Structural characterisation of oligosaccharide recognition by surfactant protein D

Jamie Reginald Littlejohn

Thesis for submission towards the degree of Doctor of Philosophy

June 2018

Keele University

Acknowledgements and thanks

I would like to thank my two supervisors, Trevor and Annette, for giving me the opportunity to complete my PhD. Thank you for your continued support, guidance and most of all, your patience throughout the last four years; I'm sure I have not been easy at times. I would also like to thank our collaborators at the University of Southampton, Dr. Howard Clark and Dr. Jens Madsen, who provided me with the purified protein that I have used throughout these projects, making the work possible. To Ian Burns, I thank you for your technical assistance and expertise throughout my project and growing the crystals I was able to use. My thanks also go to Dr. Amy Shaw and Dr. Caroline Smallcome who did some preliminary work using small glucosyl ligands and provided a foundation for my work.

In addition, I would also like to thank my family and friends. To my Mum and Dad, thank you for believing in me from the beginning and being there throughout those long, coffee fuelled nights, even when I was not at my best. Thank you to my Granma, Grandad and Nanar for inspiring me and always making me smile. Thank you to Will, Ruben, Harry (Harold), and Rob for your friendship and company on those long days and nights at Diamond and always finding the brighter side of things. Jenny, Lindsay and Pam, thank you for always finding my smile, even on the dark days, and making sure I can always see the silver-lining.

Most of all I would like to thank Claire. Without you I could not be here now. Everything I have achieved is because of you. How you have put up with me working all hours, tearing my hair out, I do not know but I thank you for every second. I look forward to doing the same for you when you start writing up.

Abstract

Human surfactant protein D (hSP-D) is a C-type lectin and member of the collectin family that is involved in pathogen recognition as part of the innate immune response. Recognition occurs through the carbohydrate arrays on the surfaces of the pathogens. This work focuses on characterising, at the atomic level, the recognition of oligosaccharide analogues of the carbohydrate arrays using a recombinant fragment of human surfactant protein D (rfhSP-D).

The crystal structures of three rfhSP-D complexes with the α(1→6)-linked isomaltotriose, β(1→4)-linked cellotriose and β(1→3)-linked laminaritriose have been successfully solved in $P2_1$ using rigid body refinement. Isomaltotriose was refined to 1.96Å with refinement statistics: R_{work} 0.1664 and R_{free} 0.2084. Cellotriose was refined to 1.59Å with refinement statistics: R_{work} 0.0.1879 and R_{free} 0.2106. Laminaritriose was refined to 1.75Å with refinement statistics: R_{work} 0.1808 and R_{free} 0.2193. In combination with two previously solved malto-N-ose complexes, the structures reveal a new binding mechanism for β-D-glucoses in SP-D and identify a new extended binding surface.

The crystal structure of rfhSP-D in complex with the disaccharide unit of peptidoglycan, muramyl disaccharide, extracted from the gram-positive bacteria, *Micrococcus luteus* has been solved in $P2_1$ and refined to 1.95Å with final refinement statistics of R_{work} 0.1708 and R_{free} 0.2169. This crystal structure provides the first insight into recognition of peptidoglycan, an important part of the bacterial cell wall.

The rfhSP-D complex with the R7 rough mutant oligosaccharide from *S. enterica* minnesota has been redetermined from previous work in the group. The structure has been solved in $P2_1$ and refined to 1.75Å with final refinement statistics of R_{work} 0.1730 and R_{free} 0.1960. The crystal structure, along with the known R5 oligosaccharide complex, builds on the

understanding of the flexibility and versatility of lipopolysaccharide recognition by hSP-D and the important role of the two flanking residues, Asp325 and Arg343.

Table of contents

Table of figures

Table of tables

Abbreviations

CCR2	C-C chemokine receptor 2
CL-43	Collectin 43
CL-46	Collectin 46
CL-K1	Collectin-kidney 1
CL-L1	Collectin-liver 1
CL-LK	Collectin liver-kidney heterocomplex
CL-P1	Collectin-placenta 1
CRD	Carbohydrate recognition domain
CTLD	C-type lectin domain
DC-SIGN	Dendritic cell-specific ICAM-grabbing non-integrin
DC-SIGNR	Dendritic cell-specific ICAM-grabbing non-integrin receptor
DNA	Deoxyribonucleic acid
Glc	Glucose
GlcNAc	N-acetyl-D-glucosamine
HA	Haemagglutinin
Hep	L-glycero-D-*manno*-heptose
HIV	Human immunodeficiency virus
IAV	Influenza A virus
ICAM	Intercellular adhesion molecule
IP_1	Inositol-1-phosphate
Kdo	3-deoxy-D-*manno*-oct-2-ulosonic acid
LPS	Lipopolysaccharide
LTA	Lipoteichoic acid
ManNAc	N-acetyl-mannnosamine
MASP	MBP-associated serine proteases
MBP	Mannose binding protein
MPD	2-methyl-2,4-pentandiol
MR	Macrophage mannose receptor
MurNAc	N-acetyl-D-muramic acid
NF-κB	Nuclear factor κB
OSCAR	Osteoclast-associated receptor
PAMPS	Pathogen-associated molecular patterns
PRR	Pattern recognition receptors
rfhSP-D	Recombinant fragment of human surfactant protein D
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SIRP-α/β	Signal inhibitory regulatory protein α/β
SP-A	Surfactant protein A
SP-D	Surfactant protein D
SRA	Scavenger receptor type-A
TA	Teichoic acid
TNF-α	Tumour necrosis factor α
TLR	Toll-like receptor

Chapter 1: Introduction to the collectins, surfactant protein D and its ligands

1.1 Introduction to the immune system

The immune system is a complex and essential part of every multicellular organism that is responsible for protecting the organism from potentially pathogenic infections from a range of bacteria, viruses, fungi and protozoa. In the phylogenetically ancient horseshoe crab, *Limulus polyphemus*, the immune system is reliant on a small number of systemic immune molecules that are able to specifically bind a wide range of different targets and effectively removing them from the system [Marchalonis and Edelman, 1968; Roche and Monsigny, 1974; Kawabata and Iwanaga, 1999]. In humans, the immune system has evolved into a large, interconnected web of immune proteins and highly specialised immune cells, concerting a response to invading pathogens and minimising the damage the pathogen can cause [Janeway, 1989; Chaplin, 2006]. This has led to the human immune system being divided into two sections: (1) the innate, protein driven, immune system; and (2) the acquired immune system.

1.1.1 The innate immune system

The innate immune system serves as the 'first line of defence' when the immune system is challenged because of its ability to recognise a huge range of pathogens by common recognition markers on the pathogens surface, referred to as pathogen-associated molecular patterns (PAMPs). Many of these markers are often integral to the pathogen which makes them good targets for pattern recognition receptors (PRRs) of the immune system as they will be expressed throughout whole genera of pathogens, leading to the innate immune system being described as non-pathogen specific [Kawai and Akira, 2010]. However, PRRs recognise specific patterns (PAMPS) allowing the immune response to

selectively target pathogens rather than structures regularly found on host cells, preventing an autoimmune activation.

1.1.2 The acquired immune system

The acquired or "adaptive" immune system responds less rapidly to an infection of a new microbe as it must be primed to elicit a response against a particular pathogen because not only is the response considerably more pathogen-specific, but also a molecular "knowledge" of the pathogen is required before an effective immune response can be mounted. Many of the molecules in the acquired immune response, such as antibodies secreted by differentiated plasma cells, are highly specific to a single epitope on the surface of a pathogen and, therefore, do not bind more than one type of pathogen [Medzhitov, 2007]. A second presentation of the microbe results in clearance of an infection more efficiently than the innate system as the previously generated antibodies are preserved on the surface of memory T cells and memory B cells in the circulation [MacLeod, Kappler and Marrack, 2010; Pape *et al.*, 2011].

1.1.3 The C-type lectin superfamily and their role in the immune system

An essential part of the innate immune system are the C-type lectins, a large family of calcium-dependent carbohydrate recognition proteins that share homology in a characteristic lectin domain. The lectin domain is formed primarily from an antiparallel beta sheet comprising five or six beta strands and two flanking alpha helices, forming a globular domain with a hydrophobic core, with the additional beta strand (β1 in Figure 1.1) forming a beta hairpin at the N-terminal of the domain [Zelensky and Gready, 2005]. Beyond the highly structured part of the domain, there are two loop regions between β3 and β4 that form the main calcium coordinating regions and form the main ligand recognition surface. The first, or short, loop region (L1, Figure 1.1) is widely conserved within C-type lectin domains (CTLDs) and has been shown to have a wide range of

topology between proteins, depending on the presence of the third calcium ion (Ca3). The second loop region (L2) is less widely conserved and has led to two distinct groups of CTLDs being identified: canonical CTLDs that have a long L2 region extending from the C-terminal of the L1 region; and compact CTLDs that do not have the long L2 region and, in some cases, lacks the L1 loop [Weis, Taylor and Drickamer, 1998; Zelensky and Gready, 2005].

As previously alluded to, the number of calcium ions varies across the C-type superfamily with some containing as few as one calcium ion, such as in the CTLD of E- and P-selectin, whilst others, including surfactant protein D, can contain up to three calcium ions [Somers *et al.*, 2000; Shrive *et al.*, 2003]. The most conserved calcium binding site is the first binding site (Ca1, Figure 1.1), formed between the C-terminal end of the L2 region (where present) and the fifth beta sheet (β5, Figure 1.1). The remaining two calcium ions bound via the loop regions, Ca2 and Ca3, are more variable than the Ca1 site, especially the Ca3 site which varies even among subfamilies of C-type lectins; such as the collectins [Weis and Drickamer, 1994; Shrive *et al.*, 2003; Feinberg, Taylor and Weis, 2007; Girija *et al.*, 2015]. Interestingly, beyond these three main calcium ion sites, a fourth alternative calcium binding site has been identified at the base of the domain, between α2 and the beta sheet, of C-type lectins that may have downstream functions in the activation of immune cell receptors [Zelensky and Gready, 2005].

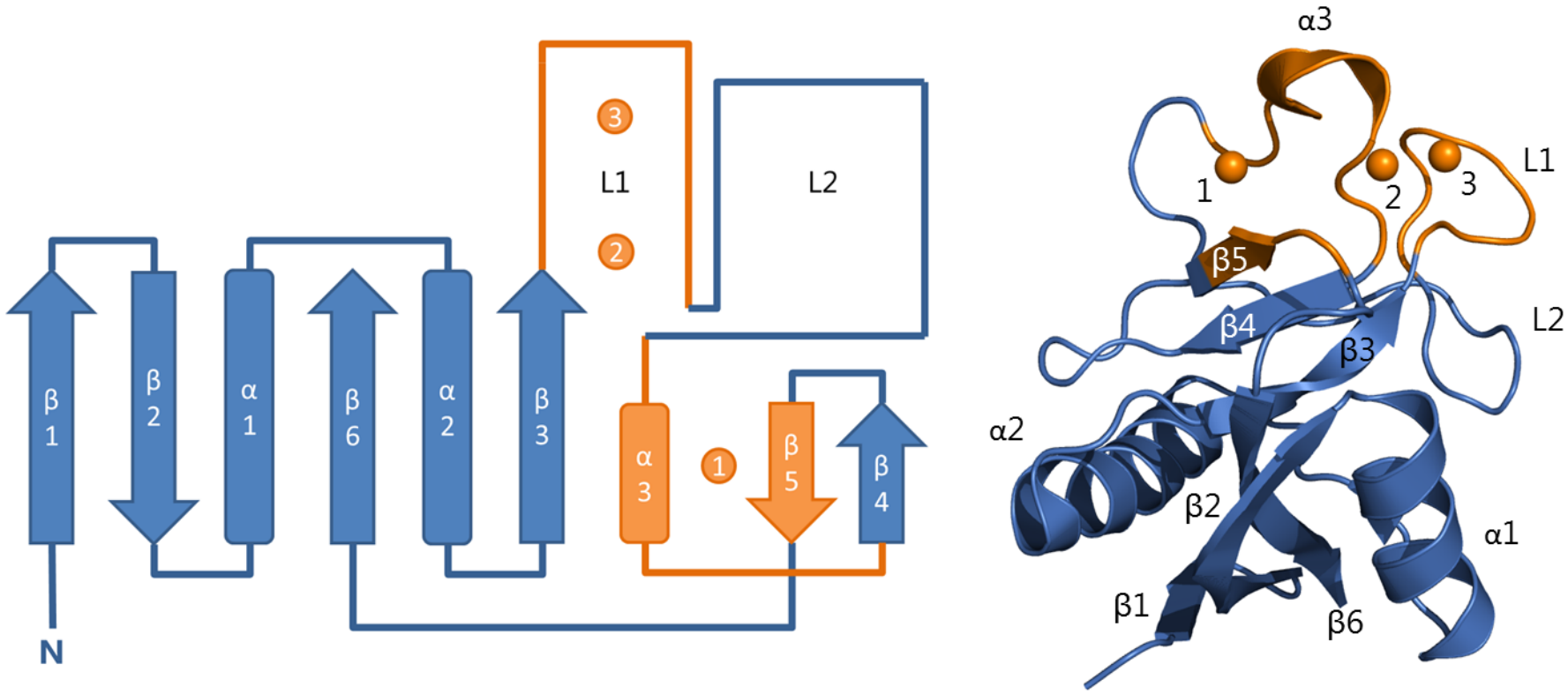


Figure 1.1: The C-type Lectin Domain. Left: 2-Dimensional, expanded representation of the C-type lectin domain (blue) and the calcium coordinating elements (orange). Right: 3-Dimensional ribbon diagram of the C-type lectin domain of human surfactant protein D, coloured to match left. The three calcium ions (represented by orange circles/spheres) are coordinated by the loop region 1 (L1), the C-terminal of loop region 2 (L2) forming an alpha-helical turn (α3), and the fifth beta strand (β5). The sixth beta strand (β6) inserts between the two alpha helices to complete the beta sheet. Residues 235 – 355 of a recombinant fragment of hSP-D (1PW9) was used to generate the ribbon diagram [Shrive et al., 2003].

The conservation of the C-type lectin fold results in members of the superfamily having numerous roles in the immune system that relate to their location in relation to the cell. This allows the family to be broadly divided into two groups that have different immunological roles. The soluble lectins, which include the collectins, have been identified as PRRs in most tissues and systemically in circulation; contributing to a significant part of the innate immune systems ability to recognise pathogenic carbohydrate arrays through interactions with the lectin carbohydrate recognition domain (CRD) [Drickamer, Dordal and Reynolds, 1986; Crouch, Smith, *et al.*, 2006; Veldhuizen, Van Eijk and Haagsman, 2011]. Following recognition of the invading microorganism, the soluble lectins then activate and regulate the cellular response to the pathogen by stimulating phagocytosis and modulating the release of inflammatory cytokines [Sano *et al.*, 1999; Jack *et al.*, 2001; Nadesalingam, Reid and Palaniyar, 2005; Wright, 2005; Ip *et al.*, 2008]. This eventually leads to the activation of the acquired immune system. The role of the soluble lectins also extends to regulating other immune systems including complement through interactions between mannose binding protein (MBP), surfactant protein A (SP-A) and, reported more recently, CL-K1 and CL-L1 and key components of the complement cascade, such as the C1 complex and serine proteases that cleave C3 [Watford *et al.*, 2001; Gingras *et al.*, 2011; Henriksen, Brandt, Andrieu, *et al.*, 2013; Ma, Skjoedt and Garred, 2013].

The second, much larger group of C-type lectins that have a role in the immune response are the membrane-bound receptors that recognise a range of ligands from a wide variety of sources. Many of the membrane-bound receptors, including mincle, dectin-1 and -2 and dendritic cell-specific ICAM-grabbing non-integrin (DC-SIGN), recognise the carbohydrate arrays of pathogens and elicit activation of the intracellular signalling pathways through cytoplasmic domains [reviewed by Drickamer and Taylor, 2015]. These signalling pathways are often shared with other immune receptors, such as toll-like receptors in the case of DC-

SIGN activation, and lead to the release of proinflammatory cytokines and trigger phagocytosis of the bound organism [Gringhuis *et al.*, 2009, 2014; Werninghaus *et al.*, 2009; Lobato-Pascual *et al.*, 2013].

The other group of membrane-bound C-type lectins function in cell adhesion and glycoprotein clearance that enable leukocyte migration and minimise the inflammatory glycoprotein components of damaged cells and microorganisms. The selectin family (E-selectin, P-selectin and L-selectin), for example, play a critical role in capturing circulating leukocytes through glycoproteins on their surface and initiate the migration of the leukocyte across the endothelium and into the infected tissue [Somers *et al.*, 2000]. Other membrane-bound C-type lectins such as CL-P1 (or scavenger receptor C-type lectin, SRCL) langerin and macrophage mannose receptor (MR) have been shown to have more versatile roles within the immune response [Weis, Taylor and Drickamer, 1998] with all three involved in clearing glycoproteins by endocytosis. Both MR and langerin can also directly recognise fungal pathogens and stimulate phagocytosis whereas CL-P1 has been shown to have a role cell adhesion comparable to the selectins [Stahl *et al.*, 1978; Taylor, Gordon and Martinez-Pomares, 2005; Feinberg, Taylor and Weis, 2007; de Jong *et al.*, 2010].

1.2 Introduction to the collectins

The collectins are a subfamily of the C-type lectin superfamily that have a characteristic collagen-like domain linked to the N-terminal of the C-type lectin domain by a short alpha-helical neck region. This section introduces the collectin family and the general structural features associated with the mammalian collectins.

