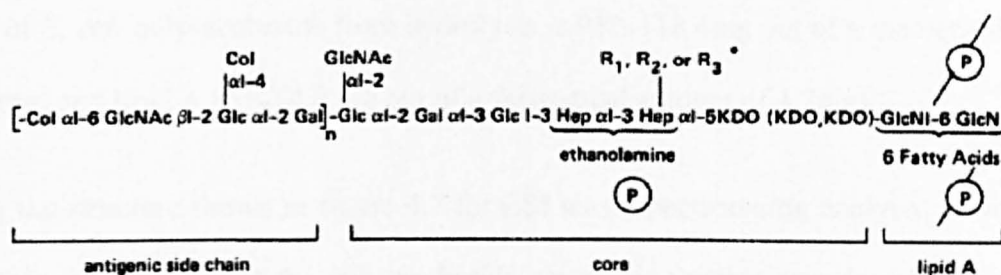


Figure 4.6 Structure of *E. coli* O111:B4 [Morrison *et al.*, 1975]. LPS I contains 11 repeat units (n) and LPS II form contains 3 repeat units of the antigenic side chain. Where R_1 , R_2 or R_3 could be Glc-GlcNAc or Hep.

According to the structure three acid labile units can be observed, the PPEtn substituted on the inner core, the Kdo-lipid A bond and the phosphate substitution on the Kdo-4 position. The crude polysaccharide fraction obtained may contain these small molecular weight molecules. The structure of the lipid A is the same with molecular weight of 1798Da [Qureshi *et al.*, 1983; Que *et al.*, 2000].

Since the LPS exist as a heterogeneous mixture, an accurate molecular weight could not be determined. The author has made an assumption that if the LPS I (12324.7Da) and LPS II (6022.3Da) are in equal amounts then taking the average weight would give a molecular weight of the native *E. coli* LPS of 9173.5 Da (these values are based on no substitution from the inner core). The yields of the products were calculated by using the formulae:

- Theoretical amount (g) = minimum number of moles of starting material x molecular weight of product
- Actual yield (%) = actual amount obtained (g) / theoretical amount (g)

According to the calculated molecular weights of the intact LPS and the polysaccharide the yield of *E. coli* polysaccharide from hydrolysis is 93% (18.4mg out of a theoretical amount of 20mg) and lipid A 85% (4.0 mg out of a theoretical amount of 4.7mg).

Using the structure shown in figure 4.7 for ESI mass spectrometric analysis, it was hoped to obtain the structure of the polysaccharide since this fraction was found to dissolve in deionised water and there would be no analytical analysis complication caused by the lipid A and the fatty acids. However no data could be obtained in either negative or positive modes and attempts at assigning structural information (Table 4.1, 4.2 and 4.3) on the *E. coli* polysaccharides obtained from ESI were not pursued further.

MS results proved very complicated due to too many variables in the native LPS. Other factors could include impurities or there may be multiple products of hydrolysis, i.e. since the LPS contains a heterogeneous mixture, cleavage of other glycosidic bonds may have occurred leading to various molecular weight fractions.

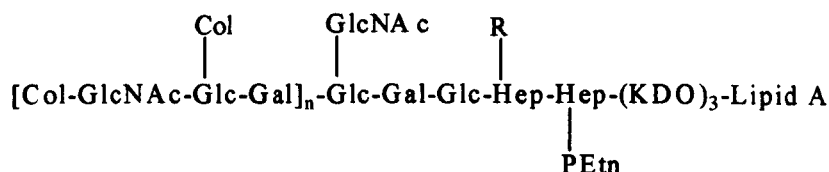


Figure 4.7 Structure of *E. coli* 0111.B4 used for ESI mass analysis. Considering R=O then the molecular weight of LPS I (n=11 subunits) is 12324.7 and LPS II (n=3 subunits) is 6022.3 Da.

Proposed composition	Calculated molecular weight (Da)
[(Col) ₂ -GlcNAc-Glc-Gal]	805.8
[(Col) ₂ -GlcNAc-Glc-Gal] ₃	2381.4
[(Col) ₂ -GlcNAc-Glc-Gal] ₃ - Glc-(GlcNAc)-Gal-Glc-(Hep) ₂ -PEtn-(KDO) ₃	4224.3
[(Col) ₂ -GlcNAc-Glc-Gal] ₃ - Glc-(GlcNAc)-Gal-Glc-(Hep) ₂ -PEtn-(KDO) ₃ (R=Hep)	4416.5
[(Col) ₂ -GlcNAc-Glc-Gal] ₃ - Glc-(GlcNAc)-Gal-Glc-(Hep) ₂ -PEtn-(KDO) ₃ (R= Glc-GlcNAc)	4589.7
Glc-Gal-Glc-Hep-Hep-PEtn	1218.3

Table 4.1 Proposed structures of *E. coli* polysaccharide possible after hydrolysis of native LPS and the possible *m/z* values scanned in the positive and negative ion mode. Molecular mass of Hex (Glc/Gal) 180.2, Hep 210.2, P 98.0, Kdo 238.2, PEtn 126.1, Col (colitose) 148.1, GalNAc/GlcNAc 221.2 and lipid A 1798.0 Da. R could be Glc-GlcNAc or Hep.

* (loss of water for glycosidic bond formation 18.0 Da).

m/z values positive ionisation	Abundance (%)	Calculated molecular weight Da ([M+H] ⁺	Calculated molecular weight Da ([M+2H] ²⁺	Calculated molecular weight Da ([M+3H] ³⁺
86.2	100	85.2	170.4	255.6
204.2	80	203.2	406.4	609.6
103.2	55	102.2	204.4	306.6
573.0	100	572.0	1144.0	1716
589.6	90	588.6	1177.2	1765.8
668.6	80	667.6	1335.2	2002.8
759.4	50	758.4	1516.8	2275.2
528.4	100	527.4	1054.8	1582.2
1212.6	100	1211.6	2423.2	3634.8
1128.9	55	1127.9	2255.8	3383.7
1284.2	50	1283.2	2566.4	3849.6
2150.7	100	2149.7	4299.4	6449.1

Table 4.2 ESI-MS in the positive ion mode. Observed m/z values of *E. coli* polysaccharide and the percentage abundance and the possible molecular weights calculated.

m/z values negative ionisation	Abundance (%)	Calculated molecular weight Da ([M-H] ⁻ *	Calculated molecular weight Da ([M-2H] ²⁻ *	Calculated molecular weight Da ([M-3H] ³⁻ *
79.1	100	80.1	160.2	240.3
97.1	60	98.1	196.2	294.3
593.5	60	594.5	1189.9	1783.5
608.3	100	609.3	1218.6	1827.9
1364.3	100	1365.3	2730.6	4095.9
2231.6	90	2232.6	4465.2	6697.8

Table 4.3 ESI negative ion mode. Observed m/z values of *E. coli* polysaccharide and the possible molecular weight calculated.

* theoretical values assigned to deduce possible ionisation that would lead to proposed structure.

Due to the complexity of the structure and the multiple fragments obtained, structural information could not be obtained based on MS data alone and would most probably require purification steps and further investigation using high field NMR if the structure of *E. coli* 0111.B4 needs to be elucidated. Since it not a strict criteria to determine the exact structure, no further work was carried out and the crude polysaccharide fraction was used in various crystal trials.

Analysis of the oily residue by MALDI-MS identified an ion at m/z 1797.65 [M-H]⁻ at 100% intensity, this is consistent with the structure of lipid A with molecular mass of 1798 Da (Figure 4.8). The structure of lipid A from the mass spec data is consistent with the literature reported structure of *E. coli* and *Salmonella typhimurium* lipid A [Qureshi *et al.*, 1983; Que *et al.*, 2000], which consists of hexa-acylated disaccharide of glucosamine that

is β , 1'-6 linked. Two phosphate groups at position 1 and 4' position and the fatty acids are two N-acyl linkages at position 2 and 2' with β -hydroxymyristoyl residues and three O-acyl linkages at position 3, 4, and 6' with 3-myristoxymyristoyl, β -hydroxymyristoyl, lauroyl and myristoyl groups attached (Figure 4.8). The molecular mass also agrees with literature values of 1798.8 Da [Qureshi *et al.*, 1983; Que *et al.*, 2000].

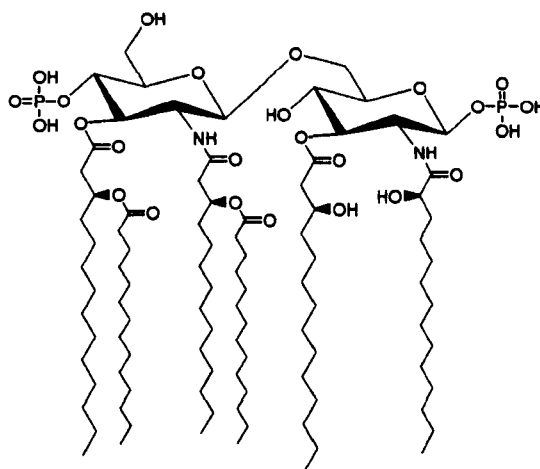


Figure 4.8 Structure of lipid A with molecular mass of 1798.8 containing two amide linked and four ester linked fatty acid residues [Qureshi *et al.*, 1983; Que *et al.*, 2000].

4.7.3 *P. aeruginosa* LPS hydrolysis and structural characterisation

Based on the results obtained from *E. coli* hydrolysis it is assumed that the hydrolysis of *P. aeruginosa* has been successful and that the polysaccharide has been obtained as well as the lipid A. it would have been complicated to assigned a structure to the *P. aeruginosa* based on MS alone because there is no reference to this strain ATCC27316 which is a serotype 10. One assumption on the possible structure which mostly involves the inner core is based on the reported structure of another serotype, serotype 10 [Bystrova *et al.*, 2004]. This particular strain is found to exist in three forms; smooth, semi smooth and

rough. The structures contain the unique sugar rhamnose and is always found substituted with a carbamoyl group substituted at the O6 of the inner core heptose unit. The inner core is also substituted with phosphates and PPEtn on the inner heptose. Hydrolysis of the inner core of *P. aeruginosa* has been found to occur at both Kdo units [Bystrova *et al.*, 2004].

4.7.4 *H. influenzae* Eagan LPS hydrolysis and structural characterisation

The structure of the *H. influenzae* Eagan is well documented from ESI MS analysis as well as high field NMR studies [Masoud *et al.*, 1997]. This is a rough chemotype with no antigenic side chains and containing a single Kdo unit attached to the lipid A containing two fatty acid residues. The Kdo is substituted at the O-4 position by either P or PPEtn, which gives rise to subpopulations of LPS. The latter is found to be extremely acid labile and is easily cleaved in very mild conditions. However the abundance of these two phosphate substituted forms is dependent on the isolation of the LPS. The subpopulation with Kdo-4-P--PEtn is found to be completely absent in isolation involving phenol-water extraction followed by extensive dialysis of the aqueous phase, suggesting this group is lost during isolation procedures. The subpopulation with this group is found to account for more than 50% of the LPS when precipitated with ethanol from aqueous phase following phenol-water extraction. The exposed portion of the LPS consists of Glc and Gal in various molar ratios. Proximal Hep I from the inner core is found substituted by α -Glc at the C₄ position of the Hep unit, although extensions through addition of PC via *licI* gene or due to the presence and activity of *lex2* gene have been described [Deadman *et al.*, 2009]. Hep II contains a PEtn at the O-6 position (Figure 1.34, chapter 1) that is not found to be acid labile, while Hep III is substituted by PC, a common feature found at O-6 of a terminal β -D-Gal residue off Hep III strain Eagan. The fully extended glycoform with Gal-Gal-Glc-

Glc extension from inner HepII (Figure 4.9) is controlled by *lic2* gene [Deadman *et al.*, 2009; Masoud *et al.*, 2003; 1997].

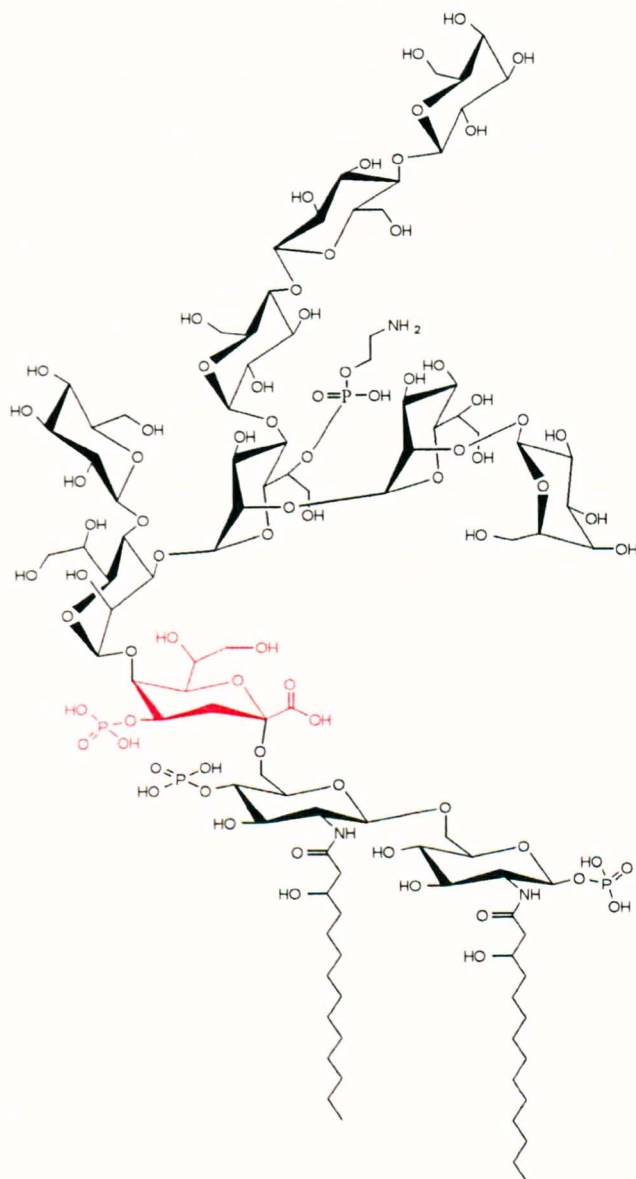


Figure 4.9 Structure of the fully extended *H. influenzae* Eagan wild type LPS deduced from NMR, NOESY connectivities [Masoud *et al.*, 1997]. The lipid A shown is attached to the core oligosaccharide via a single Kdo (red) with 1-5 glycosidic linkage. The oligosaccharide contains a PPEtn substituted from the inner core HepII and the subpopulation with phosphate substituted from Kdo-4. The molecular weight calculated is 2779.40 Da. Drawn using CS Chem. Draw Pro. Version 4.5.

Analyses of *H. influenza* Eagan by ESI MS have indicated the existence of *anhydro* Kdo from hydrolysis using the conditions used here [Masoud *et al.*, 1997; Mansson *et al.*, 2001; Mansson *et al.*, 2002; Risberg *et al.*, 1997; Yildirim *et al.*, 2003; Landerholm *et al.*, 2004; Schweda *et al.*, 2000]. In the case of R-type LPS, mild acid hydrolysis has been found to cause modifications to the Kdo at the reducing terminus. This is not the case for smooth LPS which contains high molecular weights [Masoud *et al.*, 1994; McNicholas *et al.*, 1987; Phillips *et al.*, 1990; Auzanneau *et al.*, 1991; Melaugh *et al.*, 1992].

Hydrolysis of the Eagan LPS produced a white powder which is the oligosaccharide and no further structural characterisation was carried out. The oligosaccharide after hydrolysis would contain the pentameric *anhydro* Kdo ring form (Figure 4.10).

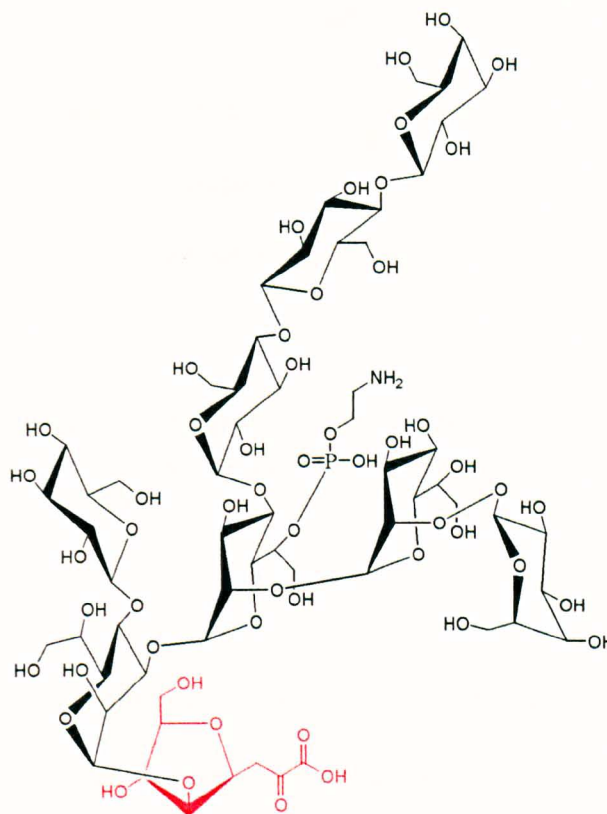


Figure 4.10 Structure of the fully extended native *H. influenzae* Eagan LPS containing the pentameric *anhydro* Kdo ring (red) with molecular weight 1746.4 Da. Drawn using CS Chem. Draw Pro. Version 4.5.

4.7.5 *H. influenzae* Eagan 4A LPS hydrolysis and structural characterisation

The Eagan 4A is much smaller than the Eagan wild type LPS containing only a single terminal Glc that is substituted from Hep I, this is coded by the *rfaF* gene (Figure 4.11). Two subpopulations of the Eagan wild type LPS [Masoud *et al.*, 1997] are identified in which the Kdo carries a 4-phosphate or its pyrophosphate analogue (P--PEtn), and hence it could well be possible that the Eagan 4A will also show this heterogeneity at this position, however this moiety will also have been lost during phenol-water extraction followed by extensive dialysis of the aqueous phase.

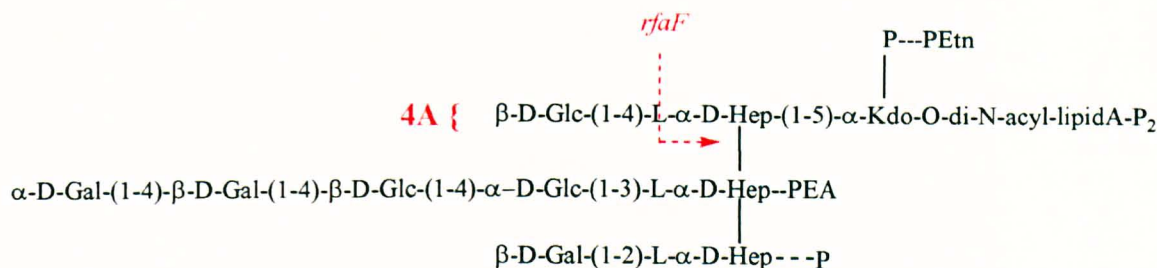


Figure 4.11 Schematic illustration of the glycosidic linkages in *H. influenzae* Eagan wild type. The Eagan 4A is a gene construct by gene *rfaF* between the inner core HepI and HepII (red arrow) [Masoud *et al.*, 1997].

Deadman and co workers (unpublished results, IMM, Oxford) who have purified and supplied the Eagan 4A, demonstrate further chain elongation from the Hep I position. Glc and Gal moieties are possible (Figure 4.12), however this extension is not normally found because the gene coding for this pathway, *lex2*, is in most cases not active or “turned off”. Substitution of PCho on the Glc position of Hep I is also described [Lysenco *et al.*, 2000], but this substitution is not described for the Eagan wild type strain [Masoud *et al.*, 1997], but has been shown to be located on the terminal Glc off Hep III [Masoud *et al.*, 1997]. Hydrolysis of this LPS also afforded the oligosaccharide (Figure 4.13) as the crude white powder and no further purification was carried out prior to MS analysis.

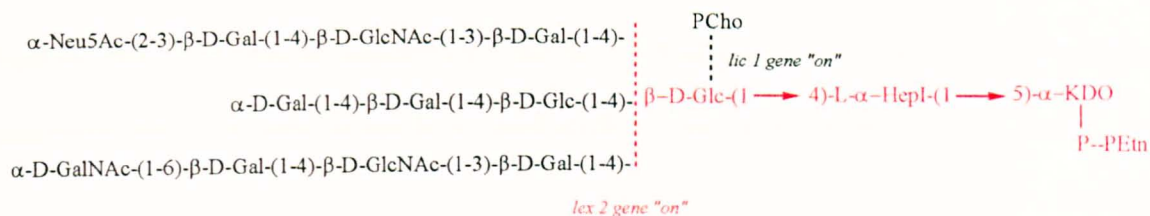


Figure 4.12 Extended glycoform of Eagan 4A (red structure), red dashed line shows extension of glycoform coded by *lex2* gene. Mary Deadman and co workers (unpublished results, IMM, Oxford).

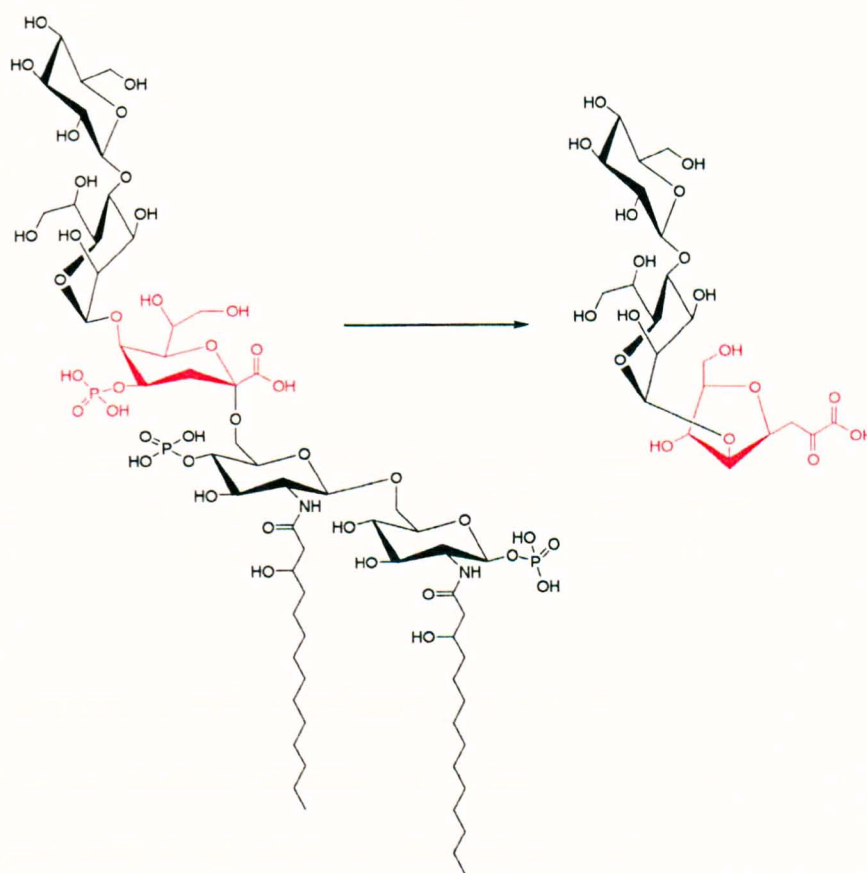


Figure 4.13 Hydrolysis of Eagan 4A using 2% aqueous acetic acid at 100°C to give oligosaccharide with the *anhydro* Kdo ring form (574.48 Da).

Mass analysis using MALDI MS in the negative ion mode on the crude oligosaccharide fraction produced several peaks. Since the 4A LPS is a small simple molecule proposed structures that could have arisen from the hydrolysis are shown in table 4.4..

Proposed composition	Molecular mass calculated (Da)	Observed ion [M-H] ⁻ and intensity (%)
lipid A-KDO-Hep-Glc	1527.4	-
KDO-Hep-Glc	592.4	590.8
<i>anhydro</i> KDO-Hep-Glc	574.40	-
Lipid A-KDO-P-Hep-Glc	1607.4 *	-
KDO-P-Hep-Glc	672.4	-
KDO-P--PEtn-Hep-Glc	780.5	-
KDO-Hep	430.3	-

* starting material.

Table. 4.4 Negative ion MALDI-MS of Eagan 4A. Molecular mass calculated using hydrated mass units, Hex (Glc) 180.2, Hep 210.2, Kdo 238.2, *anhydro* Kdo 220.2, P 98.0, PEtn 126.08, lipid A 953.0 (molecular mass determined using Chem. Draw Ultra Pro 4.5).

The mass spectrum reveals a molecular ion m/z 590.8 [M-H]⁻ at 100% intensity corresponds to the structure Kdo-Hep-Glc (Figure 4.14). Various other structures have been proposed (table 4) to which no observed ion could be related or found in the spectrum. However there are two other peaks relatively high compared to the molecular ion at 558.9 and 568.0 m/z , but no structure could be assigned to these observations.

There are no peaks corresponding to the parent molecule Glc-Hep-Kdo(P)-O-di-N-acyl-lipid A-P₂ or the subgroup Glc-Hep-Kdo(PPEA)-O-di-N-acyl-lipid A-P₂. The structure also shows no indication of Kdo-P--PEtn-Hep-Glc (or the phosphate form) that is commonly reported in the literature and LPS bearing these non carbohydrate substitution on the 4 position on Kdo are found to be extremely acid labile [Masoud *et al.*, 1997]. In this analysis mass fragments of inorganic phosphate or pyrophosphate were not scanned and would require an MS-MS analysis for such small mass fragments.

Analysis of the MS data information (table 4.4) show no evidence of peaks corresponding to the proposed structures (Figure 4.13) containing an *anhydro* Kdo form with oligosaccharide attached to Kdo-5 position. The peak at 590.80 clearly indicates the glycoform containing β -Glc-(1-4)- α -Hep-(1-5)- α -Kdo (Figure 4.14).

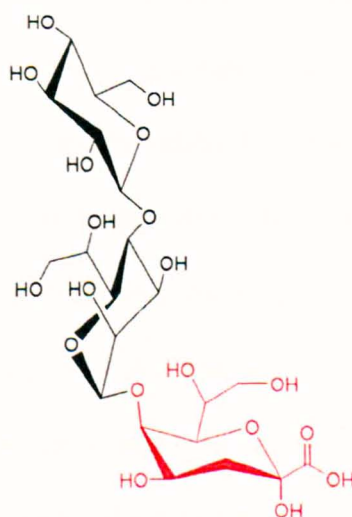


Figure 4.14 Structure of Egan 4A as deduced from the MALDI-MS after hydrolysis.

It is not clear what mechanism would have operated to produce this structure since it is well known for Kdo with phosphate substituent at Kdo-4 position to form *anhydro* Kdo by β -elimination of the phosphate group [Danan *et al.*, 1982].

In the case of Eagan 4A the mechanism involving ring opening and intra molecular rearrangement may not have happened but may have followed a different path without ring opening and without forming the *anhydro* Kdo.

4.8 Summary

Structural information of the *E. coli* and *P. aeruginosa* polysaccharide could not be determined using MALDI and ESI possibly due to impurities and no further purification or characterisation was carried out in order to determine the structures of the polysaccharide obtained from hydrolysis. MALDI-MS (Figure 4.2) on the *H. influenzae* mutant 4A oligosaccharide indicates the presence of glycoform containing β -Glc-(1-4)- α -Hep-(1-5)- α -Kdo (Figure 4.14). With reference to the structure of the Kdo, our finding based on the MALDI-MS is contradicting the work by others [Masoud *et al.*, 1997; Schweda *et al.*, 2000; 2007] who have reported the formation of the five membered *anhydro* Kdo structure obtained from the hydrolysis, which proceeds via the β -elimination of the phosphate group (Figure 4.4 and 4.5) [Danan *et al.*, 1982; Auzanneau *et al.*, 1991]. However the MALDI-MS result alone can not be conclusive and may require further purification along with other analysis such as NMR to establish the structure of the Kdo obtained.

Chapter 5

Crystallisation of bacterial LPS & polysaccharide

with *Limulus* CRP:

Results & Discussion

5.1 Isolation of *Limulus* CRP from plasma using affinity chromatography

The elution of CRP (and SAP) by affinity chromatography which isolates and purifies the proteins is judged from the elution profile obtained from the UV absorbance at 280 nm. The first elution using 10 mM PC in calcium wash buffer releases the CRP, followed by SAP which is eluted by 30 mM PE in calcium wash buffer (Figure 5.1).

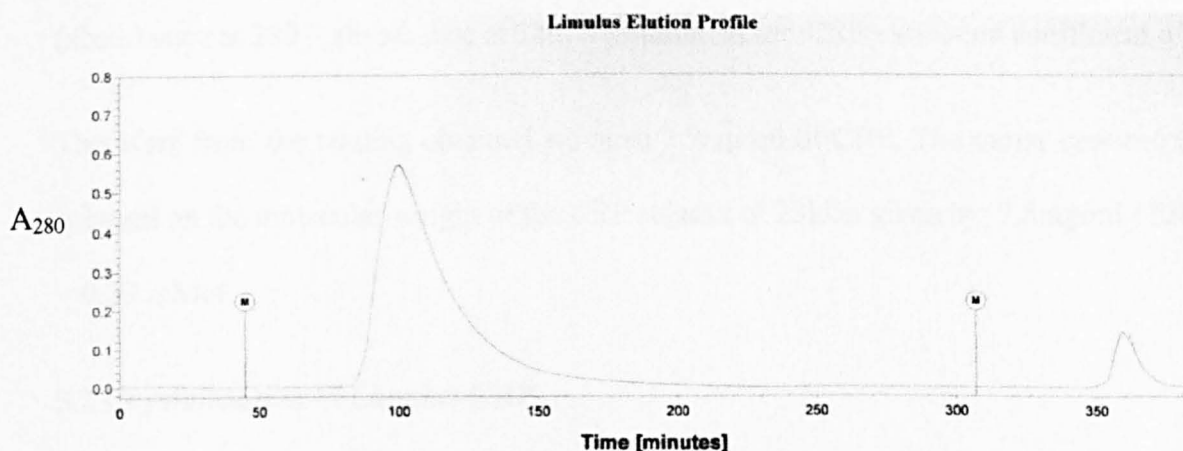


Figure 5.1 Elution profile of *Limulus* pentraxins. The first mark (M) indicates elution of CRP from approximately 83 min to 180 min from addition of 10mM PC in calcium wash buffer. Second mark indicates elution of SAP from approximately 350 min to 369 min using 30mM PE. The flow rate was at 0.5 ml/min.

The detection of CRP from the elution profile indicates it is coming off at approximately 83 minutes from washing with PC buffer and peaking at 100 minutes and then gradually decreasing. The SAP from the column takes approximately 100 minutes from the baseline of CRP elution. By looking at the two fractions on the graph (Figure 5.1) the CRP is eluted in a large quantity with a higher UV absorbance factor compared to SAP. 1ml solution of native CRP purified from 7.5 ml of *Limulus* plasma by PC column contained 4.27 mg/ml of CRP.

1ml solution containing 4.27mg/ml of CRP was concentrated to 200 μ l and a 1 in 10 dilution was analysed using a Quivet spectrophotometer and measured at UV 280 and 320 nm. The absorbance at 260nm is the correction for residual RNA from recombinant protein, and the absorbance at 320nm is the turbidity correction. The concentration of protein in the sample is calculated for 200 μ l using the formulae;

(absorbance at 280 – absorbance at 320) x dilution factor /CRP extinction coefficient of 1.5

Therefore from the reading obtained we have 7.5mg/ml of CRP. The molar concentration is based on the molecular weight of the CRP subunit of 23kDa given by; 7.5mg/ml / 23kDa = 0.33 mMol.

5.2 Crystallisation of *Limulus* CRP

Crystallisation of native *Limulus* CRP produced very small crystals after a few weeks that could be seen in wells A1, A2, B2, C1 and C3 (Table 5.1).

		Increase in CaCl ₂ concentration (mM) \longrightarrow		
Increase in PEG Concentration (%) \downarrow		1	2	3
	A	8mM CaCl ₂ 7% PEG 6K	10mM CaCl ₂ 7% PEG 6K	12mM CaCl ₂ 7% PEG 6K
	B	8mM CaCl ₂ 9% PEG 6K	10mM CaCl ₂ 9% PEG 6K	12mM CaCl ₂ 9% PEG 6K
	C	8mM CaCl ₂ 11% PEG 6K	10mM CaCl ₂ 11% PEG 6K	12mM CaCl ₂ 11% PEG 6K

Table 5.1 Conditions used in the crystallisation of *Limulus* CRP at a concentration of 7.5mg/ml. Small irregular crystals could be seen in the highlighted wells.

5.3 Crystallisation using *Limulus* CRP with *E. coli* LPS and polysaccharide

Trial 1. A single condition using 50mM MES pH7, 9% PEG 6K and 10mM CaCl₂ was investigated, based on the formation of small crystals observed from the crystallisation of native *Limulus* CRP in table 5.1. After three days the drop on the microbridge appeared opaque under the microscope and was agitated. After a further three days from agitation, the drop on the microbridge appeared to show a solid precipitate with two “toblerone” shaped crystals (Figure 5.2). The largest crystal measured 0.2 x 0.18mm using x2 magnification with the Olympus SZ-PT microscope.



Figure 5.2 Toblerone shaped crystals from trial 1 using *Limulus* CRP with *E. coli* LPS obtained using 50mM MES pH 7.0, 9% PEG 6K and 10mM CaCl₂. The largest crystal measures 0.2 x 0.18mm (x4 magnification using Olympus SZ-PT microscope)

The crystal was cryoprotected by successive addition using 2μl of 5, 10, 15 and 20% MPD. The crystal was tested for X-ray diffraction on 12th November 2005 at Daresbury SRS on station 14.1. Prior to data collection 10μl of the drop was exchanged for 10μl of cryoprotectant containing 20% MPD. Data was collected through 180° using 1°Δφ. The list of reflections was calculated from the diffraction images using MOSFLM in order to index and integrate, followed by Sortmtz, Scala and Truncate.

Trial 2. Very small irregular crystals were obtained using conditions 0.1M HEPES pH7.5, 8% Ethylene glycol, 10% PEH 8K and also using 50mM HEPES pH7.3, 10mM CaCl₂ and 9% PEG 6K.

Trial 3. Trial 3 sets out to repeat the conditions that produced the crystals from trial 1 and to obtain the same crystal form (Figure 5.2) by making 12 identical wells with the same conditions and reagents (50 mM MES pH 7.0, 9% PEG 6K and 10 mM CaCl₂). This trial proved unsuccessful at obtaining the same shape and size of crystal as trial 2.. There was a striking difference in the structures of crystals that were observed, such as needles, flat rectangles or torpedo shapes that could be seen in nearly all 12 wells, except well B3 which contained a single 4 sided rod shaped crystal that looked similar to the crystal obtained in trial 1 (Figure 5.3).