1.2.1 Known mammalian collectins

Collectins have been putatively identified in both vertebrate and invertebrate species, including phylogenetically distant members of the *chordata* phylum; all of which share structural and functional homology [Nair *et al.*, 2000; Green, Nair and Raftos, 2003; Green *et al.*, 2006]. The most well understood of the mammalian collectins. A total of nine mammalian collectins have been identified with many being present across more than one genus. A subset of these are only found in *Bovidae* species and are known as the bovine collectins.

Table 1.1: Summary of the mammalian collectins. The table summarises the number of isoforms, important phylogenetic families they have been identified in, the length of the characteristic collagen-like domain (in amino acid residues) and the total number of cysteine residues in the N-terminal that are important for oligomerisation. The data for the number of isoforms, where available, relates to the number of isoforms present in humans.

Collectin	Isoforms[1]	Species	Collagen Domain Length[2]	Total Cysteine Residues[3]	References
Surfactant Protein D	1	*H. sapiens, Suidae, Muridae, Bovidae*	26→202 (176)	2	[Håkansson *et al.*, 1999; Shrive *et al.*, 2003]
Surfactant Protein A	2	*H. sapiens, Muridae*	27→100 (73)	3[5]	[Head *et al.*, 2003; Kishore *et al.*, 2006]
Kid Binding Protein	2*	*H. sapiens, Muridae*	41→101 (60)	3	[Weis *et al.*, 1992; Gingras *et al.*, 2011]
CL-L1, Collectin 10[4], CL-Liver 1[4]	2[#]	*H. sapiens Muridae*	53→112 (59)	1	[Ohtani *et al.*, 1999; Axelgaard *et al.*, 2013]
CL-K1, Collectin 11[4], CL-Kidney 1[4]	2[#]	*H. sapiens, Muridae, Bovidae*	62→101 (39)	1	[Keshi *et al.*, 2006; Hansen *et al.*, 2010]
CL-P1, Collectin 12[4], CL-Placenta 1[4]	1	*H. sapiens, Muridae, Bovidae*	423→589 (166)	N/A	[Ohtani *et al.*, 2001; Coombs *et al.*, 2005]
Conglutinin	N/A	*Bovidae*	44→98 (54)	3[5]	[Lee *et al.*, 1991]
CL-43	N/A	*Bovidae*	48→162 (114)	2	[Hansen *et al.*, 2003]
CL-46	N/A	*Bovidae*	46→216 (170)	2	[Hansen *et al.*, 2002]

[1]Isoforms in *H. sapiens*
[2]Number of amino acids residues in the collagen-like domain
[3]Number of cysteine residues in the N-terminal domain
[4]Alternative name for collectin
[5]One of the cysteine residues are in the collagen-like domain
*In mice
[#]Including the CL-LK hybrid trimer identified by [Henriksen *et al.*, 2013a]

1.2.2 General structure and oligomerisation of collectins

Most collectins identified to date are secreted as homotrimers containing three identical subunits, with the exception of CL-P1 which is a membrane-bound scavenger receptor and hSP-A which may be secreted as a heterotrimer of two, very similar isoforms of hSP-A, SP-A1 and SP-A2 [table 1.1; DiAngelo *et al.*, 1999]. All of the collectins, however, share four major structural domains: (1) the cysteine-rich N-terminal domain; (2) the long collagen-like domain; (3) the α-coiled coil neck domain and; (4) the carbohydrate recognition domain (CRD), or lectin domain, in a distal position to the N-terminal at the C-terminus [Figure 1.2].

The N-terminal domain (1) contains a high proportion of cysteine residues (table 1.1) in all of the soluble collectins, which are involved in forming extensive sulphide bridges between monomers and trimeric units to form larger oligomeric structures [McCormack, Damodarasamy and Elhalwagi, 1999]. Disruption of the sulphide bridges at positions 15 and 20 of rat SP-D had a dramatic effect on its ability to form larger oligomers and attenuated any agglutination of viral particles or stimulation of the immune response, suggesting that the multimerisation through the N-terminal is critical for the collectins to function [Brown-Augsburger *et al.*, 1996].

The collagen-like domain (2) is characterised by a right-handed triple helical bundle formed from three, left-handed helical monomers which contain the Gly-Xaa-Yaa motif associated with collagen [Kishore *et al.*, 2006; Gingras *et al.*, 2011]. As with other collagen helices, proline at position X is commonplace in the collectins, with other amino acids substituted at position Y, resulting in torsional flexibility around the glycine residue that creates a loose coiled conformation, promoting the formation of the triple helix when two other monomers are present. Once the collagen-like domain has formed it is a highly stable structure, forming a rod-like fragment when proteolytically cleaved in SP-D, due to an

extensive network of hydrogen bonds between the backbone amide bonds and carbonyl oxygens [Crouch *et al.*, 1994]. Interestingly, a conserved cysteine residue (Cys20 in MBP) is present at the beginning of the collagen domain that forms a single interchain interaction that appears to aid the assembly of the collagen domain, showing some interplay between the cysteine-rich N-terminal and the collagen-like domain [Hartshorn *et al.*, 1994].

The role of the collagen-like domain extends beyond simple structural strength and separation of the two terminal domains, playing a more active role in the function of all of the collectins due to a number of structural differences. In SP-A and MBP, for example, there is an interruption in the Gly-X-Y motif, where X and Y are substituted with a highly conserved Gln-Gly motif in the X-Y positions, that causes the collagen domain of the trimer to kink, forcing the CRD cluster to be offset compared to the main collagen axis in a similar way to the overall architecture of the C1q hexamer [Drickamer, Dordal and Reynolds, 1986; Voss *et al.*, 1988; King, Simon and Horowitz, 1989; Tenner *et al.*, 1989; Kishore and Reid, 2000]. This is compounded by an additional cysteine present in the SP-A sequence, conserved between the two isotopes and mammalian species, that coincides with the collagen interruption and potentially forms additional disulphide bonds between the monomers of the trimer and may stabilise the larger hexatrimer [Figure 1.3; Head *et al.*, 2003].

In MBP, the kink in the collagen-like domain enables MBP-associated serine proteases (MASPs) to bind to the collagen-like helix in the space created by the kinks in a similar way to C1r and C1s proteases associated with C1q in the C1 complex, allowing the MBP-MASP complex to activate the lectin pathway in the complement system [Schwaeble *et al.*, 2002; Wallis, 2002]. However, SP-A does not activate the complement system because it does not contain the MASP-recognition motif necessary for MASPs to bind MBP, preventing the formation of a SP-A-MASP complex [Watford, Ghio and Wright, 2000; Gingras *et al.*, 2011].

Interestingly, MASPs have recently been identified in complex with CL-L1, CL-K1 and a naturally occurring heteromeric CL-LK, which contains two CL-K1 monomers and a CL-L1 in the trimeric unit [Henriksen *et al.*, 2013a; Henriksen *et al.*, 2013b]. The quaternary structure of these three trimers is not clear but there is increasing evidence that these oligomer-MASP complex have a role in activating the lectin pathway of complement [Henriksen *et al.*, 2013a; Ma, Skjoedt and Garred, 2013].

The α-coiled coil neck region is a flexible region that precedes the C-terminal CRD and is formed from three parallel α-helices in the assembled trimer, forming a left-handed coil [Hoppe, Barlow and Reid, 1994; Shrive *et al.*, 2003]. Remarkably, it appears that the neck region may be fundamental in forming the collagen-like domain as the α-coiled coil neck domain of SP-D has been shown to self-assemble and, in turn, assemble the collagen-like fragment. This property is reflected in the crystal structure of a recombinant fragment of hSP-D (containing a short collagen fragment, the neck domain and C-terminal CRD, residues 179-355), which is in its native trimeric assembly despite lacking the N-terminal and most of the collagen-like domain [Shrive *et al.*, 2003].

The final domain is the CRD (4) that shares homology with the C-type lectins, forming a large globular domain with a hydrophobic core and, in the case of collectins, containing between two and four calcium ions [Sheriff, Chang and Ezekowitz, 1994; Head *et al.*, 2003; Shrive *et al.*, 2003; Girija *et al.*, 2015]. The CRD forms the basis of pathogen recognition in a calcium dependent mechanism, through interactions with the elaborate and often conserved carbohydrate arrays on the surface of microbes and allergens. The mechanism for the recognition of carbohydrates by hSP-D, discussed later in this chapter, has been widely studied and is the focus of the work presented in this thesis. There has also been some success in deciphering the molecular mechanisms of carbohydrate recognition in MBP, SP-A and CL-K1 from a variety of species.

Collectin-P1 is the exception to the general homology. While it does have the characteristic C-type lectin domain at the C-terminus, the α-coiled coil domain and collagen-like domain are switched placing the coiled coil domain proximal to the membrane-anchor domain and the collagen-like domain connecting directly to the CRD, showing closer homology to type-A scavenger receptors (SRAs) in overall domain arrangement [Peiser, Mukhopadhyay and Gordon, 2002]. As a result, it has been suggested that CL-P1 is not a true collectin and, instead, is much more closely related to SRAs with genetic convergence producing a collectin-like CRD with a galactose specificity [Coombs *et al.*, 2005; Feinberg, Taylor and Weis, 2007].

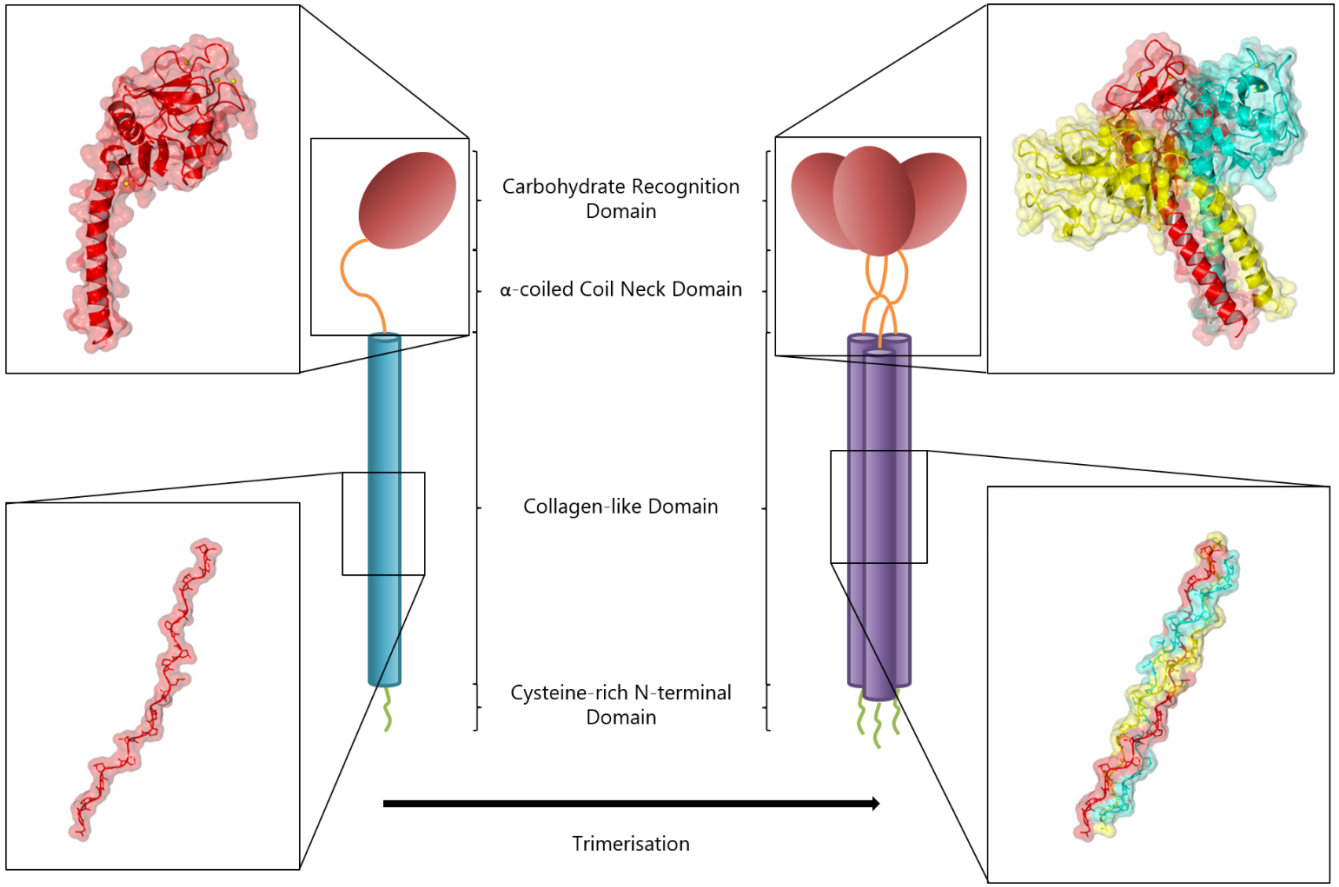


Figure 1.2: Trimerisation of Collectin Monomers. Summary of collectin trimerization, providing detail about each domain (N-terminal domain (green); collagen-like domain (cyan/purple); α-coiled coil neck domain (orange) and; carbohydrate recognition domain (red)) and the characterised structures (where available). Structure for the mannose binding protein collagen-like domain, two lower inserts, solved at 1.49Å, was generated from [3POD, Gingras et al., 2011] and the native structure of the C-terminal fragment of hSP-D, two top inserts, solved at 1.6Å, was generated from [1PW9, Shrive et al., 2003].

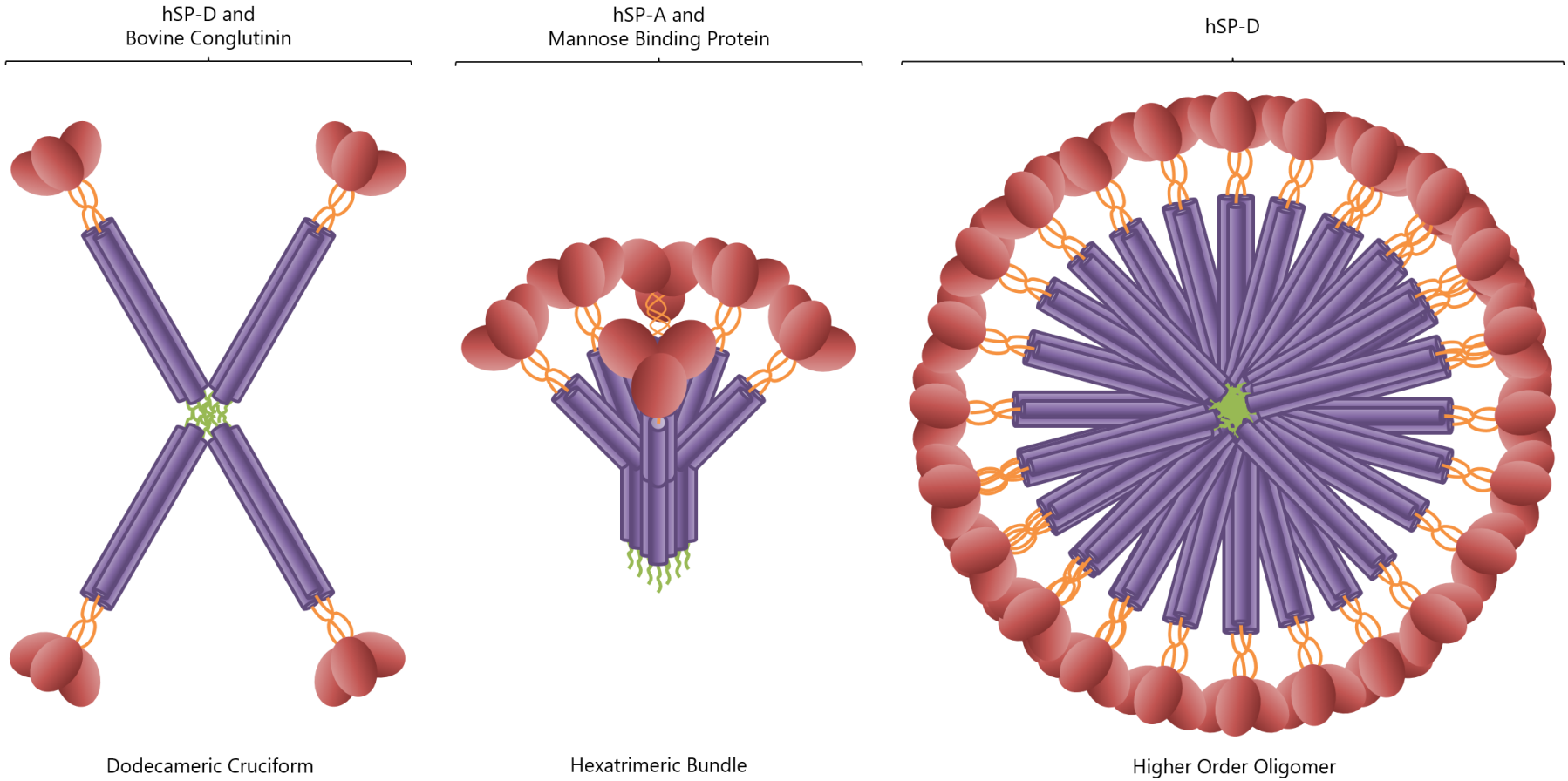


Figure 1.3: Oligomerisation of Collectin Trimers. Summary of the known oligomerisation states of collectin trimers using the same colour scheme as Figure 1.2 for reference domains. Three major oligomers have been identified; Dodecameric (tetramer of trimers) cruciform, Hexatrimeric bundles and; higher order oligomers of hSP-D cruciform. Observed in electron microscopy of bronchoalveolar lavage and blood samples.

A major characteristic of the collectins is their quaternary oligomerisation that is integral to the function of all members of the family [Palaniyar *et al.*, 2002]. Surfactant protein D and conglutinin, a bovine collectin that shares its evolutionary origins with SP-D, have both been observed as dodecameric cruciforms formed around a central 'hub' of covalently linked N-terminal domains of all four trimers [Figure 1.3; Lu *et al.*, 1993; Crouch *et al.*, 1994; Hansen and Holmskov, 2002]. Interestingly, despite the overall homology, the SP-D dodecamer appears to have further associations between the trimers, which form two trimeric pairs that restrict the overall freedom of the trimers around the N-terminal hub compared to conglutinin, which appears to have a greater degree of freedom around the hub [Lu *et al.*, 1993]. Around 10% of SP-D has also been identified in much larger, higher order oligomers containing up to 32 trimeric units arranged radially around the N-terminal hub, forming a 'cart wheel' or 'fuzzy ball' of CRDs around an N-terminal/collagen core [Crouch *et al.*, 1994]. As with the dodecamer, the higher order oligomer appears to contain pairs of trimers which may suggest the large assembly, that brings 96 individual CRDs together, may be a complex of the dodecameric cruciforms linking together through free cysteine residue in the terminal of multiple dodecamers. The high number of CRDs in both the dodecameric and oligomeric complexes is suggested to be important in agglutinating pathogens, promoting recognition by multiple CRDs and crosslinking pathogens and receptors *in vivo* [Crouch *et al.*, 1994].

SP-A and MBP also assemble into larger oligomers of the basic trimer coordinated by disulphide bridges in the N-terminal domain. However, the collagen-like domains in SP-A and MBP assemble into a parallel bundle at the N-terminal and the interruption in the collagen motif allows the recognition domains to project away from each other, in a similar way to the assembled C1q complex [Figure 1.3; Kishore and Reid, 2000]. A recent study has suggested that there may be also be additional trimeric and tetrameric oligomers of MBP

and that these may, in fact, maintain the activation of the lectin pathway in the absence of the hexatrimer, suggesting that the hexatrimeric bundle may not only improve the affinity of pathogen recognition but may improve the MASP activity of the MBP-MASP complex by bringing two MASP proteins together [Kjaer *et al.*, 2016]. This may suggest the hexameric bundle is, in fact, formed from two trimers of trimers.

Unlike the other collectins, CL-L1, CL-K1 and CL-P1 in humans and CL-46 and CL-43 in *Bovidae* have not been observed in large immunological complexes, beyond dimers and trimers, by electron microscopy however mass analysis of CL-L1 and CL-K1 has suggested that there are larger tetramers and hexamers in serum [Keshi *et al.*, 2006; Ohtani, Suzuki and Wakamiya, 2012]. The CL-LK hybrid trimer has also been shown to form trimers in serum, with some speculation that it may be able to form larger hexamers however further work is required to completely characterise CL-LK multimerisation [Henriksen *et al.*, 2013a]. CL-P1 is an exception to the other human collectins because it is membrane-bound scavenger receptor and does not undergo any further oligomerisation in the membrane, although there is evidence that trimers may cluster together at the cell-cell interface where they are active [Ohtani *et al.*, 2001; Coombs *et al.*, 2005]. Surprisingly, considering that oligomerisation of the human collectins seems important for function, the bovine collectin CL-43 has only been observed as single trimeric units and does not form any oligomeric complexes but has been shown to have a high affinity for viral particles [Holmskov *et al.*, 1995; Rothmann *et al.*, 1997; Hartshorn *et al.*, 2002]. The multimerisation of CL-46, the final bovine collectin descended from SP-D, is widely reported to form a dodecameric cruciform, similar to SP-D and conglutinin, however there is no direct evidence of this in the literature.

1.2.3 Localisation of human collectins

The human surfactant proteins, SP-D and SP-A, are both primarily expressed in the lining of the alveolar sacs in the bronchoalveolar lavage, secreted by the alveolar type II cells, clara cells and alveolar macrophages [Voorhout *et al.*, 1992]. SP-D has been largely associated with the aqueous sublayer of the lavage, closer to the cell surface, whereas SP-A more readily associates with the lipid in the surfactant, actively regulating how the surfactant is behaving and preventing the air sacs from closing by reducing surface tension during exhalation [Madsen *et al.*, 2000].

Beyond the epithelia of the lung, expression of hSP-D has been identified in a range of other tissues, including the majority of the mucosal tissues and many other important tissues such as the brain, kidneys and heart [Madsen *et al.*, 2000; Leth-Larsen *et al.*, 2004; Ujma *et al.*, 2017]. More recently, it has also been identified in the secretion from the gallbladder, where it is believed to play a role in regulating the gut flora to maintain healthy digestion, and the eye of mouse models infected with *S. aureus* [Zhang *et al.*, 2015; Sarashina-Kida *et al.*, 2017]. Similarly, expression of SP-A has been identified in the gastrointestinal tract, genital and reproductive tracts and beyond, playing an important role in host defence against a range of opportunistic pathogens, including human immunodeficiency virus and *Chlamydia trachomatis* [Oberley, Goss, *et al.*, 2004; Meschi *et al.*, 2005]. The localisation of the surfactant proteins at the interfaces between the host and the outside world, where cells are under a constant bombardment of potentially toxic and pathogenic substances and organisms has led to hSP-D and hSP-A being identified as a key part of the 'first line of defence' in the innate immune response.

By comparison, the other collectins are found more systemically and not at the epithelial interfaces where the surfactant proteins localise. Mannose binding protein and CL-L1 are primarily synthesised in hepatocytes where they are then secreted into the bloodstream

and enter general circulation, operating as an underlying surveillance mechanism once a pathogen successfully enters the bloodstream [Drickamer, Dordal and Reynolds, 1986; Ohtani *et al.*, 1999; Axelgaard *et al.*, 2013]. Similarly, CL-K1 is synthesised in renal cells and secreted back into the blood following renal processing, allowing the CL-K1 to immediately enter the circulation. Interestingly, the natural occurrence of the CL-LK hybrid trimer suggests that there is also colocalisation of CL-K1 and CL-L1 monomers as the trimers are secreted in their assembled form [Henriksen, Brandt, Andrieu, *et al.*, 2013]. This has putatively been identified as hepatocytes, however further studies are required to complete the characterisation of the synthesis. The exception to this, is CL-P1, or scavenger receptor C-type lectin, which is membrane-bound. CL-P1 was first identified in the placenta from its gene sequence homology with the other collectins but has since been identified in human umbilical vein endothelial cells (HUVECs) and the systemic endothelial cells. Wider expression in most tissues has however been observed [Ohtani *et al.*, 2001].

1.2.4 Carbohydrate recognition by the collectins

The main function of the collectins is to recognise carbohydrates via the carbohydrate recognition domains that they all share with the wider C-type lectin superfamily. The CRD in the collectins is particularly well conserved between family members with the calcium coordinating residues being 88.9% conserved or having similar functional properties, with the only exception being in the two SP-A isoforms that share a conserved proline at position 175 which would otherwise coordinate Ca2 and Ca3 in the other collectins [Figure 1.4]. Despite this, specificity for different carbohydrates has been observed for each collectin which suggests the residues surrounding the calcium-binding pocket play a significant role in selecting ligands [table 1.2].

Table 1.2: Ligand Specificity of the Collectins. Specificity based on IC_{50} values in the literature. Abbreviations: Mal, maltose; Glc, glucose; GlcN, Glucosamine; GlcNAc, N-acetylglucosamine; Man, mannose; ManNAc, N-acetyl-mannosamine; Fuc, fucose; Gal, galactose; GalNAc, N-acetyl-galactose; Lac, lactose

Collectin	Ligand Specificity	References
hSP-D	Mal, ManNAc > L-Fuc > Glc > Man >> GlcN > Gal, GlcNAc	[Ogasawara *et al.*, 1992; Lim *et al.*, 1994; Hansen *et al.*, 2002]
hSP-A	ManNAc > L-Fuc, Mal > Glc > Man/Gal	[Haurum *et al.*, 1993]
hMBP	GlcNAc > Man, L-Fuc > ManNAc >> Mal > Glc	[Haurum *et al.*, 1993]
hCL-K1	L-Fuc >> Man	[Keshi *et al.*, 2006]
hCL-L1	Man/L-Fuc > ManNAc >> Glc, GlcNAc	[Ohtani *et al.*, 2012; Axelgaard *et al.*, 2013]
hCL-P1	GalNAc >L-Fuc, D-Fuc, Gal >> Glc > GlcNAc, Man	[Yoshida *et al.*, 2003]
bCong	GlcNAc >> ManN > L-Fuc, Man > Glc > ManNAc, GlcN, Mal > Gal > D-Fuc	[Haurum *et al.*, 1993; Hansen *et al.*, 2002]
bCL-46	GlcNAc >> ManNAc, ManN > Malt > Glc, Man, L-Fuc >> Gal	[Hansen *et al.*, 2002]
bCL-43	Man, ManNAc > L-Fuc > GlcNAc > Glc, Malt > Gal > Lac >> GalNAc	[Hansen *et al.*, 2002; Hartshorn *et al.*, 2002]

In all of the collectins, there is a preference for glucose-type or mannose-type carbohydrates with some, such as hSP-D, showing higher affinities for disaccharide units compared to the monosaccharides (table 1.2). This is with the exception of CL-K1 which has high affinity for L-fucose which is common in mammalian glycosylation structures despite its irregular L-form [Keshi *et al.*, 2006; Girija *et al.*, 2015]. The ligand specificity results from small changes in the amino acid residues that form the main Ca1 binding pocket and directly influence the accessibility of Ca1, interacting with the ligands to increase binding affinity [Figure 1.4; Shrive *et al.*, 2003, 2009]. Studies using mutants of hSP-D and MBP have investigated the effects of changing or removing the influential side chains that flank the Ca1 pocket. In SP-D, mutations of Asp325 (D325A) and Arg343

(R343V) opens up the binding site, allowing for recognition of much longer mannose ligands [Goh *et al.*, 2013]. Similarly, point mutants of ligand coordinating side chains in rat MBP have identified His202, Asp212, Arg216, Pro217 and Tyr218 as important residues for ligand selection and recognition, going as far as mutating all of the side chains and changing the preference of the CRD [Feinberg *et al.*, 2000]. A similar mutation study using rat SP-A has also shown that substitutions of residues 171, 175, 197 and 203 (which form the loop that flanks the Ca1 binding site) with SP-D-like side chains produces an affinity profile similar to SP-D, confirming that the selectivity is conferred by the residues that flank the binding site [Rynkiewicz *et al.*, 2017].

```
Structural Feature                  |----------------CRD-----------------
Region
Residue Number                    235
                                   |
hSP-D           229          QHLQAAFSQY KKV-ELFPNG QSVGEKIFKT AGFVKPFTEA
hSP-A1           92          HDFRHQILQT RGALSLQGSI MTVGEKVFSS NGQSITFDAI
hSP-A2           92          HDFRHQILQT RGALSLQGSI MTVGEKVFSS NGQSITFDAI
hMBP             92          KALQTEMARI KKW-LTFSLG KQVGNKFFLT NGEIMTFEKV
hCL-L1          116          ARLKTSMKFV KNV---IAGI RETEEKFYYI VQEEKNYRES
hCL-K1          109          SQLTSELKFI KNA---VAGV RETESKIYLL VKEEKRYADA
hCL-P1          587          ---------- ------PPHW KNFTDKCYYF SVEKEIFEDA
bConglutinin    216          RRFQNAFSQY KKA-VLFPDG QAVGEKIFKT AGAVKSYSDA
bCL-46          216          QRLQNAFSRY KKA-VLFPDG QAVGKKIFKT AGAVKSYSDA
bCL-43          165          QRLQNIVTQY KKV-ELFPNG QSVGEKIFKT AGFVKPFTEA


SF      -------------------------------CRD-------------------------------
REGN                                                Ca2  Ca2/3
RESI                                                297   301
                                                      |    |
hSP-D  QLLCTQAGGQ LASPRSAAEN AALQQLV-VA KNEAAFLSMT DSKTEGKFTY PTGESL-VYS
hSP-A1 QEACARAGGR IAVPRNPEEN EAIASFV-KK YNTYAYVGLT EGPSPGDFRY SDGTPV-NYT
hSP-A2 QEACARAGGR IAVPRNPEEN EAIASFV-KK YNTYAYVGLT EGPSPGDFRY SDGTPV-NYT
hMBP   KALCVKFQAS VATPRNAAEN GAIQNLI-K- --EEAFLGIT DEKTEGQFVD LTGNRL-TYT
hCL-L1 LTHCRIRGGM LAMPKDEAAN TLIADYVAKS GFFRVFIGVN DLEREGQYMF TDNTPLQNYS
hCL-K1 QLSCQGRGGT LSMPKDEAAN GLMAAYLAQA GLARVFIGIN DLEKEGAFVY SDHSPMRTFN
hCL-P1 KLFCEDKSSH LVFINTREEQ QWIKKQM--V GRESHWIGLT DSERENEWKW LDGTSP-DYK
bCong  EQLCREAKGQ LASPRSSAEN EAVTQMV-RA QEKNAYLSMN DISTEGRFTY PTGEIL-VYS
bCL-46 QQLCREAKGQ LASPRSAAEN EAVAQLV-RA KNNDAFLSMN DISTEGKFTY PTGESL-VYS
bCL-43 QLLCTQAGGQ LASPRSAAEN AALQQLV-VA KNKHAYLSMN DISKEGKFTY PTGGSL-DYS


SF      ----------------------------CRD--------------------------|
REGN         |----Ca1/2---|3           |--|
RESI       321-324      329330       341-343         355
             | ||         ||           | ||            |
hSP-D  NWAPGEPND- ---DGGSEDC VEIFTNGKWN DRACGEKRLV VCEF------ ----   355
hSP-A1 NWYRGEPAGR -----GKEQC VEMYTDGQWN DRNCLYSRLT ICE------- ----   228
hSP-A2 NWYRGEPAGR -----GKEKC VEMYTDGQWN DRNCLYSRLT ICEF------ ----   228
hMBP   NWNEGEPNN- ---AGSDEDC VLLLKNGQWN DVPCSTSHLA VCEFPI---- ----   228
hCL-L1 NWNEGEPSD- ---PYGHEDC VEMLSSGRWN DTECHLTMYF VCEFIKKK-- ----   257
hCL-K1 KWRSGEPNN- ---AYDEEDC VEMVASGGWN DVACHTTMYF MCEFDKENM- ----   251
hCL-P1 NWKAGQPDNW GHGHGPGEDC AGLIYAGQWN DFQCEDVNNF ICEKDRETVL SSAL   722
bCong  NWADGEPNNS D--EGQPENC VEIFPDGKWN DVPCSKQLLV ICEF------ ----   351
bCL-46 NWASGEPNNN N--AGQPENC VQIYREGKWN DVPCSEPLLV ICEF------ ----   351
bCL-43 NWAPGEPNNR AK-DEGPENC LEIYSDGNWN DIECREERLV ICEF------ ----   301
```

Figure 1.4: Primary Sequence Alignment of Nine Collectins. Key calcium coordinating residues are highlighted and coloured according to similarity to human SP-D (hSP-D). Both isoforms of human SP-A (SP-A1; SP-A2) are represented for completeness. Yellow, identical; green, similar property; cyan, different property (i.e. non-polar); magenta, large side chain with different property. CRD numbering (236-355) based on the native rfhSP-D structure [Håkansson *et al.*, 1999]. Sequences from the National Center for Biotechnology Information (NCBI) protein database were manually aligned using the calcium binding amino acids to anchor the alignment.

1.3 Roles of the collectins in the immune system

The primary role of all collectins is recognising the carbohydrates on the surfaces of potentially pathogenic cells and particles, including medically important bacteria and viruses, providing an essential contribution to the innate immune response. However, the role of each collectin in the immune response is different, contributing various modes of action to disable the pathogen or stimulate further responses [Figure 1.5].