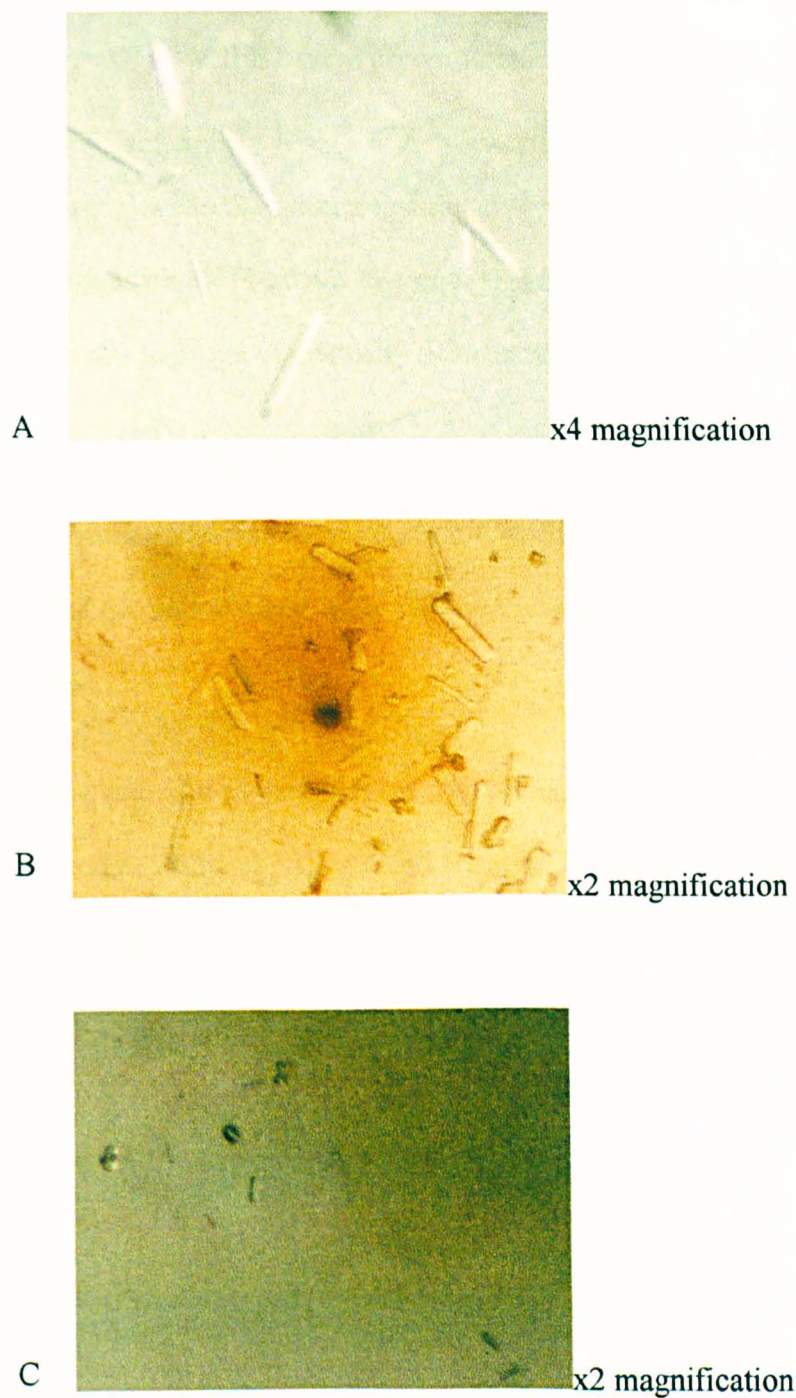


Figure 5.3 Crystals produced in trial 3 using the same conditions and reagents as trial 1 (50mM MES pH 7.0, 9% PEG 6K and 10mM CaCl_2). A) needle shaped crystals in well B1 B) flat rectangular crystals in well B2 and C) flat and tobermorite shaped crystal in well B3 (x2 magnification using Olympus SZ-PT microscope).

Crystals seen in wells B1 and B3 were taken to Daresbury SRS on 9th December 2006 for X-ray diffraction studies. Crystals in well B1 were cryoprotected using PEG 400, but this seemed to gradually decompose the crystal before any X-ray diffraction studies could be carried out, however the crystal in well B3 was successfully cryoprotected using increasing strength of MPD at 10 minutes intervals (5-20%). The crystal diffracted to 3Å but no data could be collected as the crystal was found to gradually deteriorate.

Trial 4. Trial 4 looked at crystallising the CRP (7.5mg/ml) in buffer with *E. coli* polysaccharide. The crystallisation procedure and conditions are identical to trial 1 using 50mM MES pH 7.0, 9% PEG 6K and 10mM CaCl₂. Only very small grainy crystals with no shape or size were obtained.

Trial 5. Various triangular bi-pyramidal shaped, flat rectangular and multi sided rod shaped crystals could be seen clustered in wells A1, A2, A3 and B1 after 4 weeks of setting up the tray conditions (Table 5.2 and Figure 5.4).

	1	2	3	4	5	6
A	0.17M Li ₂ SO ₄ 12% PEG 2K 0.1M Tris pH 7	0.2M (NH ₄) ₂ SO ₄ 30% PEG 8K	0.2M (NH ₄) ₂ SO ₄ 30% PEG 4K	0.5M (NH ₄) ₂ SO ₄ 0.1M Hepes pH 7.5 30% v/v MPD	1.6M (NH ₄) ₂ SO ₄ 0.1m Hepes pH 7.5 0.1M NaCl	10% PEG 8K 8% ethylene glycol 0.1M Hepes pH 7.5
B	10% PEG 6K 8% MPD 0.1M Hepes pH 7.5	10% PEG 6K 10mM CaCl ₂ 10mM NaCl 50mM MES pH 7.0	50 mM MES pH 7 9%PEG 6K 10mM CaCl ₂			

Table 5.2 Precipitant well buffer investigated in trial 5 using *Limulus* CRP with *E. coli* polysaccharide. Highlighted wells indicate growth of crystals.

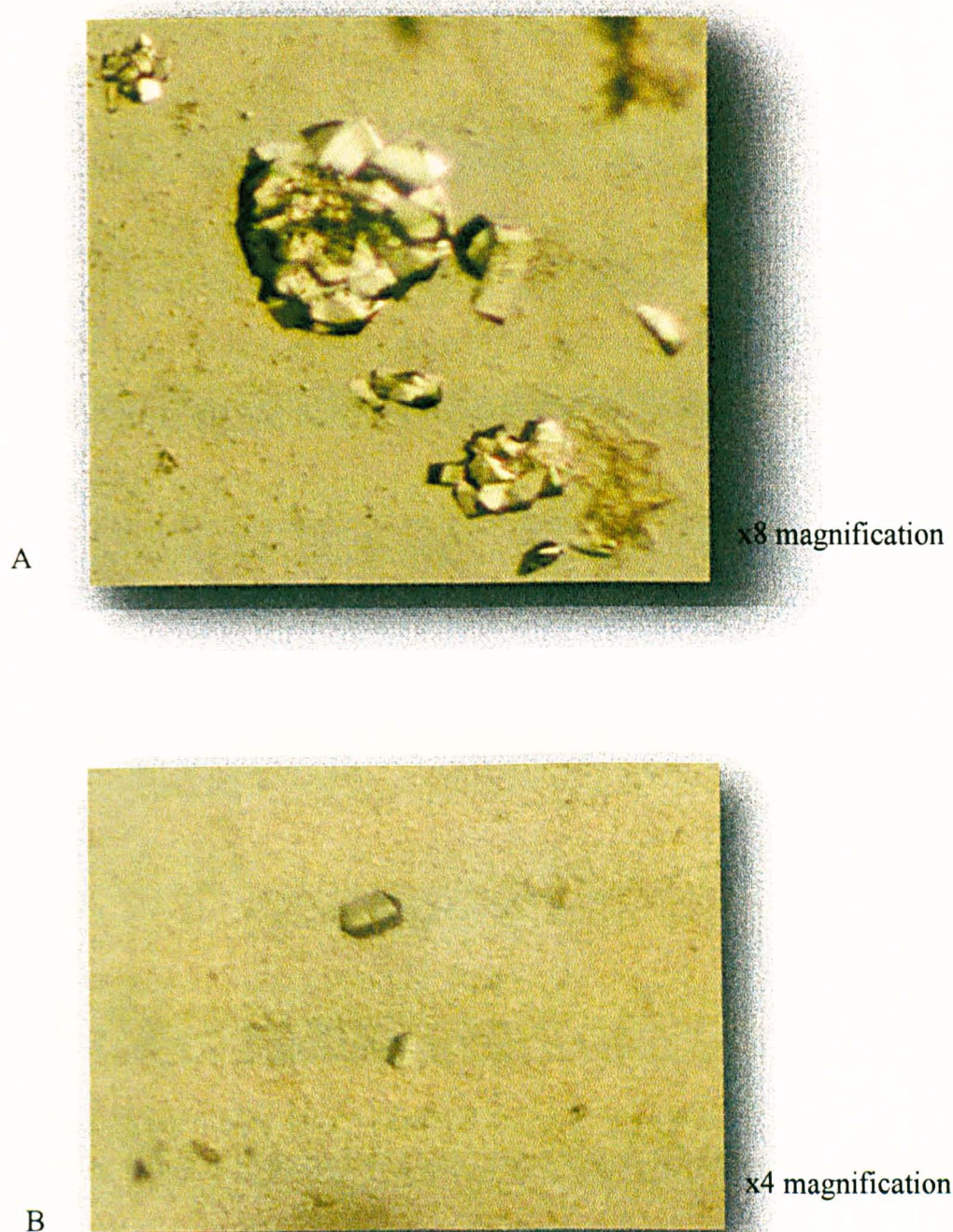


Figure 5.4 Crystals produced in trial 5 using *Limulus* CRP with *E. coli* polysaccharide. A) triangular bi-pyramidal shaped crystals seen in wells A1, A2, A3 and B) multi sided rod shaped crystal seen in well B1 (x4 magnification using Olympus SZ-PT microscope).

5.4 Crystallisation using *Limulus* CRP with *P. aeruginosa* LPS and polysaccharide

Trial 1. Very small irregular crystals could be seen using the conditions 50mM MES pH7, 9% PEG 6K and 10mM CaCl₂.

Trial 2. No crystals could be seen using *Limulus* CRP with *P. aeruginosa* LPS using the conditions shown in table 5.2.

Trial 3. Crystallisation trials using a series of conditions (outlined in the table 5.3) only produced very small irregular crystals in wells A2 and in well B3 torpedo shaped crystals (Table 5.3 and Figure 5.5).

	1	2	3	4	5	6
A	0.17M Li ₂ SO ₄ 12% PEG 2K 0.1M Tris pH 7	0.2M (NH ₄) ₂ SO ₄ 30% PEG 8K	0.2M (NH ₄) ₂ SO ₄ 30% PEG 4K	0.5M (NH ₄) ₂ SO ₄ 0.1M Hepes pH 7.5 30% v/v MPD	1.6M (NH ₄) ₂ SO ₄ 0.1m Hepes pH 7.5 0.1M NaCl	10% PEG 8K 8% ethylene glycol 0.1M Hepes pH 7.5
B	10% PEG 6K 8% MPD 0.1M Hepes pH 7.5	10% PEG 6K 10mM CaCl ₂ 10mM NaCl 50mM MES pH 7.0	50 mM MES pH 7 9% PEG 6K 10 mM CaCl ₂	50 mM MES pH 7 9% PEG 6K		

Table 5.3 Crystallisation conditions for *Limulus* CRP with *P. aeruginosa* LPS in trial 2. Highlighted conditions indicate formation of small irregular and torpedo shaped crystal from trial 3.

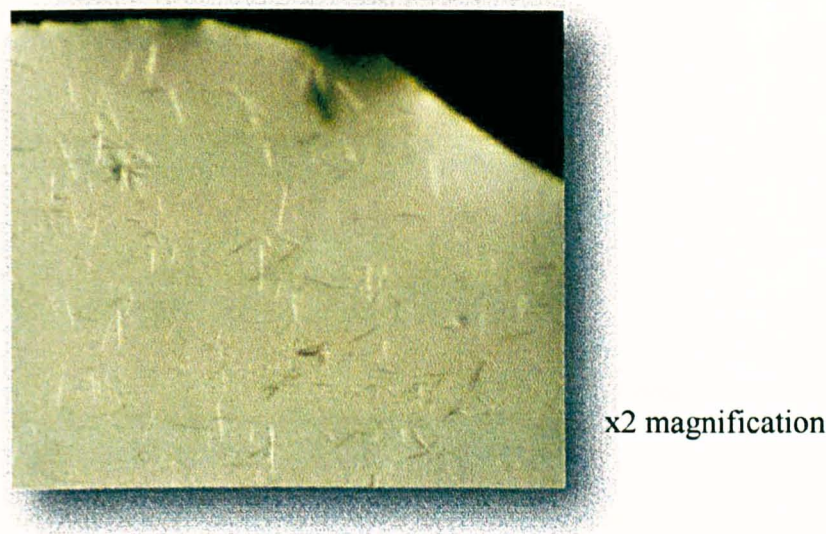


Figure 5.5 Torpedo shaped crystals seen in well B3 in trial 3 using *Limulus* CRP with *P. aeruginosa* LPS.

Trial 4. This trial produced many triangular bi-pyramidal crystals in B1 and B3 (Table 5.4 and Figure 5.6).

Condition 1 Li ₂ SO ₄ (M) 0.1 M Tris pH 7				Condition 2 (NH ₄) ₂ SO ₄				
	1	2	3	4	5	6		
% PEG 2K ↓	A	0.16M 10%	0.17M	0.18M	0.15M	0.2M	0.25M 25%	% PEG 8K ↓
	B	12%					30%	
	C	14%					35%	

Table 5.4 Crystallisation conditions in trial 4 using *Limulus* CRP with *P. aeruginosa* polysaccharide. Highlighted wells indicate growth of triangular bi-pyramidal crystals.

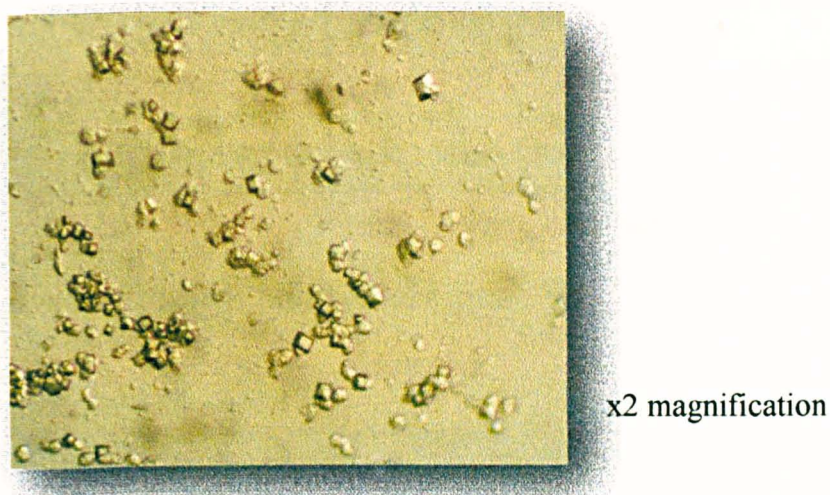


Figure 5.6 triangular bi-pyramidal crystals observed in wells B1 and B3 in trial 4 using *Limulus* CRP with *P. aeruginosa* oligosaccharide.

5.5 Crystallisation using *Limulus* CRP with *H. influenzae* Eagan oligosaccharide

Trial 1. Small triangular bi-pyramidal shaped crystals could be observed in wells A1 and A2 after approximately 5 weeks. A myriad of crystals or precipitate could also be seen in wells B2 and B4 (Table 5.5).

	1	2	3	4	5	6
A	0.17M Li ₂ SO ₄ 12% PEG 2K 0.1M Tris pH 7	0.2M (NH ₄) ₂ SO ₄ 30% PEG 8K	0.2M (NH ₄) ₂ SO ₄ 30% PEG 4K	0.5M (NH ₄) ₂ SO ₄ 0.1M Hepes pH 7.5 30% v/v MPD	1.6M (NH ₄) ₂ SO ₄ 0.1m Hepes pH 7.5 0.1M NaCl	10% PEG 8K 8% ethylene glycol 0.1M Hepes pH 7.5
B	10% PEG 6K 8% MPD 0.1M Hepes pH 7.5	10% PEG 6K 10mM CaCl ₂ 10mM NaCl 50mM MES pH 7.0	50 mM MES pH 7 9% PEG 6K 10 mM CaCl ₂	50 mM MES pH 7 9% PEG 6K		

Table 5.5 Crystallisation conditions for *H. influenzae* Eagan oligosaccharide with *Limulus* CRP. Highlighted wells indicate growth of small triangular bi-pyramidal crystals.

Trial 2. Thin glass shaped crystals in wells A2 and A3 were obtained using the conditions shown below (Table 5.6).

Condition 1 Li ₂ SO ₄ (M) 0.1 M Tris pH 7				Condition 2 (NH ₄) ₂ SO ₄		
	1	2	3	4	5	6
A	0.16M 10%	0.17M	0.18M	0.15M	0.2M	0.25 25%
B	12%					30%
C	14%					35%

% PEG 2K
↓

% PEG 8K
↓

Table 5.6 Crystallisation conditions for *H. influenzae* Eagan oligosaccharide with *Limulus* CRP. Highlighted wells indicate growth of small crystals.

Trial 3. No crystals were produced (Table 5.7).

Condition 1 %MPD				Condition 2 CaCl ₂ (mM)			% PEG 2K ↓
	1	2	3	4	5	6	
A	6% 8% [*]	8%	10%	8mM	10mM	12mM	
B	10% [*]					9%	
C	12% [*]					12%	% PEG 6K ↓

Table 5.7 Crystallisation conditions used in trial 3 for *H. influenzae* Eagan oligosaccharide with *Limulus* CRP.

Trial 4. No crystals were produced (Table 5.8).

Condition 1 Li ₂ SO ₄ (M) 0.1M Tris pH 7				Condition 2 (NH ₄) ₂ SO ₄ (M)		
	1	2	3	4	5	6
% PEG 2K ↓	A 0.16M 10%	0.17M	0.18M	0.15M	0.2M	0.25 25%
	B 12%					30%
	C 14%					35%

% PEG
8K
↓

Table 5.8 Crystallisation conditions used for trial 4 for *H. influenzae* Eagan oligosaccharide with *Limulus* CRP.

Trial 5. Small grainy crystals could be seen after approximately 4 weeks in well A1 using the following conditions (Table 5.9).

Condition 1 %MPD 0.1M Hepes pH 7.5				Condition 2 CaCl ₂ (mM) 50mM MES pH 7		
	1	2	3	4	5	6
% PEG* 2K ↓	A6% 8%*	8%	10%	8mM	10mM	12mM 6%*
	B10%*					9%*
	C12%*					12%*

%
PEG8
↓

Table 5.9 Crystallisation conditions in trial 5 for *H. influenzae* Eagan oligosaccharide with *Limulus* CRP. Grainy crystals seen in highlighted well A1.

Trial 6. Flat glass shaped crystals could be seen in wells A2 and A3 (Table 5.10).

		% MPD					
		0.1M Hepes pH 7.5					
		1	2	3	4	5	6
% PEG* 6K ↓	A	8% 6%*	10%	12%			
	B	 8%*					
	C	10%*					

Table 5.10 Crystallisation conditions in trial 6 for *H. influenzae* Eagan oligosaccharide with *Limulus* CRP. Blue highlight well indicates flat glass shaped crystals.

Trial 7. Small rectangular glass crystals could be seen in wells B2 and B3 (Table 5.11). The crystals were successfully cryoprotected without much deterioration. The crystals were tested at Daresbury SRS on the 26th May 2007 but did not produce good diffraction pattern.

		% MPD 0.1M Hepes pH 7.5					
		1	2	3	4	5	6
% PEG* 6K ↓	A	8% 6%*	10%	12%			
	B	 8%*					
	C	 10%*					

Table 5.11 Crystallisation conditions in trial 7 for *H. influenzae* Eagan oligosaccharide with *Limulus* CRP. Small rectangular glass shaped crystals seen in highlighted wells B2 and B3.

5.6 Discussion:

5.6.1 Crystallisation of *Limulus* CRP with *E. coli* LPS and polysaccharide

The initial crystal growth (trial 1 section 5.2) of *Limulus* CRP is based on previous trials that proved successful but not to the extent of producing crystals that were suitable for X-ray diffraction analysis. Various trials using simple ligands such as ribose-5-phosphate and adenosine monophosphate (Bailey, PhD thesis 2009) have been used in order to “siphon” out the various possible forms of *Limulus* CRP that would enable crystallisation of the single form, however only small grainy crystals have been obtained. The ability to replicate crystal growth of *Limulus* CRP has proved very problematic in this work and in previous work carried out by the group.

Limulus CRP, as mentioned in chapter 1, is coded by more than one gene and contains a heterogeneous mixture in a single pool. Electron microscopy has suggested that the structure is arranged as a hexamer [Fernandez-Moran *et al.*, 1968; Iwaka *et al.*, 1999], yet the three dimensional crystal structure is yet to be determined.

The initial trial (trial 1 section 5.2) focused on optimising conditions with *Limulus* CRP at a concentration of 7.5mg/ml in buffer 150mM NaCl, 50mM Tris, 10mM CaCl₂, pH 7.4 produced small irregular flat glass shaped crystal, however the crystals were found to be unsuitable for any X ray diffraction analysis. The condition in well A2 (trial 1 section 5.2) in which better crystals grew laid the starting point for investigating further growth of CRP crystals with complex ligands such as *E. coli* 0111.B4 LPS.

It was envisaged to use *E. coli* 0111. B4 strain, which contains both smooth and semi smooth LPS structures containing the CRP ligand Phosphoethanolamine [Shrive *et al.*,

2003; Robey *et al.*, 1981], with *Limulus* CRP to demonstrate the broad spectrum of recognition ability towards a diverse range of microbial pathogens [Iwaka *et al.*, 1999], and fine specificity towards certain bacterial species [Ng *et al.*, 1996]. The Japanese horseshoe crab CRP has been found not to recognise *E. coli* K12 bacteria species.

Trial 1 used a 6 fold molar excess of LPS (approximately 2mM). In order to calculate the molar excess of LPS, the molecular weight of each of the *Limulus* CRP subunit was considered to be 23kDa and using a concentration of 7.5mg/ml of protein the molar concentration is calculated to be 0.33mMol, therefore the 6 fold molar excess of LPS used in each trial is 1.95mMol. Although the concentration of LPS is in excess, the concentration would have been lower in the crystal drop because it would have been diluted by the ratio 1:2 when using equal volume 2 μ l of LPS to 2 μ l of well mixture, giving a drop concentration of 0.98mM. The concentration of the protein would also have been halved, to 0.17mM, with the concentration ratio of LPS to CRP therefore still at 6 fold molar excess.

Calculating the amount of *E. coli* required to get 1.95 mMol proved slightly problematic considering that the particular strain of *E. coli* 0111.B4 is present as a heterogeneous mixture, and a theoretical approach was applied to determine the average molecular weight of the LPS in the heterogeneous mixture. Assuming that if the LPS I (13045 Da) and LPS II (6262 Da) are in equal amounts then taking the average weight would give a molecular weight of the native *E. coli* LPS as 9653.50 Da (these values are based on no substitution from the inner core) [Morrison *et al.*, 1975].

Using the molecular weight of 9653.5 Da, 18mg of the LPS was added to deionised water only to find that the LPS did not dissolve. LPS containing large polysaccharide branches

and lipid A substituted with multiple fatty acid is found to be insoluble, and is therefore impractical for crystallisation. One method of overcoming this problem is to dissolve the LPS in a solution of detergent eg β -octylglucopyranoside which acts like a coat surrounding the fatty acid aiding solubility. Although the white fluffy/powdery LPS dissolved in this solution, it only formed a hazy solution which could indicate formation of micelles. The exact nature of this and how these micelles may affect crystallisation would need further investigation.

A 2 μ l aliquot of this solution containing intact LPS was added to the protein which produced an opaque solution/precipitate using the well condition described in trial 2 of crystallisation of *Limulus* CRP with *E. coli* LPS. Initially the precipitate was considered to be calcium precipitate and therefore the drop was agitated gently using the pipette tip, this caused the disappearance of the precipitate, however, after a further three days the solution once again precipitated but this time there were two rod shaped crystal present in the precipitate. The largest crystal measured 0.2 x 0.18mm (Figure 5.2).

The crystal was successfully cryoprotected using increasing strength of MPD and X-ray diffraction data (Figure 5.7) was collected at Daresbury on the 12th November 2005 on station 14.1. The diffraction data was collected at 2Å and showed a large unit cell indicating multiple copies of the CRP molecule, of unknown subunit aggregation, in the unit cell asymmetric unit. Structure solution was not attempted.

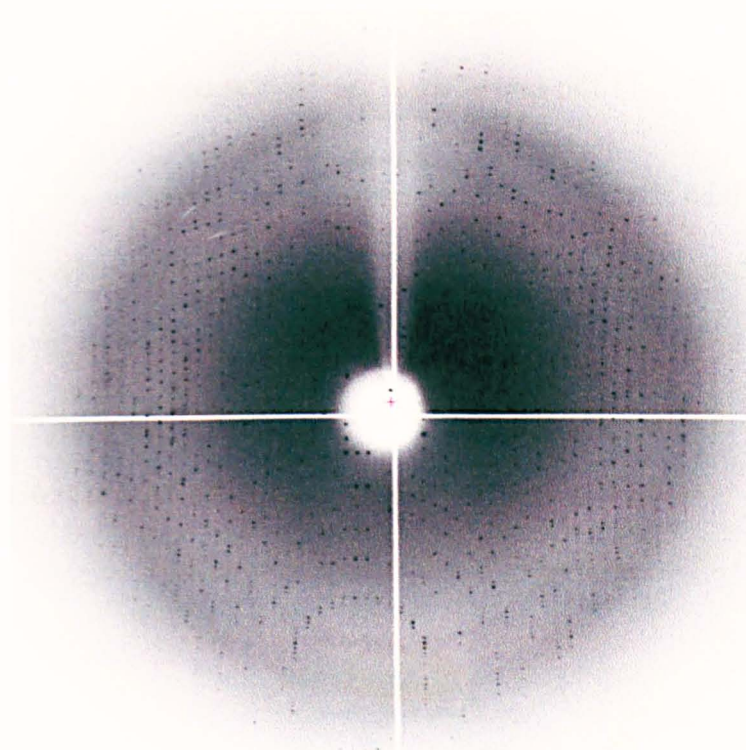


Figure 5.7 Central portion of a diffraction image of *Limulus* CRP with *E. coli* LPS taken on station SRS station 14.1. $\Delta\phi = 1^\circ$, crystal to detector distance 95mm (maximum resolution 2.0Å), exposure time 45 seconds. Maximum resolution shown here is 3Å.

The structure analysis and interpretation are still ongoing within the group in an effort to obtain the first ever, novel three dimensional structure of *Limulus* CRP, although this would possibly provide the structure of a single isoform of CRP amongst many others [Robey *et al.*, 1981]. A personal communication from A.K.Shrive reveals that the structure has subsequently been solved following significant difficulties with molecular replacement, although problems with the analysis and refinement remain and it remains unclear whether bound LPS will be revealed.

Although these results are still being interpreted in order to demonstrate the novel structure, the formation of a single crystal could well be explained in terms of multiple CRP isoforms selectively recognising “particular” forms of 0111. B4 LPS.

The observation of a white precipitate on the micro bridge could be the result of multiple CRP isoforms binding to semi smooth LPS structure. This would lead to an imbalance in the CRP concentration and crystallisation medium of the droplet on the micro bridge. Consequently a single form of CRP isoform that has the inability to recognise smooth LPS may have remained in solution and then crystallised out (Figure 5.8).

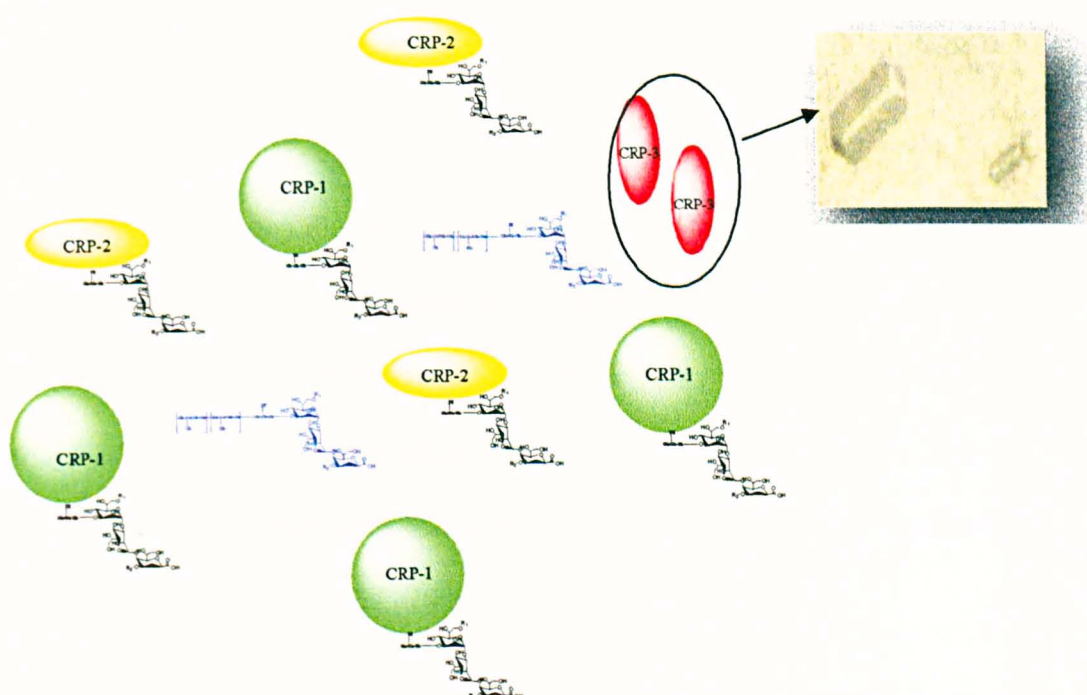


Figure 5.8 Schematic illustration demonstrating selective binding of various CRP's (here CRP-1 and 2) with semi smooth LPS (black) and unrecognised smooth LPS (blue). CRP-3 (red) is the forms that has no recognition ability with either semi or smooth LPS and hence crystallised out as shown in the inset photo 1. Drawn using CS Chem. Draw Pro. Version 4.5.

Following the success of the initial trial 1, attempts at repeating the conditions were unsuccessful (trial3) at obtaining the same type of crystals that had previously been obtained in trial 1. Upon investigating the reason for the failure to obtain the same crystal form, although the same batch of CRP has been used, the batch of commercially purchased (Sigma-Aldrich) *E. coli* 0111.B4 LPS is different (i.e. same strain 0111.B4 but the commercially purified form is of a different batch). This could mean that the ratio of the components of the heterogeneous mixture of LPS (in particular semi smooth and smooth) is altered and therefore the LPS has different binding or recognition potential for the protein. The second batch of LPS used may have contained a higher ratio of LPS I (smooth form) containing a larger number of O-antigen repeat units, which could have masked the PE ligand in the inner core. This would have hampered possible binding for most of the CRP isoforms.

The crystals from trial 3 have been found to be destroyed by cryobuffers and in some cases no diffraction data could be obtained. However to understand how this process may have worked a more thorough investigation may be needed by looking at various other forms of rough, semi-rough and smooth LPS from other strains of bacteria. It is reported that the CRP2 from the Japanese horseshoe crab was found to exhibit agglutination against the polysialic capsule of *E. coli* K1 [Tan *et al.*, 2005].

The binding of *E. coli* by CRP from the study by Tan and co workers [Tan *et al.*, 2005] may further support the possibility raised here that the observed CRP crystal is a single form that does not recognise *E. coli* 0111.B4 LPS, whilst the various remaining CRP isoforms in the pool may have bound to the LPS and precipitated out.

Trial 4 and trial 5 were carried out using the *E. coli* 0111.B4 polysaccharide fraction obtained from acid hydrolysis. Again a six-fold molar excess was used which readily dissolved in deionised water to make the required concentration (approximately 2mMol). On close examination of the crystal from trial 5 well B1 (PEG 6K, MPD and Hepes pH7.5), there are similarities with the crystal obtained in trial 1, but the size was far too small to be suitable for X-ray diffraction studies. New microfocus beamlines at Diamond and elsewhere could mean that these crystals are now suitable for analysis.

5.6.2 Crystallisation of *Limulus* CRP with *P. aeruginosa* LPS and polysaccharide

Work by Tan and co workers (2005) and Ng and co workers (2004) has shown that *P. aeruginosa* is recognised by CRP and has antimicrobial activity and is involved in effective bacterial clearance. The strain used in their studies is different from that investigated in this work. The LPS used in this investigation is a serotype 10 strain ATCC27316. Since the structure of this strain is not documented, the structure of serotype 10 strain 170034 [Bystrova *et al.*, 2004] is used as a model to derive a proposed structure for strain ATCC27316.

The lipid A component determined by Kulshin and co workers (1991) shows that the major structure consists of a penta acyl species with a molecular mass of 1327 Da. The remaining part of the oligosaccharide can exist as rough, semi rough and smooth type (approximate molecular weights of 2459, 5284 and 9836 Da respectively) [Bystrova *et al.*, 2004]. The inner core contains a single PPEtn ligand substituted from the first heptose. As mentioned above, the ligand may be “hidden” or “exposed” to the CRP depending on the type of LPS, i.e. smooth form or rough form respectively.

Crystallisation trials did not provide any suitable crystals for diffraction studies, but the crystallisation drop containing a mirage of very small sandy grains indicating that the CRP may have recognised the LPS and precipitated out. Unlike the previous trials using *E. coli*, which appear to have selectively caused a single form of *Limulus* CRP to crystallise out, the same method of CRP selectivity using *P. aeruginosa* did not occur suggesting that the heterogeneous mixture of *Limulus* CRP has recognised whatever form of *P. aeruginosa* LPS may be present and hence precipitated out without forming any suitable crystal.

The trials in section 5.5 produced many unique crystals, but all were found to be unsuitable for X-ray diffraction analysis.

5.6.3 Crystallisation of *Limulus* CRP with *H. influenzae* Eagan oligosaccharide.

The ability of *Limulus* CRP to recognise the human pathogen *H. influenzae* would have proven a novel finding considering that *Limulus* is an ancient creature with a demonstrated ability to survive. The trials produced very good crystals with mostly three different forms triangular bi-pyramidal, rectangular and flat glass shaped crystals. Successful cryoprotection was achieved using 5-20% MPD without any sign of crystal deterioration, but useful X-ray diffraction data could not be obtained. *H. influenzae* LPS provides two ideal targets, PC and PEtn at two substituents. The PC is exposed on the outer glucose and PEtn is off the inner core heptose according to the stereochemistry obtained from Nuclear Overhauser Effect (NOE) NMR studies [Masoud *et al.*, 1997]. Both of these ligands may be inaccessible for CRP binding due to the branching oligosaccharide “hiding” the ligands from CRP recognition.

5.7 Summary

5.7.1 *Limulus* CRP with *E. coli* LPS and polysaccharide

The initial crystal trial using CaCl_2 and PEG 6K as precipitant buffers were based on previous screening conditions that produced numerous crystals, which were found not to be of good diffraction quality (A. K. Shrive. personal communication). In trial 1 the intact LPS was added to deionised water in the presence of the detergent β -octyl glucopyranoside which caused the LPS to dissolve but produced a hazy solution. This was used in crystallisation trials using 50mM MES pH7, 9% PEG 6K and 10mM CaCl_2 . These conditions produced a single large crystal (Figure 5.7). The crystal was successfully cryoprotected using 20% MPD prior to X-ray diffraction data collection at Daresbury SRS on station 14.1. The diffraction data obtained (Figure 5.7) is novel and following significant molecular replacement difficulties has subsequently provided a solution leading to the structure of *Limulus* CRP although problems with the analysis and refinement remain (A.K.Shrive, Personal Communication). Repeating the conditions used in trial 1 failed to produce identical crystals and no further crystals suitable for X-ray diffraction were produced using the *E. coli* polysaccharide obtained from hydrolysis.

5.7.2 *Limulus* CRP with *P. aeruginosa* LPS and polysaccharide

Intact LPS dissolved in β -octyl glucopyranoside producing a hazy solution, and crystal trials (Table 5.3) did not produce any crystals. The cleaved polysaccharide produced some very interesting crystals described as torpedoes (Figure 5.5) and triangular bi-pyramidal crystals (Figure 5.6), but did not produce any useful diffraction at Daresbury SRS.

5.7.3 *Limulus* CRP with *H. influenzae* Eagan polysaccharide

Small triangular bi-pyramidal crystals were only obtained using the conditions Li_2SO_4 , PEG 2K, Tris pH 7 and condition $(\text{NH}_4)_2\text{SO}_4$, PEG 8K (Table 5.5). These crystals also failed to produce any X-ray diffraction and exploring further conditions (table 5.6) screened around these initial conditions only produced a precipitate with small grainy crystals.

Chapter 6

Structural Studies of rfhSP-D with *H. influenzae* Eagan

and Eagan 4A oligosaccharide

Results & Discussion

6.1 Crystallisation and X ray diffraction of rfhSP-D with *H. influenzae* Eagan oligosaccharide

Flat glass shaped crystals were obtained using 16% PEG 4K in 0.1M Tris pH 7.5. Addition of cryoprotectant using MPD has always been found to be used successfully with rfhSP-D (Shrive *et al.*, 2009). There appeared to be no deterioration of the crystals even with an exchange of 20% MPD containing 30mM *H. influenzae* wild type Eagan. The drop solution on the microbridge still remained clear. X-ray diffraction was carried out at Daresbury SRS station 14.1 on the 26th of May 2007. No ice rings could be seen suggesting successful cryoprotection. The crystal diffracted beautifully to at least 1.3Å. The electron density map from wild type Eagan oligosaccharide ligand soaks of rfhSP-D crystals failed to reveal bound ligand. (A.K. Shrive, personal communication).