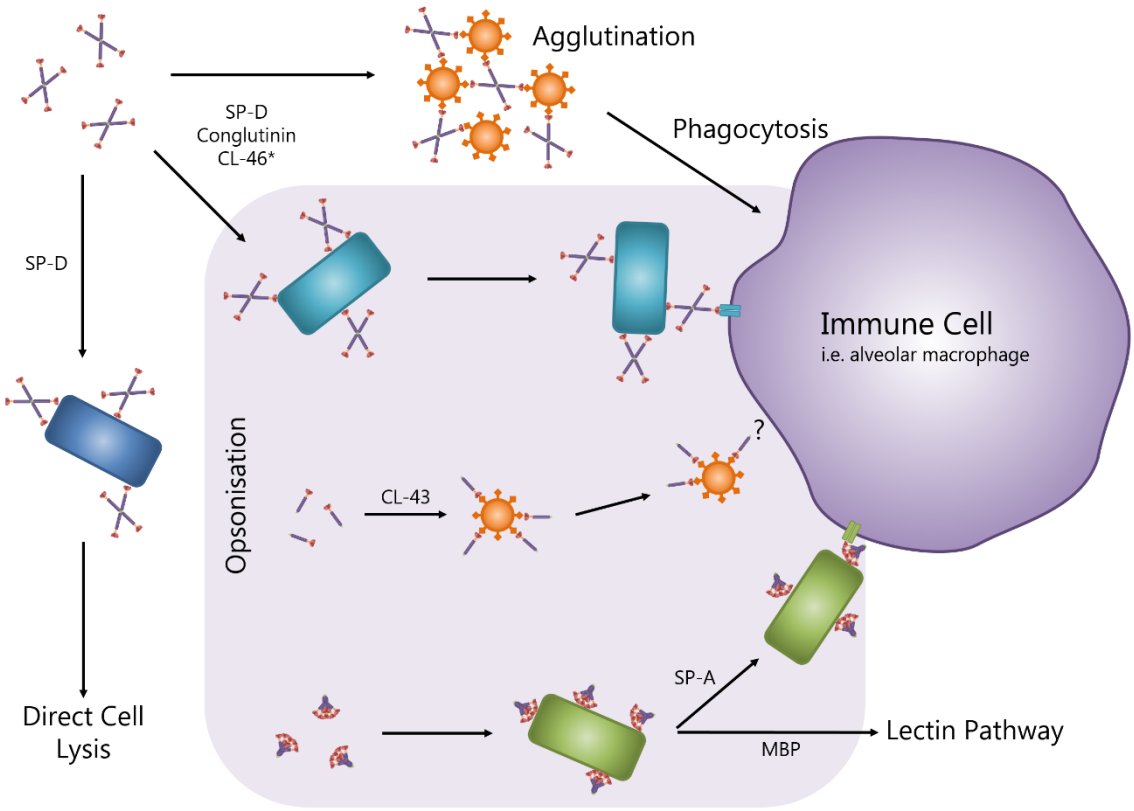


Figure 1.5: Cell Clearance by the Collectins. Summary of pathogen clearance by the collectins through a number of pathways. The four major pathways are opsonisation (highlighted by pale purple); agglutination and deactivation of virulence factors; direct cell lysis; and activation of the lectin pathway in complement, leading to cell clearance. (*) no published evidence that CL-46 forms tetratrimeric cruciforms. (?) unclear how CL-43 upregulates phagocytosis

1.3.1 Opsonisation and phagocytosis

Opsonisation, the process by which a pathogen or immunogenic material is 'tagged' for destruction, is a common function for all of the soluble collectins [Malhotra and Sim, 1995; Pikaar *et al.*, 1995]. MBP, SP-A and SP-D have all been shown to opsonise strains of bacteria, virus particles and fungi, with SP-D being identified as an opsonin (molecular 'tag') for further species including the protozoa *Schistosoma mansoni*, allergens and carbon nanotubes [Van De Wetering *et al.*, 2004; Pondman *et al.*, 2017]. The bovine collectins also opsonise a range of pathogens associated with disease in cattle, including *M. bovis* and Nebraska Calf Diarrhoea Virus [Reading, Holmskov and Anders, 1998]. The function of CL-43, in particular, is primarily as an opsonin as it does not multimerise but has been shown to effectively upregulate phagocytosis of influenza A virus (IAV) [? in Figure 1.5].

The surfactant collectins opsonise gram-positive and -negative bacteria, such as *E. coli* and *S. aureus*, viral particles from human immunodeficiency virus (HIV) and IAV and fungal pathogens including *Aspergillus fumigatus* [McNeely and Coonrod, 1993; Hartshorn *et al.*, 1996, 1998; Madan *et al.*, 1997; Madsen *et al.*, 2013]. A number of receptors can then be activated through interactions between SP-A and SP-D and the extracellular domains, eliciting cell activation and phagocytosis [reviewed by Jakel *et al.*, 2013]. The interaction between the surfactant proteins and the calreticulin/CD91 complex stimulates phagocytosis and release of tumour necrosis factor α (TNF-α) from macrophages and alveolar type II cells following activation of nuclear factor-κB (NF-κB) and p38 signalling pathways [Gardai *et al.*, 2003]. Recognition of hSP-D and hSP-A by calreticulin/CD91 is reported to occur through the collagen-like domains [Gardai *et al.*, 2003], enabling them to bind a surface carbohydrate by the CRD whilst being bound by the receptor, potentially anchoring the pathogen to the surface of the macrophage during phagocytosis.

A similar binding mechanism has also been observed more recently during the recognition of hSP-D by the $CCR2^+$ monocyte receptor, osteoclast-associated receptor (OSCAR) [Barrow *et al.*, 2015; Zhou *et al.*, 2016]. As with the calreticulin/CD91 interaction in macrophages, binding of SP-D to OSCAR elicited a release of TNF-α from the monocytes leading to propagation of the inflammatory response [Barrow *et al.*, 2015]. Interestingly, the recognition of the SP-D collagen domain was further characterised, identifying a nine-amino acid motif (GPOGPAGFO) as the main recognition domain which was ligated by two separate binding sites in the F_c region of OSCAR [Zhou *et al.*, 2016].

Other receptors involved in phagocytosis have also been implicated in SP-A and SP-D stimulated cellular uptake, including toll-like receptors (TLRs), CD14 and, tentatively, the soluble receptor gp340. TLR2 and TLR4 both recognise SP-A and SP-D following bacterial recognition, regulating phagocytosis following activation of either receptor [Murakami *et al.*, 2002; Sato *et al.*, 2003; Ohya *et al.*, 2006]. In SP-D, the interactions with TLR2 and TLR4 appear to be via the CRD but by a secondary binding site as the rough lipopolysaccharide from *E. coli* and inositol phosphate were still able to bind SP-D in a calcium dependent manner. Interactions with the neck domain were ruled out using neck-domain specific monoclonal antibodies which were able to bind both native SP-D and rfhSP-D following incubation with the extracellular domains of TLR2 and TL4 [Ohya *et al.*, 2006]. It has been reported that the interaction with soluble CD14 and gp340 are both CRD mediated, however, in the case of gp340 where the interaction is calcium-dependent, maltose was still able to bind in the main binding pocket, suggesting gp340 recognition occurs through an alternative site and is a protein-protein interaction [Holmskov *et al.*, 1997]. Activation of both soluble receptors leads to activation of alveolar macrophages and amplifies opsonisation and phagocytosis. Further CRD-dependent interactions between SP-A, SP-D,

macrophages and dendritic cells have been identified, some of which were calcium-dependent, however the receptors still require characterisation [Jakel and Sim, 2012].

1.3.2 Agglutination and deactivation

In addition to opsonising pathogens, the soluble collectins also agglutinate bacteria, fungi and viruses, increasing the chemokine signal and inactivating the pathogen to limit virulence. Agglutination occurs through crosslinking of distant binding sites on different cells by the oligomeric collectins, causing the cells to aggregate into larger, collectin-mediated clumps that are easier to phagocytose [Kishore *et al.*, 2006]. SP-A, SP-D, MBL and conglutinin have been observed agglutinating *E. coli*, *Aspergillus fumigatus* and Influenza A Virus, among others, where SP-D and conglutinin agglutinated the susceptible cells more effectively than SP-A and MBL [Madan *et al.*, 1997; Hartshorn *et al.*, 1998; LeVine *et al.*, 2001]. It is thought that this is primarily due to the oligomeric forms of SP-D and conglutinin which allow crosslinking of targets more than 100nm from each other, stabilised by the N-terminal cysteine-rich core formed in the centre of the dodecameric cruciform. This may suggest that the same pattern of agglutination is also possible for CL-46 which may also form the same dodecameric quarternary structure as SP-D and conglutinin.

How agglutination inactivates the pathogens at the molecular level remains unclear, but two main models have been suggested: (1) agglutination sequesters the virulence factors or; (2) inactivation is a result of the virulence factors being directly blocked by the collectins. The latter, model 2, appears to be the case in virus particles as in both IAV and HIV, the virulent haemagglutinin and gp120, respectively, are the primary targets for SP-D and SP-A and effectively block viral entry by interrupting the binding mechanism between the virus particles and the target cells [Hartshorn *et al.*, 1994; Meschi *et al.*, 2005]. However, the virulence of bacteria and fungi is less well defined and likely involves a more complex

interaction with the host. Sequestration of virulence factors is an obvious side effect of agglutination as the cells embedded in the aggregated cell mass are unable to interact with the host, leaving only the external layer and free pathogens to invade and affect the host. The likelihood is that inactivation of the agglutinated pathogens is a result of both models, whereby the virulence factors are bound by the collectin and the agglutination sequesters any remaining available factors.

1.3.3 Complement activation and regulation

Activation of the complement system by the collectins is the preserve of MBP and, putatively, CL-K1 and CL-LK heterocomplex, due to their ability to associate with serine proteases (MASPs) and form an activating complex [Schwaeble *et al.*, 2002; Hansen *et al.*, 2010; Gingras *et al.*, 2011; Henriksen, Brandt, Andrieu, *et al.*, 2013]. Activation of the complement cascade culminates in the attachment of the complement membrane attack complex to the cell causing an increase in permeability resulting in the eventual death of the cell [Esser, 1994]. Mannose binding protein was the first collectin to be identified as a complement activation molecule, through what then became known as the lectin pathway because of the lectin activity of MBP. Activation proceeds through assembly with MASPs (MBP associated serine proteases) which bind in two sites in the collagen region and share overall homology with the C1 complex which activates the classical pathway [Kishore *et al.*, 2004; Gingras *et al.*, 2011]. More recently, MASPs have also been identified associated with trimers of trimeric CL-K1 and the putative heterotrimer CL-LK, and have been shown to activate the lectin pathway in synergy with MBP [Henriksen, Brandt, Andrieu, *et al.*, 2013; Ma, Skjoedt and Garred, 2013].

Other collectins have also been implicated in the complement system but in a regulatory role. SP-A recognises and binds the C1 complex in a calcium-dependent manner through interactions with C1q that prevent the activation of the C1r and C1s serine protease

components [Watford *et al.*, 2001]. The function of inhibiting the complement pathway remains unclear, however it may play a role in minimising the immune response to low titres of pathogens that the lungs are regularly exposed to as a result of inhalation, preventing complete activation of the immune system and avoiding unnecessary damage to the linings of important tissues. This may suggest that SP-A and SP-D behave as a threshold mechanism, beyond which the infection is too great for surfactant mediated immunity to combat and other immune system mechanisms are required to fight the infection.

1.3.4 Direct cell killing

Surfactant protein A and surfactant protein D have both been shown to clear bacterial and fungal cells by direct cell killing, however the mechanism for this is poorly understood. A study by *Wu* and co-workers suggested that adherence of the surfactant collectins to the cell wall of rough strains of *E. coli* and strains of *Klebsiella*, *Enterobacter* and *E. coli* isolated from patients, increased the permeability of the cell wall thus limiting cell growth and eventually causing cell death [Wu *et al.*, 2003]. The study identified intracellular components in the external environment of the strains following localisation of SP-A and SP-D on their surface, suggesting that pores were induced in the bacterial surface.

In a similar study, *Histoplasma capsulatum*, an intracellular fungal pathogen, was shown to be directly killed by the pulmonary collectins by increasing fungal cell wall permeability [McCormack *et al.*, 2003]. Again, the results suggested that binding of SP-A or SP-D to cell wall created pores that not only caused the release of proteins from the cells, but also allowed for otherwise impermeant substrates to enter the cell, suggesting that the pore did not have any selectivity. This may suggest that an influx of water into the cells may, in part, lyse the cells as a means of cell clearance, leaving the collectins and other innate receptors to clear the debris. Interestingly, however, this study also showed that once *H. capsulatum*

is internalised by a macrophage, neither SP-A or SP-D is able to inhibit the growth of the fungus, suggesting that the SP-A and SP-D do not have an intracellular immune role [McCormack *et al.*, 2003].

In both cases, the mechanism for pore formation requires further investigation as there is little understanding as to how SP-A and SP-D are able to directly kill pathogens. However, these two studies show that direct killing may be a key role for the collectins when immunocytes are not localised to the infection in the early stages, providing some insight into the role of the collectins in defending against intracellular pathogens before they enter the cell.

1.3.5 Other roles in the immune response

The roles of the collectins beyond cell clearance by the four main pathways (sections 1.3.1 – 1.3.4) are extensive and wide ranging. The best understood of these is the regulation of the inflammatory response to infection by the surfactant proteins. SP-A and SP-D both have roles in regulating the expression and release of pro-inflammatory cytokines and downregulating the release of TNF-α [Murakami *et al.*, 2002; Gardai *et al.*, 2003; Kawai and Akira, 2010; Fournier *et al.*, 2012]. SP-D is able to recognise and activate the signal inhibitory regulatory protein α (SIRP-α) and SIRP-β through calcium-dependent interactions between the CRD and the membrane proximal domain, causing a downstream blockade of the p38 cascade when SIRP-α is activated [Fournier *et al.*, 2012]. Interestingly, the p38 cascade is activated by SP-D and SP-A when the collagen-like domain interacts with the calreticulin/CD91 complex to stimulate TNF-α expression, suggesting that recognition of a pathogen acts as a switch mechanism by competing with the CRD for SIRP-α recognition [Gardai *et al.*, 2003; Fournier *et al.*, 2012]. This interaction is further complicated by an interaction between the SIRP-α-SP-D complex and CD47 which has been shown to further regulate the downstream cascades [Takizawa and Manz, 2007; Fournier *et al.*, 2012].

1.4 Human surfactant protein D

Human surfactant protein D, along with MBP and SP-A, is one of the best understood collectins in terms of both structure and function. This section intends to provide a brief summary of some of that research, focussing on the structure and interactions that SP-D is capable of making with both microorganisms and cell surface receptors.

1.4.1 Native structure of a rfhSP-D

The unbound, native structure of rfhSP-D, a recombinant fragment of hSP-D containing amino acid residues 179 through to the C-terminal Phe335, includes a small portion of collagen domain, the α-helical coiled coil neck domain and the CRD [Håkansson *et al.*, 1999; Shrive *et al.*, 2003]. Overall, the structure is similar to MBP and rat SP-A forming a trimeric fragment containing three CRDs arranged around a central trimeric axis formed by the α-coiled coil neck domain [Sheriff, Chang and Ezekowitz, 1994; Weis and Drickamer, 1994; Håkansson *et al.*, 1999; Head *et al.*, 2003; Shrive *et al.*, 2003]. However, in contrast to rat SP-A which forms a "T"-shaped fragment, the CRDs in rfhSP-D form a "Y"-shaped fragment more similar to MBP. This is, in part, due to the CRDs in rfhSP-D making comparatively few interactions either between subunits or with the neck domain, allowing the CRDs to be more flexible around a short semi-rigid hinge-like region that centres around Pro235 [Weis and Drickamer, 1994; Head *et al.*, 2003; Shrive *et al.*, 2003]. The interactions present include a small set of hydrophobic interactions between the neck region side chains of Val231 and Phe234 and the CRD of the adjacent subunits (A to C, B to A and C to B) [Håkansson *et al.*, 1999]. These hydrophobic interactions are supplemented by interactions between the neck domain and the CRD of the other adjacent subunit (A to B, B to C and C to A) comprising hydrogen bonds (Phe234-Lys246; Gln227-Glu242; Gln227-Phe355) and intra-subunit long salt bridges between Glu232 and Lys246.

The CRD is characteristic of other C-type lectins containing a twisted β-sheet flanked by two α-helices that form one of two large insertions into the β-sheet; the second insertion creates the primary calcium binding pocket, where Ca1 is bound, between the α-helical turn (α3) and β-strand (β5). The primary calcium binding pocket (Ca1) of the CRD is in a distal position to the alpha-helical neck domain in all three CRDs of the trimeric structure and the three sites form a plane across the molecule which, *in vivo*, would facilitate recognition of carbohydrate arrays [Weis and Drickamer, 1994; Shrive *et al.*, 2003]. There are a further two calcium ions in each CRD, spatially separated from Ca1 by the helical turn [Figure 1.6].

It has been suggested that the second calcium ion, Ca2, is involved in stabilising the helical turn though direct interactions with the carbonyl of Glu329, reducing the flexibility in the side chain and maintaining the contact between Glu329 and Ca1 [Shrive *et al.*, 2003]. The third calcium is closer to the surface of the CRD and makes fewer interactions with the protein, potentially providing a site for cell surface receptors to recognise the CRD or, alternatively, acting as a secondary binding site for longer polysaccharide ligands that occur naturally. The proximity of the Ca2-Ca3 binding sites to the main binding pocket may also allow for cell receptors to recognise both the bound ligand and the CRD.

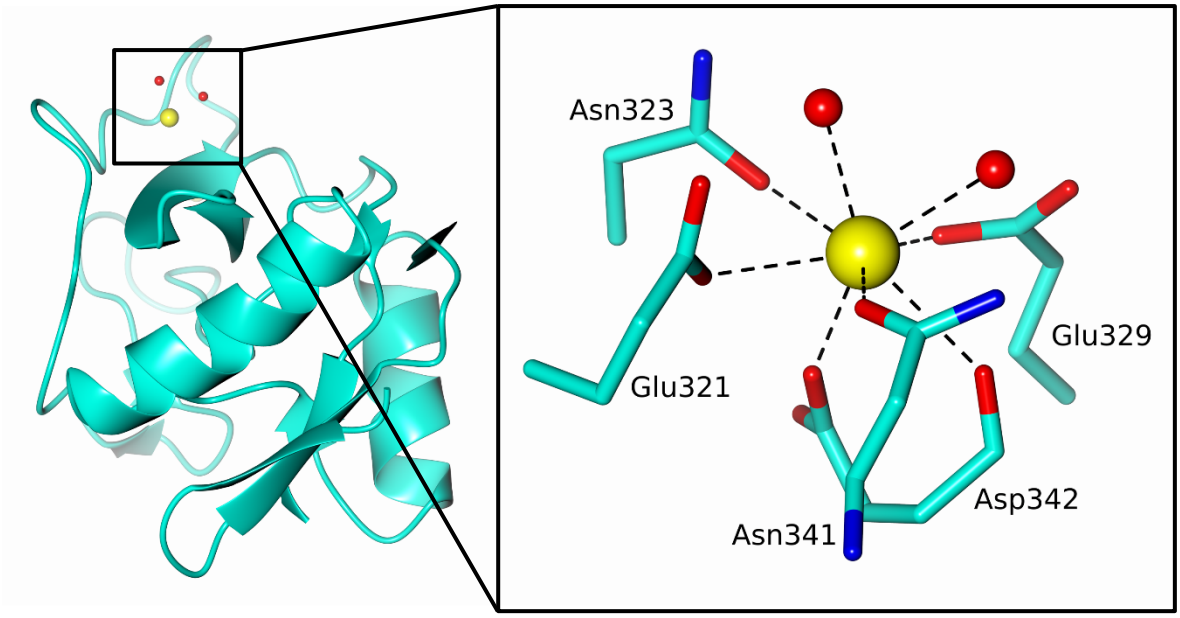


Figure 1.6: Native carbohydrate recognition domain of rfhSP-D with the calcium coordination (right). Calcium coordination is completed by Glu321 OE1, Asn323 OD1, Glu329 OE1, Asn341 OD1 and Asp342 OD1 and main-chain O. Figures generated from PDB accession code 1PW9 [Shrive et al., 2003].

Coordination of the calcium ion in the main binding pocket, Ca1, is completed through six coordinate covalent bonds between the calcium ion (Ca^{2+}) from five acidic side chains and a main chain carbonyl group from Asp342. Glu321, Asn323 and Glu329 from the α3 region and Asn341 from the β5-strand form four of the main interactions through carboxyl and carbonyl groups which donate electron pairs to the calcium [Håkansson *et al.*, 1999; Shrive *et al.*, 2003]. The final two coordinate bonds with Ca1 are completed by Asp342 which, along with the carbonyl group, coordinates through the carboxyl group, forming the base of the calcium binding pocket. The protein thus coordinates six of the eight calcium interactions and the final two are fulfilled by water molecules in the native structure.

1.4.2 Ca4 and the asymmetric tyrosine

The high resolution structure of rfhSP-D revealed a fourth calcium ion (Ca4) at the pseudosymmetric three-fold axis of the trimer, at the base of an elongated funnel created by the three CRDs, coordinated by the side chains of Glu323 from all three subunits [Figure

1.7a; Shrive *et al.*, 2003]. However, in the maltose-bound structure, the Ca4 binding site is depleted causing the side chain of Glu323 in subunit B to rearrange and interact with the asymmetric tyrosine in subunit C. The asymmetric tyrosine 228 in subunit C is present in all the published structures and fundamentally changes the interface between the neck and CRD domains of adjacent subunits, allowing LysA229 to move into the space vacated by TyrC228 [Shrive *et al.*, 2003].

The role of Ca4 is unclear as the binding site is unoccupied in the maltose-bound structure suggesting that the calcium described in the native structure may be a remnant of crystallisation or demonstrate a calcium concentration dependent occupancy of the fourth binding site [Figure 1.7b]. However, a subsequent study using a range of different ligand bound complexes revealed that Ca4 is present in the structure at physiological calcium concentrations [Shrive *et al.*, 2009]. The conformational changes associated with the occupancy of the Ca4 binding site [Figure 1.7b] may suggest that Ca4 occupancy could act as a switch mechanism in response to environmental calcium, potentially allowing for receptor binding in the downstream immune response.

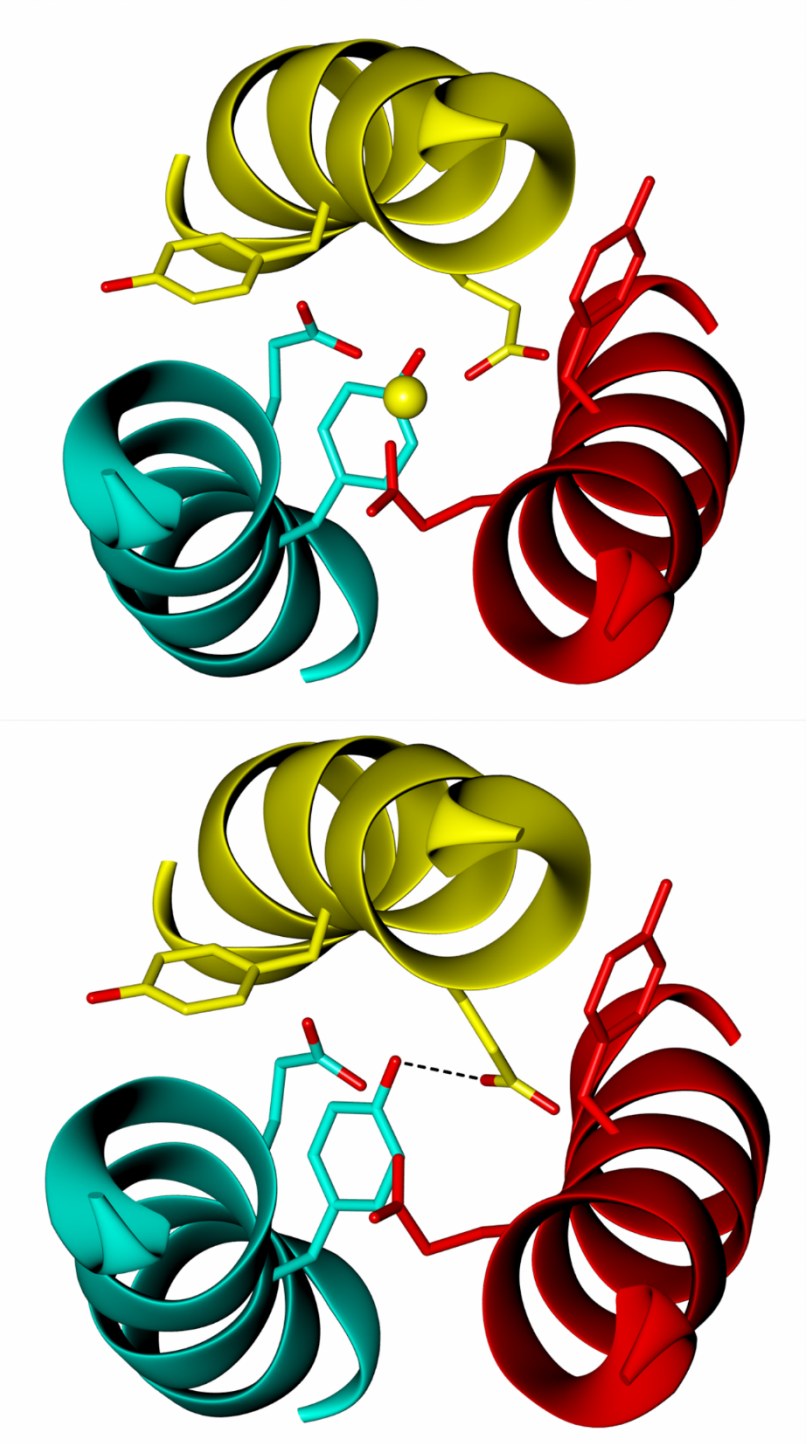


Figure 1.7: Comparison of Ca4 binding site in the native rfhSP-D (above) and maltose-bound complex (below). Calcium (yellow) depletion causes GluB232 to move away from the calcium site and interact with asymmetric TyrC228.

1.4.3 hSP-D recognition of pathogenic species

The carbohydrate recognition domain of SP-D enables it to recognise a wide range of organisms from almost all phylogenetic groups [table 1.3]. Recognition occurs through the glycosylation structures on the surface of the pathogen, including lipopolysaccharides on gram-negative bacteria, lipoteichoic acids on gram-positive bacteria and glycosaminoglycan, lipoarabinomannans and glycoproteins from other groups of pathogens [Kuan, Rust and Crouch, 1992; van de Wetering *et al.*, 2001; Crouch *et al.*, 2009; Ferguson *et al.*, 2010]. It has also been shown that peptidoglycan, a core component of the cell wall of all bacteria, is recognised by hSP-D [van de Wetering *et al.*, 2001].

Rough LPS mutants of gram-negative bacteria have been shown to be bound with higher affinity than smooth LPS mutants, such as *S. enterica* minnesota R7 and *E. coli* J5 [Kuan, Rust and Crouch, 1992]. Smooth mutants differ from rough mutants by the presence of the repeating O-antigen at the terminal of the core polysaccharide of the LPS [Holst and Brade, 1991, 1992; Haishima, Holst and Brade, 1992]. The difference in affinity was particularly evident in three strains of *S. enterica* minnesota LPS where the smooth mutant was not bound by radiolabelled SP-D but two rough mutants, R7 and R5, were both strongly bound by the same radiolabelled SP-D [Kuan, Rust and Crouch, 1992].

A recent study has shown that a rfhSP-D has a protective role in mice challenged with the main component of the fungal cell wall, a β(1→3)-linked polymer of glucose, β-glucan, attenuating the allergic response and suggesting that β-glucan is recognised by rfhSP-D [Fakih *et al.*, 2015]. This provides some indication of how hSP-D binds to fungal pathogens such as *A. fumiugatus* and *C. albincans*. Similarly, the lipoarabinomannan component of the mycobacterial cell wall of *M. tuberculosis* was bound by hSP-D from recombinant and BAL sources via the terminal mannose in a calcium-dependent, lectin domain interaction [Ferguson *et al.*, 2010]. *Mycoplasma pneumoniae* was also shown to be recognised in this

way [Chiba *et al.*, 2002]. Peptidoglycan and lipoteichoic acids from gram-positive bacteria have been successfully identified as ligands for hSP-D, providing a mechanism for recognition of gram-positive bacteria [van de Wetering *et al.*, 2001].

There have been extensive studies of hSP-D recognition and interaction with the medically important viruses, influenza A virus (IAV), responsible for annual flu outbreaks, and human immunodeficiency virus (HIV), responsible for causing systemic breakdown of the immune system leading to acquired immune deficiency syndrome [LeVine *et al.*, 2001; Meschi *et al.*, 2005; Crouch *et al.*, 2009; Madsen *et al.*, 2013]. IAV is recognised by the highly mannosylated viral glycoprotein heamagglutinin (HA) through calcium dependent CRD interactions with the mannosylation on the surface of HA [Crouch *et al.*, 2009; Goh *et al.*, 2013]. gp120 of HIV, a 120kDa glycoprotein in the lipid bilayer, is also highly mannosylated in a comparable way to HA and is also recognised by the CRD of hSP-D through a calcium-dependent recognition of the terminal mannose, preventing cellular entry [Meschi *et al.*, 2005; Madsen *et al.*, 2013]. SP-D recognition was conclusively shown to be dependent on the mannosylation of gp120 by deglycosylation of the protein and exposure to recombinant SP-D. The deglycosylation resulted in complete abrogation of SP-D binding and agglutination of HIV, confirming that SP-D does not complete any protein-protein interactions [Meschi *et al.*, 2005].

Table 1.3: Comparison of SP-A and SP-D Recognition of Pathogenic Species. Table shows which pathogenic species surfactant protein A (SP-A) and surfactant protein D (SP-D) recognise, using + to indicate evidence of surfactant protein binding. (rf) recombinant fragment SP-D binds pathogen when (+).

Pathogen	SP-A	SP-D (rf)	References
Gram-Negative Bacteria			
Chlamydia pneumoniae	+	+	[Oberley *et al.*, 2004b]
Chlamydia trachomatis	+	+	[Oberley *et al.*, 2004a]
Escherichia coli	+	+ (+)	[Kuan *et al.*, 1992; Van Iwaarden *et al.*, 1994]
Haemophilus influenzae	+	+ (+)	[LeVine *et al.*, 2000]
Helicobacter pylori		+	[Murray *et al.*, 2002]
Klebsiella pneumoniae	+	+	[Ofek *et al.*, 2001; Sahly *et al.*, 2002]
Pseudomonas aeruginosa	+	+	[Restrepo *et al.*, 1999]
Salmonella enterica		+ (+)	[Kuan *et al.*, 1992]
Gram-Positive Bacteria			
Staphlococcus aureus	+	+	[McNeely *et al.*, 1993; Zhang *et al.*, 2015]
Streptococcus pneumoniae	+	+ (+)	[McNeely *et al.*, 1993; Jounblat *et al.*, 2004]
Mycobacteria			
Mycobacterium avium	+	+	[Kudo *et al.*, 2004]
Myocbacterium tuberculosis	+	+	[Ferguson *et al.*, 2010]
Mycoplasma			
Mycoplasma pneumoniae	+	+ (+)	[Chiba *et al.*, 2002]
Viruses			
Herpes Simplex Virus	+		[Van Iwaarden *et al.*, 1992]
Human Immunodeficiency Virus	+	+	[Meschi *et al.*, 2005; Gaiha *et al.*, 2008]
Influenza A Virus	+	+	[Benne *et al.*, 1995; LeVine *et al.*, 2001]
Respiratory Syncytial Virus	+	+	[Hickling *et al.*, 1999; LeVine *et al.*, 2004]
Fungi			
Alternaria alternata (tenuis)		+	[Ooi *et al.*, 2007]
Aspergillus fumigatus	+	+	[Madan *et al.*, 1997]
Candida albicans	+	+	[Rosseau *et al.*, 1997; van Rozendaal *et al.*, 2000]
Histoplasma capsulatum	+	+	[McCormack *et al.*, 2003]
Pneumocystis jiroveci (carinii)	+	+	[O'Riordan *et al.*, 1995]
Saccharomyces cerevisiae		+	[Allen *et al.*, 2001]
Protozoa			
Nippostrongylus brasiliensis		+	[Thawer *et al.*, 2016]
Schistosoma mansoni		+	[Van De Wetering *et al.*, 2004]

1.5 Small ligands and surfactant protein D

There have been several studies using rfhSP-D, a recombinant fragment of hSP-D containing the neck and carbohydrate domains in complex with mono-, di- and trisaccharides. This has revealed some key aspects about the binding mechanism that SP-D uses to recognise its ligands and provided insights into what elements of the CRD are responsible for selecting and recognising those ligands.

1.5.1 Importance of small ligand studies

Understanding the molecular mechanisms which proteins employ to recognise their targets is an essential part of understanding the function of the protein. To do this often requires using crystallography to determine, at high resolution, the atomic structure of the protein and identify in the finest detail how a ligand interacts with the surface or binding site of the protein. Small ligands that are representative of putative binding epitopes of large physiological ligands not only provide a convenient route to crystallographically characterise ligand binding, but where the environment around the binding sites in the crystal is spatially restricted due to the proximity of other molecules, they may provide the only opportunity for understanding some of the fundamental recognition mechanisms that provide insight into how the protein may recognise its natural ligands *in vivo*.

Many physiological hSP-D ligands are long, complex oligosaccharides on the surface of pathogens that are often membrane bound or form part of insoluble structures on the surface of cells. Smaller oligosaccharide subunits that are often more soluble than the larger polysaccharides that they assemble into provide a way of accessing the binding sites in the crystal systems, enabling the binding mechanisms to be elucidated. The native recombinant hSP-D fragment rfhSP-D first characterised by *Håkansson* and colleagues has provided invaluable insight into how SP-D is able to recognise a myriad of ligands using this technique, allowing for the primary binding mechanism to be characterised.

1.5.2 Mono and disaccharides: establishing the recognition mechanism

The mechanism of carbohydrate recognition by human surfactant protein D was first elucidated by *Shrive* and colleagues (2003) using both native rfhSP-D and a maltose-bound complex of the biologically and therapeutically active recombinant fragment [Shrive *et al.*, 2003]. Recognition of maltose, a small disaccharide of α(1→4)-linked glucose, occurred by the non-reducing terminal glucose primarily through interactions with Ca1, allowing the second glucose to project out of the binding pocket. In subunits B and C, the second glucose is not defined in the structure suggesting that there is free rotation around the α(1→4) glycosidic bond. However, in subunit A the crystal system restricts the rotation, allowing for the second glucose to be defined [Figure 1.8].

The calcium in the main binding pocket primarily coordinates the two equatorial hydroxyl groups on the third (C3, O3′) and fourth (C4, O4′) carbons of the non-reducing terminal glucose which displace the two water molecules from their coordinating positions in the native structure [Shrive *et al.*, 2003]. The coordinate bonds between the two hydroxyls and Ca1 form the basis of the recognition mechanism, supplemented by protein-ligand hydrogen bonds from the side chains of Glu321, Asn323, Glu329 and Asn341. Glu321 and Asn323 hydrogen bond to O3′ through the carboxylic acid and amine, respectively. Similarly, O4′ interacts with Glu329 and Asn341. The second glucose in the disaccharide does not interact directly with the CRD in the crystal structure which is at first sight surprising as maltose has a higher affinity for SP-D than glucose. Interestingly, the two flanking residues, Asp325 and Arg343, do appear to influence ligand binding, forming long range interactions with the calcium coordinating glucose and, in the case of Arg343, undergo conformational changes in response to maltose binding [Shrive *et al.*, 2003]. This is particularly evident in subunit A where Arg343 has two conformations, presumably due to the proximity of the second glucose residue, caused by a 10° rotation in ligand position.

A subsequent study also revealed that two alternative isomers of mannobiose are recognised by surfactant protein D in a comparable mechanism to maltose [Shrive *et al.*, 2009]. Both the α(1→2)-linked and α(1→4)-linked mannobiose disaccharides were recognised by the non-reducing terminal mannose residue, coordinating the calcium ion in the main binding pocket by the two equatorial hydroxyls synonymous with the maltose complex (O3' and O4'). In both structures, the O3' and O4' coordinate Ca1 and complete the protein interactions with Glu321, Asn323, Glu329 and Asn341 however the orientation of the mannose residue visible in the binding site differs between the two isoforms.