6.2 Crystallisation and X ray diffraction of rfhSP-D with *H. influenzae* Eagan 4A oligosaccharide

Addition of 30mM *H. influenzae* Eagan 4A oligosaccharide in the presence of 20% MPD did not cause any deterioration of the flat rectangular crystals of rfhSP-D and the nature of the mother liquor was unchanged. Similar to the method of data collection for wild type Eagan, X-ray diffraction data were collected at Daresbury on 26th May 2007. The crystal to detector distance was initially set at 110mm and good diffraction pattern to at least 1.7Å was observed (Figure 6.1)

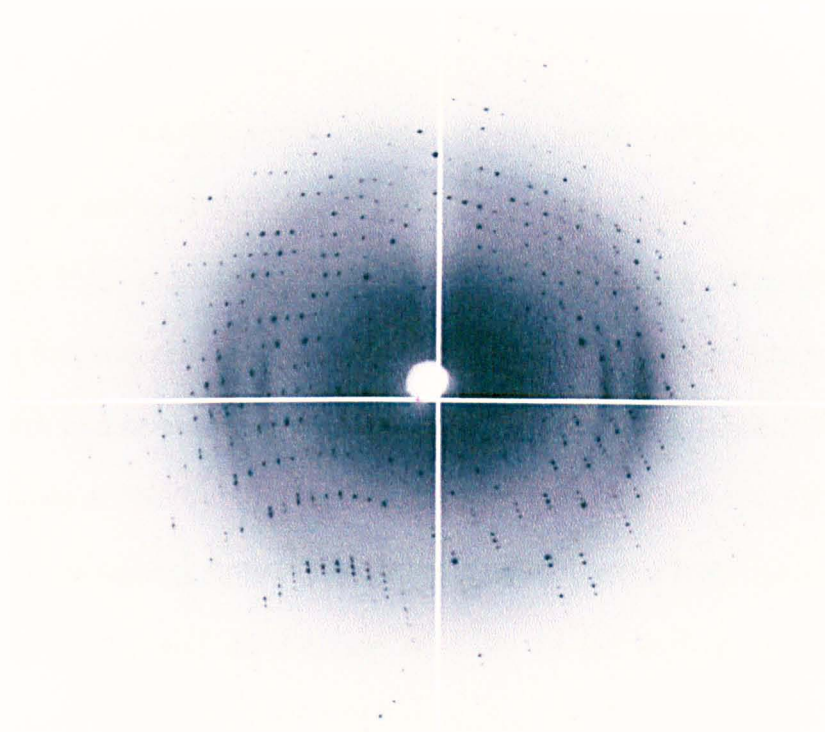


Figure 6.1 X-ray diffraction pattern of rfhSP-D with *H. influenzae* Eagan 4A oligosaccharide soak at 30mM. Data was collected on station 14.1 on ADSC Quantum 4R CCD, Daresbury SRS.

6.3 Data collection and processing of rfhSP-D with *H. influenzae* Eagan 4A oligosaccharide

All data was processed and interpreted using the CCP4 programme suite (CCP4, 1994). A single data set was collected at high resolution and a total number of 159 images were collected through $1^\circ \Delta\phi$ with crystal to detector distance of 75mm. Space group $P2_1$ was selected at auto indexing with crystal to detector distance of 72.8mm which allowed the prediction of spots to match with experimentally obtained diffraction spots. The crystal has unit cell dimensions of 55.35Å, 107.99Å, 55.65Å and 92.14° for a , b , c and β respectively. Integration of all images was followed by SCALA using various high resolution cut-offs. 1.70Å gave $I/\sigma I$ of 2.4 and an R-factor of 0.305 in the highest shell with 99.9% completeness overall. The output from 1.75Å resolution was accepted with $I/\sigma I = 3.0$ in the highest resolution shell, with 65639 reflections indexed.

6.3.1 Rigid body

Isomorphism was sufficient to allow the coordinates of the previously determined 1.6Å native rfhSP-D structure [Shrive *et al.*, 2003] to be used as a starting model. For rigid body refinement, the coordinates for the three chain of rfhSP-D were downloaded from the model (pdb id 1pwb) to calculate against the observed intensity. Since the model is the same molecule in the same unit cell as the structure under investigation, only minor rotation and translational operations allowed the model to fit in the new unit cell to match with the observed diffraction data. 31 cycles of refinement were carried out using a high resolution cut off of 1.75Å. The R-factor obtained was 0.327. Crystallographic data and refinement statistics are shown in table 6.1.

	Native rfhSP-D ^a	Eagan 4A
Data collection		
Wavelength (Å)	1.488	1.488
Temperature (K)	100	100
Space group	P2 ₁	P2 ₁
Cell dimensions		
a (Å)	55.55	55.35
b (Å)	108.52	107.99
c (Å)	55.69	55.65
β (°)	91.20	92.14
Maximal resolution (Å)	1.6	1.7
Observations	236,479	217,593 (31,047)
Unique reflections	84,773	71,591 (10,454)
Completeness (%)	97.8	99.9
R _{merge}	0.054	0.075 (0.295)
Mean I/σI	9.1	5.1 (2.4)
Highest resolution bin (Å)		
observations	1.69-1.60	1.84-1.75
Unique reflections	22,539	28,736
Completeness (%)	11,823	9588
R _{merge} ^a	93.3	100
Mean I/σI	0.167	0.214
	3.3	3.0
Refinement		
		Rigid body only
Protein atoms (chains A, B, C)	3457	3421
Other atoms	10 calcium ions	10 calcium ions
Resolution range (Å)	55.5-1.60	55.9-1.75
R _{conv} ^e (%)	22.7	32.7
R _{free} ^f (%)	24.9	32.0

^a $R_{\text{merge}} = \sum_h \sum_j |I_{h,j} - I_h| / \sum_h \sum_j I_{h,j}$ where $I_{h,j}$ is the j th observation of reflection h

^e $R_{\text{conc}} = \sum_h \|F_{\text{ob}} - F_{\text{cal}}\| / \sum_h \|F_{\text{ob}}\|$ where F_{ob} and F_{cal} are the observed and calculated structure factor amplitudes, respectively, for the reflection h .

^f R_{free} is equivalent to R_{conv} for a 5% subset of reflection not used in the refinement.

Table 6.1 Crystallographic data collection and rigid body refinement statistics for *H. influenzae* Eagan wild type with rfhSP-D compared with statistics obtained previously for native rfhSP-D^a by [Shrive *et al.*, 2003].

6.3.2 Map calculation, model building and electron density map

Using the programmes SFALL and SigmaA from the CCP4 suit (CCP4, 1994) structure factors and estimated phases were obtained. For a total of 71,591 reflections the reliability index (figure of merit) was 0.69232.

6.3.3 Electron density map overall structure

The electron density map and the superimposed model displayed in the graphics package *O* (Jones *et al.*, 1991) corresponded well. The overall structure seen is the trimeric aggregate with the three C-terminal globular domains (residues 236-355) and the three chains linked through an extended α -helical coiled-coil neck region (residues 203-235). The short N-terminal collagen-like region of rfhSP-D is not visible in the electron density map. The electron density map clearly shows the three chains A, B, C of the rfhSP-D with the three calcium ions and the fourth calcium ion present in the trimeric axis between the three chains (Figure 6.2 and 6.3).

The principal Ca^{2+} ion in all three chains is found coordinated by the side chains of Glu321, Asn323, Glu329, Asn341, and Asp342, and by the main chain carbonyl of Asp342. The remaining two Ca^{2+} ions are coordinated by amino acid residues previously described by Shrive and co workers (Shrive *et al.*, 2003, 2009).

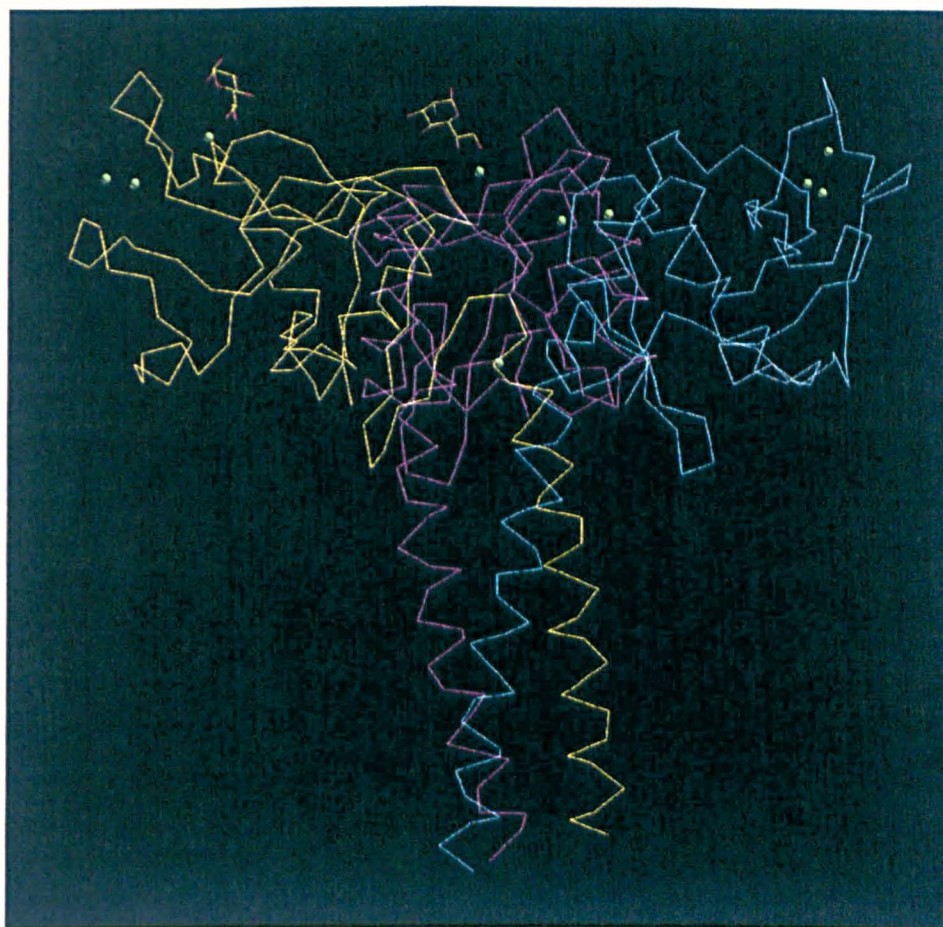


Figure 6.2 Backbone of the rfhSP-D trimer (chain A blue; chain B magenta and chain C yellow) with three calcium ions in the CRD and the fourth calcium ion located in the central axis of the trimeric neck. Ligand is seen bound in chains B and C only. Image displayed in *O* and taken using Snapshot.

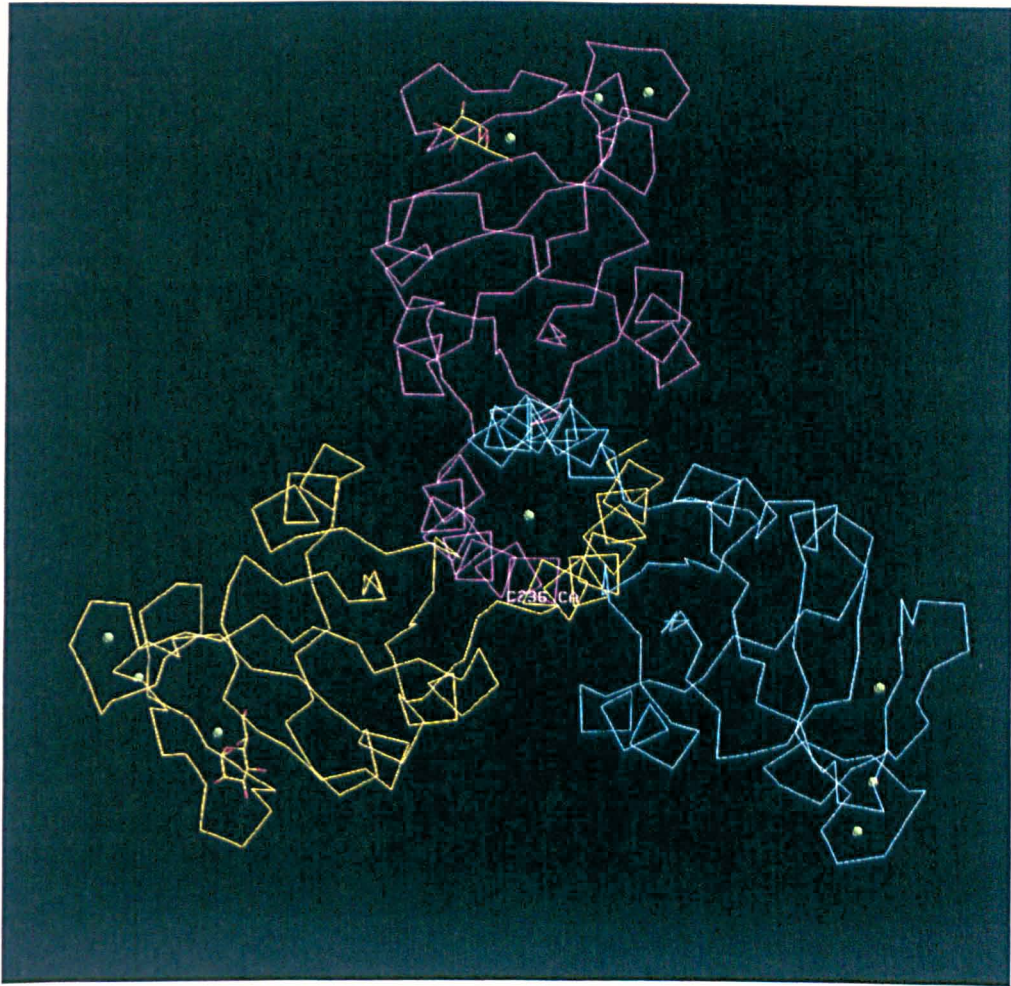


Figure 6.3 View down the trimeric axis of rfhSP-D. Chains A, B and C are shown in blue, magenta and yellow respectively. Calcium ions are shown as green spheres with ligand binding seen in chains B and C. Image displayed in *O* and taken using Snapshot.

6.3.4 Electron density of Egan 4A oligosaccharide in the binding region

The crystal structure of rfhSP-D complexed with the Egan 4A oligosaccharide reveals Ca-dependent binding of the ligand via the Kdo-linked inner core heptose (Hep1), utilising the O6' and O7' hydroxyl groups of Hep1 (Figure 6.4 and 6.5) as well as forming coordination bonds between the protein side chains. The electron density reveals bound saccharide density in only two of the three CRDs with ligand being absent in chain A. The electron density map shows no density for the β 1-4 glucose attached to Hep1 and while there is clear electron density for the α 1-5 Hep-Kdo glycosidic bond (Figure 6.4), the density of the attached sugar suggest a closed 5-membered ring rather than the expected Kdo pyranose ring. The best fit to this density (Figure 6.6) is a modified ribose ring with side groups placed according to the density, agreeing with the proposed β -elimination of the phosphate group [Danan *et al.*, 1982; Auzanneau *et al.*, 1991] forming the enantiomeric furanoid derivative resulting from mild acid hydrolysis (Figure 4.4, 4.5 and 4.13).

The Kdo C4 enantiomeric substitution is not visible in the electron density although density resembling the acidic C1 group (COOH) is clearly evident in a position and at a distance which suggests interaction with Asp325 and Arg343, with hydrogen bonds to Asp325 being formed by one of the hydroxyl groups, while Arg343 on the opposite side is tilted towards the Hep1 in a manner similar to that only seen in the galactose-bound structure where a pair of hydrogen bonds between the galactose and Arg343 are formed [Shrive *et al.*, 2009].

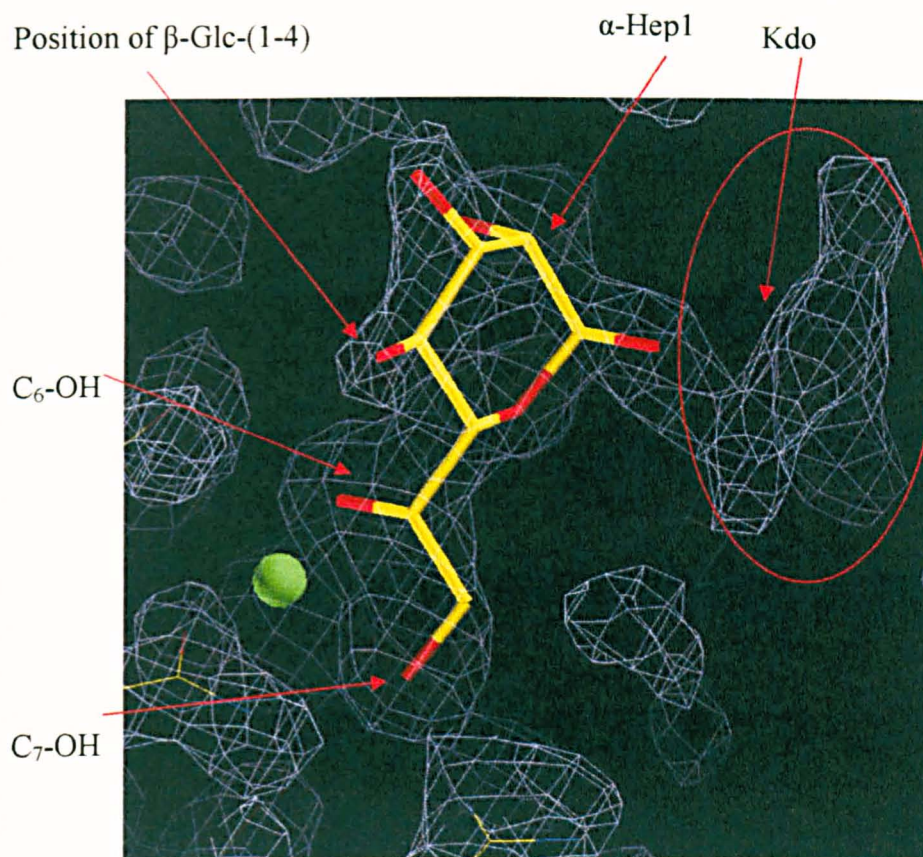


Figure 6.4 Electron density of the bound ligand in chain B with its associated linkages. Electron density of the Hep1 unit can be seen clearly coordinated to the Ca^{2+} ion (green sphere) by C_6 and C_7 -OH. Substitution of β -Glc at Hep C_4 -OH is not present in the density. The pyranose ring structure representing Kdo is distorted at α -Hep-(1-5)- α -Kdo. Map visualised in the O graphics package and contoured at 1 rms.

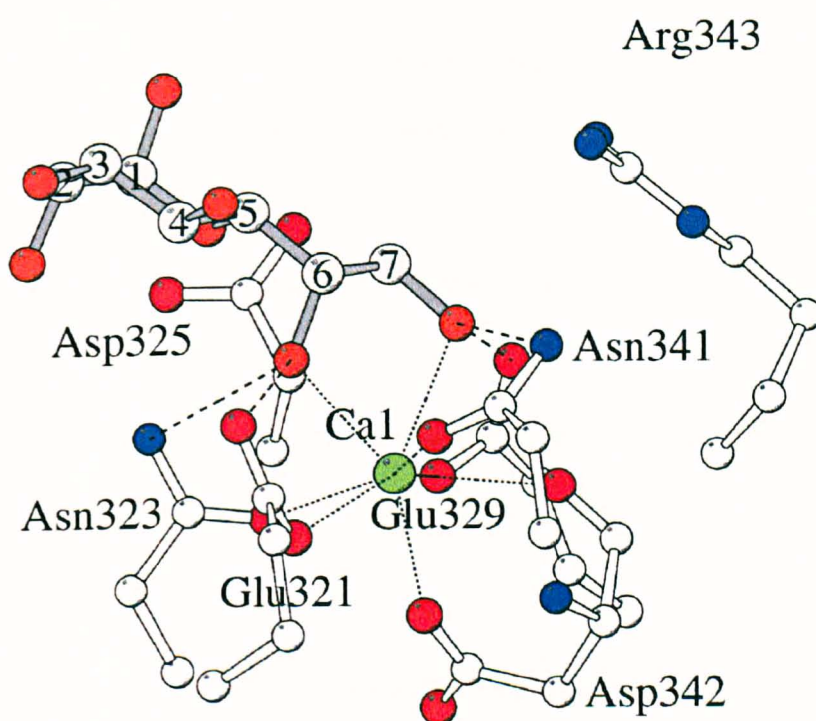


Figure 6.5 Bound Egan 4A in subunit B of rfhSP-D. The ligand is bound via coordination of Hep1 to the calcium ion Ca1 and protein side chains. Figure generated using MOLSCRIPT (Kraulis, 1991)

6.4 Discussion

Structural studies of the interaction between hSP-D and LPS have utilised a recombinant homotrimeric fragment rfhSP-D, which consists of a CRD, a trimeric neck region and short segment of the collagen domain. The well characterised *H. influenzae* Eagan oligosaccharide (structure 4.9 and 4.10) provides three potential structural recognition sites for the calcium dependent interaction of rfhSP-D which recognises the vicinal equatorial hydroxyl groups that are located on:

- The exposed portion of the terminal glucose (C₃-OH and C₄-OH) substituted from the inner core Hep I,
- The C₆-OH and C₇-OH located on the inner core Hep I and Hep III and
- The C₆-OH and C₇-OH from the KDO attached to the lipid A of lipopolysaccharide.

It is possible that SP-D carbohydrate binding interactions with LPS occurs with certain strains of gram negative bacteria. Kuan and co workers (1992) and Wang and co workers (2008) have demonstrated that SP-D recognises the terminal core region and inner core region of *E. coli* LPS, namely rough chemotype of LPS and not smooth chemotype LPS, that contain O-polysaccharide. If we consider the *H. influenzae*, which can be classed as the rough chemotype, that lacks the O-polysaccharide but contains Glc on the terminal core oligosaccharide (figure 4.9), then this structure should provide interactions for SP-D recognition, X-ray diffraction studies using rfhSP-D with wild type *H. Influenzae* Eagan oligosaccharide at an excess concentration of 30mM revealed only the structure of native rfhSP-D as obtained previously [Shrive *et al.*, 2003; 2009]. The electron density map reveals the trimeric aggregate with three C-terminal globular CRD domains and no bound

ligand. These findings could be because the branching chain sugars stemming from the inner core heptose units in the wild type Eagan LPS somehow protect this core region of the bacteria from recognition by SP-D. Binding studies using native hSP-D with wild type Eagan also failed to show any binding between the protein and LPS [Clark *et al.*, 2011].

The Eagan 4A is a simple much smaller molecule consisting only of Glc-Hep-Kdo compared to the Eagan wild type. Unlike the Eagan wild type, the 4A has no branched oligosaccharide and hence the inner core is exposed for recognition by SP-D. The electron density map of the *H. influenzae* Eagan 4A oligosaccharide soaked at 30mM concentration with rfhSP-D is identical to that of the native rfhSP-D, but this time there is a large region of well defined electron density in the vicinity of Ca1, implicating that the Eagan 4A oligosaccharide is recognised and bound by the protein. There also appears to be very little conformation change between the native rfhSP-D and Eagan 4A bound structure.

Recognition of the Eagan 4A LPS core by the SP-D CRD is mediated through calcium binding of Hep1 enhanced by additional interactions between the Hep1-Kdo core and the ligand-binding determinants Asp325 and Arg343, which flank the ligand binding pocket of the protein. The combined crystal studies between the wild type Eagan and the Eagan 4A suggest that the CRD of SP-D is interacting with the inner core region of the LPS and that the more extended Eagan wild type LPS structure interfere with this recognition. Despite the availability of the equatorial hydroxyl pair O3' and O4' on the terminal Glc of Hep1, rfhSP-D complexed with the Eagan 4A oligosaccharide shows Ca-dependent binding of the ligand via the Kdo-linked inner core heptose Hep1. The binding utilises the Hep1 O6' and O7' hydroxyl pair rather than the Hep1 equatorial mannose O3' and O4' pair commonly associated with hSP-D recognition of carbohydrate [Shrive *et al.*, 2003;

2009]. The general disposition of Hep1 here is similar to that of the terminal bound Hep in the heptose disaccharide bound rfhSP-D structure [Wang *et al.*, 2008], except that here there is significant rotation of the bound Hep1 away from Asp325. The glucose attached to Hep1 (β 1-4) on the O4 hydroxyl group (Figure 6.5) is not visible in the electron density, presumably due to the freedom of conformational states about the glycosidic linkage and a lack of constraining protein or crystal contacts.

Bound LPS is not present in the chain A binding site, which is far more restricted, due to crystal contacts, than that in chain B and C (Shrive *et al.*, 2009). The crystal contact in chain A is in the form of hydrogen bond between Asp325 and Tyr314 in chain B of a neighbouring trimer with different orientations of the side chain of Asp325. This crystal contact interface in chain A therefore causes a much more restricted ligand binding site compared to chains B and C. In chains B and C the CRD crystal contacts are very similar to each other but not identical (Figure 6.6), with charge-charge interactions between Asp325 (OD1) and Lys230 (NZ) in a neighbouring trimer. These crystal contacts provide a much more open ligand binding pocket and hence it is more accessible for ligand binding [Shrive *et al.*, 2009]. The major contributing factor which determines orientation of bound ligand has been attributed to Asp325 which in turn will influence the side chain of Arg343 located on the binding pocket opposite Asp325. This region in other collectins such as rat MBP-A which has His189 and Ile207 and MBP-C which has Val194 and Val212 produced two different binding orientation differing by 180°. The difference in these orientations have been attributed to energy minimisation, Van der Waals interaction between the sugar and the protein involving the two residues and also the solvation energy between the pyranose ring form and the protein [Ng *et al.*, 1996; 2002].

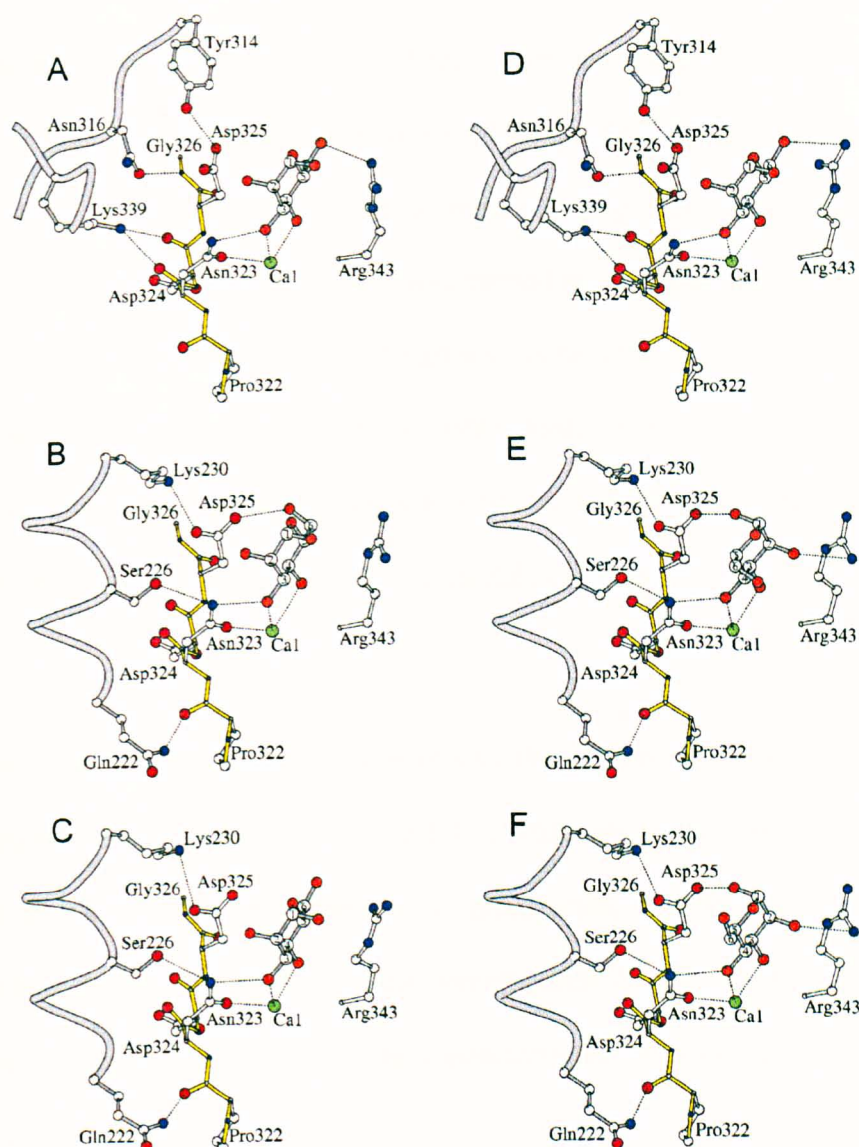


Figure 6.6 The influence of the variable crystal contacts on the orientation and interactions of bound manno-oligosaccharide and Arg343. A, B and C: man α 1-2man chains A, B and C respectively. D, E and F: man α 1-4man chains A, B and C respectively. In all figures the symmetry contact protein molecule and the bound terminal mannose are represented in grey, with selected residues of the hSP-D monomer represented by yellow main chain bonds and white side chain bonds. The crystal contacts with chains B and C, in both structures, are very similar but not identical. Of the six subunit structures, the only pairs with similar ligand orientation and similar ligand non-bonded contacts are chains B and C in man α 1-4man and chain A in both man α 1-2man and man α 1-4man, although in the latter case the two orientations of the Arg343 side chain are significantly different. Image taken from Shrive *et al.*, 2009.

There is clear electron density for the 1-5 Hep-Kdo glycosidic bond but the striking observation is that the Kdo electron density does not resemble a pyranose ring structure (Figure 6.7). Previous analysis of hydrolysed Egan 4A LPS by MALDI-MS (see chapter 4, Fig. 4.14) indicated the presence of pyranose ring Kdo, which may well be due to experimental errors since the analysis was carried out without further purification. ESI-MS analysis have consistently indicated various uncharacterised *anhydro* forms of the Kdo [Deadman *et al.*, 2009; Masoud *et al.*, 1997; Manson *et al.*, 2002; Yildirim *et al.* 2003] arising from a rearrangement of the Kdo with the loss of water after the β -elimination of the phosphate or pyrophosphate group from Kdo O4 [Danan *et al.*, 1982; Auzanneau *et al.*, 1991].

The best fit to the electron density is indeed the predominant 5 membered *anhydro* Kdo ring structure with the absence of the C4 substituents in the electron density, except for density which strongly resembles the terminal acidic COOH group, consistent with the presence of the two (Figure 6.7). This demonstrates for the first time the structure of the dominant *anhydro* Kdo formed by the characterisation of a variety of LPS from a number of bacterial species.

The *anhydro* Kdo makes contact with Asp325 through hydrogen-bonding involving the OH group off Kdo C6 which would be present in a similar position in a normal Kdo oriented similarly with respect to Hep1. The movement of Hep1 away from Asp325 compared to the bound heptose disaccharide structure [Wang *et al.*, 2008] is consistent with the proximity of the furanoid ring to Asp325. Crucially, the position of C4 in the *anhydro* Kdo is such that normal Kdo would present the phosphate carried by C4 to Arg343. As an alternative to the Asp325-Kdo O6 and Arg343-Kdo P interactions, rotation

of the normal Kdo about the glycosidic α 1-5 Hep1-Kdo bond would direct the OH groups off Kdo C7 and C8 towards Asp325 and orient the acidic group off Kdo C2 to interact with Ar343 (Figure 6.7 and 6.8). The position of this normal Kdo acidic group could then align with the isolated density seen here which suggests an acidic group and is in a position consistent with the extended acidic group of both enantiomers off the *anhydro* Kdo C4. Thus, recognition of the Egan 4A inner core appears to involve not only coordination to the primary calcium ion, but also interaction with both binding site flanking residues [Shrive *et al.*, 2003; 2009; Wang *et al.*, 2008].

The optimal binding requiring the Hep-Kdo pair shown in this study is consistent with studies of a *B. pertussis* [Schaeffer *et al.*, 2004] mutant which contained a single heptose core and maximised aggregation and membrane permeability by SP-D. For *Salmonella Minnesota*, only strains with incomplete cores, including strains Rd (triheptosyl core only) and Rc (additional terminal glucose) (Figure 1.20, chapter 1), exhibit significant binding to SP-D [Kuan *et al.*, 1992]. Interestingly, early studies on mannose binding protein to rough strains of *E. coli* proposed binding to Hep and GlcNAc in the rough core oligosaccharide of the *E. coli* K-12 cell wall, and to Hep in the incomplete core of *E. coli* B. [Kawasaki *et al.*, 1989]

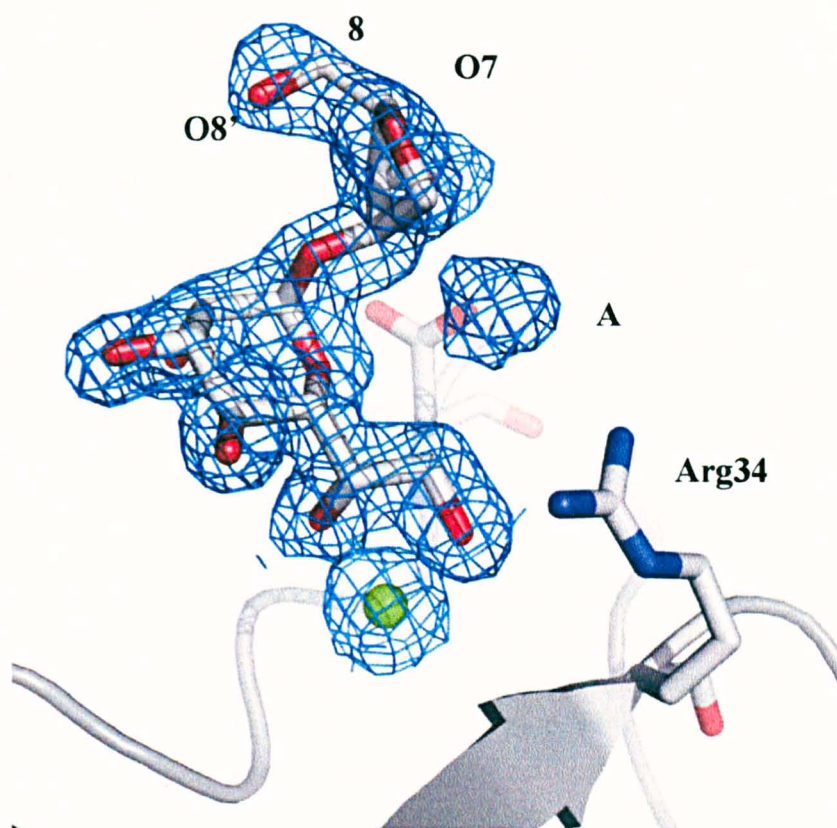


Figure 6.7 Electron density for the bound Glc-Hep and the putative *anhydro* Kdo ligand. Original Kdo numbering retained. Density labelled A corresponds to the putative position of O2' (original Kdo numbering) in either or both of the enantiomeric substituent off KdoC4. The glucose is not visible in the electron density. Figure generated using the PyMOL Molecular Graphics System Version 1.4 (Schrödinger, LLC, 2011).

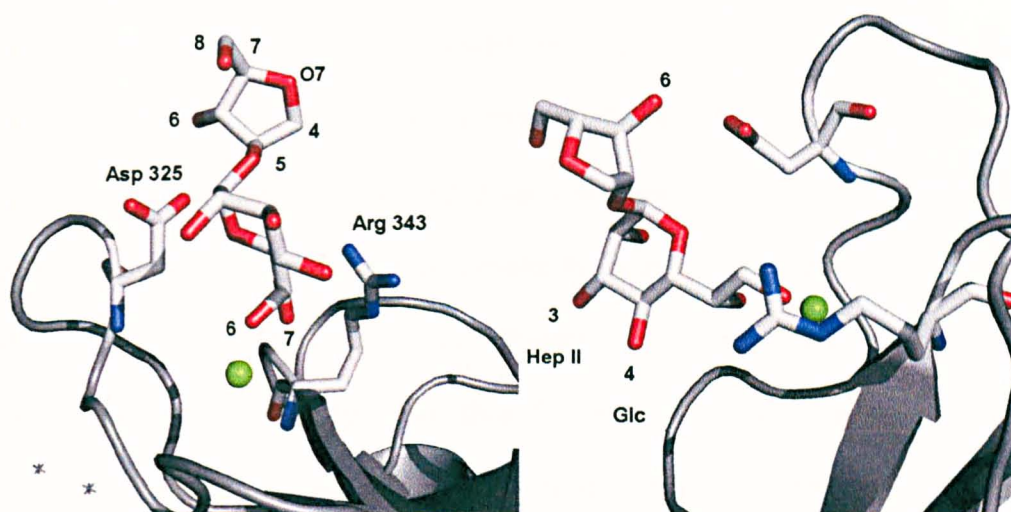


Figure 6.8. The Hep-Kdo(*anhydro*) disaccharide in the ligand binding site. Kdo O6' is positioned to interact with Asp325, while the enantiomeric C4 substituent is not identifiable in the electron density except for an indication of O2' (original Kdo numbering), directed towards Arg343, which may be similarly positioned in both enantiomers. Original Kdo numbering retained. Figure generated using the PyMOL Molecular Graphics System Version 1.4 (Schrödinger, LLC, 2011).

6.5 Summary

Electron density maps from wild type Eagan oligosaccharide ligand soak of rfhSP-D crystals failed to reveal bound ligand, whilst the crystal structure of rfhSP-D complexed with the Eagan 4A core saccharide isolated from the Eagan 4A LPS revealed Ca-dependent binding of the ligand via the Kdo-linked inner core heptose (Hep1) utilising the O6' and O7' equatorial hydroxyl pair of the Hep1 despite the availability of the equatorial hydroxyl pair O3' and O4' on the terminal of Hep1 (Figure 6.2 and 6.5). Electron density reveals bound saccharide in only the two of the three CRDs with ligand being absent in chain A because the binding site in chain A is far more restricted, due to crystal contacts, than that in Chain B and C.

There is clear electron density for the 1-5 Hep-Kdo glycosidic bond but the striking observation is that Kdo electron density does not resemble a pyranose ring, but the best fit to the density is a 5-membered *anhydro* Kdo ring arising from a rearranged Kdo with the loss of water after the β elimination of the phosphate from Kdo O4 after mild acid hydrolysis. The glucose attached to Hep O4 (β 1-4) (Figure 6.5 and 6.6) is not visible in the electron density, presumably due to the freedom of conformational states about the glycosidic linkage and a lack of constraining contact.

The crystal study using Eagan LPS with rfhSP-D suggests that the CRD of SP-D is interacting with the inner core region of the LPS and that the more extended wild type Eagan LPS of the bacteria interfere with this recognition, thus allowing the bacteria to evade detection and killing by innate immune proteins.

Chapter 7

Conclusions and future work

7.1 Conclusions

Work carried out in the structural Biology laboratory at Keele University before the start of this project focused on the understanding of how innate immune proteins such as C-reactive protein and human lung surfactant protein D recognise and bind to natural ligands such as LPS. Innate immune proteins are at the forefront of providing immune protection from birth onwards when the adaptive immune system is still in its infancy of development. Research in this field provides crucial information on how microbial targets are recognised by these proteins by recognising specific target ligands that are present on the surface and how this information can be manipulated for designing therapeutic agents. The structures of rfhSP-D and human CRP in both native and ligand-bound forms have been obtained previously by X-ray crystallography. The novel X-ray diffraction from *Limulus* CRP described here has subsequently provided the much sought after crystal structure of *Limulus* CRP, while the rfhSP-D structure in complex with *H. influenzae* Eagan 4A oligosaccharide provides the first structural description of a collectin bound to a microbial ligand.

7.1.1 Crystallisation studies using *Limulus* CRP with LPS.

At the start of the project numerous crystallisation trials were conducted to obtain the three dimensional crystal structure of *Limulus* CRP, a remarkable creature residing in the Atlantic Ocean that has a wide spectrum of defence mechanisms against a plethora of pathogenic microbes. Much of this defence is provided by the phylogenetically ancient C-reactive protein. The structure of human CRP, the prototypical acute phase protein, has

been solved [Shrive *et al.*, 1996; Thompson *et al.*, 1999] but the structure of *Limulus* CRP remains a mystery. Amino acid sequences have shown that the binding region is conserved in the pentraxin family, with binding to the PC and PE moieties which are present in complex structures such as lipopolysaccharides. The ligand PC was identified from teichoic acid, a surface component of the gram positive bacteria *Streptococcus pneumonia* [Tillet and Francis 1930; Volanakis *et al.*, 1971]. The LPS used in this study is from gram negative bacteria and contains PC and PEtn. Binding to *E. coli* and *P. aeruginosa* LPS by *Limulus* CRP has been described [Tan *et al.*, 2005; Ng *et al.*, 2005; Iwaka *et al.*, 1999] but not for *H. influenzae*. While the extensive *Limulus* CRP crystallisation trials using *P. aeruginosa* oligosaccharide and *H. influenzae* Eagan oligosaccharide produced a variety of crystals, none diffracted sufficiently well for structure analysis. No crystals were obtained using intact *P. aeruginosa* LPS. The *Limulus* CRP natural ligand complex studies using native *P. aeruginosa*, and *H. influenzae* LPS/oligosaccharide proved disappointing in terms of their potential to reveal details of binding. However, attempts at looking at the recognition of *E. coli* LPS/oligosaccharide by *Limulus* CRP has provided the opportunity to define the crystal structure of *Limulus* CRP [Shrive and co workers, unpublished data] and provides the foundation for further work. Whether the structure will reveal bound *E. coli* LPS is not yet known.

7.1.2 Crystallisation studies using rfhSP-D and *H. influenzae* Eagan and Eagan 4A

X-ray diffraction data was collected to 1.7Å resolution for rfhSP-D soaked with *H. influenzae* Eagan 4A, a mutant strain with well defined structure consisting of Glc-Hep-Kdo. The electron density map, which is a construct of 2F_O-F_C, shows quite clearly the

structure of rfhSP-D with the Eagan 4A ligand in the CRD binding region coordinated to the principal Ca1 ion in only two of the three CRDs with ligand being absent in chain A. This absence of ligand in chain A is due to the restricted crystal contact symmetry impeding entry of ligand. Although the structure of SP-D is found to be flexible, allowing multiple binding to pathogen ligands, little structural deviation is seen between the rfhSP-D with bound 4A oligosaccharide and the native rfhSP-D structure [Shrive *et al.*, 2003, 2009].

The electron density map reveals that the binding of the Eagan 4A ligand takes place via the C₆ and C₇-OH groups of the glycerol side chain of the heptose moiety in a calcium dependent manner. The spatial distribution between the side chain of heptose C₆ and C₇-OH groups and vicinal C₃ and C₄-OH groups of pyranose rings seen in previous structures [Shrive *et al.*, 2003; 2009] are found to be identical. Several conclusions can be drawn from this work; as we know that SP-D has a preference for binding to a mannose unit via the vicinal equatorial C₃- and C₄-OH groups. Eagan 4A LPS is recognised by SP-D via the inner core heptose unit despite the availability of terminal Glc bearing the vicinal equatorial C₃- and C₄-OH. The electron density map also reveals the first ever novel three dimensional structure of the rearranged five membered *anhydro* Kdo obtained from the β -elimination of the phosphate group, a widely accepted and known mechanism for acid hydrolysis of *H. influenzae* LPS.

The electron density of the Eagan 4A clearly shows the Hep-Kdo(*anhydro*) unit and absence of the Glc unit, but this does not necessarily imply that Glc is not attached to the structure, but could indicate that there is the lack of steric hindrance in the CRD binding pocket and the Glc has enough freedom of rotation and thus does not show up in the

electron density. The work presented here shows that the rfhSP-D ligand interaction is via the CRD and is dependent on the accessibility of the inner core heptose. The native Eagan oligosaccharide is a much larger molecule with extended oligosaccharide (and a terminal Glc), which may interfere with this recognition.

This demonstrates how important the mechanism of binding of the CRDs, revealed by the crystal structure is for efficient recognition by the CRD of SP-D. Strategies designed to disrupt LPS synthesis in *Haemophilus* and other Gram negative organisms could thus be very effective in rendering bacteria susceptible to host innate immune defences mediated by collectins.

7.2 Future work

One of the main obstacles in crystallisation trials using LPS is the ability to get the LPS homogeneous in crystallisation conditions; the method employed in this work is dilipidation of the lipid A-Kdo linkage using acetic acid hydrolysis, which causes rearrangement of the Kdo into the five membered *anhydro* Kdo. One of the disadvantages of this rearrangement in terms of ligand binding studies is that this rearrangement mechanism causes the loss of the C₆ and C₇-OH groups of the Kdo glycerol side chain that is stereochemically identical to the heptose C₆ and C₇-OH groups of the glycerol side chain. Retention of these hydroxyls would enable an investigation to determine if there is a preference for binding by SP-D with the inner core heptose or Kdo unit. Several approaches can be applied such as O-deacylation with anhydrous hydrazine that will remove ester linked fatty acids responsible for the insolubility from the lipid A moiety.

This approach leads to a number of avenues for structural analysis; O-deacylation allows retention of sialic acid residues that are recognised by innate immune proteins and secondly, the LPS could be made soluble in crystallisation trials. Using chemotypes of LPS such as Ra, Rb, Rc, Rd, Re, feasibility studies can be carried out to see how the heterogeneous mixture of *Limulus* CRP recognises any particular forms of LPS and as a result filters out a particular form of CRP by precipitation or crystallisation.

Pathogen recognition by innate immune proteins has focused on using the recombinant hSP-D fragment rfhSP-D. In the physiological conditions in the human body environment, however, recognition of such pathogens may not only be by the CRD as demonstrated in this work, but may involve other aspects of the protein such as the lipid A with other parts of the pathogen such as capsular polysaccharide or teichoic acid.

References

- Adderson E. E., Byington C., Spencer L., Kimbal A., Hindiyeh M., Carroll K., *et al.* (2001) Invasive serotype *Haemophilus influenzae* infections with a virulence genotype resembling *Haemophilus influenzae* type b: emerging pathogen in the vaccine era. *Paediatrics* **108**, 1-6.
- Agrawal A. and Volanakis J. E. (1994) Probing the C1q-binding site on human C reactive protein by site directed mutagenesis. *J. Immunol.* **152**, 5404-5410.
- Agrawal A., Shrive A. K., Greenhough T. J. and Volanakis J. E. (2001) Topology and structure of the C1q-binding site on C-reactive protein. *J. Immunol.* **166**, 3998-4004.
- Alexander C. and Rietschel E.T. (2001) Bacterial lipopolysaccharide and innate immunity. *J. Endotoxin Res.* **7**, 167-202.
- Allen M J., Laederach A., Reilly P J. and Mason R J. (2001) Polysaccharide recognition by surfactant protein D: Novel interactions of a C-type Lectin with nonterminal glucosyl residues. *Biochem.* **40**, 7789-7798.
- Allen M J., Laederach A., Reilly P J. Mason R J. and Voelker R. (2004) Arg343 in human surfactant protein D governs discrimination between glucose and N-acetylglucosamine ligands. *Glycobiol.* **14**, 693-700.
- Allen M. J., Voelker D. R. and Mason R. J. (2001) Interactions of surfactant protein D with *Saccharomyces cerevisiae* and *Aspergillus fumigates*. *Infect. Immunol.* **69**, 2037-2044.
- Al-Tahhan R. A., Sandrin T. R., Bodour A. A. and Maier R. M. (2000) Rhamnolipid induced removal of lipopolysaccharide from *Pseudomonas aeruginosa*: Effect on cell

surface properties and interaction with hydrophobic substrates. *Applied and Environ. Microbiol.* **66**, 3262-3268.

Amor K., Heinrichs D. E., Fridrich E., Ziebell K., Johnson R. P. and Whitfield C. (2000) Distribution of core oligosaccharide types in lipopolysaccharides from *Escherichia coli*. *Infection and Immun.* **68**, 1116-1124.

Anderson M. S., and Raetz C. R. H. (1987) Biosynthesis of lipid A precursor in *Escherichia coli*: a cytoplasmic acyltransferase that converts UDP-acetylglucosamine to UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamine. *J. Biol. Chem.* **262**, 5159-516.

Apicella M. A., Dudas K. C., Campagnari A., Rice P., Mylotte J. M. and Murphy T. F. (1985) Antigenic heterogeneity of lipid A of *Haemophilus influenzae*. *Infect. Immun.* **50**, 9-14.

Apicella M. A., Griffis J. M. and Schneider H. (1994) Isolation and characterisation of lipopolysaccharides, lipooligosaccharides, and lipid A. *Methods in Enzymol.* **235**, 242-252.

Aplin J. and Hall L D. (1980) Sepharose 4B as a matrix for affinity chromatography: A spin-labelling investigation using nitroxides as model ligands. *Eur. J. Biochem.* **110**, 295-309.

Auzanneau F. I., Charon D. and Szabo L. (1991) Phosphorylated sugars. Part 27. Synthesis and reactions, in acid medium, of 5-O-substituted methyl 3-deoxy- α -D-manno-oct-2-ulopyranosidonic acid 4-phosphates. *J. Chem. Soc. Perkin Trans. 1*, 509-517.

- Berman S., Gewurz H. and Mold C. (1986) Binding of C-reactive protein to nucleated cells leads to complement activation without cytolysis. *J. Immunol.* **136**, 1354-1359.
- Billal S. D., Hotomi M., Suzumoto M., Yamauchi K., Kobayashi I., Fujihara K. and Yamanaka N. (2007) Rapid identification of nontypeable and serotype d *Haemophilus influenzae* from nasopharyngeal secretions by multiplex PCR. *Int. J. Pediatr. Otorhinolaryngology* **71**, 269-274.
- Blow D. M. Outline of crystallography for Biologists. Oxford University press 2002.
- Bone R. C. (1991) The pathogenesis of sepsis. *Ann Intern. Med.* **115**, 457-469
- Brown-Ausberger P., Hartshorn K., Chang D., Rust K., Flisher C., Welgus H. G. and Crouch E. C (1996) Site-directed mutagenesis of Cys15 and Cys20 of pulmonary surfactant protein D. Expression of a trimeric protein with altered anti-viral properties. *J. Biol. Chem.* **271**, 13724-13730.
- Brouwer N., Dolman K. M., van Zwieten R., *et al.*, (2006) Mannan-binding lectin (MBL)-mediated opsonisation is enhanced by the alternative pathway amplification loop. *Mol. Immunol.* **43**, 2051-2060.
- Brunger A. T. (1992) Free R-value-a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475
- Brunger A.T., Adams P.D., Clore G.M., DeLano W.L., Gros P., Grosse-Kunstleve R.W., Jiang J.S., Kuszewski J., Nilges M, Pannu N.S., Read R.J., Rice L.M., Simonson T. &

- Warren G.L. (1998) Crystallography and NMR System: a new software suite for macromolecular structure determination. *Acta Cryst.* **D54**, 905-921.
- Bulawa C. E. and Raetz C. R. H. (1984) The biosynthesis of gram-negative endotoxin: identification and function of UDP-2,3-diacylglucosamine in *Escherichia coli*. *J. Biol. Chem.* **259**, 4846-4851.
- Burrows L. I. and Lam J. S. (1999) Effects of *wzx* (*rfbX*) mutations on A-band and B-band lipopolysaccharide biosynthesis in *Pseudomonas aeruginosa* 05. *J. Bacteriol.* **181**, 973-980.
- Buttery J. And Moxon E. R. (2002) Capsulate bacteria and the lung. *Br. Med. Bull.* **61**, 63-80.
- Bystrova O. V., Linder B., Moll H., Kocharova N. A., Zahringer U. and Pier, G. B. (2004) Full structure of the lipopolysaccharide of *Pseudomonas aeruginosa* immunotype 5. *Biochem. (Mosc)* **69**, 170-175.
- Calandra T., Baumgartner J. D., Grau G. E., Wu M. M., Lambert P. H., Schellekens J., Verhoef J. and Glauser M. P. (1990) Prognostic values of tumor necrosis factor/cachectin, interleukin 1, interferon-alpha and interferon gamma in the serum of patients with septic shock. Swiss-Ditch J5 immunoglobulin study group. *J. Infect. Dis.* **161**, 982-987
- Campagnari A. A., Gupta M. R., Dudas K. C., Murphy T. F. and Apicella M. A. (1987) Antigenic diversity of lipopolysaccharide of nontypeable *Haemophilus influenzae*. *Infect. Immun.* **55**, 882-887.

Collaborative Computational Project Number 4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760-763.

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. and Reid K. (2001) Structural requirements for SP-D function in vitro and in vivo: Therapeutic potential of recombinant SP-D. *Immunobiol.* **205**, 619-631.

Clark H. and Reid K. (2003) The potential of recombinant surfactant protein D therapy to reduce inflammation in neonatal chronic lung disease, cystic fibrosis, and emphysema. *Arc. Dis. Child.* **88**, 981-984.

Claus D. R., Osmand A. P. and Gewurz H. (1976) Radioimmunoassay of human C-reactive protein and levels in normal sera. *J. Lab. Clin. Med.* **87**, 120-128.

Cohen S. L. and Chait B. T. (1996) Influence of matrix solution condition on the MALDI-MS analysis of peptide and proteins. *Anal. Chem.* **68**, 31-37.

Colley K. J. and Baenziger J. U. (1987) Identification of the post translational modifications of the core specific lectin. The core specific lectin contains hydroxyproline, hydroxylysine, and glucosylgalactosylhydroxylysine residues. *J. Biol. Chem.* **262**, 10290-10295.

Cox A. D., Hood D. W., Martin A., Makepeace K. M., Deadman M., Li, J., Brisson J.-R., Moxon E. R. and Richards J. C. (2001) Structural analysis of the lipopolysaccharide from the non-typeable *Haemophilus influenzae* strain SB 33. *Eur. J. Biochem.* **268**, 5278-5286.

- Crouch E. C. (2000) Surfactant protein-D and pulmonary host defense. *Respir. Res.* **1**, 93-108.
- Crouch E. C., Persson A., Chang D. and Heuser J. (1994a) Molecular structure of pulmonary surfactant protein D (SP-D). *J. Biol. Chem.* **269**, 17311-17319.
- Crouch E. C., Chang D., Rust K., Persson A. and Heuser J. (1994b) Recombinant pulmonary surfactant protein D. Post translational modification and molecular assembly. *J. Biol. Chem.* **269**, 15808-15813.
- Crouch E. C., Hartshorn K., Horlacher T., McDonald B., Smith K., Carafella T., *et al.* (2009) Recognition of mannosylated ligands and influenzae a virus by human lung surfactant protein D: contributions of an extended site and residue 343. *Biochem.* **48**, 3335-3345.
- Crouch E. C., Parghi D., Kuan S. F. and Persson A. (1992) Surfactant protein D: sub cellular localisation in nonciliated bronchiolar epithelial cells. *Am. J. Physiol.* **263** (pt. 1): L60-L66.
- Crouch E. C., Rust K., Mariencheck W., Parghi D., Chang D. and Persson A. (1991) Developmental expression of pulmonary surfactant protein D (SP-D). *Am. J. Respir. Cell Mol. Biol.* **5**, 13-8.
- Crouch E. C., Rust K., Veile R., Donis-Keller H. and Grosso L. (1993) Genomic organisation of human surfactant protein D (SP-D). SP-D is encoded on chromosome 10q22.2-23.1. *J. Biol. Chem.* **268**, 2976-2983.

- Crouch E C., Tu Y., Briner D., McDonald B., Smith K., Holmskov U. and Hartshorn K. (2005) Ligand Specificity of human surfactant protein: Expression of a mutant trimeric collectin that shows enhanced interactions with influenzae A virus. *J. Biol. Chem.* **280**, 17046-17056.
- Crouch E. C. and Wright J. R. (2001) Surfactant proteins A and D and pulmonary host defense. *Annu. Rev. Physiol.* **63**, 521-554.
- Cryz S. J. Jr., Pitt T., Furer E. and Germanier R. (1984) Role of lipopolysaccharide in virulence of *Pseudomonas aeruginosa*. *Infect. Immun.* **44**, 508-513.
- Cuatrecasas P. (1970) Protein purification by affinity chromatography: derivatization of agarose and polyacrylamide beads. *J. Biol Chem.* **245**, 3059-3065.
- Cuatrecasas P. and Anfinsen C B. (1971) Affinity chromatography. *Ann. Rev. Biochem.* **40**, 259-278.
- Cuatrecasas P., Wilcheck M. and Anfinsen C B. (1968) Selective enzyme purification by affinity chromatography. *Biochem.* **61**, 636-643.
- Dahl M., Juvonen P. O., Holmskov U. and Husby S. (2005) Surfactant protein D in new born infants: Factors influencing surfactant protein D levels in umbilical cord blood and capillary blood. *Pediatr. Res.* **58**, 908-912.
- Deadman M. E., Hermant P., Engskog M., Makepeace K., Moxon E. R., Schweda E. K. and Hood D. W. (2009) Lex2B, a phase-variable glycotransferase, adds either a glucose or

a galactose to *Haemophilus influenzae* lipopolysaccharide. *Infection Immunity* **77**, 2376-2384.

De Kievit T. R. and Lam J. S. (1994) Monoclonal antibodies that distinguish inner core, outer core, and lipid A regions of *Pseudomonas aeruginosa* lipopolysaccharide. *J. Bacteriol.* **176**, 7129-7139.

Distler A. M. and Allison J. (2001) 5-Methoxysalicylic acid and spermine: a new matrix for the matrix assisted laser desorption/ionization mass spectrometry analysis of oligonucleotide. *J. Am. Soc mass spectrum* **12**, 456-462.

Dong A., Caughey W. S. and DuClos T. W. (1994) Effects of calcium, magnesium, and phosphorylcholine on secondary structures of human C-reactive protein and serum amyloid P component observed by infrared spectroscopy. *J. Biol. Chem.* **269**, 6424-6430.

Dong Q. and Wright J. R. (1996) Expression of C-reactive protein by alveolar macrophages. *J. Immunol.* **156**, 4815-4820.

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

Drickamer K., Dordal M. S. and Reynolds L. (1986) Mannose binding protein isolated from rat liver containing carbohydrate recognition domains linked to collagenous tails. Complete primary structures and homology with pulmonary surfactant apoprotein. *J. Biol. Chem.* **261**, 6878-6887.

- Dulkerian S. J., Gonzales L. W., Ning Y. and Ballard P. L. (1996) Regulation of surfactant protein D in human fetal lung. *Am. J. Respir. Cell Mol. Biol.* **15**, 781-786.
- Ealick S. and Walter R. (1993) Synchrotron beamlines for macromolecular crystallography. *Curr. Opin. Struct. Biol.* **3**, 725-736.
- El-Aneed A., Cohen A. and Banoub J. (2009) Mass spectrometry, review of the basics: Electrospray, MALDI and commonly used mass analysers. *Appli. Spectros. Rev.* **44**, 210-230.
- Emsley J., White H. E., O'Hara B. P., Oliva G., Srinivasan N., Tickle I. J., Blundell T. L., Pepys M. B. and Wood S. P. (1994) Structure of pentameric human serum amyloid P component. *Nature* **367**, 338-345.
- Faden H., Duffy L., Williams A., Krystofik D. A. and Wolf J. (1995) Epidemiology of nasopharyngeal colonisation with *Haemophilus influenzae* in the first 2 years of life. *J. Infect. Dis.* **172**, 132-135.
- Falla T. J., Crook D. W. M., Brophy L. N., Maskel D., Kroll, J. S. and Moxon E. R. (1994) PCR for capsular typing of *Haemophilus influenzae*. *J. Clin. Microbiol.* **32**, 2382-2386.
- 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. *J. Immunol.* **163**, 312-321.

- Ferguson J., Voelker D. R., Ufnar J. A., Dawson A. J. and Schlesinger L. S. (2002) Surfactant protein D inhibition of human macrophage uptake of *mycobacterium tuberculosis* is independent of bacterial agglutination. *J. Immunol.* **168**, 1309-1314.
- Fernandez-Moran H., Marchalonis J. J. and Edelman G. M. (1968) Electron microscopy of a hemagglutinin from *Limulus polyphemus*. *J. Mol. Biol.* **32**, 467-469.
- Fischer W., Markwitz S. and Labischinski J. (1997) Small-angle X-ray scattering analysis of pneumococcal lipoteichoic acid phase structure. *Eur. J. Biochem.* **244**, 913-917.
- Fleischmann R. D., Adams M. D., White O., Clayton R. A., Kirkness E. F., Kerlavage A. R., Bult C. J., Tomb J. F and *et al.* (1995) Whole genome random sequencing and assembly of *Haemophilus influenzae* rd. *Science* **269**, 496-512.
- Floros J. and Hoover R. R. (1998) Genetics of the hydrophilic surfactant proteins A and D. *Biochim. Biophys. Acta.* **1408**, 312-322.
- Foxwell A. R., Kyd J. M. and Cripps A. W. (1998) Nontypeable *Haemophilus influenzae*: Pathogenesis and prevention. *Microbiol Mol. Biol. Rev.* **62**, 294-308.
- Friedrich E. and Whitfield C. (2005) Lipopolysaccharide inner core oligosaccharide structure and outer core stability in human pathogens belonging to the Enterobacteriaceae. *J. Endotoxin Res.* **11**, 133-144.
- Fujita T. (2002) Evolution of the lectin-complement pathway and its role in innate immunity. *Nat.Rev. Immunol.* **2**, 346-353.

- Gabay C. and Kushner I. (1999) Acute phase proteins and other systemic response to inflammation. *N. Engl. J. Med.* **340**, 448-454.
- Galanos C., Luderritz O. and Westphal O. (1969) A new method for the extraction of R lipopolysaccharide. *Eur. J. Biochem.* **9**, 245-249.
- Ganapathi M. K., Rzewnicki D., Samols D., Jiang S. L., Kushner I. (1991) Effects of combinations of cytokines and hormones on synthesis of serum amyloid A and C-reactive protein in hep 3B cells. *J. Immunol.* **147**, 1261-1265.
- Gernert K. N., Smith R. and Carter D. (1988) A simple apparatus for controlling nucleation and size in protein crystal growth. *Anal. Biochem.* **168**, 141-147.
- Gilsdorf J. R. (1998) Minireview: Antigenic diversity and gene polymorphism in *Haemophilus influenzae*. *Infect. Immun.* **66**, 5053-5059.
- Glusker J. P. and Trueblood K. N. Crystal Structure Analysis A primer. Oxford University Press, 1985,
- Goldsby R A., Kindt T. A., Osborne B. A. and Kubay J. (2003) *Immunology*. W. H. Freeman and Company, New York.
- Gotschlich E. C. and Edelman G. M. (1965) C-reactive protein: A molecule composed of subunits. *Proc. Natl. Acad. Sci. USA*, **54**, 566.
- Griffin R. Cox, A. D., Makepeace K., Richards, J. C., Moxon E. R. and Hood D.W. (2003) The role of *lex2* in lipopolysaccharide biosynthesis in *Haemophilus influenzae* strain RM7004 and RM153. *Microbiol.* **149**, 3165-3175.

- Groenveld K., van Alpen L., Eijk P. P., Visschers G., Jansen H. M. and Zanen H. C. (1990) Endogenous and exogenous reinfection by *Haemophilus influenzae* in patients with chronic obstructive pulmonary disease: The effect of antibiotic treatment on persistence. *J. Infect. Dis.* **161**, 512-517.
- Gruner S. M. (1994) X-ray detectors for macromolecular crystallography. *Curr. Opin. Struct. Biol.* **4**, 765-769.
- Hailman E., Lichenstein H. S., Wurfel M. M. *et al.* (1994) Lipopolysaccharide (LPS) binding protein accelerates the binding of LPS to CD14. *J. Exp. Med.* **179**, 269-277.
- Hakansson K. and Reid K. B. M. (2000) Collectin structure: A review. *Protein Science*, 1607-1717
- Hakansson K., Lim N. K., Hoppe H. J. and Reid K. B. (1999) Crystal structure of the trimeric alpha-helical coiled coil and the three lectin domains of human lung surfactant protein D. *Struct. Fold. Des.*, **7**, 255-264.
- Hanayama R., Tanaka M., Miyasaka K., Aozasa K., Koike M., Uchiyama Y. and Nagata S. (2004) Autoimmune disease and impaired uptake of apoptotic cells in MFG-E8 deficient mice. *Science*, **304**, 1147-1150.
- Hancock R. E. W., Mutharia L. M., Chan L., Darveau R. P., Speert D. P. and Pier G. B. (1983) *Pseudomonas aeruginosa* isolates from patients with cystic fibrosis: A class of serum sensitive, nontypeable strains deficient in lipopolysaccharide O side chains. *Infect. Immun.*, **42**, 170-177.

Hartshorn K. L., Crouch E. C., White M. R., Eggleton P., Tauber A. J., Chang D., Sastry N. (1994) Evidence for protective role of pulmonary surfactant protein D (SP-D) against influenza A viruses. *J. Clin. Invest.* **94**, 311-319.

Hartshorn K., Crouch E., White M R., *et al.*, (1998) Pulmonary surfactant proteins A and D enhance neutrophil uptake of bacteria. *Am. J. Physiol.* **274**, L958-L969.

Hartshorn K. L., White M. R., Shepherd V. L., Reid K. B. M., Jensenius J. C., Crouch E. C. (1997) Mechanism of anti influenza activity of surfactant protein A and D: comparison with serum collectins. *Am. J. Physiol.* **273**, L1156-L1166.

Hartshorn K., White M R., Voelker D., Coburn J P., Zaner K S. and Crouch E C. (2000) Mechanism of binding of Surfactant Protein D to influenza A viruses: Importance of binding to haemagglutinin to antiviral activity. *Biochem. J.* **351**, 449-458.

Hartshorn K., White M R., Tecle T., *et al.*, (2007) reduced influenza viral neutralizing activity of natural human trimers of surfactant protein D. *Resp. Res.* **8**, 1-10.

Harvey D. J. (2003) Matrix assisted laser desorption/ionization mass spectrometry of carbohydrates and glycoconjugates. *Int. J. Mass Spectrom.* **226**, 1-35.

Haziot A., Chen S., Ferrero E., Low M. G., Silber R. and Goyert S. M. (1988) The monocyte differentiation antigen, CD14, is anchored to the cell membrane by phosphatidylinositol linkage. *J. Immunol.* **141**, 547-552.

- Head J. F., Mealy T. R., McCormack F. X. and Seaton B. A. (2003) Crystal structure of trimeric carbohydrate recognition and neck domains of surfactant protein A. *J. Biol. Chem.* **278**, 43254-43260.
- Heindrichs D. E., Yethon J. A. and Whitfield C. (1998) Molecular basis for structural diversity in the regions of lipopolysaccharides of *Escherichia coli* and *Salmonella enteric*. *Mol. Microbiol.*, **30**, 221-232.
- Helliwell J. R. (1992) Macromolecular crystallography with synchrotron radiation. Cambridge University Press.
- 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*. *Eur. J. Immunol.* **29**, 3478-3848.
- Hickling T. P., Clark H., Malhotra R., and Sim R. B. (2004) Collectins and their role in lung immunity. *J. Leukocyt Biol.* **75**, 27-33.
- High N. J., Deadman M. E. and Moxon E. R. (1993) The role of a repetitive DNA motif (5'-CAAT-3') in the variable expression of the *Haemophilus influenzae* lipopolysaccharide epitope gal- α (1-4) β Gal. *Mol. Microbiol.* **9**, 1275-1282.
- Hind C. R. K., Collins P. M., Baltz M. L. and Pepys M. B. (1985) Human serum amyloid P component, a circulating lectin with specificity for the cyclic pyruvate acetal of galactose. *Biochem. J.* **225**, 107-111.

Hirschfield G. M. and Pepys M. B. (2003) C-reactive protein and cardiovascular disease: New insights from an old molecule. *Q. J. Med.*, **97**, 797-807.

Hoffman E and Stroobant U. (2007) Mass spectrometry principals and applications. 3rd Ed. John Wiley & Sons.

Holmskov U., Mollenhauer J., Madsen J., Vitvel L., Gronlund J., Torne I., Kliem A., Reid K. B. M., Poustka A. and Skojdt L. (1999) Cloning of gp340 a putative opsonin receptor for lung surfactant protein D. *Proc. Natl. Acad. Sci. USA*. **96**, 10794-10799

Holmskov, U., Thiel S. and Jensenius J. C. (2003) Collectins and ficolins: Humoral lectins of the innate immune defence. *Annu. Rev. Immunol.* **21**, 547-578.

Hood D. W., Cox A. D., Gilbert M., Makepeace K., Walsh S., Deadman M. E., Cody A., Martin A., Mansson M., Schweda E. K., Brisson J. R., Richards J. C., Moxon E. R. and Wakarchuk W. W. (2001) Identification of a lipopolysaccharide alpha-2,3-sialyltransferase from *Haemophilus influenzae*. *Mol. Microbiol.* **39**, 341-350.

Hood D. W., Deadman M. E., Jennings M. P., Bisercic M., Fleischmann R. D., Venter J. C. and Moxon E. R. (1996a) DNA repeats identify novel virulence genes in *Haemophilus influenzae*. *Proc. Natl. Acad. Sci. USA*. **93**, 11121-11125.

Hood D. W., Deadman M. E., Allen T., Masoud H., Martin A., Brisson J. R., Fleischmann R., Venter J. C., Richards J. C. and Moxon E.R. (1996b) Use of the complete genome sequence information of *Haemophilus influenzae* strain Rd to investigate lipopolysaccharide biosynthesis. *Mol. Microbiol.* **22**, 951-965.

- Hope H. (1990) Crystallography of biological macromolecules at ultra-low temperature. *Annual Rev. Biophys. Biophysical Chem.* **19**, 107-126.
- Hurlimann J., Thorbecke G. J. and Hochwald G. M. (1966) The liver as the site of C-reactive protein formation. *Experimental Medicine*, **123**, 365-378.
- Inzana T. J. (1983) Electrophoretic heterogeneity and interstrain variation of the lipopolysaccharide of *Haemophilus influenzae*. *J. Infect. Dis.* **148**, 492-499.
- Iwaka D., Osaki T., Mizunoe Y., Wai S. N., Iwanaga S. and Kawabata S. -I. (1999) Functional and structural diversities of C-reactive protein present in horseshoe crab hemolymph plasma. *Eur. J. Biochem.* **264**, 314-326.
- Jackman J. E., Fierke C. A., Tunny L. N., Pirrung M., Uchiyama T., Tahir S. H., Hindsgaul O., and Raetz C. R. H. (2000) Antimicrobial agents that target lipid A biosynthesis in gram negative bacteria: Inhibition of diverse UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosaminideacetylases by substrate analogs containing zinc binding motifs. *J. Biol. Chem.* **275**, 11002-11009.
- Jancarik J., Kim S-H. (1991) Sparse matrix sampling-a screening method for crystallisation of proteins. *J. Appl. Cryst.* **24**, 409-411.
- Janeway C. A Jr, Travers P, Walport M, *et al.* Immunobiology: The Immune System in Health and Disease. 5th edition. New York: Garland Science; 2001.
- Janeway C. A Jr. and Medzhitov R. (2002) Innate Immune Recognition. *Annu. Rev. Immunol.* **20**, 197-216

- Jones T. A., Zou J-Y., Cowan S. W. Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.* **A47**, 110-119.
- Kandil A. A., Chan N., Klein M. and Chang P. (1997) Chemical synthesis of *Haemophilus influenzae* glycopeptide conjugate. *Glycoconjugate J.* **14**, 13-17.
- Kaplan M. H. and Volanakis J. E. (1974) Interactions of C-reactive protein with the complement system: I. consumption of human complement associated with the reaction of C-reactive protein with pneumococcal polysaccharide and with the choline phosphatides, lecithin, and sphingomyelin. *Immunology* **112**, 2135-2147.
- Katyal S. L., Singh G., Locker J. (1992) Characterisation of a second human pulmonary surfactant-associated protein SP-A gene. *Am. J. Respir. Cell Mol. Biol.* **6**, 446-452.
- Kawasaki N., Kawasaki T., and Yamashina I. (1989) A serum lectin (mannan-binding protein) has complement-dependent bacterial activity. *J. Biochem.* **106**, 483-489
- Kim S. J., Gershov D., Ma X., Brot N. and Elkon K. B. (2003) Opsonisation of apoptotic cells and its effect on macrophage and T cell immune responses. *Ann. N. Y. Acad. Sci.* **987**, 68-78.
- Kindmark C. O. (1971) Stimulating effect of C-reactive protein on phagocytosis of various species of pathogenic bacteria. *Clin. Exp. Immunol.* **8**, 941-948.