In the α(1→2)-linked disaccharide, the ring is orientated in the same plane as maltose in all three subunits allowing O3' to hydrogen bond with Glu321 and Asn323 and O4' to interact with Glu329 and Asn341, placing the undefined second mannose residue in space distal to the flanking residues Asp325 and Arg343 in subunits A and B [Shrive *et al.*, 2003]. In subunit C, it is less clear which of the mannose residues are being coordinated by the calcium and the density for the second mannose residue is very poor. The α(1→4)-linked mannobiose, on the other hand, binds in two distinct mechanisms as a result of a 180° rotation about the axis of the binding mannose residue. In subunit A, it is reported that the orientation is comparable to the maltose-bound and α(1→2)-mannobiose-bound complexes however, in subunit C where there is less restriction of the binding site due to a weaker interaction with the symmetry related molecule, the coordinating mannose residue is rotated such that O3' interacts with Glu329 and Asn341 and O4' interacts with Glu321 and Asn323 [Figure 1.8; Shrive *et al.*, 2003, 2009]. Interestingly, the density for the α(1→4) mannobiose in subunit B shows evidence for both binding mechanisms with higher occupancy in the rotated conformation, most similar to subunit C [Shrive *et al.*, 2009].

The rotations in the binding mannose also revealed a key role for Arg343 and Asp325 in the recognition of ligands by surfactant protein D. In the subunits where the ring was in an

equivalent position to maltose, the O6' formed contacts with NH1 or NH2 of Arg343 with varying strength due to rearrangements of the side chain allowing Arg343 to move closer to the binding site, in part aided by the stabilising interactions between Arg343 and Glu333 [Shrive *et al.*, 2009]. The contact with Arg343 is also important in the recognition mechanism for the subunits where the mannobiose is rotated, interacting weakly with the hydroxyl on carbon 2 (O2'). In this position, the O1' hydroxyl also interacts with Asp325 on the opposing flank [Figure 1.8] suggesting that Asp325 and Arg343 play important roles in ligand recognition and orientation in the calcium binding site.

The influence of the undefined mannose residue on the orientation of the ligand is unclear as it is not defined in either mannobiose isoform. However, a study using a recombinant fragment of porcine SP-D in complex with mannose revealed a mechanism more similar to the α(1→2)-linked mannobiose suggesting that the α(1→4)-linkage is affecting the recognition mechanism, at least in the crystal structure [Van Eijk *et al.*, 2012]. How this affects longer mannosyl ligands that SP-D is known to recognise, such as mannan and the N-linked mannosylation of viral proteins, is unclear although it may provide some insight into why the affinities for mannosyl ligands differ.

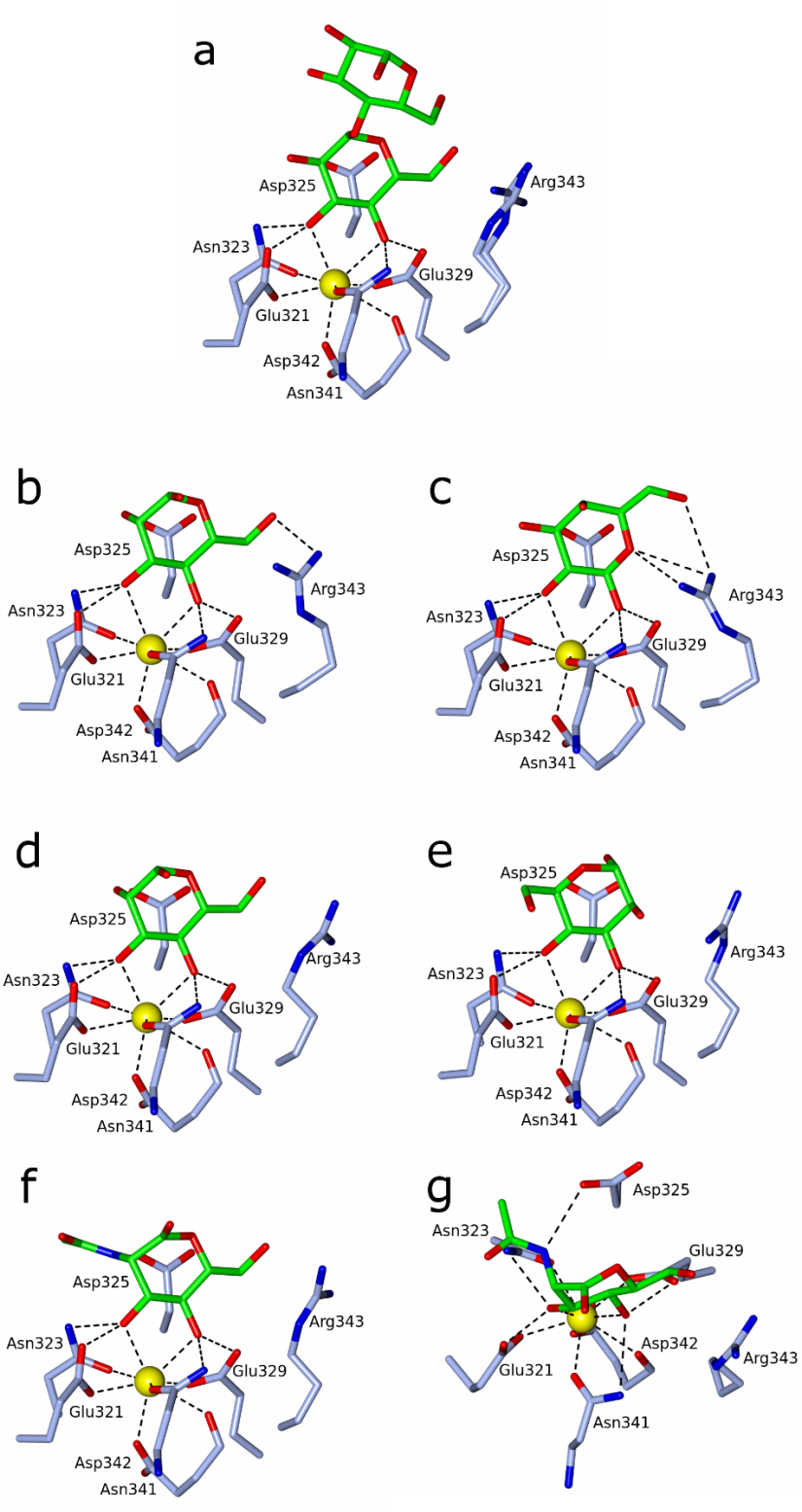


Figure 1.8: Calcium coordination in the mono- and disaccharide bound complexes. (a) maltose complex; (b) α(1→2)-linked mannobiose complex; (c) galactose complex; (d) α(1→4)-linked mannobiose complex (subunit A); (e) α(1→4)-linked mannobiose complex (subunit B); (f) N-acetylmannosamine (ManNAc) complex; (g) 90° rotation of ManNAc complex to illustrate Asp325 interaction. Figure generate from PDB accession codes 1PWB, 3IKQ, 3IKN, 3IKR and 2ORJ [Shrive et al., 2003, 2009; Crouch et al., 2007]

N-acetyl-mannosamine (ManNAc) has also been complexed with rfhSP-D and successfully crystallised, showing that recognition of the ManNAc occurs in a similar way to α(1→2)-linked mannobiose, the A subunit of α(1→4)-linked mannobiose and maltose with O3′ and O4′ being the primary points of recognition [Shrive *et al.*, 2003, 2009; Crouch *et al.*, 2007]. However, in the ManNAc-bound complex, the orientation of the ligand is thought to be limited by the position of Arg343, forcing the N-linked acetyl group to extend away from the two flanking residues and above Glu321 and Asn323 [Figure 1.8; Crouch *et al.*, 2007]. As with the other ligand-bound structures discussed so far, coordination of the calcium also allows for Glu321, Asn323, Glu329 and Asn341 to interact with the calcium coordinating O3′ and O4′ hydroxyls, further characterising the overall binding mechanism. As seen in the mannobiose structures, ManNAc O6′ interacts with Arg343 to complete the protein-ligand interactions in the ManNAc complex and may also limit the orientation of the ManNAc in the binding site, elaborating on the important role of Arg343 in ligand selection and recognition [Allen *et al.*, 2004].

Interestingly, the low affinity ligand galactose, relative to maltose and ManNAc, has been shown to have a novel binding mechanism not seen in the other oligosaccharide ligands studied. In the galactose/rfhSP-D complex, the two water molecules that complete the calcium coordination are displaced by the O1′ and O2′ hydroxyls of the galactosyl ring instead of the O3′ and O4′ of the glucosyl and mannosyl type rings [Shrive *et al.*, 2009]. Importantly the two coordinating hydroxyls are in the same equatorial positions in galactose as the O3′ and O4′, sharing the same geometry in relation to the main ring, allowing the hydroxyls to coordinate Ca1 in approximately the same positions. O1′ forms additional protein-ligand interactions with Glu329 and Asn341, in a similar way to O4′ in the other structures, whilst O2′ interacts with Glu321 and Asn323. This orientation presents the ring-bound oxygen (O5′) to Arg343, allowing it to directly interact with the oxygen

atom, forming an Arg343 mediated bridge between O5′ and the O6′ on the galactose [Figure 1.8].

The physiological relevance of the galactose-bound complex is unclear. While galactose is a common component of glycosylated structures in Gram-negative bacteria and viruses, O1′ and O2′ are regularly unavailable for recognition by surfactant protein D [Holst, 2011]. In many of the complex oligosaccharides, O1′ is involved in glycosidic linkages between monosaccharide residues where galactose is in a terminal or non-terminal position, such as in haemagglutinin subtype 5 where O1′ and O3′ form the basis of glycosidic bonds [Gamblin and Skehel, 2010].

The small oligosaccharide studies reveal a conserved binding mechanism in hSP-D that is dependent on calcium coordination by the two equatorial hydroxyl groups, equivalent to the O3′ and O4′ positions in glucose or mannose. The small ligand studies have also identified Asp325 and Arg343 as essential determinants in ligand binding, directly involved in recognition of some of the monosaccharides and playing a selective role in the small ligands where they do no directly interact with the ligand.

1.5.3 Carbohydrate analogue recognition: elaborating the binding mechanism

Carbohydrate analogues form an important part of cell signalling and the structural component of many lipid moieties. Two of these analogues are *myo*-inositol and inositol phosphate. Structures for *myo*-inositol and inositol-1-phosphate in complex with recombinant fragment human surfactant protein D have been solved at high resolution [Crouch *et al.*, 2007; Shrive *et al.*, 2009]. In both structures, the inositol ring was primarily recognised in a calcium dependent mechanism coordinating two equatorial hydroxyls on adjacent carbon atoms in a similar way to the other carbohydrate ligands. *Myo*-inositol was coordinated by the hydroxyls at positions O1′ and O6′ of the ring which assume the

positions of the O3′ and O4′ hdyroxyls of mannose-type ligands in relation to the calcium ion [Crouch *et al.*, 2007]. Recognition of *myo*-inositol was completed by the familiar amino acid residues: Glu321 and Asn323 interact with O1′ and; Glu329 and Asn341 interact with O6′. The orientation of the *myo*-inositol in the binding site was similar to the other monosaccharides including the mannose in porcine SP-D but the density reported for *myo*-inositol did suggest that the ring could rotate 180° in the binding site in the same way as α(1→4)-mannobiose showed binding heterogeneity [Crouch *et al.*, 2007; Shrive *et al.*, 2009; Van Eijk *et al.*, 2012].

There are two structures for inositol-1-phosphate (IP_1) in complex with rfhSP-D which show agreement in the binding mechanism [Crouch *et al.*, 2007; Shrive *et al.*, 2009]. The calcium ion recognises the O4′ and O5′ hydroxyl groups of the inositol ring where O4′ interacts with Glu321 and Asn323 and O5′ makes contacts with Glu329 and Asn341 by the same mechanism as previously identified with an additional hydrogen bond between O6′ of IP_1 and Arg343 [Figure 1.9]. The phosphate, substituted at position C1, extends in the direction of Asp325 and Arg343, causing small shifts in the conformations of the two side chains, with a more pronounced change in Arg343 in subunit C compared to subunit B. *Shrive* and colleagues also noted a small change in the orientation of IP_1 in the binding site consistent with the differences of side chain positions, further suggesting an important role of these two side chains in ligand orientation [Shrive *et al.*, 2009].

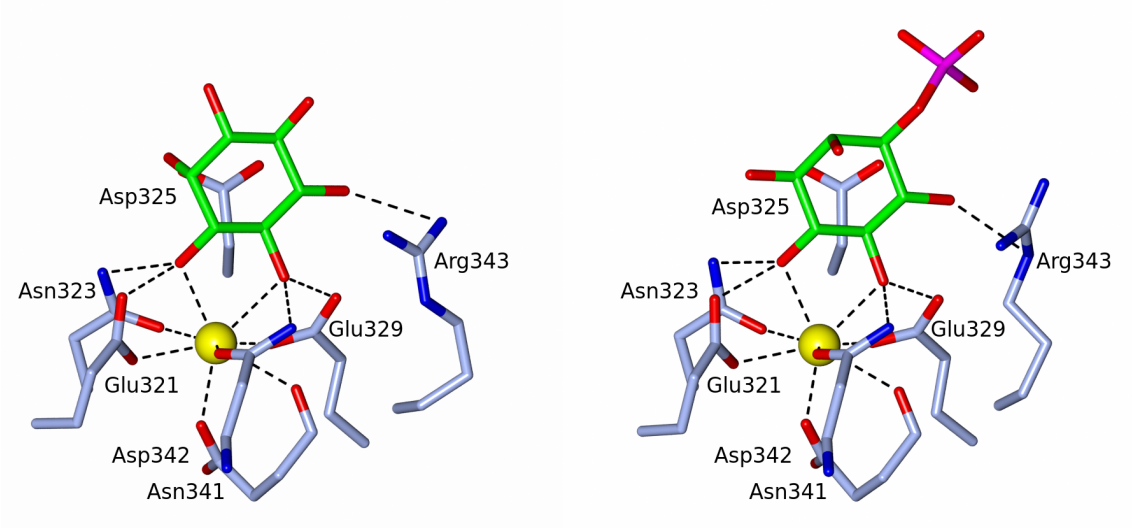


Figure 1.9: Calcium coordination of *myo*-inositol (left) and inositol-1-phosphate (right). Coordination of the calcium ion in the *myo*-inositol complex is completed by O1′ and O6′, and O4′ and O5′ in the inositol-1-phosphate complex. Arg343 is fundamental in recognition of both ligands, forming hydrogen bonds to the main ring in both complexes.

In the crystal structures of the *myo*-inositol complex and IP_1 complex the ligand is absent in subunit A as a result of crystal packing, which is surprising as maltose is comfortably accommodated by subunit A in the same crystal form. Simple fitting of IP_1 in subunit A suggested that this may be because of a clash between the phosphate moiety and the symmetry related molecule as a result of IP_1 favouring the geometry in the other subunits [Shrive *et al.*, 2009]. The favoured geometry may be a result of the proximity of the negatively charged phosphate limiting ligand orientation as a point mutation of Arg343 (R343K) revealed a greatly improved affinity for IP_1 as a result additional flexibility [Crouch *et al.*, 2007].

These two structures demonstrate that the calcium-dependent mechanism identified in mono- and disaccharide complexes with rfhSP-D is capable of recognising a diverse number of ligands with little alteration to the CRD or the immediate binding site. They also elaborate on the importance of Asp325 and Arg343, the two flanking residues, and their role in selecting and orientating ligands in the binding site through their interactions with

the inositol ring through small conformational changes in their side chain positions, even when the only interactions are water mediated [Shrive *et al.*, 2009].

1.5.4 Maltotriose: extending the binding surface

While the mono- and disaccharide-bound complexes reveal important details of the binding mechanism in rfhSP-D, many of the ligands that SP-D and the other collectins recognise are long polysaccharides at the cell surface and contain many more saccharide residues [Haishima, Holst and Brade, 1992; Vollmer, Blanot and De Pedro, 2008; Fakih *et al.*, 2015]. Trisaccharides are the next step towards identifying how longer oligosaccharides may be recognised by an extended binding surface in the CRD of SP-D, as in the scavenger receptor CL-P1 CRD where a lewis^x acid containing a Gal-GlcNAc-Fuc triose was recognised by both terminal residues in a bidentate fashion [Feinberg, Taylor and Weis, 2007]. To explore this, maltotriose (three α(1→4)-linked glucoses) was complexed with rfhSP-D revealing an extending binding surface was also present in hSP-D for this ligand [Crouch, McDonald, *et al.*, 2006].

Primary recognition of the non-reducing terminal glucose was comparable to the maltose-bound complex published by *Shrive* and colleagues [Shrive *et al.*, 2003; Crouch *et al.*, 2007]. Binding of the ligand in the Ca1 pocket was between the calcium ion and the two equatorial hydroxyl groups, O3' and O4' with further hydrogen bonds to the calcium coordinating side chains Glu321, Asn323, Glu329 and Asn341. As with maltose, O3' hydrogen bonds to the carboxylic acid and amine groups of Glu321 and Asn323, respectively, and O4' to Glu329 and Asn341 in the same way. However, due to the increased binding angle, in relation to the protein, in the maltotriose-bound complex, the distances from the binding glucose residue to the flanking residues are larger than the maltose structure as a result of a π-stacking interaction between the reducing terminal glucose and Phe335 [Figure 1.10; Crouch *et al.*, 2006].