- Kirpekar F., Nordhoff E., Arsen L. K., Kristiansen K., Roepstarff P. and Hillenkamp F. (1998) DNA sequence analysis by MALDI mass spectrometry. *Nucleic Acid Res.* **26**, 2554-2559.
- Kishore U., Greenhough T. J., Waters P., Shrive A. K., *et al.*, (2006) Surfactant proteins SP-A and SP-D: Structure, function and receptors. *Mol. Immunol.* **43**, 1293-1315.
- Kishore U. and Reid K. B. M. (2000) C1q: Structure, function and receptors. *Immunopharmacol.* **49**, 159-170.
- Kleywegt G. J. and Brunger A. T. (1996) Checking your imagination: applications of the free R value. *Structure* **4**, 897-904.
- Kleywegt G.J., Henrick K., Dodson E.J., & van Aalten D.M.F. (2003) Pound-wise but penny-foolish - How well do micromolecules fare in macromolecular refinement? *Structure* **11**, 1051-1059.
- Klosterhalfen B., Hortsman-Jungemann K., Vogel, P., Flohe S. and Offner F. (1992) Time course of various inflammatory mediators during recurrent endotoxemia. *Biochem. Pharmacol.* **43**, 2103-2109
- Knirel Y. A., Bystrova O., Shashkov A. S., Linder B., Kocharova N. A., Senchenkova S. N., Moll H., Zahringer U., Hatano K. and Pier G. B. (2001) Structural analysis of the lipopolysaccharide core of a rough, cystic fibrosis isolate of *Pseudomonas aeruginosa*. *Eur. J. Biochem.* **268**, 4708-4719.

- Koenraad van de Wetering J., van Eijk M., van Golde L. M. G., Hartung T., van Strijp J. A. G. and Batenberg J. J. (2001) Characterisation of surfactant protein A and D binding to lipoteichoic acid and peptidoglycan, 2 major cell wall components of gram positive bacteria. *J. Infect. Dis.* **184**, 1143-1151.
- Koenraad van de Wetering J., van Golde L. M. G. and Batenburg J. J. (2004) Collectins: Players of the innate immune system. *Eur. J. Biochem.* **271**, 1229-1249.
- Kolatkar A. R. and Weis W. I. (1996) Structural basis of galactose recognition by C-type animal lectins. *J. Biol. Chem.* **271**, 6679-6685.
- Kolbe K., Lu, J., Mole S. E., Kaluz S. and Reid K. B. (1993) Assignment of human pulmonary surfactant protein D gene (SFTP4) to 10q22-q23 close to the surfactant protein A gene cluster. *Genomics* **17**, 294-298.
- Korfhagen T. R., Sheftelyevich V., Burhans M S., Bruno M D., et al., (1998) Surfactant protein-D regulates surfactant phospholipid homeostasis *in vivo*. *J. Biol. Chem.* **273**, 28438-28443.
- Kraulis, P. J. (1991) MOLSCRIPT: A programme package to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946.
- Kroll J. S. (1992) The genetics of encapsulated *Haemophilus influenzae*. *J. Infect. Dis.* **165**, S93-S96.

- Kuan S. F., Rust K., Crouch E. (1992) Interaction of surfactant protein D with bacterial lipopolysaccharides. Surfactant protein D is an *Escherichia coli*-binding protein in bronchoalveolar lavage. *J. Clin. Invest.* **90**, 97-106.
- Kulshin V. A., Zahringer U., Lindner B., Jager K., Dmitriev B. A. and Rietschel E. T. (1991) Structural characterisation of the lipid A component of *Pseudomonas aeruginosa* wild type and rough mutant lipopolysaccharides. *Eur. J. Biochem.* **198**, 697-704.
- Kushner I and Kaokab M. H. (1961) Studies on acute phase protein. I. an immunohistochemical method for the localisation of C_x-reactive protein in rabbits: Association with necrosis in local inflammatory lesions. *J. Exp. Med.* **114**, 961-973.
- LaClair L. L., Tondella M. L. C., Beall D. S., Nople C. A., Raghunathan P. L. and Rosenstein N. E. (2003) The active bacterial core surveillance team members. identification of *Haemophilus influenzae* serotypes by standard slide agglutination serotyping and PCR-based capsular typing. *J. Clin. Microbiol.* **41**, 393-396.
- Lairson L. L., and Withers S. G. (2004) Mechanistic analogies amongst carbohydrate modifying enzymes. *Chem. Commun.* **20**, 2243-2248.
- Landerholm M. K., Li, J., Richards J. C., Hood D. W., Moxon E. R. and Schweda E. K. H. (2004) Characterisation of novel structural features in the lipopolysaccharide of nondisease associated nontypeable *Haemophilus influenzae*. *Eur. J. Biochem.* **271**, 941-953.

- Laskowski R., MacArthur M.W., Moss D.S. & Thornton J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.
- Lawson P. R. and Reid K. B. M. (2000) The roles of surfactant proteins A and D in innate immunity. *Immunol. Rev.* **173**, 66-78.
- Lee R. T., Ichikawa Y., Fay M., Drickamer K., Shao M. -C. and Lee Y. C. (1991) Ligand-binding characteristics of a rat serum type mannose-binding protein (MBP-A). homology of binding site architecture with mammalian and chicken hepatic lectins. *J. Biol. Chem.* **266**, 4810-4815.
- Lei K. J., Liu T., Zon G., Soravia E., Liu T.-Y. and Goldman, N. D. (1985) Genomic DNA sequence for human C-reactive protein. *J. Biol. Chem.* **260**, 13377-13383.
- Leslie A. G. W. (1993) Data collection and processing. in: Sawyer L., Isaac, N., Bailey, S. eds. proceedings of the CCP4 study weekend. Warrington, UK: SERC Daresbury laboratory, 44-51.
- Leslie, A.G.W. (1992) Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 and ESF-EACMB Newsletter on Protein. Crystallography No. 26. Daresbury Laboratory, Warrington, UK.
- Leslie A. G. W. (2006) The integration of macromolecular diffraction data. *Acta Cryst.* **D62**, 48-57.

- LeVine A. M., Whitsett J. A., Gwozdz J. A., Richardson T. R., Fisher J. H., Burhans M. S. and Korfhagen T. R. (2000) Distinct effects of surfactant protein A and D deficiency during bacterial infection on the lung. *J. Immunol.* **165**, 3934-3940.
- Li Y. P., Mold C. and Du Clos T. W. (1994) Sublytic complement attack exposes C-reactive protein binding sites on cell membranes. *J. Immunol.* **152**, 2995-3005.
- Li J. and Richards J. C. (2007) Application of capillary electrophoresis mass spectrometry to the characterization of bacterial lipopolysaccharide. *Mass Spectrom. Rev.* **26**, 35-50.
- Lim B. L., Willis A. C., Reid K. B. M., Lu, J., Laursen S. B., Jensinius J. C. and Holmskov U. (1994). Primary structure of bovine collectin-43 (CL-43) Comparison with conglutinin and lung surfactant protein D. *J. Biol. Chem.* **269**, 11820-11824.
- Lin Z. and Floros J. (2002) Heterogeneous allele expression of pulmonary SP-D gene in rat large intestine and other tissues. *Physiol. Genomics* **11**, 235-243.
- Liu P. V., Matsumoto H., Kusama H. and Bergan T. (1983) Survey of heat-stable, major somatic antigens of *Pseudomonas aeruginosa*. *Int. J. Syst. Bacteriol.* **33**, 256-264.
- Lovell S.C., Davis I.W., Bryan Arendall III W., Paul I. W., de Bakker J., Word M., Michael G., Prisant M.G., Richardson J.S. and Richardson D.C. (2003) Structure validation by C-alpha geometry: phi, psi, and C-beta deviation. *Proteins: Structure, Function, and Genetics.* **50**, 437-450.
- Luft J. R., Arakail S. V., Kirisits J. *et al.*, (1994) A macromolecular crystallisation procedure employing diffusion cells of varying depths as reservoirs to tailor the time

course of equilibration in hanging drop and sitting drop vapour diffusion and micro dialysis experiments. *J. Appl. Cryst.* **27**, 443-453.

Luong D. C., Ishiwada N., Takeda N. and Kohno Y. (2004) Serotypes of *Haemophilus influenzae* strains isolated from pediatric patients with respiratory tract infections. *J. Exp. Med.* **202**, 245-254.

Lysenko E., Richards J. C., Cox A. D., Stewart A., Martin A., Kapoor M. and Weiser J. N. (2000) The position of phosphorylcholine on the lipopolysaccharide of *Haemophilus influenzae* effects binding and sensitivity to C reactive protein-mediated killing. *Mol. Microbiol.* **35**, 234-245.

Mackiewicz A., Speroff T., Ganapathi M. K. and Kushner I. (1991) Effects of cytokine combinations on acute phase protein production in two human hepatoma cell lines. *J. Immunol.* **146**, 3032-3037.

Macleod C. M. and Avery O. T. (1941) Occurrence during acute phase infection of a protein not normally present in the blood (II). *J. Exp. Med.* **73**, 183-190.

Madan T., Kishore U., Singh M., Strong P., Hussain E. M., Reid, K. B. M. and Sarma P. U. (2001a) Protective role of lung surfactant protein D in a murine model of invasive pulmonary Aspergillosis. *Infect. Immunity* **69**, 2728-2731

Madan T., Kishore U., Singh M., Strong P., Clark H., Hussain E. M., Reid K. B. M. and Sarma P. U (2001b) Surfactant proteins A and D protect mice against hypersensitivity induced by *Aspergillus fumigatus* and allergens. *J. Clin. Invest.* **107**, 467-475.

Madan T., Kishore U., Shah A., Eggleton P., Strong P., *et al.*, (1997) Lung surfactant proteins A and D can inhibit specific IgE binding to the allergens of *Aspergillus fumigatus* and block allergen induced histamine release from human basophils. *Clin. Exp. Immunol.* **110**, 241-249.

Madsen J., Kliem A., Tornøe I., Skjodt K., Koch C. and Holmskov U. (2000) Localisation of lung surfactant protein D (SP-D) on mucosal surfaces in human tissues. *J. Immunol.* **164**, 5866-5870.

Mansson M., Bauer S. H. J., Hood D. W., Richards J. C., Moxon E. R. and Schweda E. K. H. (2001) A new structural type for *Haemophilus influenzae* lipopolysaccharide: Structural analysis of the lipopolysaccharide from nontypeable *Haemophilus influenzae* strain 486. *Eur. J. Biochem.* **268**, 2148-2159.

Mansson M., Hood D. W., Li, L., Richards J. C., Moxon E. R. and Schweda E. K. H. (2002) Structural analysis of the lipopolysaccharide from nontypeable *Haemophilus influenzae* strain 1003. *Eur. J. Biochem.* **269**, 808-818.

Mansson M., Hood D. W., Moxon E. R. and Schweda E. K. H. (2003) Structural diversity in lipopolysaccharide expression in nontypeable *Haemophilus influenzae*. Identification of L-glycero-D-manno-heptose in the outer core region in three clinical isolates. *Eur. J. Biochem.* **270**, 610-624.

Marchalonis J. J. and Edelman G. M. (1968) Isolation and characterisation of a hemagglutinin from *limulus polyphemus*. *J. Mol. Biol.* **32**, 453-465.

- Marnell L., Mold C. and Du Clos T. W. (2005) C-reactive protein: Ligands, receptors and role in inflammation. *Clinical Immunology* **117**, 104-111.
- Marolda C. L., Vicarioli J. and Valvano M. A. (2004) Wzx proteins involved in biosynthesis of O-antigen function in association with the first sugar of the O-specific polysaccharide subunit. *Microbiology* **150**, 4095-4105.
- Martich, G. D., Boujoukos, A. J. and Suffredini, A. F. (1993) Response of man to endotoxin. *Immunobiol.* **187**, 403-416
- Mason R. J., Nielson L. D., Kuroki Y., Matsuura E., Freed J. H. and Shannon J. M. (1998) A 50 kDa variant form of human surfactant protein D. *Eur. Respir. J.* **12**, 1147-1155.
- Masoud H., Martin A., Thibault P., Moxon E. R and Richardson J. C. (2003) Structure of extended lipopolysaccharide glycoform containing two globotriose units in the *Haemophilus influenzae* serotype b strain, RM7004. *Biochem.* **42**, 4463-4475.
- Masoud H., Moxon E. R., Martin A., Krajacarski D. and Richards J. C. (1997) Structure of the variable and conserved lipopolysaccharide oligosaccharide epitopes expressed by *Haemophilus influenzae* serotype b strain Eagan. *Biochem.* **36**, 2091-2103.
- Masoud H., Sadovskaya I., De-Kievit T., Altman E., Richards J. C. and Lam J. S. (1995) Structural elucidation of the lipopolysaccharide core region of the O-chain deficient mutant strain A28 from *Pseudomonas aeruginosa* serotype 06 (international antigenic typing scheme). *J. Bacteriol.* **177**, 6718-6726.

- Matsushita M. and Fijita T. (1992) Activation of the classical complement pathway by mannose binding protein in association with a novel C1s-like serine protease. *J. Exp. Med.* **176**, 1497-1502.
- Maudsley S., Rowe I. F., DeBeer F. C., Munn E. A., Herbert J., Feinstein A. and Pepys M. B. (1987) Identification and isolation of two pentraxins from bovine serum. *Clin. Exp. Med.* **67**, 662-673.
- McCormack F. X., Damodarasamy M. and Elhalwagi B. M. (1999) Deletion mapping of N-terminal domains of surfactant protein A. the N-terminal segment is required for phospholipid aggregation and specific inhibition of surfactant secretion. *J. Biol. Chem.* **274**, 3173-3181.
- McCormack F. X., Pattanajitvilai S., Stewart J., Possmayer F., Inchley K. and Voelker D. R. (1997) The Cys6 intermolecular disulphide bond and the collagen-like region of rat SP-A play critical roles in interactions with alveolar type II cells and surfactant lipids. *J. Biol. Chem.* **272**, 27971-27979.
- Meir M., Bider M. D., Malashkevich V. N., Spiess M. and Burkard P. (2000) Crystal structure of the carbohydrate recognition domain of the H1 subunit of the asialoglycoprotein receptor. *J. Mol. Biol.* **300**, 857-865.
- Meschi J., Crouch E. C., Skolnik P., Yaha K., Holmskov U., Leth-Larson R., Tornøe I., Tecle T., White M. R. and Harsthorn K. L. (2005) Surfactant protein D binds to human immunodeficiency virus (HIV) replication. *J. Gen. Virol.* **86**, 3097-3107.

- Mikhail I., Yildirim H. H., Lindahl E. C. H. and Schweda E. K. H. (2005) Structural characterisation of lipid A from non-typeable and type f *Haemophilus influenzae*: Variability of fatty acid substitution. *Analytical Biochem.* **340**, 303-316.
- Miyamura K., Malhotra R., Hoppe H. J., Reid K. B., Phizackerley P. J., Macpherson, P. and Lopez Bernal I. (1994) Surfactant proteins A (SP-A) and D (SP-D): Levels in human amniotic fluid and localisation in the fetal membranes. *Biochem. Biophys. Acta.* **1210**, 303-307.
- Mizuno Y., Kozutsumi Y., Kawasaki T., and Yamashina I. (1981) Isolation and characterisation of a mannan-binding protein from rat liver. *J. Biol. Chem.* **256**, 4247-4252.
- Mold C., Gewurz H., Du Clos T. W. (1999) Regulation of complement activation by C-reactive protein. *Immunopharmacology* **42**, 23-30.
- Mold C., Kingzette M. and Gewurz H. (1984) C-reactive protein inhibits activation of the alternative pathway by increasing the interaction between factor H and C3b. *J. Immunol.* **133**, 882-885.
- Mold C., Nakayama S., Holzer T. J., Gewurz H and Du clos T. W. (1981) C-reactive protein is protective against *Streptococcus pneumonia* infection in mice. *J. Exp. Med.* **154**, 1703-1708.
- Moxon E R. (1985) *Haemophilus influenzae*. In Principles and Practice of Infectious Disease, pp. 1274-1279. Edited by G. Mandrell, R. Douglas & J. Bennet. New York: Wiley.

Moxon E. R. and Rappuoli R. (1990) *Haemophilus influenzae* infections and whooping cough. *Lancet* **335**, 1324-1329.

Moxon E R., Smith A. L., Averill D R. and Smith D H. (1974) *Haemophilus influenzae* meningitis in infant rats after intranasal inoculation. *J. Infect. Dis.* **129**, 154-162.

Muhlemann K., Balz M., Aebi S. and Schofer K. (1996) Molecular characteristics of *Haemophilus influenzae* causing invasive disease during the period of vaccination in Switzerland: Analysis of strains isolated between 1986-1993. *J. Clin. Microbiol.* **34**, 560-563.

Murphy T. F. and Apicilla M. A. (1987) Nontypable *Haemophilus influenzae*: A review of clinical aspects, surface antigens, and the human immune response to infection. *Rev. Infect. Dis.* **9**, 1-15.

Nadesalingham J., Bernal A. L., Dodds A. W., Willis A. C., Mahoney D. J., Day A. J., Reid K. B. and Palaniyer N. (2003) Identification and characterisation of a novel interaction between pulmonary surfactant protein D and decorin. *J. Biol. Chem.* **278**, 25678-25687.

Nadesalingham J., Reid K. B. and Palaniyer N. (2005) Collectin surfactant protein D binds antibodies and interlinks innate and adaptive immune system. *FEBS Lett.* **579**, 4449-4453.

Narkates A. J. and Volanakis J. E. (1982) C-reactive protein binding specificities: Artificial and natural phospholipid bilayers. *Ann. N. Y. Acad. Sci.* **389**, 172-181.

- Nigam V. N. (1975) Effects of core lipopolysaccharide from *Salmonella Minnesota* R mutants on the survival times of mice bearing Ehrlich tumours. *Cancer Res.* **35**, 628-633.
- Ng K. K. S., Drickamer K. and Weis W. I. (1996) Structural analysis of monosaccharide recognition by rat liver mannose-binding protein. *J. Biol. Chem.* **271**, 663-674.
- Ng P. M. L., Jin Z., Tan S. S. H., Ho B. and Ding J. L. (2004) The C-reactive protein: A predominant LPS-binding acute phase protein responsive to pseudomonas infection. *J. Endotoxin Res.* **10**, 1-12.
- Ng K. K., Kolatkar A. R., Park-Snyder S., Feinberg H., Clark D. A., Drickamer K. and Weis W. I. (2002) Orientation of bound ligands in mannose-binding proteins. Implications for multivalent ligand recognition. *J. Biol. Chem.* **277**, 16088-16095.
- Nguyen N. Y., Suzuki A., Boykins R. A. and Liu T. Y. (1986a) The amino acid sequence of *Limulus* C-reactive protein: Evidence of polymorphism. *J. Biol. Chem.* **261**, 10456-10465.
- Nguyen N. Y., Suzuki A., Cheng S. M., Zon G. and Liu T. Y. (1986b) Isolation and characterisation of *Limulus* C-reactive protein genes. *J. Biol. Chem.* **261**, 10450-10455.
- Nikaido H. (2003) Molecular basis of bacterial outer membrane permeability revisited. *Microbiol. Mol. Biol. Rev.* **67**, 593-656.
- Oberley R. E., Goss K. L., Ault K. A., Crouch E. C. and Snyder J. M. (2004) Surfactant protein D is present in the human female reproductive tract and inhibits *Chlamydia trachomatis* infection. *Molecular Human Reproduction* **10**, 861-870.

- Odegaard T. J., Kaltashov I. A., Cotter R. J., Kaltashov I. A., Cotter R. J., Steeghs L., Vander Ley P., Khan S., Maskel D. J., and Raetz C. R. H. (1997) Shortened hydroxyl chains on lipid A of *Escherichia coli* cells expressing a foreign UDP-N-acetylglucosamine O-acetyltransferase. *J. Biol. Chem.* **272**, 19688-19696.
- Ofek I., Mesika A., Kalina M., Keisari Y., Podschun R., Sahly H., Chang D., McGregor D. and Crouch E. C. (2001) Surfactant protein D enhances phagocytosis and killing of unencapsulated phase variants of *Klebsiella pneumoniae*. *Infect. Immun.* **69**, 24-33.
- Ogasawara Y., Kuroki Y. and Akino T. (1992) Pulmonary surfactant protein D specifically binds to phosphatidylinositol. *J. Biol. Chem.* **267**, 21244-21249.
- Ogasawara Y. and Voelker D. R. (1995) Altered carbohydrate recognition specificity engineered into surfactant protein D reveals different binding mechanism for phosphatidylinositol and glucosylceramide. *J. Biol. Chem.* **270**, 14725-14732.
- Ohtani K., Suzuki Y., Eda S., Kawai T., Kase T., Keshi H., Sakai Y., Fukouh A., Sakamoto T., Itabe H., Suzutani T., Ogasawara M., Yoshida I, and Wakamiya N. (2001) The membrane type collectin CP-P1 is a scavenger receptor on vascular endothelial cells. *J. Biol. Chem.* **276**, 44222-44228.
- Ohtani K., Suzuki Y., Eda S., Kawai T., Kase T., Yamazaki H., Shimada T., Keshi H., Sakai Y., Fukuoh A., Sakamoto T., Wakamiya N. (1999) Molecular cloning of a novel human collectin from liver (CL-L1). *J. Biol. Chem.* **274**, 13681-13689.
- Oliveira E. B. Gotschlich C. and Liu T. Y. (1979). Primary structure of human C-reactive protein. *J. Biol. Chem.* **25**, 489-502.

- Osmand A. P., Friedensen B., Gewurz H., Painter R. H., Hofman T. and Sheneton E. (1977) Characterization of C-reactive protein and the complement subcomponent C1t as homologous proteins displaying cyclic pentameric symmetry (pentraxins). *Proc. Natl. Acad. Sci. USA.* **74**, 739-743.
- Palaniyer N., Clark H., Nadesalingham J., Shih M. J., Hawgood S. and Reid K. B. (2005) Innate immune collectin surfactant protein D enhances the clearance of DNA by macrophages and minimises anti-DNA antibody generation. *J. Immunol.* **174**, 7352-7358.
- Palsson-McDermott E. V. and O'Neil L. A. J. (2004) Signal transduction by the lipopolysaccharide receptor, Toll-like receptor 4. *Immunol.* **113**, 153-162
- Parrilo J. E. (1993) Pathogenic mechanism of septic shock. *N. Engl. J. Med.* **328**, 1471-1477.
- Pattle R. E. (1955). Properties and origins of the alveolar lining layer. *Nature* **175**, 1125-1126.
- Pepys M. B., Dash A. C., Fletcher T. C., Richardson N., Munn E. A. and Feinstein A. (1978) Analogues in other mammals and in fish of human plasma proteins C-reactive protein and amyloid P component. *Nature* **273**, 168-170.
- Persson A. V, Gibbons B. J., Shoemaker J. D., Moxley M. A. and Longmore W. J. (1992) The major glycolipid recognised by SP-D in surfactant is phosphatidylinositol. *Biochem.* **31**, 12183-12189.

- Poltorak A. He X., Smirnova I. *et al.* (1998) Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in the Tlr4 gene. *Science* **282**, 2085-2088.
- Price N. J., Kelly T. M., Raetz C. R. H. and Carlson R. W. (1994) Biosynthesis of a structurally novel lipid A in *Rhizobium leguminosarum*; identification and characterization of six metabolic steps leading to form UDP-GlcNAc to 3-Deoxy-manno-2-octulonic acid lipid IVA. *J. Bacteriol.* **176**, 4646-4655.
- Quigley J., Misquith S., Surolia A., Srimal S. and Armstrong P. (1994) Preliminary investigation of the molecular basis for the functional differences between the two pentraxins limulin and C-reactive protein from the plasma of the American horseshoe crab, *Limulus polyphemus*. *Biol. Bull.* **187**, 229-230
- Qureshi N., Takayama K., Mascagni P., Honovich J., Wong R. Cotter R. J. (1988) Complete structural determination of lipopolysaccharide obtained from deep rough mutant of *Escherichia coli*. *J. Biol. Chem.* **263**, 11971-11976.
- Raetz C. R., and Whitfield C. (2002) Lipopolysaccharide endotoxins. *Ann. Rev. Biochem.* **71**, 635-700.
- Rahman M. M., Gu X., Tsai C. M., Kolli V. S. K. and Carlson R. W. (1999) The structural heterogeneity of the lipoolysaccharide (LOS) expressed by pathogenic non-typeable *Haemophilus influenzae* strain NTHi 9274. *Glycobiol.* **9**, 1371-1380.
- Reinhold V. N., Reinhold B. B. and Chan S. (1996) Carbohydrate sequence analysis by electrospray ionization spectrometry. *Methods Enzymol.* **271**, 377-402

- Restrepo C. I., Dong Q., Savov J., Mariencheck W. I. and Wright J. R. (1999) Surfactant protein D stimulates phagocytosis of *Pseudomonas aeruginosa* by alveolar macrophages. *Am. J. Respir. Cell Mol. Biol.* **21**, 576-585.
- Reyes C. L. and Chang G. (2005) Lipopolysaccharide stabilizes the crystal packing of the ABC transporter MsbA. *Acta Cryst.* **F61**, 655-658
- Rhodes G. Crystallography made crystal clear, a guide for users of macromolecular models. San Diego academic press 2000.
- Ridker P. M. (2003) Clinical application of C-reactive protein for cardiovascular disease detection and prevention. *Circulation* **107**, 363-369.
- Risberg A., Schweda E.K. H. and Jansson P. (1997) Structural studies of the cell envelope oligosaccharide from the lipopolysaccharide of *Haemophilus influenzae* strain RM.118-28. *Eur. J. Biochem.* **243**, 701-707.
- Robey F. A and Liu T. Y. (1981) Limulin: A C-reactive protein from *Limulus polyphemus*. *J. Biol. Chem.* **256**, 969-975.
- Robey F. A. and Liu T. -Y. (1983) Isolation and characterisation of two major serum proteins from the dog fish, *Mustelus canis*, C-reactive protein and amyloid P component. *J. Biol. Chem.* **258**, 3889-3894.
- Robey F. A., Jones K. D. and Steinberg A. D. (1985) C-reactive protein mediates the solubilisation of nuclear DNA by complement *in vitro*. *J. Exp. Med.* **161**, 1344-1356.

- Rossman, M. G. and Blow, D. M. (1962) The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst.* **15**, 24-32
- Rothman A. B., Mortensen H. D., Holmskov U. and Hojrup P. (1997) Structural characterisation of bovine collectin-43. *Eur. J. Biochem.* **243**, 630-635.
- Sadovskya I., Brisson J., Thibault P., Richards J. C., Lam J. S. and Altman E. (2000) Structural characterisation of the outer core and the O-chain linkage region of lipopolysaccharide from *Pseudomonas aeruginosa* serotype 05. *Eur. J. Biochem.* **267**, 1640-1650.
- Sahly H., Ofek I., Podschun R., Brade H., He Y., Ullmann U. and Crouch E. C. (2002) Surfactant protein D binds selectively to *Klebsiella pneumonia* lipopolysaccharides containing mannose-rich O-antigen. *J. Immunol.* **169**, 3267-3274.
- Sano H., Chiba H., Iwaka D., Sohma H., Voelker D. R., and Kuroki Y. (2000) Surfactant proteins A and D bind CD14 by different mechanism. *J. Biol. Chem.* **275**, 22442-22451.
- Sano H., Kuronuma K., Kudo K., Mitsuzawa H., Sato M., Murakami S., Kuroki Y. (2006) Regulation of inflammation and bacterial clearance by lung collectins. *Respirology* **11** (Suppl.), S46-S50.
- Sato T., Endo Y., Matsushita M. and Fujita T. (1994) Molecular characterisation of a novel serine protease involved in activation of the complement system by mannose-binding protein. *Int. Immunol.* **6**, 665-669.

- Saville J., Dransfield I., Gregory C. and Haslett C. (2002) A blast from the past: Clearance of apoptotic cells regulates immune responses. *Nat. Rev. Immunol.* **2**, 965-975.
- Schaeffer L. M., McCormack F. X., Wu H. and Weiss A. A. (2000) Interactions of pulmonary collectins with *Bordetella bronchiseptica* and *Bordetella pertussis* lipopolysaccharides elucidate the structural basis of their antimicrobial activities. *Infect. Immunity* **72**, 7124-7130.
- Schiller N. L. (1988) Characterisation of the susceptibility of *Pseudomonas aeruginosa* to complement killing: Role of antibodies to the rough lipopolysaccharide on serum sensitive strains. *Infect. Immunity*. **56**, 632-639.
- Schweda E. K. H., Brisson J., Alvelius G., Martin A., Weiser J. N., Hood D. W., Moxon E. R. and Richards J. C. (2000) Characterisation of the phosphocholine-substituted oligosaccharide in lipopolysaccharides of type b *Haemophilus influenzae*. *Eur. J. Biochem.* **267**, 3902-3913.
- Schweda E. K. H., Richards J. C., Hood D. W. and Moxon E. R. (2007) Expression and structural diversity of the lipopolysaccharide of *Haemophilus influenzae*: Implication in virulence. *Int. J. Microbiol.* **297**, 297-306.
- Shapira L., Aubrey Soskoline W., Houri Y., Barak V., Halabi A. and Stabholz A. (1996) Protection against endotoxic shock and lipopolysaccharide induced local inflammation by tetracycline: correlation with inhibition of cytokine secretion. *Infect. Immunity* **64**, 825-828.

- Sheriff S., Chang C. Y. and Ezekowitz R. A. (1994) Human mannose-binding protein carbohydrate recognition domain trimerises through a triple alpha-helical coiled coil. *Nat. Struct. Biol.* **1**, 789-794.
- Shively R. G., Shigi J. T., Peterson E. M. and De La Maza L. M. (1981) Typing of *Haemophilus influenzae* by coagglutination and conventional slide agglutination. *J. Clin. Microbiol.* **14**, 706-708.
- Shrive A. K., Burns I., Chou H., Stahlberg H., Armstrong P. B. and Greenhough T. J. (2009) Crystal structure of Limulus SAP-like pentraxin reveal two molecular aggregations. *J. Mol. Biol.* **386**, 1240-1254.
- Shrive A. K., Cheetham G. M., Holden D., Myles D. A., Turnell W. G., Volanakis J. E., Pepys M. B., Bloomer A. C. and Greenhough T. J. (1996) Three dimensional structure of human C-reactive protein. *Nat. Struct. Biol.* **3**, 346-354
- Shrive A. K., Metcalfe A. M., Cartwright R. R. and Greenhough T. J. (1999) C-reactive protein and SAP-like pentraxin are both present in *Limulus polyphemus* haemolymph: Crystal structure of *Limulus* SAP. *J. Mol. Biol.* **290**, 997-1008.
- Shrive A. K., Tharia H. A., Strong P., Kishore U., Burns, I., Rizkallah P. J., Reid K. B. M. and Greenhough T. J. (2003) High resolution structural insights into ligand binding and immune cell recognition by human lung surfactant protein D. *J. Mol. Biol.* **331**, 509-523.
- Sjoberg A. P., Trouw L. A. and Blom A. M. (2008) Complement activation and inhibition: a delicate balance. *Trends in immunol.* **30**, 83-89.

Smyth M. S and Martin, J. H. J. (2000) X ray crystallography. *J. Clin. Path. Mol. Path.* **53**, 8-14.

Solomons organic chemistry (1992) 5th edition, John Wiley and Sons.

Sorensen G. L., Husby S. and Holmskov U. (2007) Surfactant protein A and surfactant protein D variations in pulmonary disease. *Immunobiol.* **212**, 381-416.

Spinola S. M., Peacock J., Denny F. W., Smith D. L. and Cannon J. G. (1986) Epidemiology of colonisation by nontypeable *Haemophilus influenzae* in children: A longitudinal study. *J. Infect. Dis.* **154**, 100-109.

Srinivasen N., White H. E., Emsley J., Wood S. P., Pepys M. B. and Blundell T. L. (1994) Comparative analysis of pentraxins: Implications for protomer assembly and ligand binding. *Structure* **2**, 1017-1027.

Staff M., Urbina F., Weintraub A. and Widmalm G. (1999) Structure elucidation of the O-antigenic polysaccharide from the enteroaggregative *Escherichia coli* strain 62D₁. *Eur. J. Biochem.* **262**, 56-62.

Steel D. M. and Whitehead A. S. (1994) The major acute phase reactant: C-reactive protein, serum amyloid P component and serum amyloid A protein. *Immunology* **15**, 81-88.

Stevenson G. Neal B., Liu D., Hobbs M., Packer N. H., Batley M., Redmond J. W., Lindquist L. and Reeves P. (1994). Structure of the O-antigen of *Escherichia coli* K-12 and the sequence of its rfb gene cluster. *J. Bacteriol.* **176**, 4144-4156.

- Swords W. E., Buscher B. A., Li K. V., Preston A., Nicholas W. A., Weiser J. N., Gibson B. W. and Apicella M. A. (2000) Non typeable *Haemophilus influenzae* adhere to and invade human bronchial epithelial cells via an interaction of lipooligosaccharide with the PAF receptor. *Mol. Microbiol.* **37**, 13-27.
- Szalai A. J., Briles D. E. and Volanakis J. E. (1995) Human C-reactive protein is protective against fatal *Streptococcus pneumonia* infection in transgenic mice. *J. Immunol.* **155**, 2557-2563.
- Szalai A. J., Briles D. E. and Volanakis J. E. (1996) The role of complement in C-reactive protein mediated protection in mice from *Streptococcus pneumonia*. *Infect. Immunol.* **64**, 4850-4853.
- Szalai A. J., van Ginkel F. W., Dalrymple S. A., Murray R., McGhee J. R. and Volanakis J. E. (1998) Testosterone and IL-6 requirements for human C-reactive protein gene expression in transgenic mice. *J. Immunol.* **160**, 5294-5299.
- Szalai A. J., van Ginkel F. W., Wang Y., McGhee J. R., Volanakis J. E. (2000) Complement dependent acute phase expression of C-reactive protein and serum amyloid P-component. *J. Immunol.* **165**, 1030-1035.
- Takahashi M., Iwaki D., Kanno K., Ishida Y., Xiong J., Matsushita M., *et al.*, (2008) Mannose-Binding Lectin (MBL)-Associated Serine Protease (MASP)-1 Contributes to Activation of the Lectin Complement Pathway. *J. Immunol.* **180**, 6132-6138.
- Tan S. S. H., Ng P. M. L., Ho B., Ding J. L. (2005) The antimicrobial properties of C-reactive protein (CRP). *J. Endotoxin Res.* **11**, 249-256.

- Tang H. B., Dimango E., Bryan R., Gambello M., Iglewski B. H., Goldberg J. B. and Prince A. (1996) Contribution of specific *Pseudomonas aeruginosa* virulence factors to pathogenesis of pneumonia in a neonatal mouse model of infection. *Infect. Immun.* **64**, 37-43.
- Tauber A. I. (2003) Metchnikoff and the phagocytosis theory. *Nature* **4**, 897-901.
- Tennent G. A., Butler P. J. G., Hutton T., Woolfitt A. R., Harvey D. J., Radmacher T. W. and Pepys M. B. (1993) Molecular characterization of *Limulus polyphemus* C-reactive protein: I. Subunit composition. *Eur. J. Biochem.* **214**, 91-97.
- Thiel S., Vorup-Jensen T., Stover C. M., Schweable W., Laurensen S. B., Poulsen K., *et al.* (1997) A second serine protease associated with mannan-binding lectin that activates complement. *Nature* **386**, 506-510.
- Thiel S. *et al.*, (2000) Interaction of C1q and mannan-binding lectin (MBL) with C1r, C1s MBL-associated serine proteases 1 and 2, and the MBL-associated protein MAp19. *J. Immunol.* **165**, 878-887.
- Thompson D., Pepys M. B. and Wood S. P. (1999) The physiological structure of human C-reactive protein and its complex with phosphocholine. *Structure* **7**, 169-177.
- Tillet W. S. and Francis T. J. (1930) Serological reactions in pneumonia with a non-protein fraction of pneumonia. *J. Exp. Med.* **52**, 561-517.
- Tinnert A., Mansson M., Yildirim H. H., Hood D. W. and Schweda E. K. H. (2005) Structural investigation of lipopolysaccharide from nontypeable *Haemophilus influenzae*:

Investigation of inner-core phosphoethanolamine addition in NTHi strain 981. *Carbohydrate Res.* **340**, 1900-1907.

Tobias P. S., Soldau K., Gegner J. A., Mintz D., Ulevitch R. J. (1995) Lipopolysaccharide binding protein mediated complexation of lipopolysaccharide with soluble CD14. *J. Biol. Chem.* **270**, 10482-10488.

Toniatti C., Arcone R., Majello B., Ganter U., Arpaia G. and Ciliberto G. (1990) Regulation of the human C-reactive protein gene, a major marker of inflammation and cancer. *Mol. Biol. Med.* **7**, 199-212.

Trent M. S. (2004) Biosynthesis, transport and modifications of lipid A. *Biochem. Cell Biol.* **82**, 71-86.

Turner M. W. (2003) The role of mannose-binding lectin in health and disease. *Mol. Immunol.* **40**, 423-429.

Van Alpen L., Klein M., Geelen-van den Broek L., Riemens T., Eijk P. and Kamerling J. P. (1990) Biochemical characterisation and world wide distribution of serologically distinct lipopolysaccharides of *Haemophilus influenzae* type b. *J. J. Infect. Dis.* **162**, 659-663.

Van Eijk M., Haagsman H. P., Skinner T., Archibold A., Reid K. B. M. and Lawson P. R. (2000) Porcine lung surfactant protein D: Complementary DNA cloning, chromosomal localisation, and tissue distribution. *J. Immunol.* **164**, 1442-1450.

- Vandivier R. W., Ogden C. A., Fadok V. A., Hoffmann P. R., Brown K. K., Botto M., Walport M. J., Fisher J. H., Henson P. M. and Greene K. E. (2002) Role of surfactant proteins A, D and C1q in the clearance of apoptotic cells in vivo and in vitro: Calreticulin and CD91 as a common collectin receptor complex. *J. Immunol.* **169**, 3978-3986.
- Verez-Bencomo V., Fernandez-Santana V., Hardy, E., Toledo M. E., *et al.* (2004) A synthetic conjugate polysaccharide vaccine against *Haemophilus influenzae* type b. *Science* **305**, 522-525.
- Viseux N., de Hoffman E. and Domon B. (1998) Structural assignment of permethylated oligosaccharide subunits using sequential tandem mass spectrometry. *Anal. Chem.* **70**, 4951-4959.
- Volanakis J. E. (2001) Human C-reactive protein: expression, structure, and function. *Molecular Immunology* **38**, 189-197.
- Volanakis J. E. and Kaplan M. H. (1971) Specificity of C-reactive protein for choline phosphate residues of pneumococcal C-polysaccharide. *Proc. Soc. Exp. Biol. Med.* **136**, 612-614.
- Voorhout W. F., Veenendal T., Kuroki Y., Ogasawara Y., Van Golde L. M. and Geuze H. J. (1992) Immunocytochemical localisation of surfactant protein D (SP-D) in type II cells, clara cells, and alveolar macrophages of rat lungs. *J. Histochem. Cytochem.* **40**, 1589-1597.
- Wallis R. and Drickamer K. (1999) Molecular determinants of oligomer formation and complement fixation in mannose binding proteins. *J. Biol. Chem.* **274**, 3580-3589.

- Walport M. J. (2001) Complement. *N. Eng. J. Med.* **344**, 1058-1066.
- Wang H., Head J., Kosma P., Brade H., Miller-Loennies S., Sheikh S., McDonald B., Smith K., Cafarella T., Seaton B. and Crouch E. (2008) recognition of heptose and the inner core of bacterial lipopolysaccharides by surfactant protein D. *Biochem.* **47**, 710-720.
- Wang J. Y., Kishore U., Lim B L., Strong P. and Reid K B. (1996) Interaction of human lung surfactant protein A and D with mite (*Dermatophagoides pteronyssinus*) allergens. *Clin. Exp. Immunol.* **106**, 367-373.
- Weis W. I. and Drickamer K. (1994) Trimeric structure of a C-type mannose-binding protein. *Structure* **2**, 1227-1240.
- Weis W. I., Drickamer K. and Hendrickson, W. A. (1992) Structure of a C-type mannose binding protein complexed with an oligosaccharide. *Nature* **360**, 127-134.
- Weis W. I., Kahan R., Fourne R., Drickamer K. and Hendrickson W. A. (1991) Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. *Science* **254**, 1608-1615.
- Weiser J. N., Love J. and Moxon E. R. (1989) The molecular mechanism of phase variation in *Haemophilus influenzae* lipopolysaccharide. *Cell* **59**, 657-665.
- Weiser J. N., Maskell D. J., Butler P. D. and Moxon E. R. (1990) Characterisation of repetitive sequences controlling phase variation of *Haemophilus influenzae* lipopolysaccharide. *J. Bacteriol.* **172**, 3304-3309.

- Weiser J. N., Pan N., McGowan K. L., Musher D., Martin A. and Richards J. (1998) Phosphorylcholine on the lipopolysaccharide of *Haemophilus influenzae* contributes to serum killing mediated by C-reactive protein. *J. Exp. Med.* **187**, 631-640.
- White M. R., Crouch E. C., Chang D. and Harsthorn K. L. (2001) Increased antiviral and opsonic activity of a highly multimerised collectin chimera. *Biochem. Biophys. Res. Commun.* **286**, 206-213.
- White R.T., Damm D., Miller J., Spratt K., Schilling J., Hawgood S., Benson B. and Cordell B. (1985) Isolation and characterisation of the human pulmonary surfactant apoprotein gene. *Nature* **317**, 361-363.
- Wilson F. J., Bray T. L. and Suddath F. L. (1991) Crystallisation of proteins by dynamic control of evaporation. *J. Crystal Growth* **110**, 142-147.
- Woo P., Korenberg J. R. and Whitehead A. S. (1985) Characterisation of genomic and complementary DNA sequence of human C-reactive protein, and comparison with the complementary DNA sequence of serum amyloid P component. *J. Biol. Chem.* **260**, 13384-13388.
- Wright S. D., Tobias P. S., Ulivetch R. J., Ramos R. A. (1989) Lipopolysaccharide (LPS) binding protein opsonizes LPS-bearing particles for recognition by a novel receptor on macrophages. *J. Exp. Med.* **170**, 1231-1241.
- Yamazoe M., Nishitani C., Takahashi M., Katoh T., Arikis S., et al., (2008) Pulmonary surfactant protein D inhibits lipopolysaccharide (LPS)-induced inflammatory cell response by altering LPS binding to its receptors. *J. Biol. Chem.* **283**, 35878-35888.

- Yi W., Bystricky P., Yao Q., Guo H., Zhu L., Li H., Shen J., Li, M., Ganguly S., Bush A. and Wang P. G. (2006) Two different O-polysaccharides from *Escherichia coli* 086 are produced by different polymerisation of the same O-repeating unit. *Carbohydrate Res.* **341**, 100-108.
- Yildirim H. H., Hood D. W., Moxon E. R. and Schweda E. K. H. (2003) Structural analysis of lipopolysaccharides from *Haemophilus influenzae* serotype f: Structural diversity observed in three strains. *Eur. J. Biochem.* **270**, 3153-3167.
- Yildirim H. H., Li J., Richards J. C., Hood D. W., Moxon E. R. and Schweda E. K.H. (2005) Complex O-acetylation in non-typeable *Haemophilus influenzae* lipopolysaccharide: Evidence for a novel site of O-acetylation. *Carbohydrate Res.* **340**, 2598-2611.
- Zai J. (2004) Mass spectroscopy of oligosaccharides. *Mass Spectrom. Rev.* **23**, 161-227.
- Zelensky A. N. and Gready J. E. (2005) Review article: The C-type lectin-like domain superfamily. *FEBS J.* **272**, 6179-6217.
- Zhu L., Parr G. R., Fitzgerald M. C., Nebon M. and Smith L. M. (1995) Oligonucleotide fragmentation in MALDI/TOF mass spectrometry using 355nm radiation. *J. Am. Chem. Soc.* **117**, 6048-6056.

Appendices

Appendix 1

Publication: *Haemophilus influenzae* evades innate immune defence mediated by surfactant protein D by blocking multiple interactions with the LPS core

REFERENCES

1. Holmskov, U.L. Collectins and collectin receptors in innate immunity. *APMIS Supplement* **100**, 1-59 (2000).
2. Sano, H. & Kuroki, Y. The lung collectins, SP-A and SP-D, modulate pulmonary innate immunity. *Molecular Immunology* **42**, 279-287 (2005).
3. Kuan S. F., Rust K. & Crouch E. Interactions of Surfactant Protein-D with Bacterial Lipopolysaccharides - Surfactant Protein-D Is an Escherichia-Coli Binding-Protein in Bronchoalveolar Lavage. *Journal of Clinical Investigation* **90**, 97-106 (1992).
4. Khamri W., Moran A. P., Worku M. L., Karim Q. N., Walker M. M., Annuk H., Ferris J. A., Appelmelk B. J., Eggleton P., Reid K. B. M. & Thursz M. R. Variations in *Helicobacter pylori* lipopolysaccharide to evade the innate immune component surfactant protein D. *Infection and Immunity* **73**, 7677-7686 (2005).
5. Sahly H., Ofek I., Podschun R., Brade H., He Y., Ullmann U. and Crouch E. Surfactant Protein D Binds Selectively to *Klebsiella pneumoniae* Lipopolysaccharides Containing Mannose-Rich O-Antigens. *J. Immunol.* **169**, 3267-3274 (2002).
6. LeVine A. M., Whitsett J. A., Gwozdz J. A., Richardson T. R., Fisher J. H., Burhans M. S. & Korfhagen T. R. Distinct Effects of Surfactant Protein A or D Deficiency During Bacterial Infection on the Lung. *J. Immunol.* **165**, 3934-3940 (2000).
7. Restrepo C. I., Dong Q., Savov J., Mariencheck W. I. & Wright J. R. Surfactant Protein D Stimulates Phagocytosis of *Pseudomonas aeruginosa* by Alveolar Macrophages. *Am. J. Respir. Cell Mol. Biol.* **21**, 576-585 (1999).
8. Schaeffer, L.M., McCormack, F.X., Wu, H. & Weiss, A.A. Interactions of pulmonary collectins with *Bordetella bronchiseptica* and *Bordetella pertussis* lipopolysaccharides elucidate the structural basis of their antimicrobial activities. *Infection and Immunity* **72**, 7124-7130 (2004).
9. Giannoni E., Sawa T., Allen L., Wiener-Kronish J. & Hawgood S. Surfactant Proteins A and D Enhance Pulmonary Clearance of *Pseudomonas aeruginosa*. *Am. J. Respir. Cell Mol. Biol.* **34**, 704-710 (2006).
10. Moxon, R.E. & Vaughn, K.A. The Type b Capsular Polysaccharide as a Virulence Determinant of *Haemophilus influenzae*: Studies Using Clinical Isolates and Laboratory Transformants. *J. Infectious Diseases* **143**, 517-524 (1981).
11. Campagnari, A.A., Gupta, M.R., Dudas, K.C., Murphy, T.F. & Apicella, M.A. Antigenic diversity of lipooligosaccharides of nontypable *Haemophilus influenzae*. *Infect. Immun.* **55**, 882-887 (1987).
12. Moxon, E.R. & Maskell, D. *Haemophilus influenzae* lipopolysaccharide: the biochemistry and biology of a virulence factor. In *Molecular Biology of Bacterial Infection: Current Status and Future Perspectives* (Harmache, C.E., Penn, C.W. & Smyth, C.J., eds), pp. 75-96. Cambridge University Press, Cambridge, UK (1992).

13. Murphy, T.F. & Apicella, M.A. Nontypable *Haemophilus influenzae*: A Review of Clinical Aspects, Surface Antigens, and the Human Immune Response to Infection. *Reviews of Infectious Diseases* **9**, 1-15 (1987).
14. Moxon, E.R. The molecular basis of *Haemophilus influenzae* virulence. *J. R. Coll. Physicians London* **19**, 174-178 (1985).
15. Malhotra, R., Willis, A.C., Lopez Bernal, A., Thiel, S. & Sim, R.B. Mannan-binding protein levels in human amniotic fluid during gestation and its interaction with collectin receptor from amnion cells. *Immunology* **82**, 439-444 (1994).
16. Masoud, H., Moxon, E. R., Martin, A., Krajcarski, D. & Richards, J. C. Structure of the variable and conserved lipopolysaccharides oligosaccharide epitopes expressed by *Haemophilus influenzae* serotype b strain Eagan. *Biochemistry* **36**, 2091-2103 (1997).
17. Hood, D.W., Richards, J.C. & Moxon, E.R. *Haemophilus influenzae* lipopolysaccharide. *Biochem. Soc. Trans.* **4**, 493-498 (1999).
18. 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. Use of the complete genome sequence information of *Haemophilus influenzae* strain Rd to investigate lipopolysaccharides biosynthesis. *Mol. Microbiol.* **22**, 951-965 (1996).
19. Håkansson, K. & Reid, K.B.M. Collectin structure: A review. *Prot. Sci.* **9**, 1607-1617 (2000).
20. Crouch, E.C. Structure, biologic properties, and expression of surfactant protein D (SP-D). *Biochimica et Biophysica Acta* **1408**, 278-289 (1998).
21. Lawson, P.R. & Reid, K.B.M. The roles of surfactant proteins A and D in innate immunity. *Immunol Rev.* **173**, 66-78 (2000).
22. Weiss, W.I. & Drickamer, K. Structural basis of lectin-carbohydrate recognition. *Annual Review of Biochemistry* **65**, 441-473 (1996).
23. Madan, T., Kishore, U., Singh, M., Strong, P., Clark, H.W., Hussain, E.M., Reid, K.B.M. & Sarma, P.U. Surfactant proteins A and D protect mice against pulmonary hypersensitivity induced by *Aspergillus fumigatus* antigens and allergens. *J. Clin. Invest.* **107**, 467-475 (2001).
24. Madan, T., Reid, K.B.M., Clark, H.W., Singh, M., Nayak, A., Sarma, P.U., Hawgood, S. & Kishore, U. Susceptibility of mice genetically deficient in SP-A or SP-D gene to invasive pulmonary aspergillosis. *Mol. Immunol.* **47**(10), 1923-1930 (2010).
25. Clark, H.W. & Reid, K.B.M. Structural requirements for SP-D function in vitro and in vivo: Therapeutic potential of recombinant SP-D. *Immunobiology* **205**, 619-631 (2002).
26. Singh, M., Madan, T., Waters, P., Parida, S.K., Sarma, P.U. & Kishore, U. Protective effects of a recombinant fragment of human surfactant protein D in a murine model of pulmonary hypersensitivity induced by dust mite allergens. *Immunology Letters* **86**, 299-307 (2003).
27. Crouch, E., Hartshorn, K., Horlacher, T., McDonald, B., Smith, K., Carafella, T., Seaton, B., Seeberger, P.H. & Head, J. 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 (2009).
28. Crouch, E., McDonald, B., Smith, K., Cafarella, T., Seaton, B. & Head, J. Contributions of phenylalanine 335 to ligand recognition by human surfactant protein D. *J. Biol. Chem.* **281**, 18008–18014 (2006).
 29. Crouch, E., McDonald, B., Smith, K., Roberts, M., Mealy, T., Seaton, B. & Head, J. Critical role of Arg/Lys343 in the species-dependent recognition of phosphatidylinositol by pulmonary surfactant protein D. *Biochemistry* **46**, 5160–5169 (2007).
 30. Wang, H., Head, J., Kosma, P., Brade, H., Muller-Loennies, S., Sheikh, S., McDonald, B., Smith, K., Cafarella, T., Seaton, B. & Crouch, E. Recognition of heptoses and the inner core of bacterial lipopolysaccharides by surfactant protein D. *Biochemistry* **47**, 710–720 (2008).
 31. Shrive, A.K., Tharia, H.A., Strong, P., Kishore, U., Burns, I., Rizkallah, P.J., Reid, K.B.M. & Greenhough, T.J. High resolution structural insights into ligand binding and immune cell recognition by human lung surfactant protein D. *J. Mol. Biol.* **331**, 509–523 (2003).
 32. 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. Structural characterisation of ligand-binding determinants in human lung surfactant protein D: influence of Asp325. *J. Mol. Biol.* **394**, 776–788 (2009).
 33. Auzanneau, F.-I., Charon, D. & Szabó, L. Phosphorylated sugars. Part 27. Synthesis and reactions, in acid medium, of 5-*O*-substituted methyl 3-deoxy- α -D-manno-oct-2-ulopyranosidonic acid 4-phosphates. *J. Chem. Soc., Perkin Trans. 1*, 509–517 (1991).
 34. Crouch, E.C., Smith, K., McDonald, B., Briner, D., Linders, B., McDonald, J., Holmskov, U., Head, J., & Hartshorn, K. Species differences in the carbohydrate binding preferences of surfactant protein D. *Am. J. Respir. Cell Mol. Biol.* **35**, 84–94 (2006).
 35. Deadman, M.E. Hermant, P. Engskog, M. Makepeace, K. Moxon, E. R. Schweda, E.K.H. & Hood, D.W. Lex2B, a phase-variable glycosyltransferase, adds either a glucose or a galactose to *Haemophilus influenzae* Lipopolysaccharide. *Infection and Immunity* **77**, 2376–2384 (2009).
 36. Mansson, M., Hood, D. W., Li, L., Richards, J. C., Moxon, E. R. & Schweda, E. K. H. Structural analysis of the lipopolysaccharide from nontypeable *Haemophilus influenzae* strain 1003. *Eur. J. Biochem.* **269**, 808–818 (2002).
 37. Yildirim, H. H., Hood, D. W., Moxon, E. R. & Schweda, E. K. H. Structural analysis of lipopolysaccharides from *Haemophilus influenzae* serotype f: structural diversity observed in three strains. *Eur. J. Biochem.* **270**, 3153–3167 (2003).
 38. Danan, A., Mondange, M., Sarfati, S. R. & Szabo, P. Synthesis and behaviour under acidic conditions of 2-deoxy-D-arabino-hexopyranose and 3-deoxy-2-ketoaldonic acids bearing O-phosphono or O-glucosyl substituents at position β to the carbonyl function. *J. Chem. Soc. Perkin Trans. 1*, 1275–1282 (1982).

39. Caroff, M., Brisson, J., Martin, A. & Karibian, D. Structure of the *Bordetella pertussis* 1414 endotoxin. *FEBS Lett.* **477**, 8-14 (2000).
40. Caroff, M. & Karibian, D. Structure of bacterial lipopolysaccharides. *Carbohydrate Research* **338**, 2431-2447 (2003).
41. Kawasaki, N., Kawasaki, T. & Yamashina, I. A serum lectin (mannan-binding protein) has complement-dependent bactericidal activity. *J. Biochem.* **106**, 483-489 (1989).
42. Bezouska, K., Crichlow, G.V., Rose, J.M., Taylor, M.E. & Drickamer, K. Evolutionary conservation of intron position in a subfamily of genes encoding carbohydrate-recognition domains. *J. Biol Chem.* **266**, 11604-11609 (1991).
43. Strong, P., Kishore, U., Morgan, C., Lopez Bernal, A., Singh, M. & Reid, K.B.M. A novel method of purifying lung surfactant proteins A and D from the lung lavage of alveolar proteinosis patients and from pooled amniotic fluid. *J. Immunol. Methods* **1**, 139-149 (1998).
44. Leslie, A.G.W. Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography No. 26. Daresbury Laboratory, Warrington, UK (1992).
45. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. & Warren, G.L. Crystallography and NMR System: a new software suite for macromolecular structure determination. *Acta Cryst.* **D54**, 905-921 (1998).
46. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760-763 (1994).
47. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.* **A47**, 110-119 (1991).
48. Kleywegt, G.J., Henrick, K., Dodson, E.J., & van Aalten, D.M.F. Pound-wise but penny-foolish - How well do micromolecules fare in macromolecular refinement? *Structure* **11**, 1051-1059 (2003).
49. Chen, V.B., Arendall III, W.B., Headd, J.B., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. & Richardson, D.C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst.* **D66**, 12-21 (2010).
50. Laskowski, R., MacArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-291 (1993).
51. Kraulis, P.J. MOLSCRIPT: A program package to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946 (1991).

Haemophilus influenzae RM153 (Eagan)

CA7 { 4A {

..... Glc β 1,4 Hep_I α 1,5 Kdo — LipidA

| α 1,3 ← *rfaF*

Gal α 1,4 Gal β 1,4 Glc β 1,4 Glc α 1,3 Hep_{II} — PE

| α 1,2 ← *orfH*

Gal β 1,2 Hep_{III} — P

|

PC

- 30 -

Figure 2.

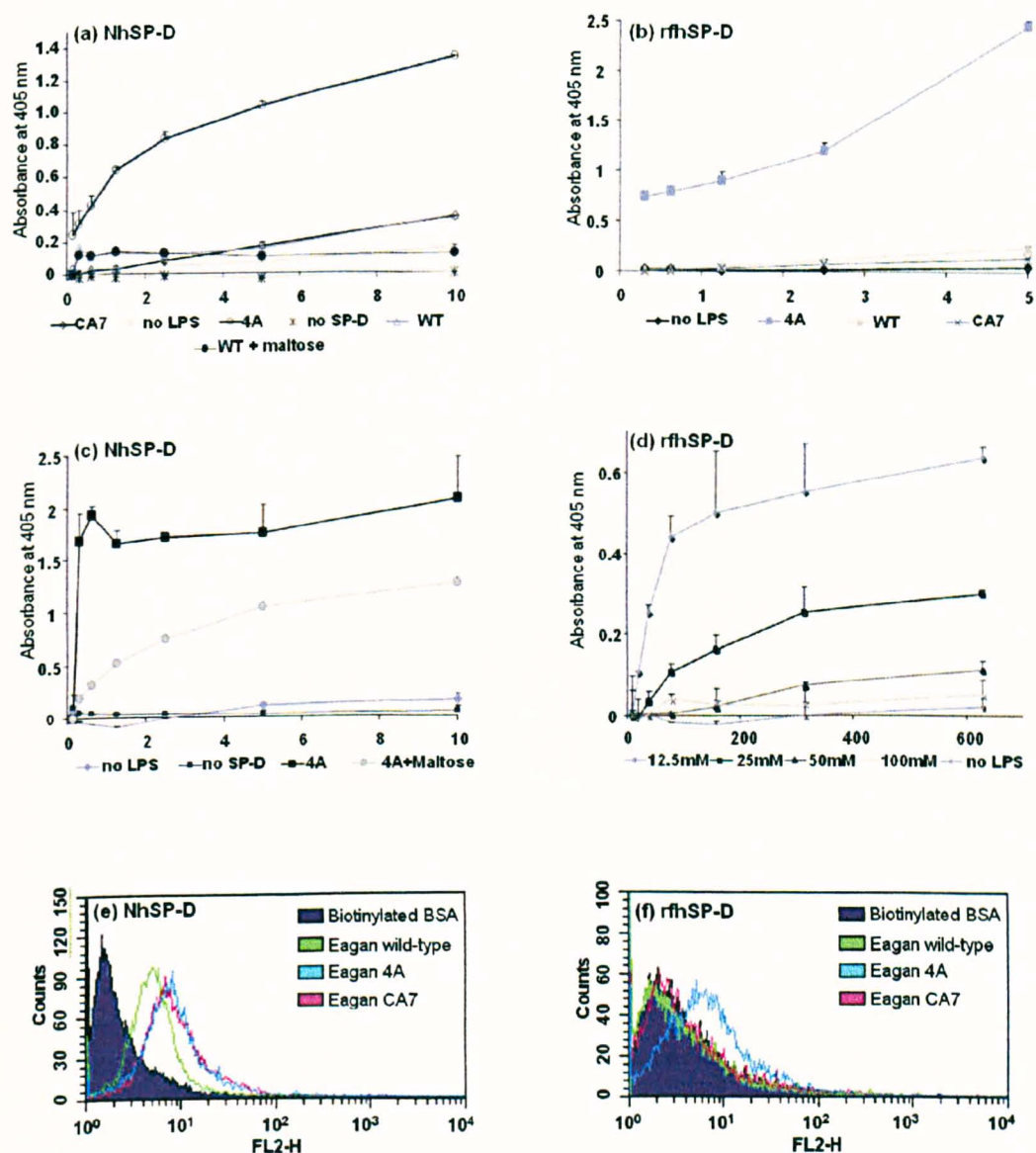


Figure 2. SP-D interactions with *H. influenzae* Eagan wild-type and mutant strains of isolated LPS and whole bacteria. (a), (b): interactions with *H. influenzae* LPS of (a) native hSP-D ($\mu\text{g/ml}$), (b) recombinant SP-D ($\mu\text{g/ml}$). (c), (d): Sugar competition assays with Eagan 4A LPS (c) native hSP-D ($\mu\text{g/ml}$), (d) recombinant SP-D (ng/ml). (e), (f) FACS analysis of binding of whole bacteria by (e) native hSP-D (f) recombinant SP-D.

Figure 3.

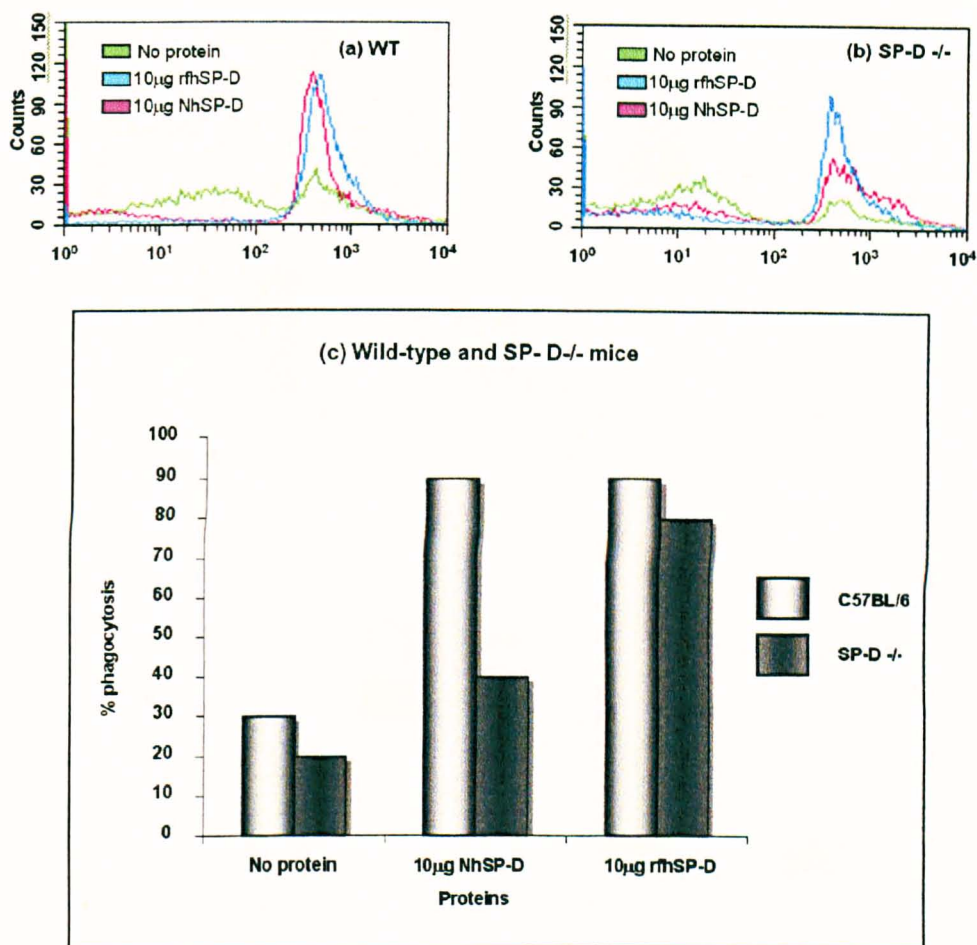


Figure 3. Uptake of FITC labelled Eagan 4A bacteria by alveolar macrophages. (a) Bacteria uptake by macrophages isolated from C57BL/6 mice. (b) bacteria uptake by macrophages isolated from SP-D knock-out mice and (c) graph representing the mean fluorescence of bacterial uptake in the isolated alveolar macrophages, shown as percentage phagocytosis.

Figure 4.

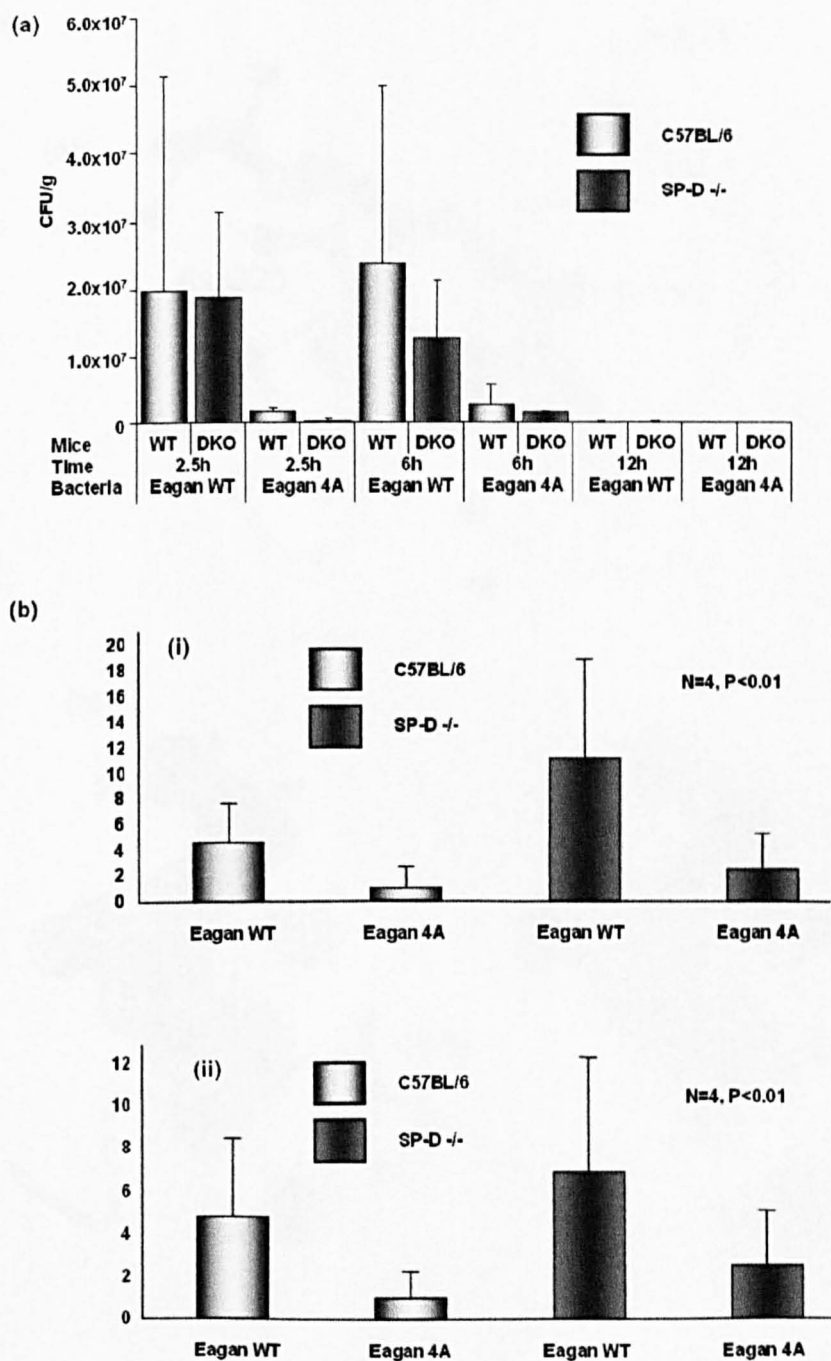


Figure 4. Response to *H. influenzae* Eagan infection in wild-type and SD-D knockout mice. (a) Clearance of *H. influenzae* Eagan wild-type and 4A bacteria from C57BL/6 mice and SP-D knock-out mice. Mice were infected with 10⁷ CFU *H. influenzae* then were sacrificed at 2.5, 6 and 12 hour time points (n=4-6) to determine bacterial numbers by plating. (b) Pro-inflammatory cell infiltration in response to *H. influenzae* Eagan wild-type and mutant strain 4A infection. (i) alveolar macrophage and (ii) neutrophil counts in lavage from infected mice were determined per high powered field (hpf) (24 counts per high powered field per group).

Figure 5.

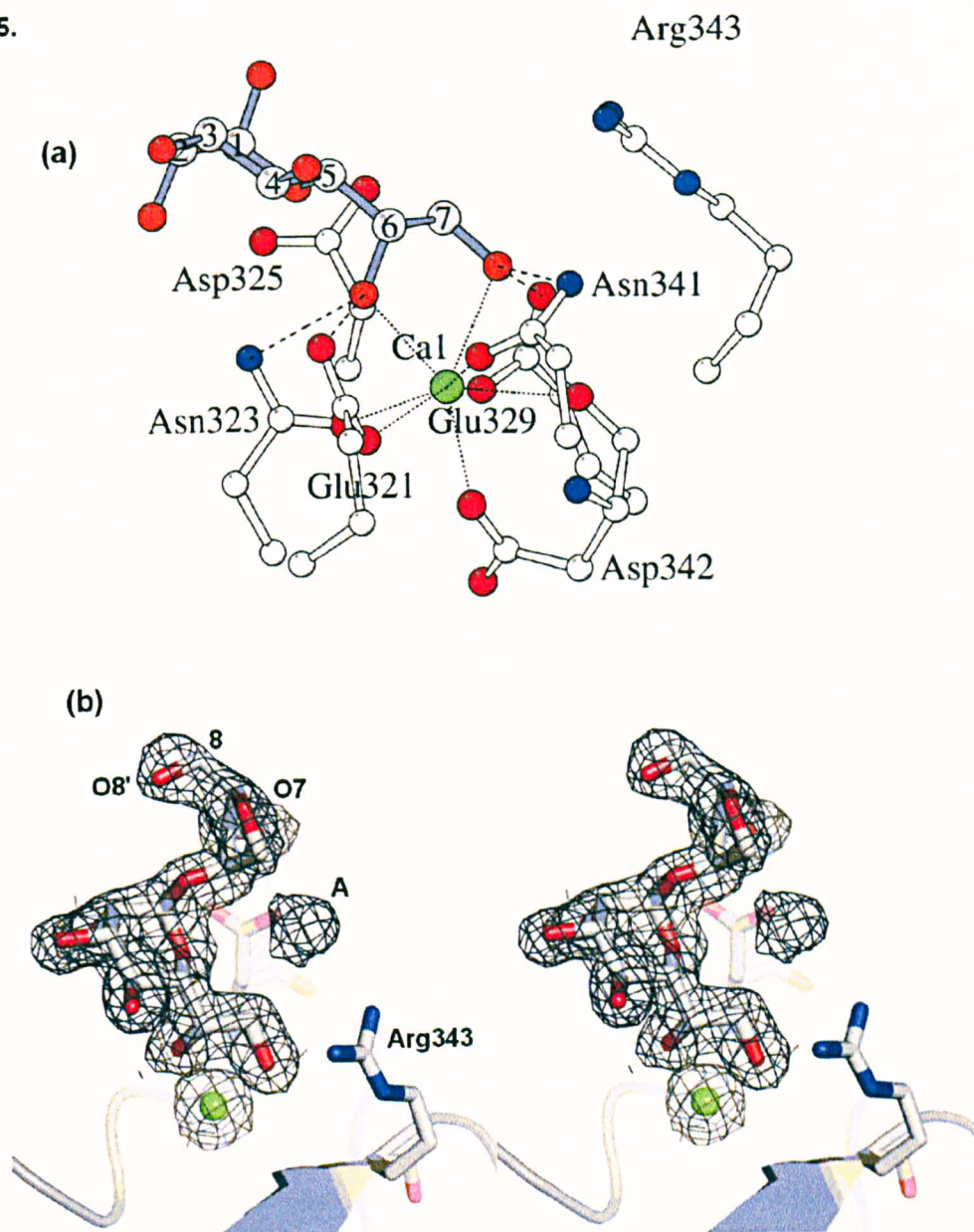


Figure 5. Bound Egan 4A in subunit B of the rfhSP-D – ligand complex. (a) The coordination of the calcium ion Ca1 and HepI (b) Stereo electron density for the bound Glc-Hep and the putative Kdo(anhydro) ligand. Original Kdo numbering retained. Density labelled A corresponds to the putative position of O2' (original Kdo numbering) in either or both of the enantiomeric substituents off KdoC4. The glucose is not visible in the electron density.

Figure 6.