The π-stacking interaction was present in all three subunits of the rfhSP-D complex and appears to stabilise the ligand across the binding surface, providing an explanation for the increased affinity for maltotriose over other glucosyl ligands [Crouch, McDonald, *et al.*, 2006]. However, the physiological relevance of this mechanism is unclear as longer malt-N-oses, such as those naturally occurring on pathogens, may prevent the formation of the π-stacking system because of steric clashes with the surrounding main chain of the CRD. In addition, the third glucose in the maltotriose structure appears to be in a β-enantiomeric form which results in an overall "flatter" ring that is essential for the two rings to stack together. This was demonstrated in the same study which revealed a p-nitrophenyl-maltoside, with the reducing terminal glucose of maltotriose substituted by a nitrophenyl ring, in the same orientation as maltotriose and showed a higher affinity for rfhSP-D due to the more complete π-π stacking system [Crouch, McDonald, *et al.*, 2006].

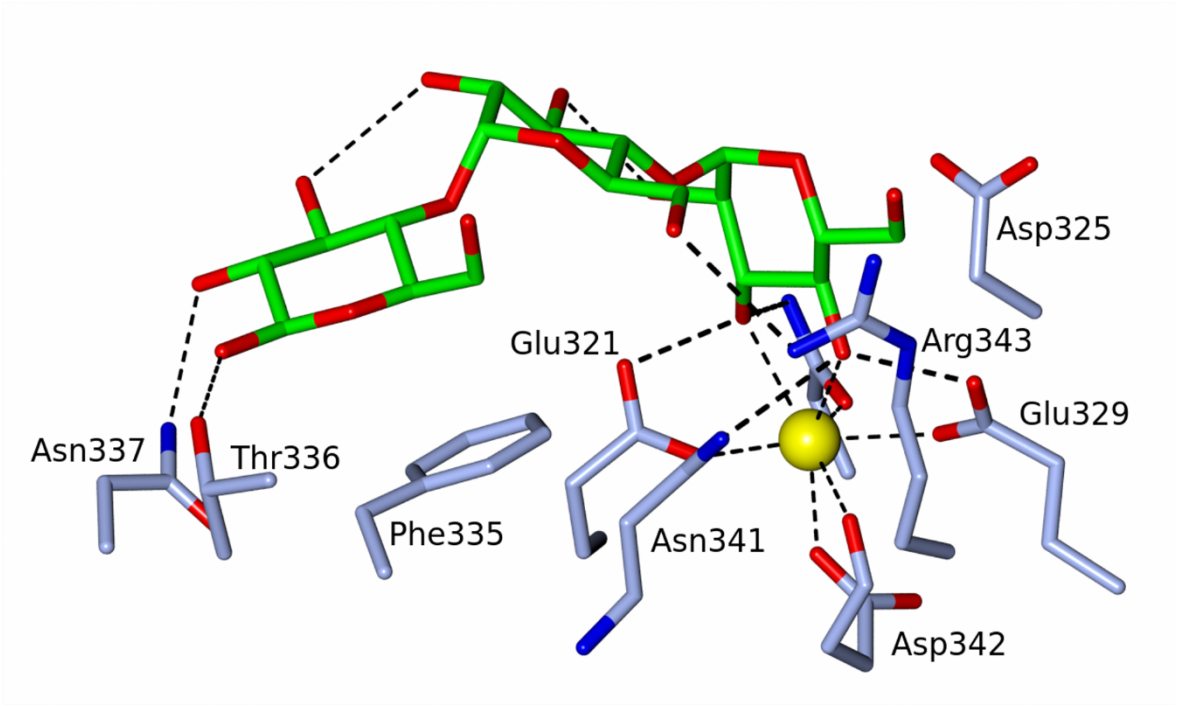


Figure 1.10: Calcium coordination of maltotriose and additional interactions with Phe335, Thr336 and Asn337. Calcium coordination is completed by O3′ and O4′ hydroxyls by the first glucose in the trisaccharide with additional interactions between Glu321, Asn323, Glu329 and Asn341. Recognition of maltotriose is completed by interactions between Arg343 and the nonterminal glucose; π-stacking interactions with Phe335 and further hydrogen bonds with Thr336 and Asn337. Figure generated from PDB accession code 2GGU [Crouch, McDonald, *et al.*, 2006]

1.6 Natural targets for SP-D and early mechanisms for recognition

Small oligosaccharide ligands provide important information about the binding mechanism of hSP-D, however many of the ligands recognised by SP-D are longer oligosaccharides embedded in the cell envelope, either forming structurally important parts of the cell well or integral components of the outer cell membrane. A number of studies have explored how hSP-D recognises natural targets and structural insight into these binding mechanisms are beginning to emerge. This section introduces the components of the pathogen cell wall that hSP-D is able to recognise and the structural data that provides insight into the binding mechanisms employed by hSP-D.

1.6.1 Carbohydrate arrays of pathogens

The surfaces of all pathogens are decorated with a complex array of long carbohydrate moieties that are integral to the structure and function of the cell wall with many of them exposed to the environment. They play an important role as virulence factors in viruses and, in the case of lipopolysaccharides (LPS) from gram-negative bacteria, are able to act as endotoxins and directly attack the host cells. However, the integral nature of many of the carbohydrates and their exposure to the surroundings make them ideal targets for the innate immune system with many of them forming the basis of the pathogen-associated molecular patterns that hSP-D and other immune lectins are able to recognise and thus lead to cell clearance. Extensive work has been done to characterise the cell wall of many pathogenic species with much still to do but the nature of the carbohydrate arrays is introduced here.

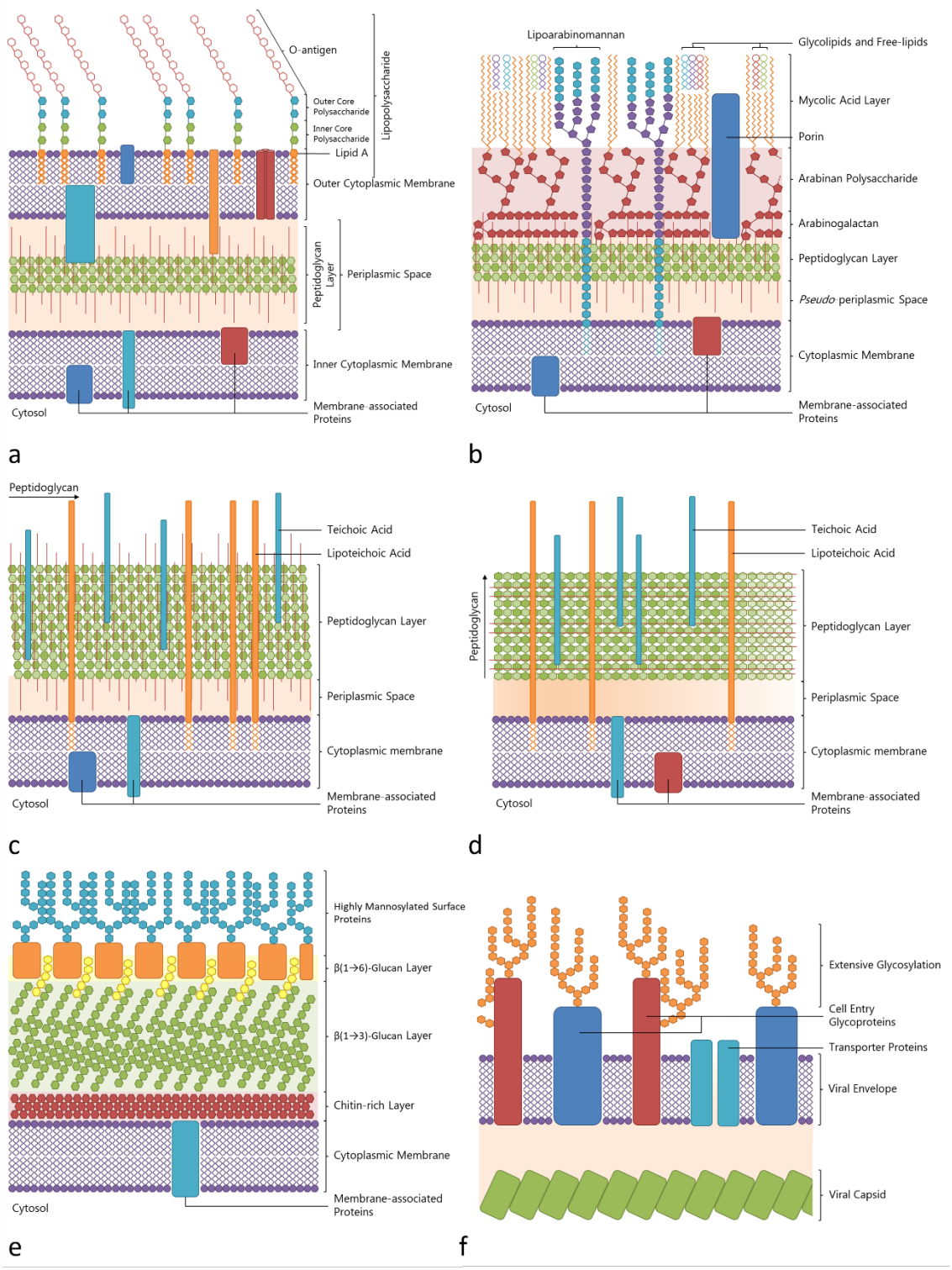


Figure 1.11: Overview of the Carbohydrate Arrays on Various Classes of Pathogen. Brief schematics of the carbohydrate arrays of (a) Gram-negative bacteria; (b) mycobacteria; (c,d) Gram-positive bacteria, both models differentiated by peptidoglycan direction; (e) fungi ; (f) enveloped virus particles. Based on data from references and others [Brennan and Nikaido, 1995; Caroff and Karibian, 2003; Vollmer, Blanot and De Pedro, 2008].

The cell wall of gram-negative bacteria is characterised by: an inner cytoplasmic membrane, composed mainly of phospholipids, transport proteins and membrane-associated synthetic enzymes; a thin peptidoglycan layer surrounded by a fluid filled periplasmic space that provides structural stability; and an outer cytoplasmic membrane containing complex array of porins and LPS [Figure 1.11a]. The LPS molecules are localised to the outer leaflet of the outer membrane and contain four distinct domains that allow them to engage with the host: the membrane-anchor lipid A, the inner core oligosaccharide, the outer core oligosaccharide and the long repeating O-antigen [Figure 1.12; Holst, 2016].

Lipid A is composed of a glucosamine-1-phosphate α(1→6)-linked to glucosamine-4-phosphate with four primary acyl chains linked by two amide bonds (N-linked) at carbon 2 (C2) and two ester bonds (O-linked) at carbon 3 (C3) of each glucosamine [Luderitz *et al.*, 1973]. There are additional O-linked secondary acyl chains bound to the primary acyl chains in some lipid A moieties. The inner core oligosaccharide is a short, conserved chain of bacteria-specific monosaccharides – L-glycero-D-*manno*-heptose and 3-deoxy-D-*manno*-oct-2-ulosonic acid – forming the backbone of the chain with further glucose and N-acetylglucosamine residues [Haishima, Holst and Brade, 1992; Holst and Brade, 1992; Rund *et al.*, 1999; Holst, 2007]. The outer core is less well conserved, containing a larger variety of monosaccharides and linking the inner core oligosaccharide to the O-antigen. The O-antigen is a long, repeating oligosaccharide that adds complexity to the LPS that varies between strains of the same species and a level of resistance to the immune system [Lüderitz, Staub and Westphal, 1966; Forsberg, Bhat and Carlson, 2000; Li *et al.*, 2017].

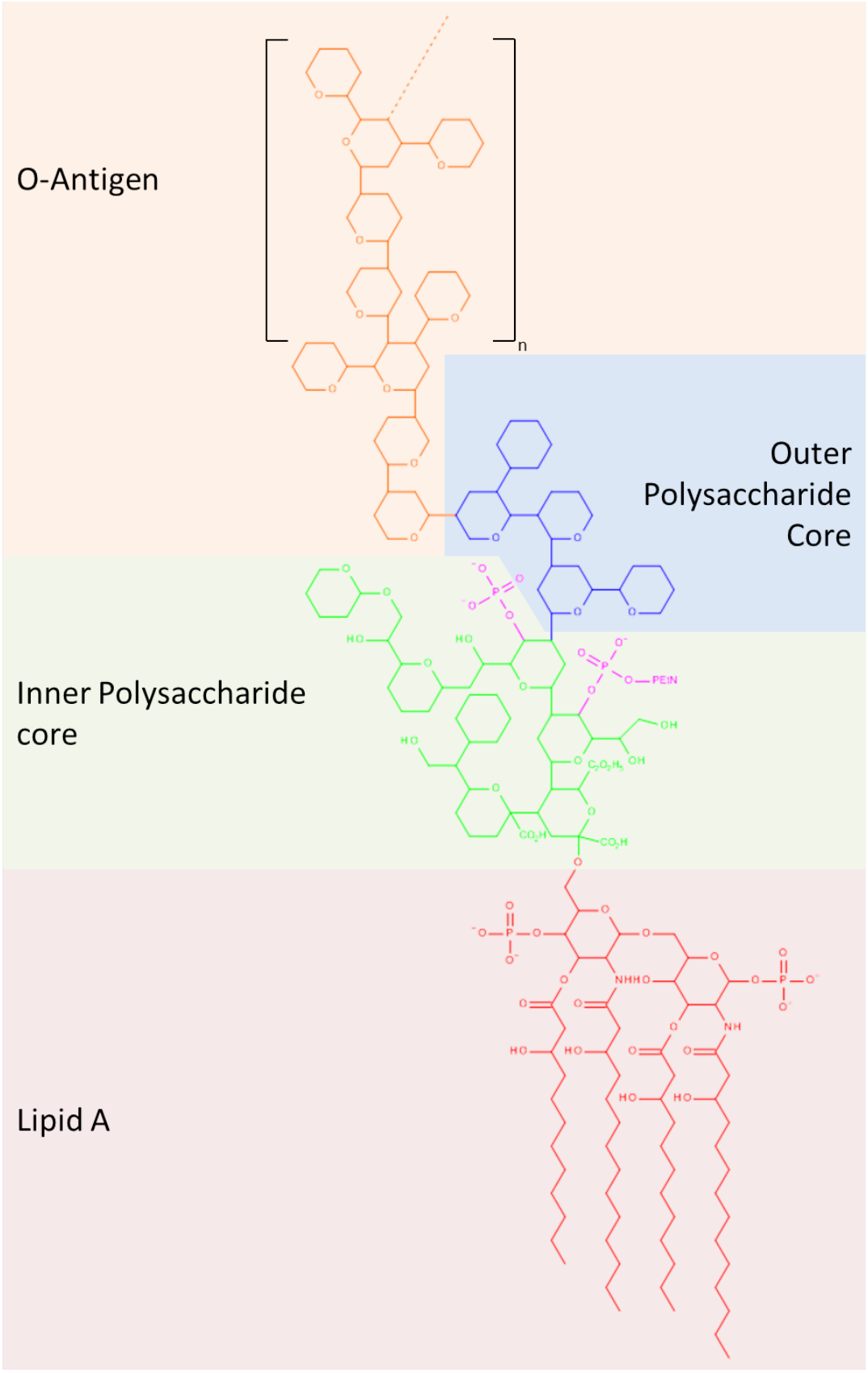


Figure 1.12: General structure of a lipopolysaccharide. Lipid A (red); inner polysaccharide core (green); outer polysaccharide core (blue); O-antigen (orange). Rough mutant LPS lack the O-antigen.

The presence of the O-antigen in the LPS varies between species and mutants. When the O-antigen is present, the LPS is described as a smooth mutant and where the O-antigen is not present, the LPS are described as rough mutants [Haishima, Holst and Brade, 1992; Olsthoorn *et al.*, 1998]. The extent of the rough mutant is dependent on the mutations associated with the LPS synthesis and results in varying degrees of the core polysaccharide, culminating in deep rough mutants which lack all but the core heptose residues in the inner core. Surfactant protein D has been shown to have a higher affinity for rough mutants than smooth mutants.

The cell wall of gram-positive bacteria is structurally less complex than gram-negative bacteria but retains many of the physical properties [Figure 1.11c and 1.11d]. This is principally due to a considerably thicker layer of peptidoglycan surrounding the cell membrane, forming a semi-impermeable barrier to the environment and maintaining the cell shape and stability. Peptidoglycan is a large polymer of muramyl disaccharide, an α (1→4)-linked disaccharide of N-acetylglucosamine and N-acetylmuramic acid, that is crosslinked by a network of short oligopeptides and penta-alanine bridges to form a multi-layered structure of polysaccharide cores [de Pedro and Cava, 2015; Kim, Chang and Singh, 2015]. The peptidoglycan layer is crossed by transport proteins and lipoteichoic acids that anchor into the cell membrane and additional teichoic acids anchor into the peptidoglycan layer.

The orientation of the peptidoglycan in the cell wall of gram-positive still remains unclear despite extensive work in the area [reviewed by Vollmer, Blanot and De Pedro, 2008; Vollmer and Seligman, 2010]. This has led to two distinct models of peptidoglycan architecture. The first model, summarised in Figure 1.11c, describes the polysaccharide cores of peptidoglycan extending along the surface of the cell membrane, either following the axis of cell fission or wrapping around the poles of the cell, encapsulating the cell in

long, crosslinked carbohydrate chains. The second model suggests the peptidoglycan extends away from the cell membrane in a perpendicular direction, shown in Figure 1.11d, forming shorter chains that stack together with more extensive peptide crosslinking, forming a hexagonal matrix. Interestingly, there is evidence for both models in different species, with the sacculi of *Bascillus subtilis* being shown to have rings of peptidoglycan, like in the first model, whereas sacculi from *S. aureus* revealed a pitted, or net-like surface more similar to the second model [Hayhurst *et al.*, 2008; Saar Dover *et al.*, 2015]. This suggests that there may not be a single answer for the peptidoglycan arrangement in the cell wall and may differ between species; changing the way the immune system detects different species.