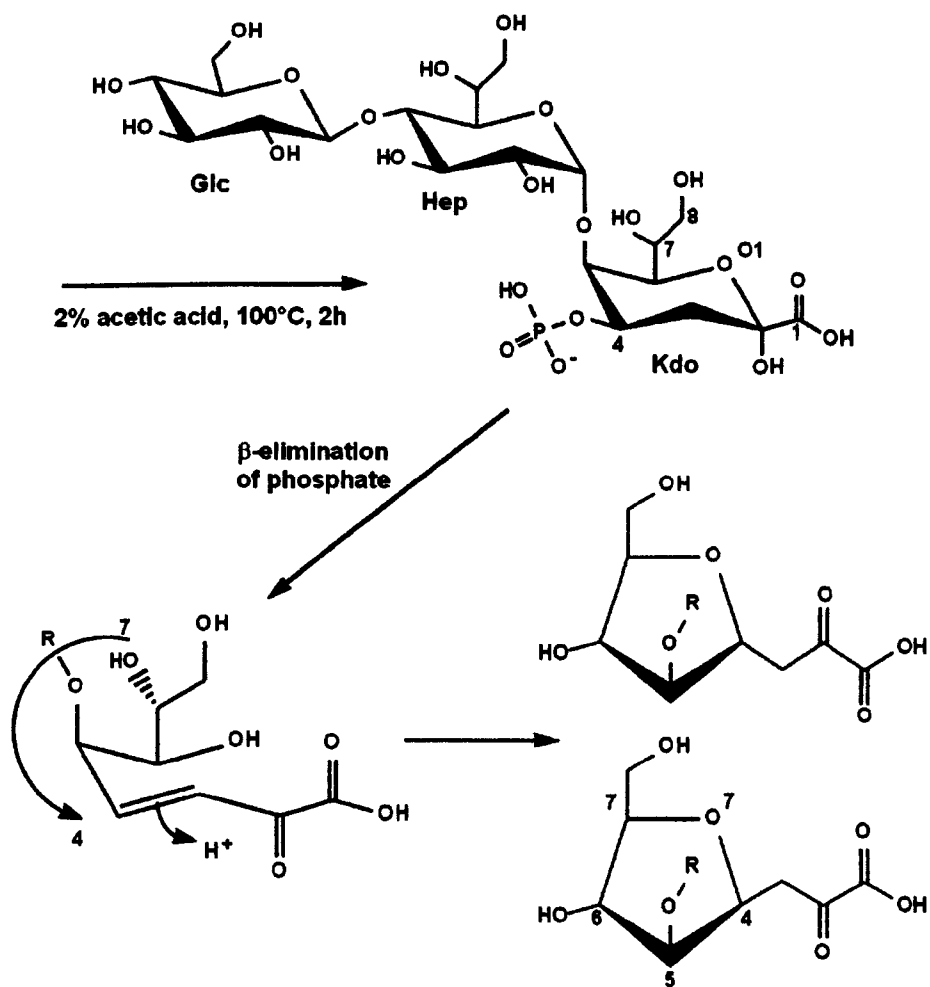


Figure 6. Formation of the 4,7 closure furanoid derivative (anhydro Kdo) following mild acid hydrolysis and β -elimination of the phosphate from O4 of Kdo. Original Kdo numbering retained. Adapted from ref. 33.

Figure 7.

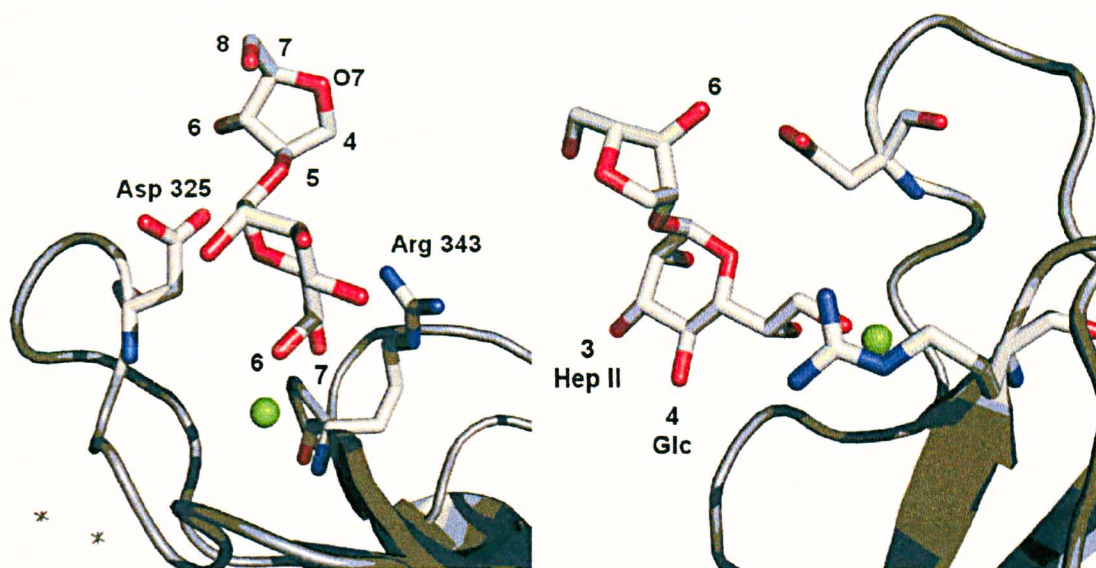


Figure 7. The Hep-Kdo(anhydro) disaccharide in the ligand binding site. Kdo O6' is positioned to interact with Asp325, while the enantiomeric C4 substituent is not identifiable in the electron density except for an indication of O2' (original Kdo numbering), directed towards Arg343, which may be similarly positioned in both enantiomers. Original Kdo numbering retained.

Appendix 2

MALDI MS of *Haemophilus influenzae* Eagan 4A oligosaccharide

MALDI-TOF analysis of *Haemophilus Influenzae*
Eagan 4A oligosaccharide

Sample was spotted 1:1 with matrix and analysed in reflector mode.

CHCA matrix – 10mg/ml in 70% MeCN / 0.1% TFA

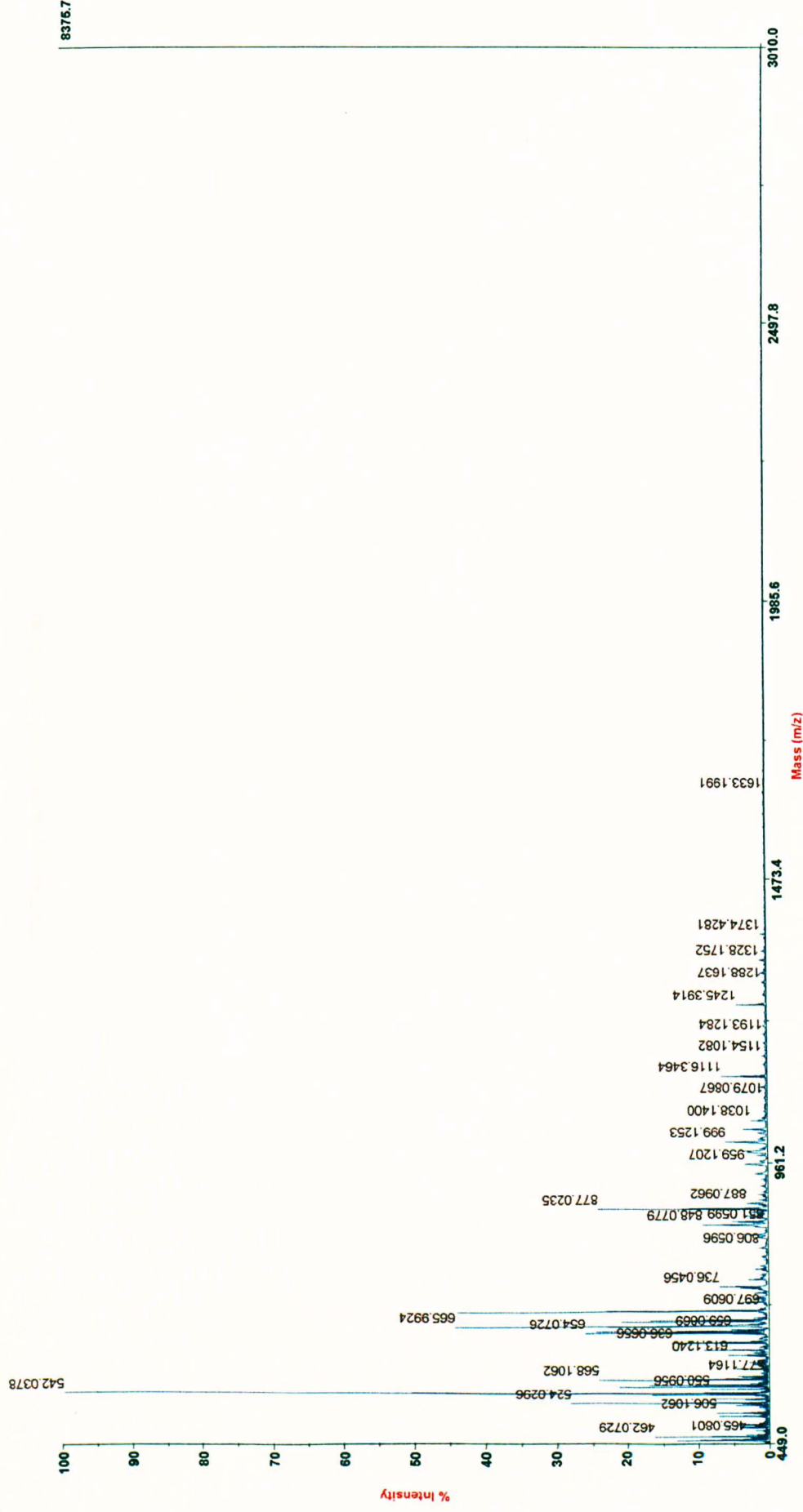
DHB matrix – 10mg/ml in 100% Methanol

The instrument was calibrated using close external peptide standards.

CHCA matrix – positive mode

Polysaccharide

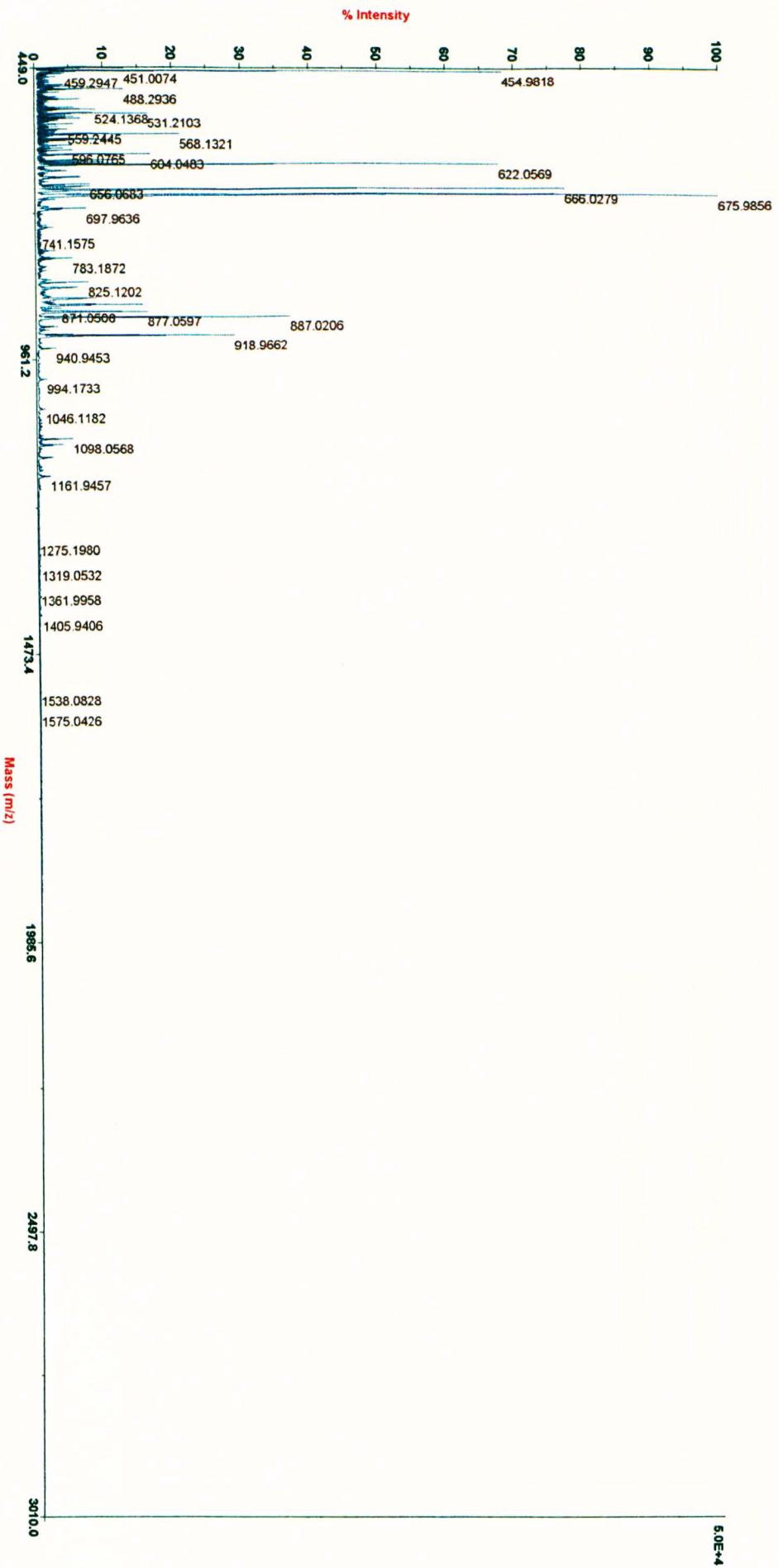
4700 Reflector Spec #1 MC[BP = 542.0, 8376]



CHCA matrix – positive mode

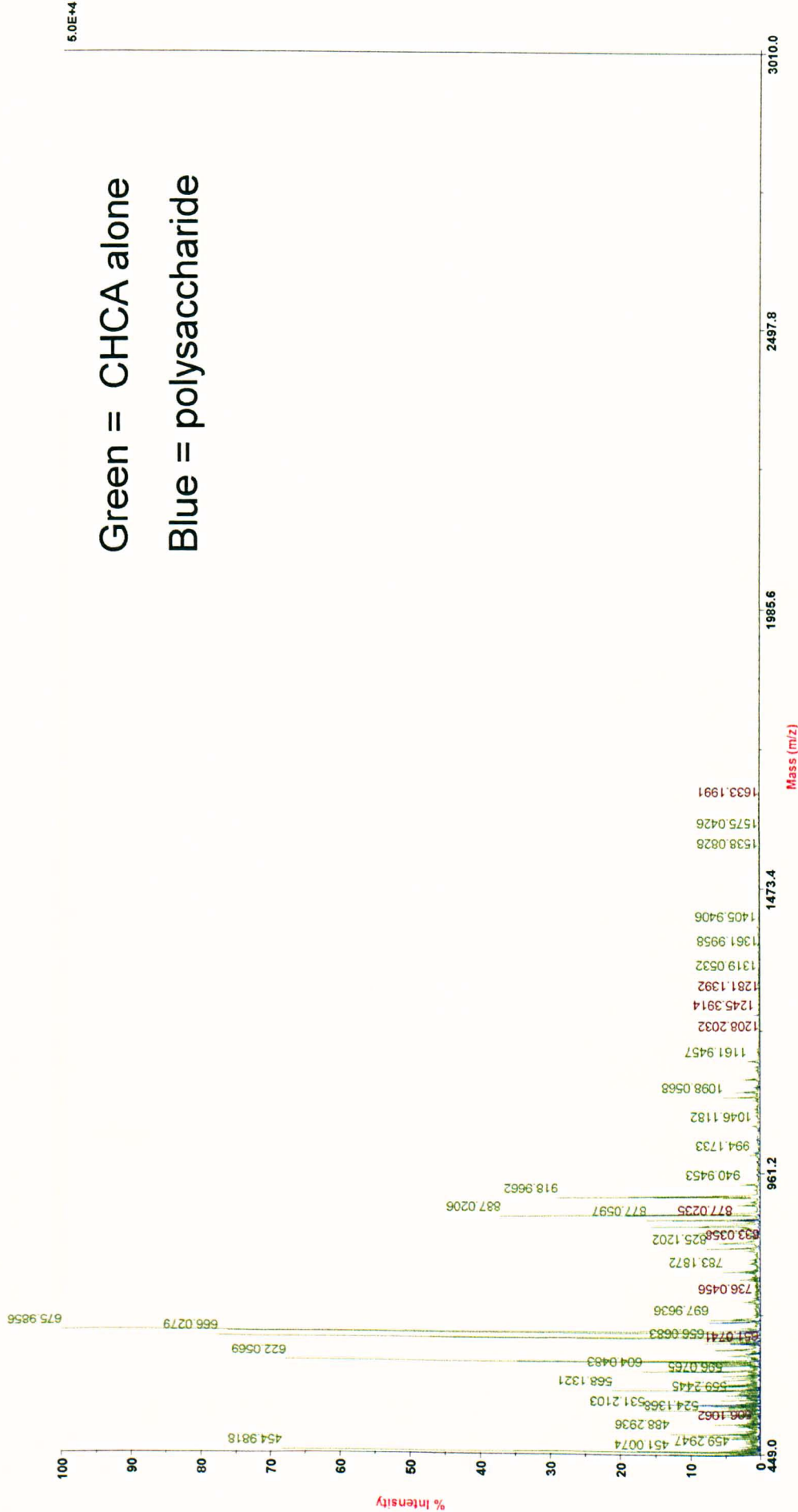
CHCA alone control

4700 Reflector Spec #1 MC[BP = 676.0, 50148]



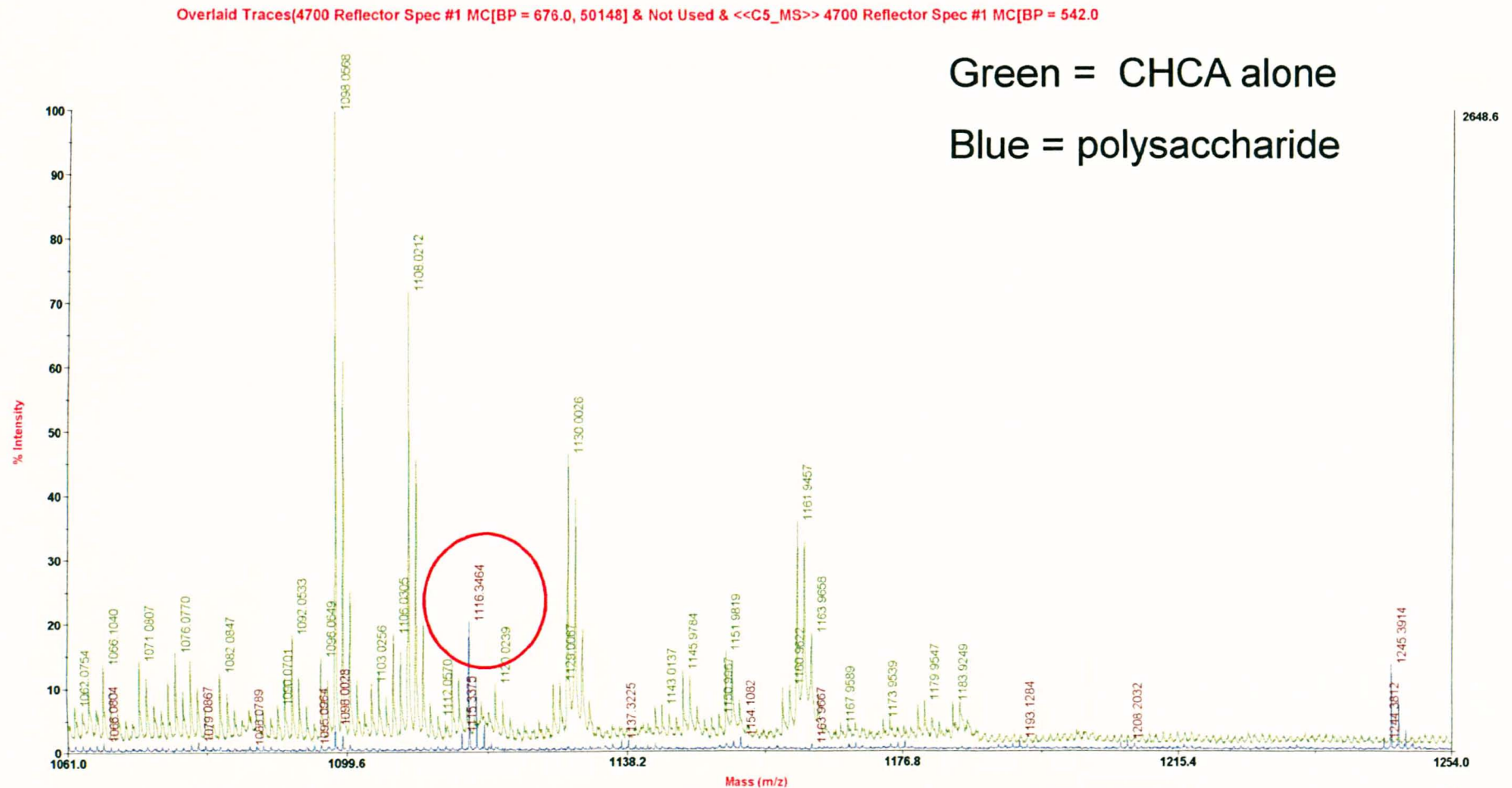
Polysaccharide and CHCA alone overlay (positive mode)

Overlaid Traces(4700 Reflector Spec #1 MC[BP = 676.0, 50148] & Not Used <<C5_MS>> 4700 Reflector Spec #1 MC[BP = 542.0



Polysaccharide and CHCA alone overlay (positive mode)

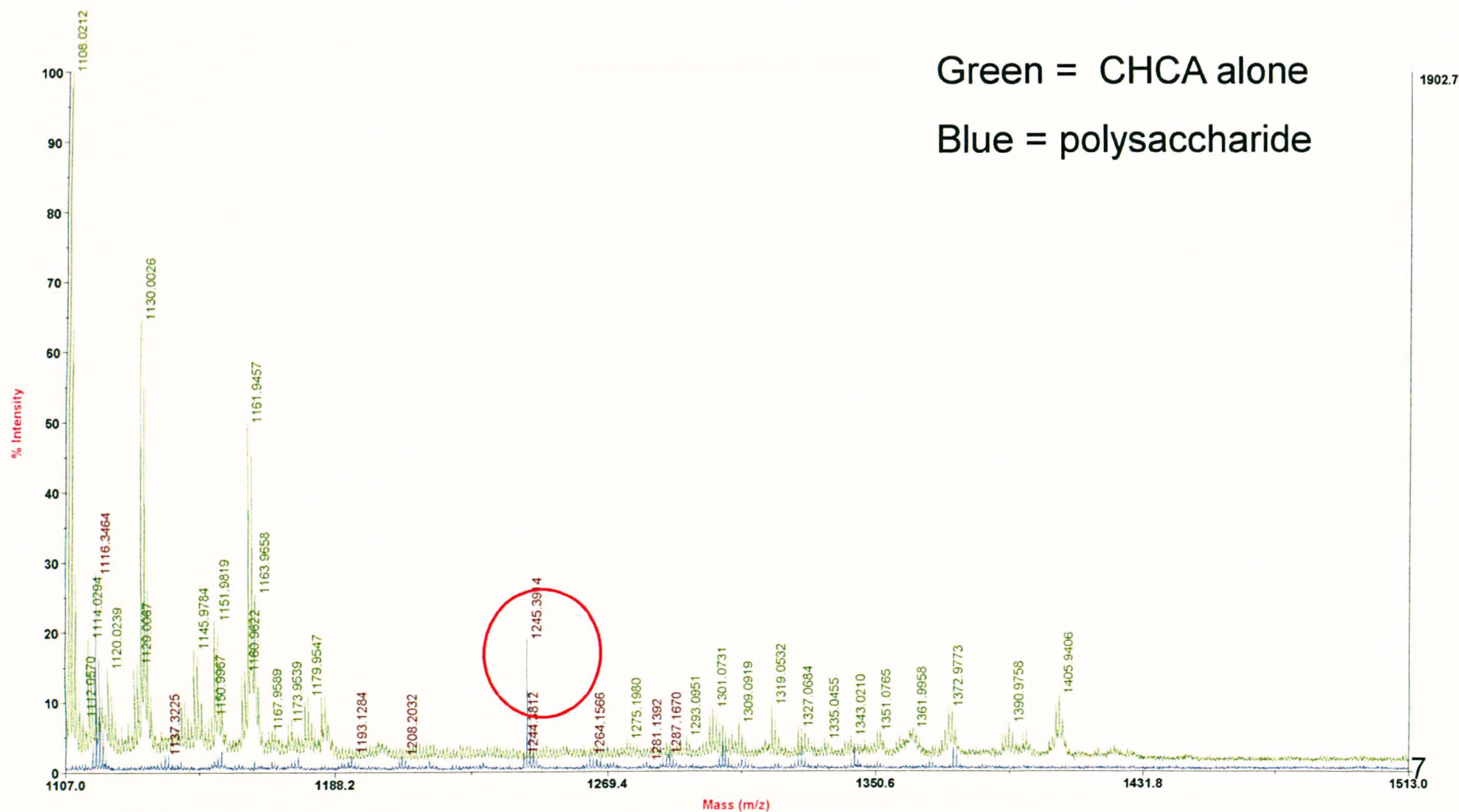
Zoomed in to show where the main differences are 1



Polysaccharide and CHCA alone overlay (positive mode)

Zoomed in to show where the main differences are 2

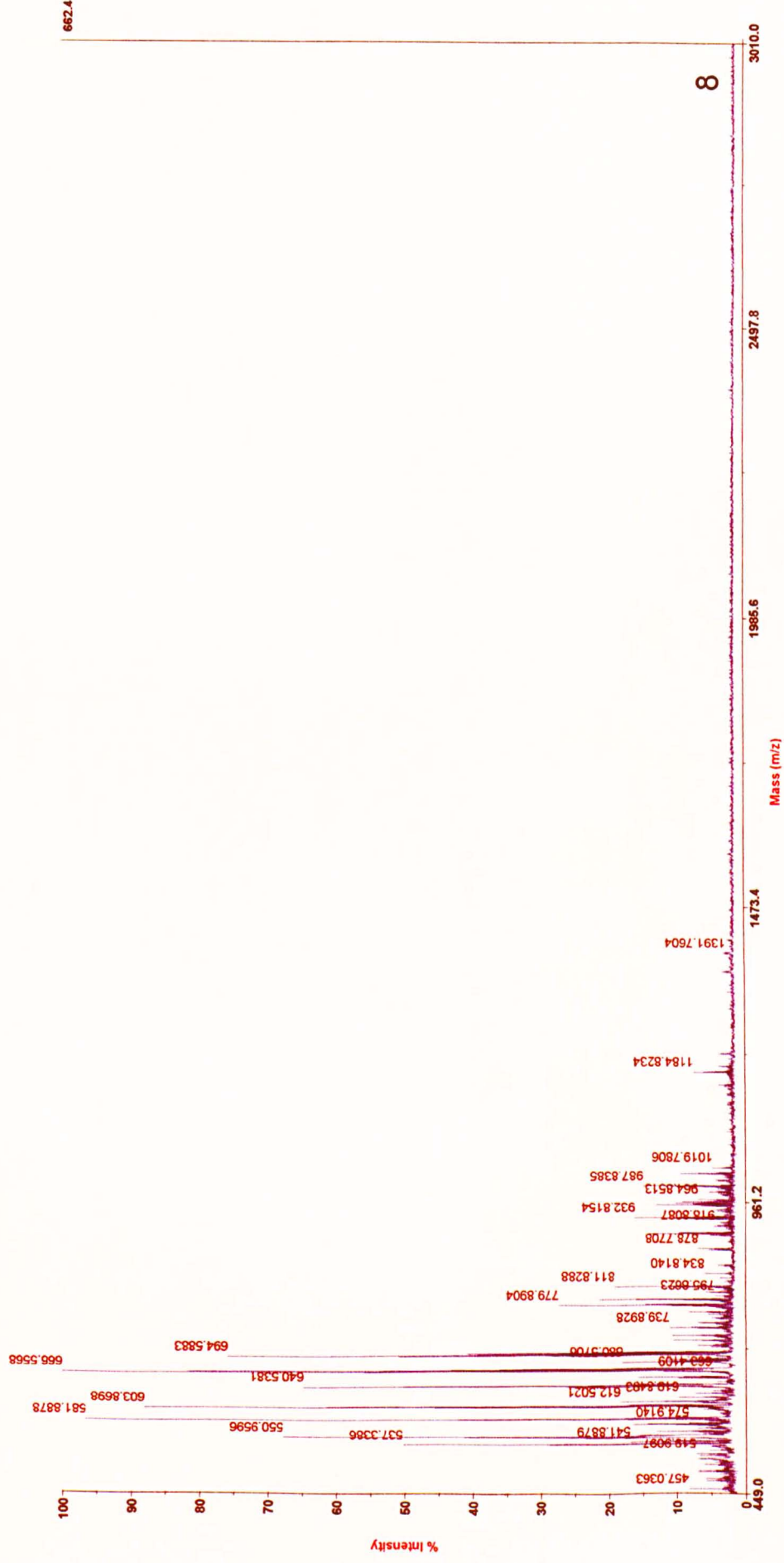
Overlaid Traces(4700 Reflector Spec #1 MC[BP = 676.0, 50148] & Not Used & <<C5_MS>> 4700 Reflector Spec #1 MC[BP = 542.0



DHB matrix – positive mode

Polysaccharide

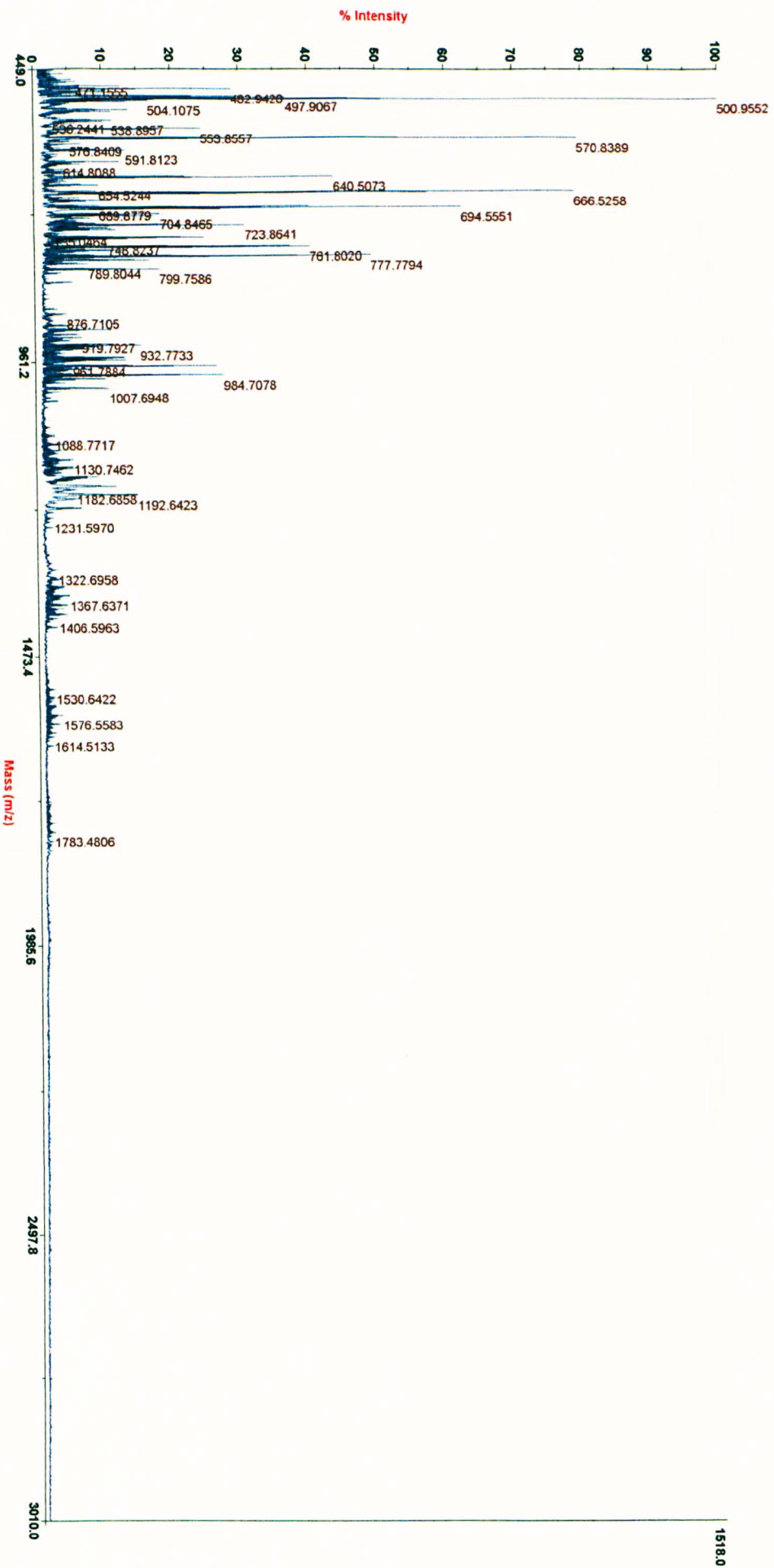
4700 Reflecter Spec #1 MC[BP = 666.6, 662]



DHB matrix – positive mode

DHB alone control

4700 Reflecter Spec #1 MCIBP = 501.0, 1518]



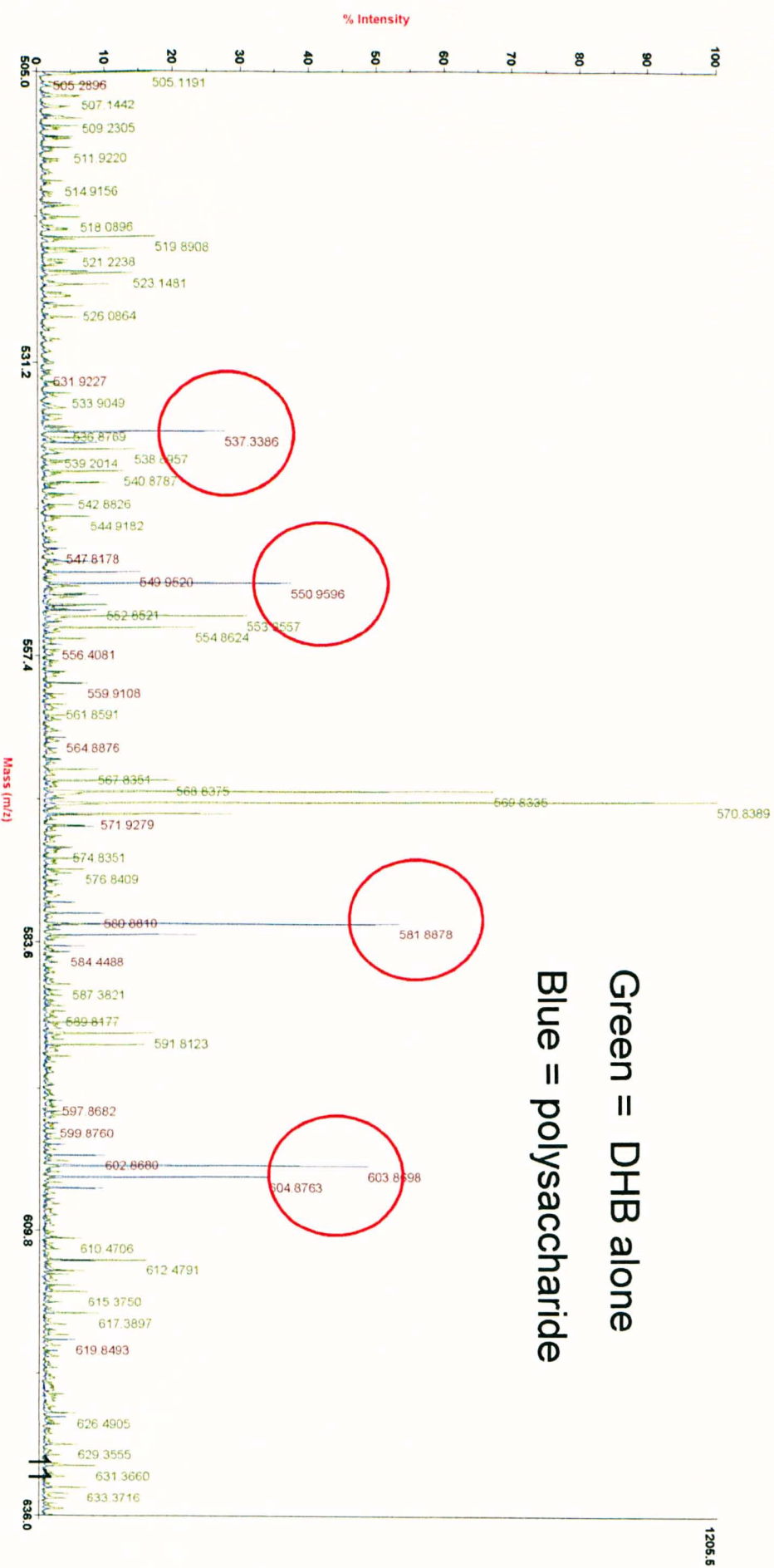
Polysaccharide and DHB alone overlay (positive mode)



Polysaccharide and DHB alone overlay (positive mode)

Zoomed in to show where the main differences are

Overlaid Traces(4700 Reflector Spec #1 MC[BP = 501.0, 1518] & Not Used & <<C3_MS>> 4700 Reflector Spec #1 MC[BP = 666.6,



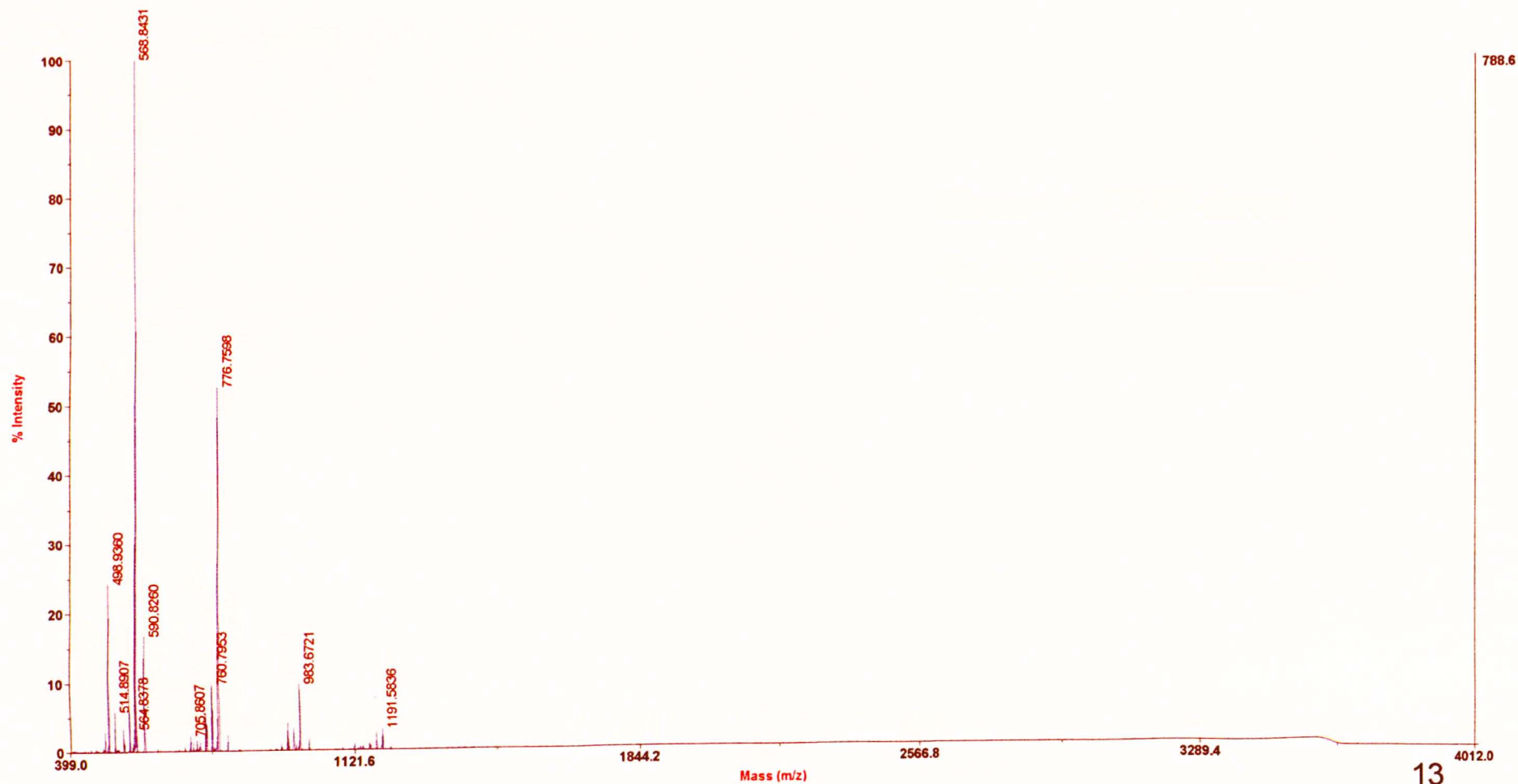
DHB matrix – negative mode Polysaccharide

4700 Reflector Spec #1 MC[BP = 590.9, 1931]



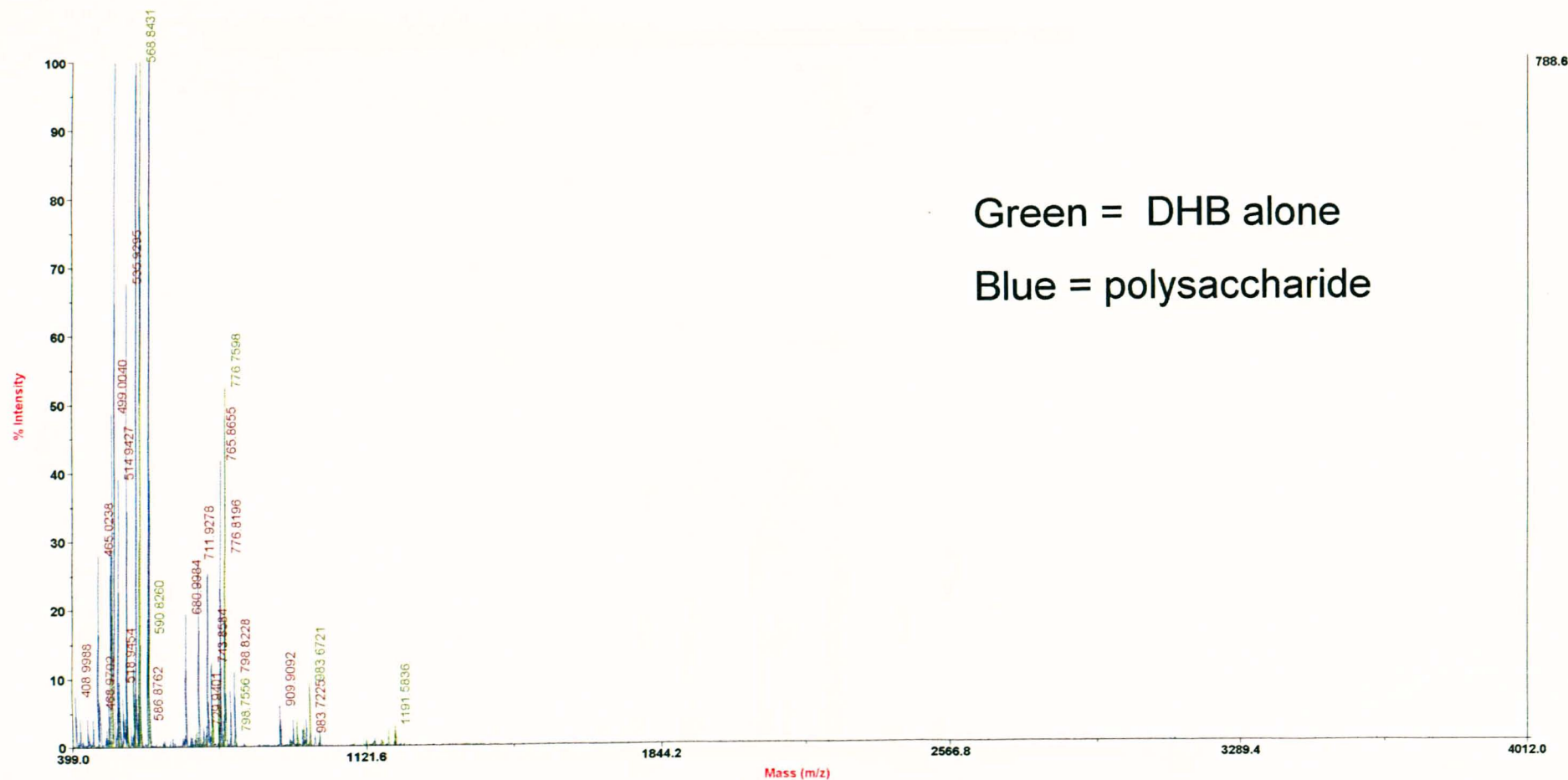
DHB matrix alone – negative mode matrix control

4700 Reflector Spec #1 MC[BP = 568.9, 789]



Polysaccharide and DHB alone overlay (negative mode)

Overlaid Traces(4700 Reflector Spec #1 MC[BP = 568.9, 789] & Not Used & <<C2_MS>> 4700 Reflector Spec #1 MC[BP = 590.9,



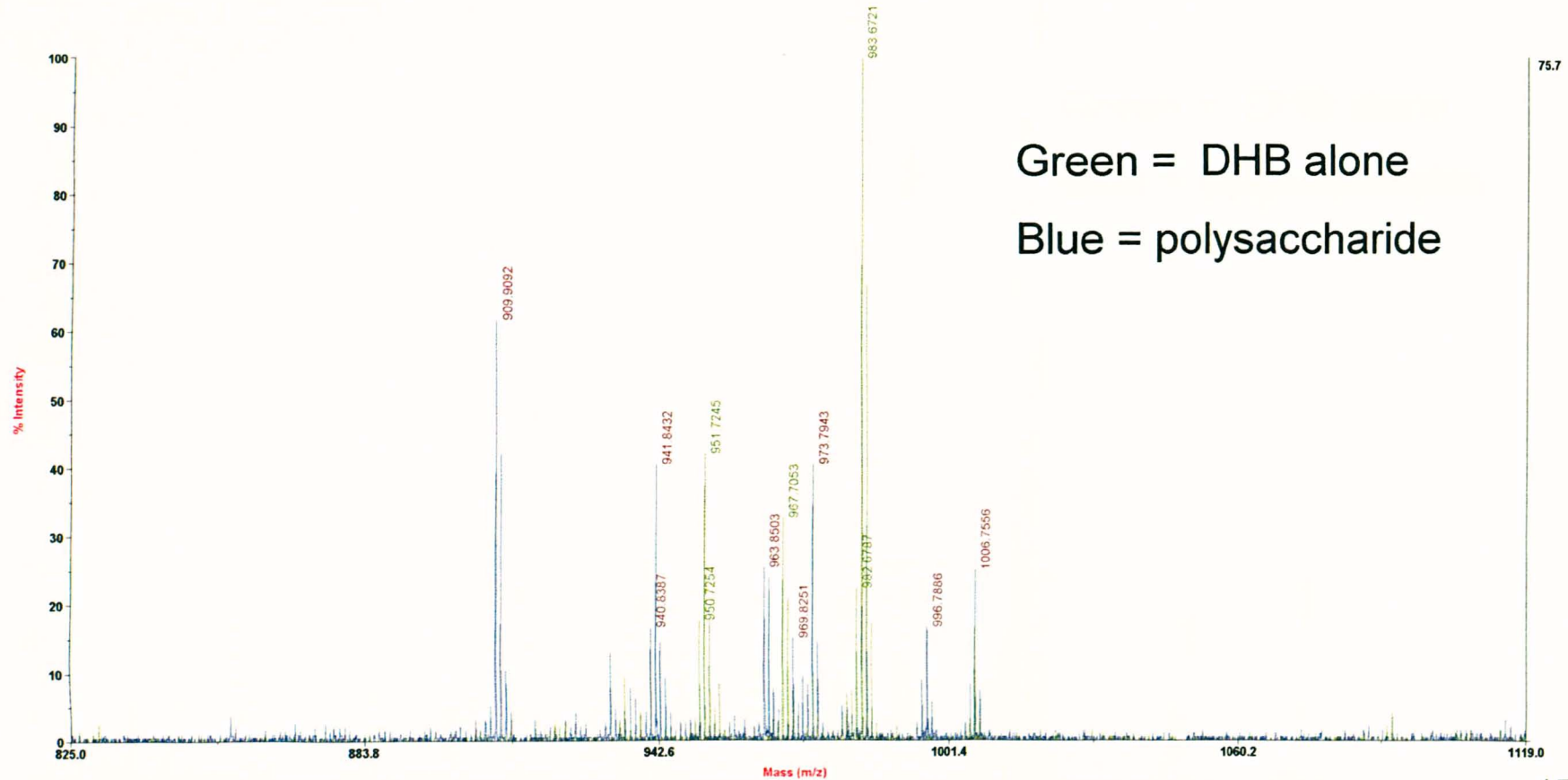
Green = DHB alone

Blue = polysaccharide

Polysaccharide and DHB alone overlay (negative mode)

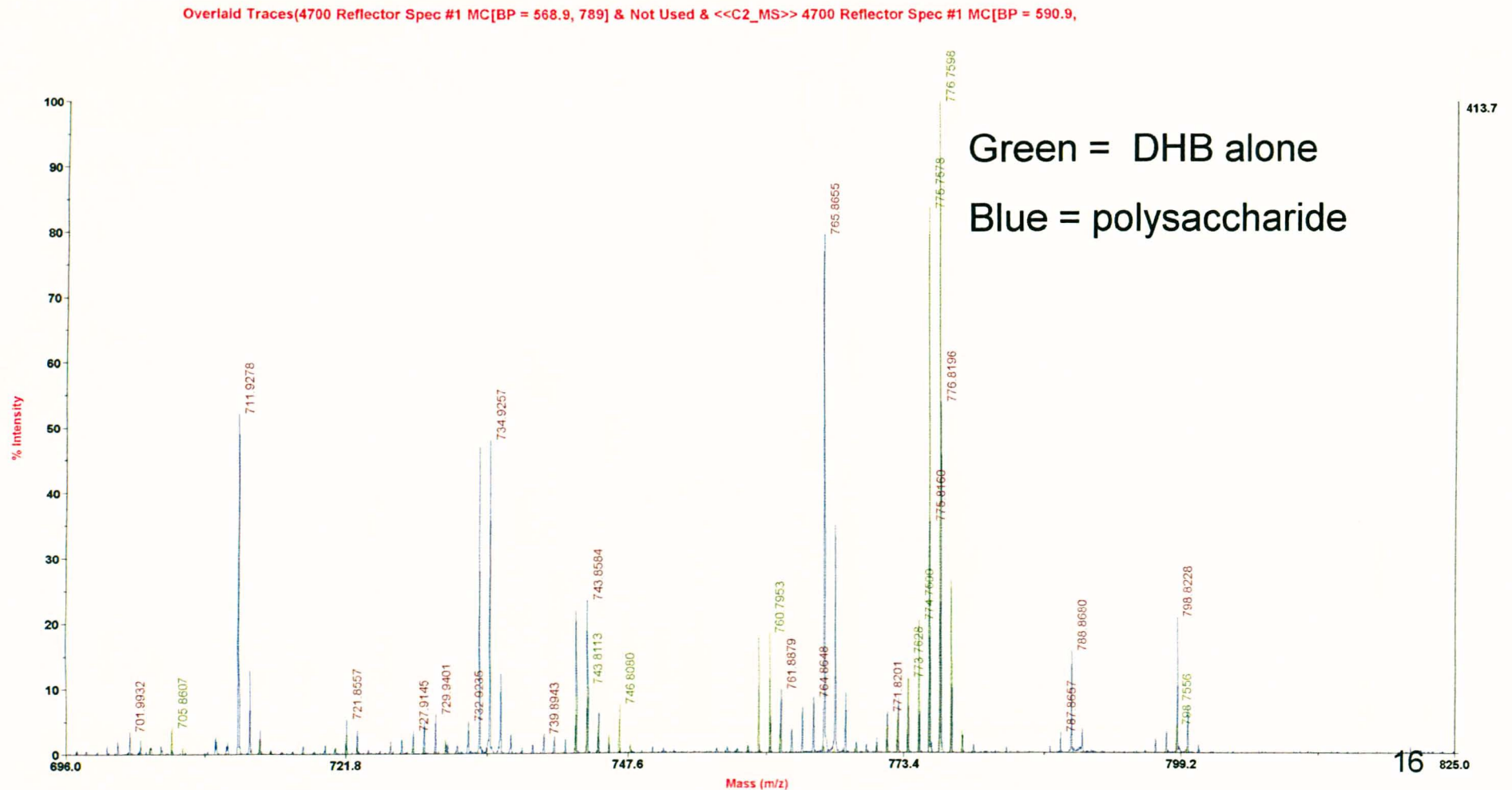
Zoomed in to show where the main differences are

Overlaid Traces(4700 Reflector Spec #1 MC[BP = 568.9, 789] & Not Used & <<C2_MS>> 4700 Reflector Spec #1 MC[BP = 590.9,



Polysaccharide and DHB alone overlay (negative mode)

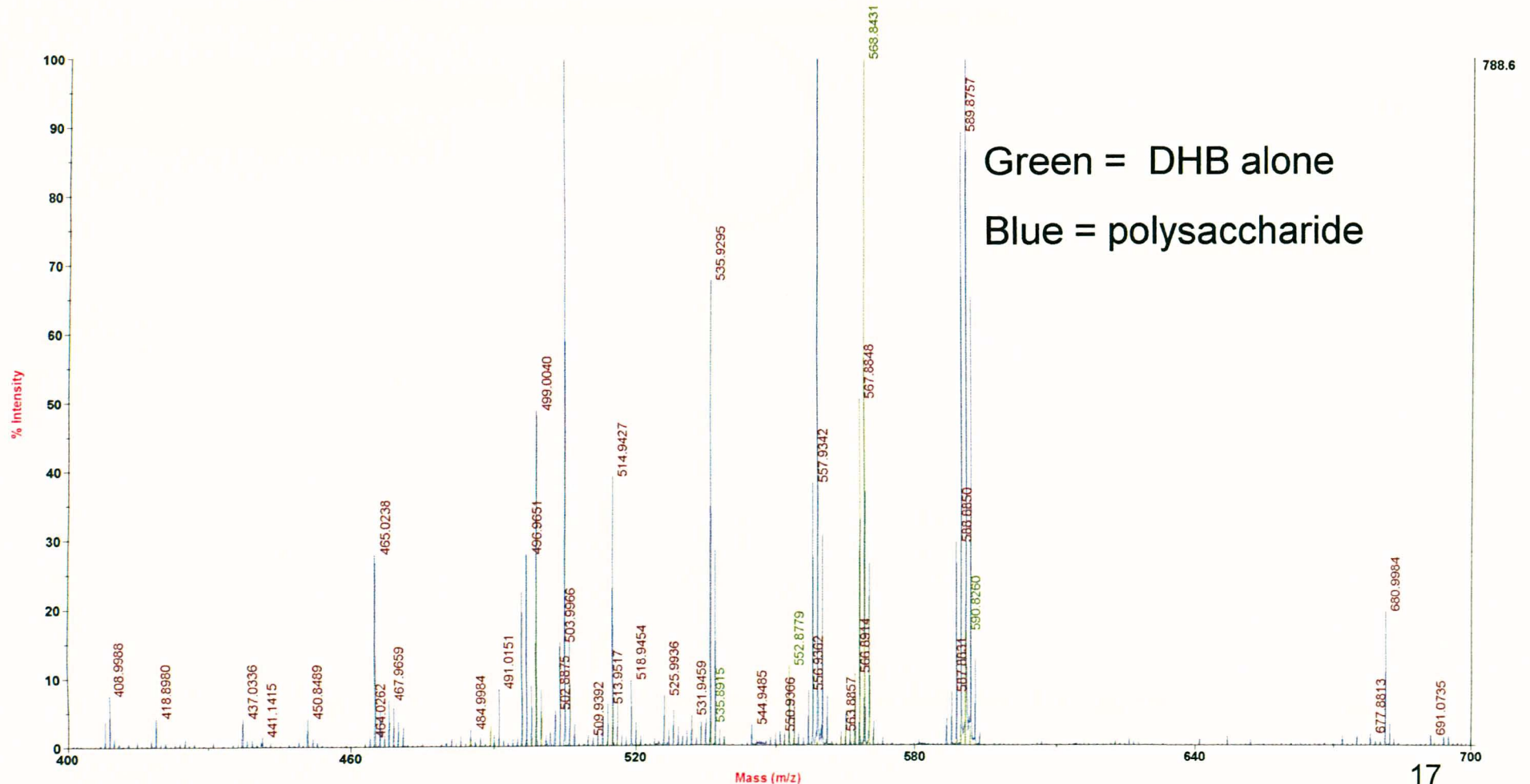
Zoomed in to show where the main differences are



Polysaccharide and DHB alone overlay (negative mode)

Zoomed in to show where the main differences are

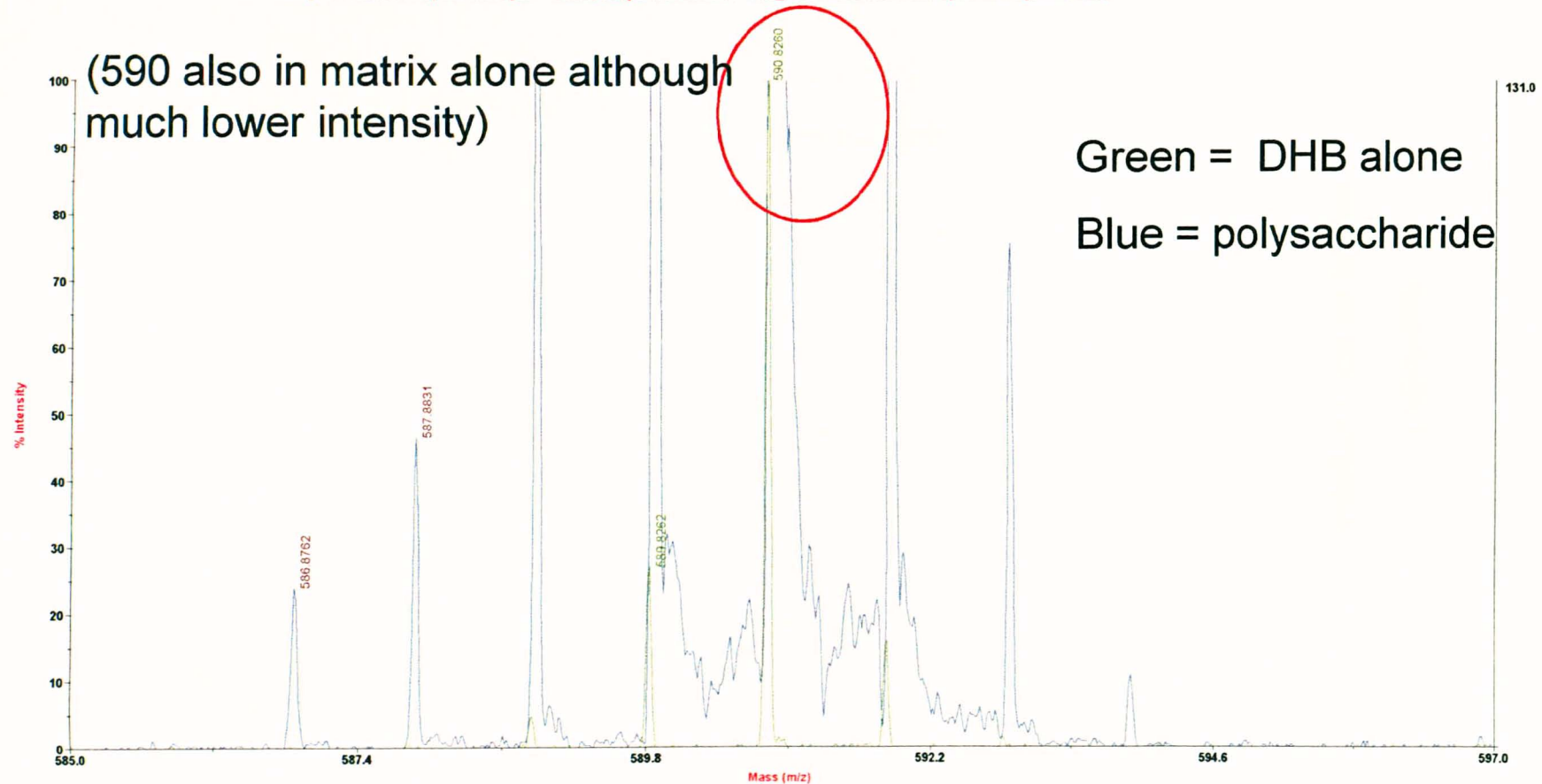
Overlaid Traces(4700 Reflector Spec #1 MC[BP = 568.9, 789] & Not Used & <<C2_MS>> 4700 Reflector Spec #1 MC[BP = 590.9,



Polysaccharide and DHB alone overlay (negative mode)

Zoomed in to show where the main differences are

Overlaid Traces(4700 Reflector Spec #1 MC[BP = 568.9, 789] & Not Used & <<C2_MS>> 4700 Reflector Spec #1 MC[BP = 590.9,



Appendix 3

MALDI MS of *E. coli* lipid A

**MALDI-TOF analysis of *E. coli*
0111.B4 lipid A**

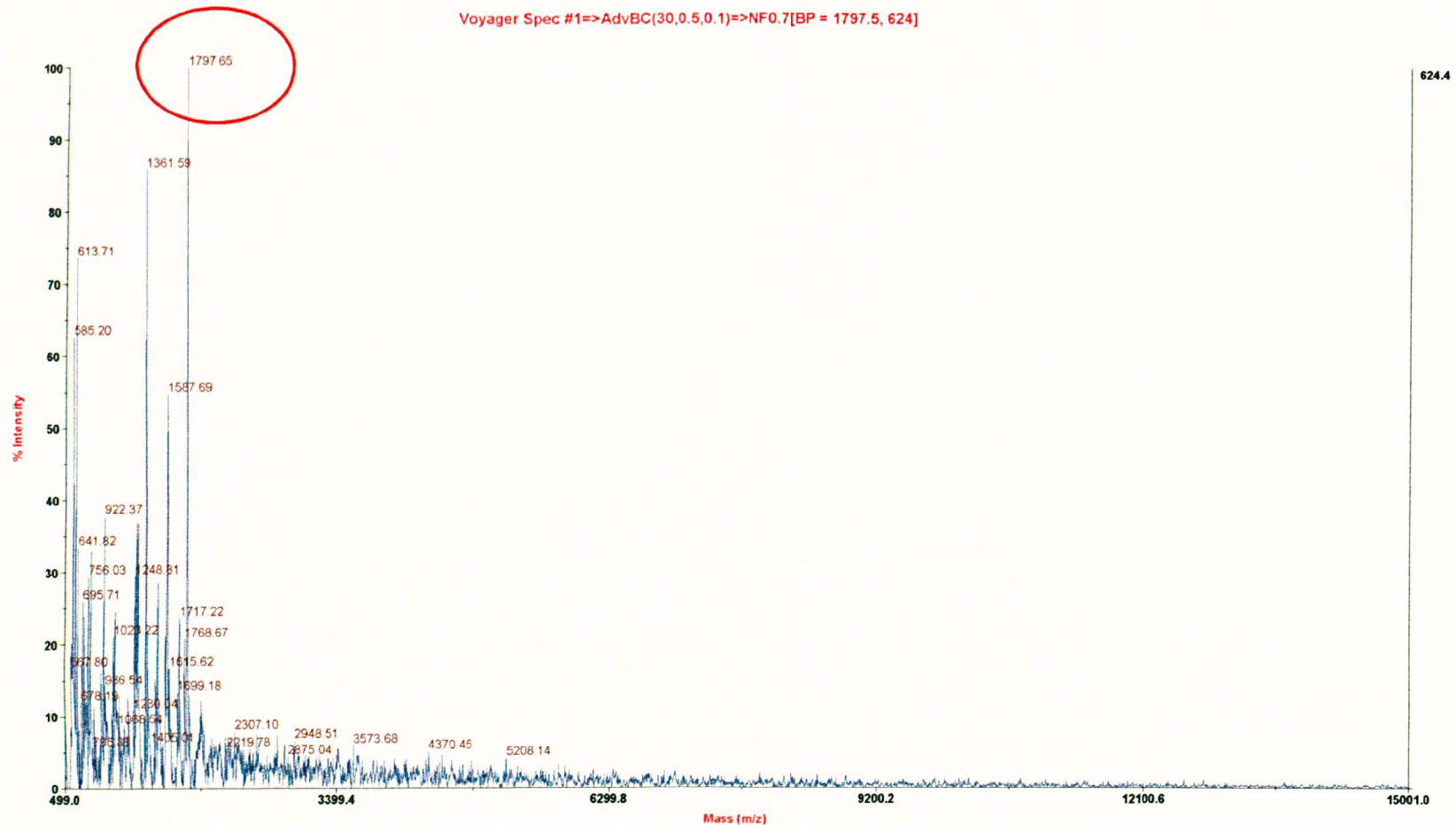
Instrument settings:

Applied Biosystems Voyager DE Pro

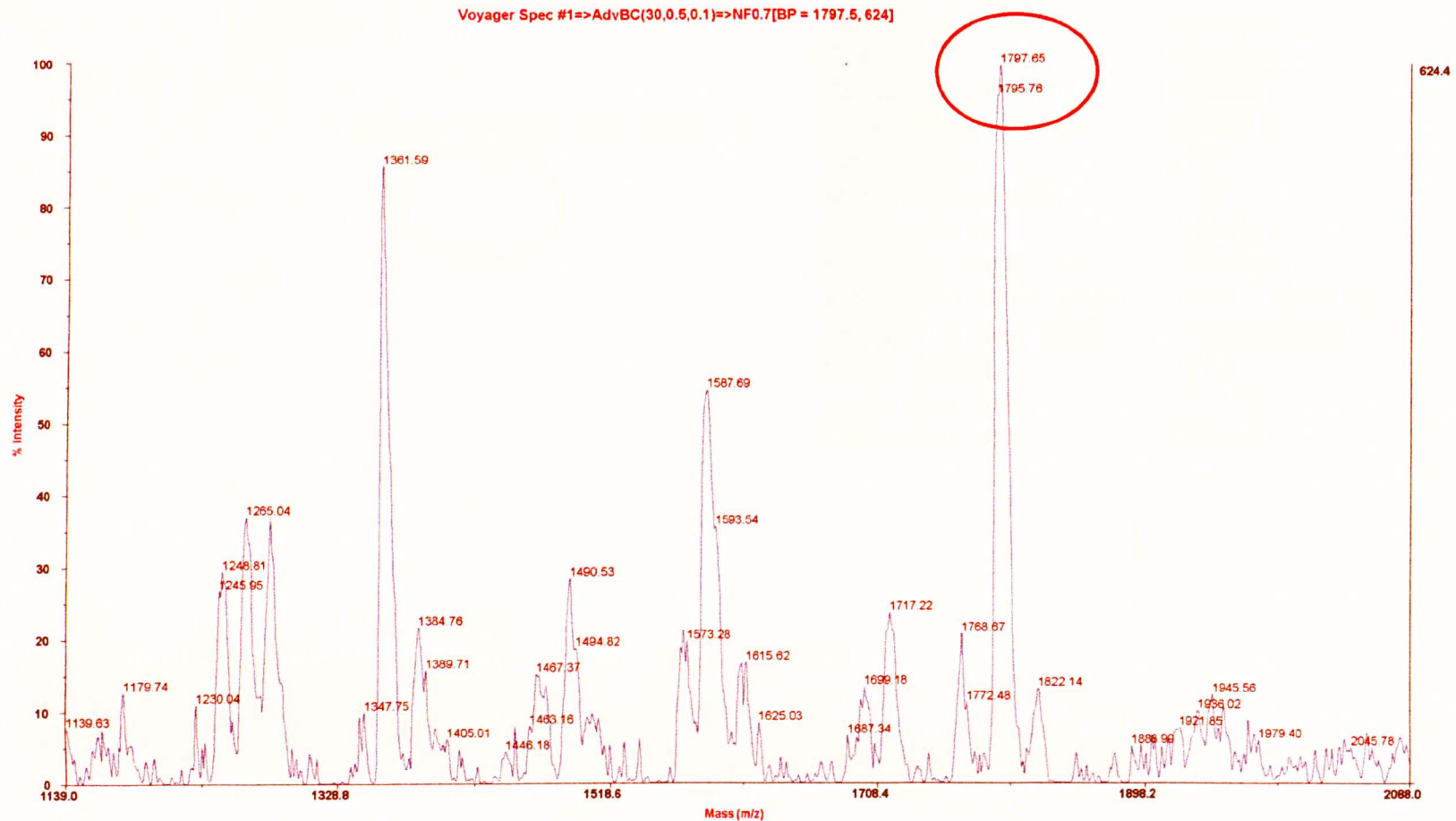
- Mode of operation: Linear
 - Extraction mode: Negative
 - Acquisition control: Manual
 - Polarity: Delayed
-
- Accelerating voltage: 25000V
 - Grid voltage: 94%
 - Guide wire 0: 0.05%
 - Extraction delay time: 800 nsec
-
- Laser intensity: 2499
 - Laser Rep Rate: 20.0Hz
 - Calibration type: Default
 - Calibration matrix: 2,5-Dihydroxybenzoic acid
 - Low mass gate: 500 Da

Lipid A – wide mass range

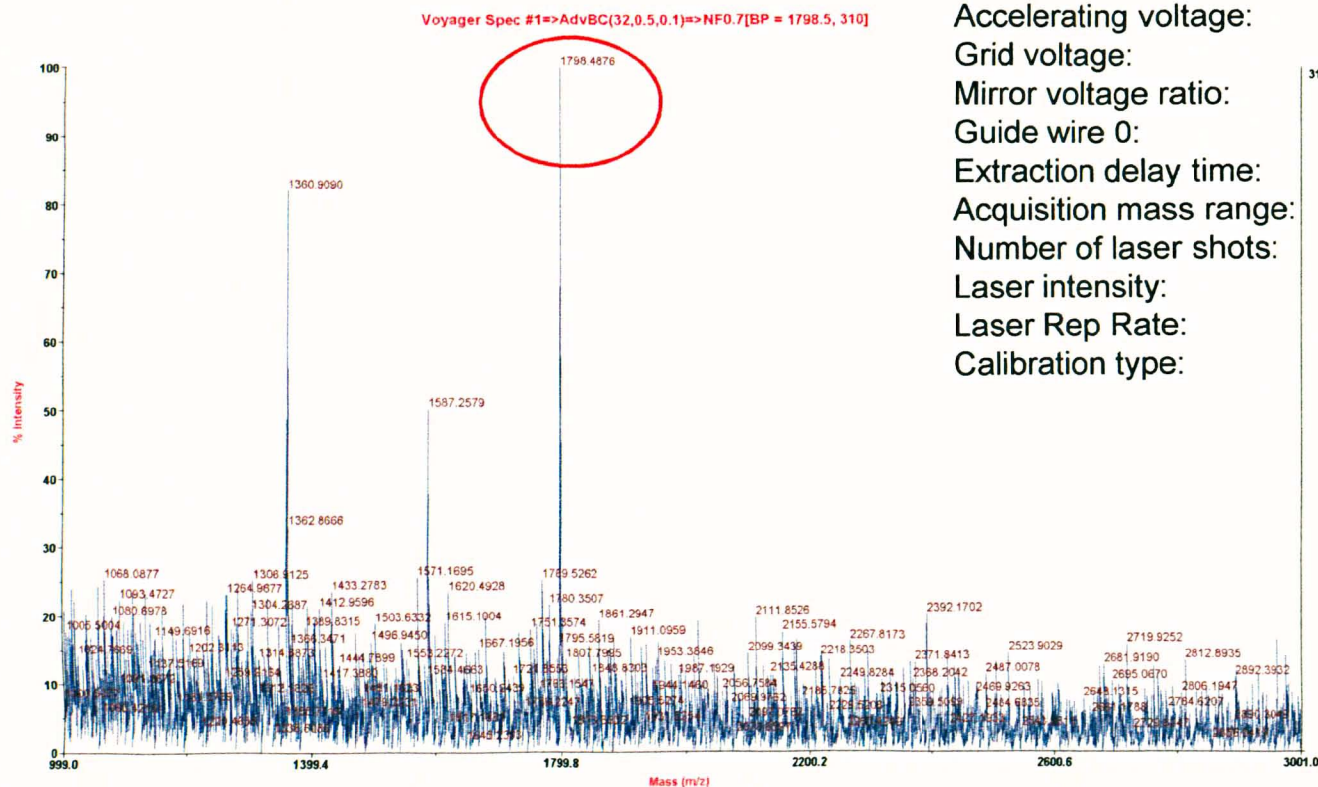
Mass accuracy is not great since it is over a wide mass range and also I had nothing suitable for calibration. Lower molecular weight peaks are probably due to matrix (see slide 5).



Lipid A – wide mass range (zoomed in image – again mass accuracy not great)



Better mass accuracy (spectrum acquired over smaller mass range)



Mode of operation:

Extraction mode:

Polarity:

Acquisition control:

Accelerating voltage:

Grid voltage:

Mirror voltage ratio:

Guide wire 0:

Extraction delay time:

Acquisition mass range:

Number of laser shots:

Laser intensity:

Laser Rep Rate:

Calibration type:

Reflector

Delayed

Negative

Manual

25000 V

75%

1.12

0.003%

100 nsec

1000 -- 3000 Da

150/spectrum

2618

20.0 Hz

Default

2,5 DBH only

Voyager Spec #1=>AdvBC(32,0.5,0.1)=>NF0.7[BP = 697.5, 21414]