Mycobacteria have a much more complex cell wall, containing a number of additional layers compared to other types of bacteria that includes a lipid-based capsule containing mycolic acids [Figure 1.11b; Brennan and Nikaido, 1995]. Many of the additional layers are composed primarily of polysaccharides such as arabinan and arabinogalactan which encapsulates a thin layer of modified peptidoglycan and covalently links to the mycolic acids on the outer surface of the cell envelope [Hunter, Gaylor and Brennan, 1986]. The mycolic acid layer also contains a variety of glycolipids and lipooligosaccharides, which resemble LPS, and contribute to reducing the permeability of the cell wall. The only interruptions to the lipid layer are large transport proteins which extend to the peptidoglycan layer and the large, branched, membrane-anchored lipoarabinomannan that decorate the surface of the cells and play an active role in the virulence of the mycobacteria [Chan *et al.*, 1991; Briken *et al.*, 2004; Kang *et al.*, 2005]. Interestingly, it is these structures that have been identified as targets for the C-type lectins and hSP-D [Maeda *et al.*, 2003; Kudo *et al.*, 2004; Ferguson *et al.*, 2010].

Fungal cell walls [Figure 1.11e] differ from bacterial cell walls significantly, with a higher proportion of protein and no peptidoglycan layer to withstand changes to the turgor pressure within the cell. Instead, there is a thick layer of the immunogenic β(1→3)-glucan (a β(1→3)-linked glucose oligosaccharide) and chitin close to the cell membrane that provide strength and shape to the fungal cell [Bowman and Free, 2006; Lenardon, Munro and Gow, 2010; Fakih *et al.*, 2015]. The cell wall is completed by highly mannosylated proteins and β(1→6)-glucan (β(1→6)-linked glucose oligosaccharide) that interweaves with the β(1→3)-glucan matrix [Gander, 1974; Bowman and Free, 2006; Fesel and Zuccaro, 2016].

Unlike the surface of other pathogens, the external surface of an encapsulated virus is derived from the host cell membrane and forms a viral envelope surrounding the central viral particle [Figure 1.11f]. The membranous layer contains a mixture of host-derived proteins and numerous copies of a small set of viral glycoproteins that decorate the envelope, allowing for cell invasion and, in some cases, immune escape [Sodeik and Krijnse-Locker, 2002; Van Kooyk and Geijtenbeek, 2003; Shirato *et al.*, 2004; Welsch *et al.*, 2007]. In most encapsulated viruses, the exposed glycoproteins are heavily glycosylated with complex arrays of N-linked oligosaccharides derived from their mammalian host [Mir-shekari *et al.*, 1997; Vigerust and Shepherd, 2007]. The extensive glycosylation forms the basis for host cell targeting through interactions with extracellular lectin-like domains, anchoring the virus to the host and enabling cell invasion.

In human immunodeficiency virus (HIV), for example, the variable glycosylation of gp120 facilitates binding to dendritic cells through interactions with the C-type lectin DC-SIGN (dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin) and its homolog DC-SIGNR which results in transfer to CD4+ T-cells in the lymph nodes [Geijtenbeek *et al.*, 2000; Mitchell, Fadden and Drickamer, 2001; Snyder *et al.*, 2005; Raska *et al.*, 2010]. Similarly, the high mannose post translational modification of haemagglutinin

and neuraminidase in influenza A virus has also been recognised as major virulence factors during cell invasion; demonstrated by the inhibition of cell entry in the presence of red alga-derived lectins [Blok *et al.*, 1982; Varghese, Laver and Colman, 1983; Jayaraman *et al.*, 2012; Sato *et al.*, 2015].

Presence of the viral envelope leads to two classifications of virus: encapsulated viruses where the viral envelope is present and non-encapsulated viruses that lack the membranous envelope. In both cases, the viral capsid, at the centre of the virus particle, is assembled from a series of interlocking, symmetry related proteins that form a protein-rich 'case' around the genomic nucleic acids that code for all of the viral proteins. This gives rise to viral capsids being described by their symmetry properties. However, as the capsid is enclosed by the viral envelope in encapsulated viruses, the proteins that make up the capsid are not exposed to the host defences. Instead, the glycosylation of envelope glycoproteins form the major recognition sites for the innate immune system and surfactant protein D [Benne *et al.*, 1995; Meschi *et al.*, 2005; Van Eijk *et al.*, 2012].

1.6.2 Recognition of mannosylated structures by a mutant rfhSP-D

These studies used a mutant recombinant fragment of human surfactant protein D that has a single point mutation in the flanking residue Arg343 [Crouch *et al.*, 2009; Goh *et al.*, 2013]. By mutating the Arg343 to valine (R343V), it was shown that the affinity increased for branched mannosyl ligands, haemagglutinin from influenza A virus (IAV) and intact IAV particles compared to native hSP-D which was shown to bind all of the ligands but with significantly lower affinity [Crouch *et al.*, 2009]. The largest increase in binding affinity (in the R343V mutant) was observed for a highly branched nonamannose which had the highest affinity of all the oligosaccharides used in the study. The authors concluded that this may be a result of improved access to the binding site when Arg343 is not present and confirms the important role of Arg343 in ligand selection in the natural SP-D molecule, first

suggested by the earlier small oligosaccharide-bound complexes [Shrive *et al.*, 2003, 2009; Allen *et al.*, 2004; Crouch *et al.*, 2007]. This was confirmed by crystallography using an α(1→2)-linked dimannose in complex with the R343V rfhSP-D mutant which successfully illustrated the effects of reducing the size of the 343 side chain as the second mannose residue in the chain accommodated the space left by the side chain, allowing it to interact with Asp325 and the more distant Arg349 side chain [Figure 1.13a, Crouch *et al.*, 2009].

It was also suggested that the R343V mutant may be recognising a non-terminal epitope in the longer hexa- and nonamannose ligands, enabled the by more accessible binding site [Crouch *et al.*, 2009]. The mechanism for this was later elucidated by the same group who successfully complexed the mannosylation from HA (nonasaccharide made up of α(1→2), α(1→3) and α(1→6)-linked mannose residues, man9) with a mutant rfhSP-D containing two point mutations that reduced the size of both flanking residues [Goh *et al.*, 2013]. In this mutant the R343V mutation, as previously mentioned, had an additional D325A mutation, reducing Asp325 to a minimum size without introducing additional flexibility to the α3 region. This resulted in a further increase in affinity for the HA and the intact virus compared to both the native structure and the R343V single point mutant. The D325A+R343V mutant rfhSP-D crystallised in a different space group, $P2_12_12_1$, to previous structures as a result of Asp325 not being available to make the key crystal contact. In the binding site, the structure revealed three α(1→2) mannose residues through calcium recognition via the non-terminal mannose O3′ and O4′ hydroxyls. The reducing terminal mannose was bound in a similar orientation to that seen in the dimannose/R343V complex and the third mannose was positioned in approximately the position of Asp325 in the native structure [Figure 1.13; Shrive *et al.*, 2003; Goh *et al.*, 2013].

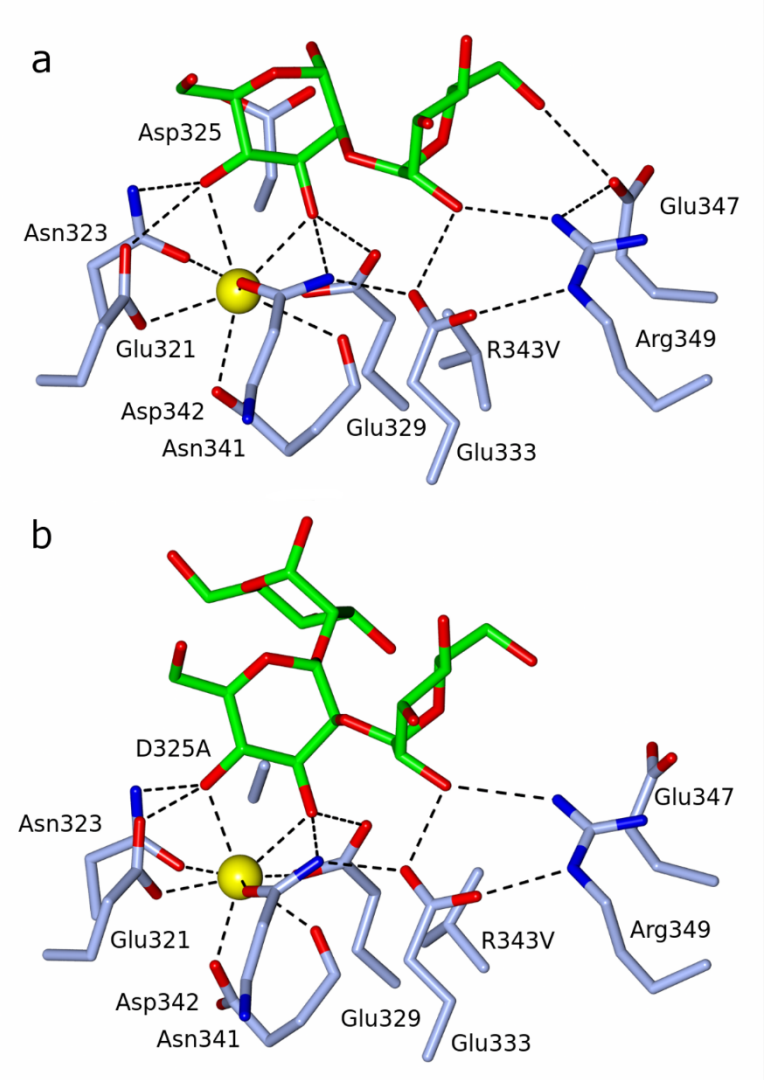


Figure 1.13: Calcium coordination of α(1→2)-linked mannobiose (a) and nonamannose (b) in two mutants of rfhSP-D. (a) mannobiose coordinates the calcium in the R343V mutant through O3′ and O4′ hydroxyls. (b) nonamannose coordinates the calcium in the same way in a D325A/R343V double mutant of rfhSP-D. Figures generated from PDB accession codes 3G84 and 4M18 [Crouch *et al.*, 2009; Goh *et al.*, 2013].

The study confirms the importance of Asp325 in ligand selection and orientation in the calcium binding pocket, first proposed by *Shrive* and colleagues in 2009, as the presence of Asp325 occludes the binding site and prevent man9 from being recognised by Ca1, which is supported by the binding studies comparing the double mutant to native in recognising the H2N3 strain of IAV [Shrive *et al.*, 2009; Goh *et al.*, 2013]. This selection mechanism has been seen in other collectins, including MBP where the two amino acids in equivalent

positions to D325 and R343, His189 and Ile207 respectively, dramatically affect the binding specificity to monosaccharides, identifying His189 (325 in human SP-D) conferring the major ligand selection activity [Iobst and Drickamer, 1994; Iobst *et al.*, 1994].

1.6.3 Recognition of lipopolysaccharides by rfhSP-D

Recognition of Gram-negative bacteria by surfactant protein D is a well-established phenomenon, first recognised when laboratory and clinical bacterial samples of *E. coli*, *S. paratyphi*, and *K. pneumonia* were successfully bound and agglutinated by full length SP-D collected from BAL [Kuan, Rust and Crouch, 1992]. Since then, a range of Gram-negative bacteria including *S. enterica* and *H. influenzae* have been shown to be bound by hSP-D by the polysaccharide portion of the lipopolysaccharide through different regions in the polysaccharide [Kuan, Rust and Crouch, 1992; LeVine *et al.*, 2000]. In *Klebsiella pneumonia*, the highly mannosylated O-antigens appears to be selectively targeted by SP-D, mediated through CRD interactions with the mannose residues, whereas in most cases, the inner core polysaccharide provides the recognition motif for SP-D, shown by the propensity of hSP-D to recognise rough LPS mutants [Kuan, Rust and Crouch, 1992; Sahly *et al.*, 2002]. More recently, it has been suggested that the increased affinity for rough LPS mutants may be a result of heptose recognition in the widely conserved inner core polysaccharide with a recent study confirming that a rfhSP-D has a higher affinity for heptose-rich synthetic oligosaccharide cores compared to cores with no heptose residues [Wang *et al.*, 2008; Reinhardt *et al.*, 2016].

The nature of the interaction between heptose and a rfhSP-D was previously investigated using a representative L-glycero-α-D-*manno*-heptose-bound complex, revealing a novel binding mechanism between the O6′ and O7′ hydroxyls of the glyceryl side chain and Ca1 in the main binding pocket [Wang *et al.*, 2008]. The interactions between the O6′ and O7′, which assumed a comparable position to the O3′ and O4′ hydroxyls of maltose and

mannobiose, and rfhSP-D were completed by the surrounding side chains associated with the recognition of carbohydrate ligands; namely Glu321 and Asn323 which interact with O6′ and Glu329 and Asn341 which interact with O7′ [Shrive *et al.*, 2003, 2009; Crouch *et al.*, 2009]. The mannose-type ring makes no direct interactions with the protein in subunits A and C but makes a single interaction with the amine group of Asn323 in subunit B [Figure 1.14; Wang *et al.*, 2008]. The interaction in subunit B appears to be a consequence of the crystal system and the direct interactions made between the mannosyl ring and the symmetry related molecule which differ between the three subunits.

In the same study, a synthetic analogue of the heptose disaccharide found in the core of most LPS structures of *E. coli*, *S. enterica* and *H. influenzae* was also successfully crystallised with a rfhSP-D [Holst and Brade, 1991; Haishima, Holst and Brade, 1992; Phillips *et al.*, 1992; Wang *et al.*, 2008]. Calcium coordination of the binding heptose occurred in the same way as the L-D-heptose-bound complex, primarily through coordinate bonds with the O6′ and O7′ hydroxyls of the glyceryl side chain and completed by the same protein-ligand interactions with the calcium coordinating site chains [Figure 1.14]. The second heptose, linked by an α(1→3) glycosidic bond, was defined in the structure as a result of crystal contacts and makes no direct interaction with the protein.

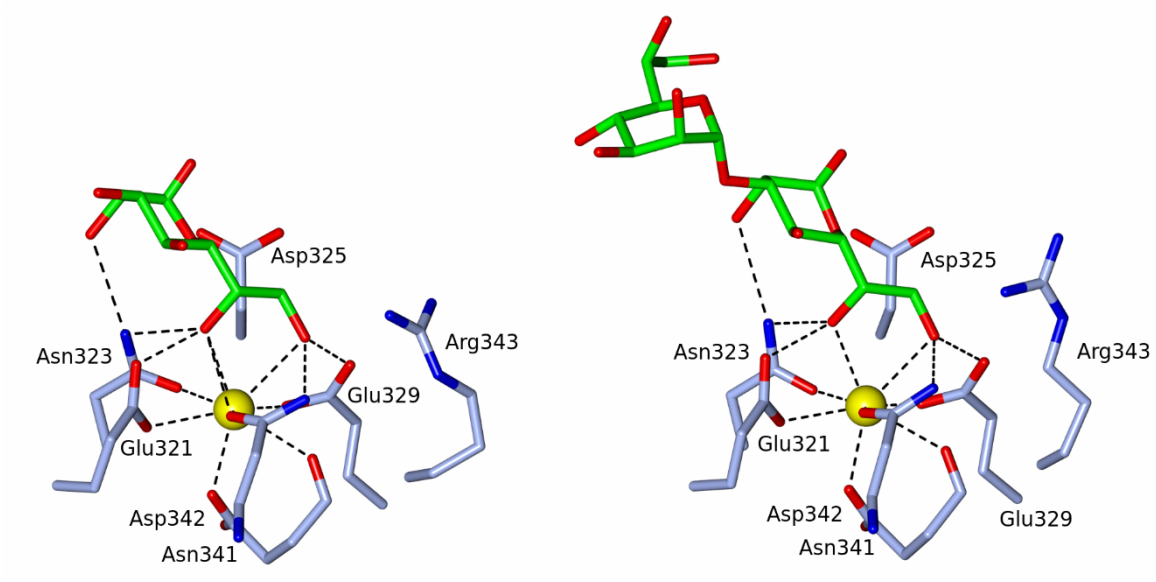


Figure 1.14: Calcium coordination of L-glycero-D-manno-heptose (left) and α(1→3)-linked diheptose (right). Calcium coordination is completed through O6′ and O7′ hydroxyls in both complexes, with recognition completed by hydrogen bonds between Glu321, Asn323, Glu329 and Asn341. Figure generated from the PDB accession codes 2RIB (left) and 2RIC (right) [Wang *et al.*, 2008]

A significant step in the understanding of how SP-D recognises rough LPS mutants was made following the elucidation of the binding mechanism for a natural rough mutant of *H. influenzae* Eagan 4A [Clark *et al.*, 2016]. *Clark* and colleagues used a delipidated oligosaccharide from Eagan 4A, containing the lipid A proximal Kdo-Heptose-β-D-Glucose, to successfully crystallise the first LPS polysaccharide-bound complex. The structure revealed ligand bound in subunits B and C where the primary recognition occurred through the calcium coordination from O6′ and O7′ of the heptose. These two hydroxyl groups also complete the core protein-ligand interactions with Glu321, Asn323, Glu329 and Asn341. The overall position of the heptose is very similar to the heptose-bound complexes, however there are additional interactions with the Kdo residue that was successfully defined in both subunits where ligand was bound [Figure 1.15]. Interestingly, the Kdo was in a furanose derivatised form following a β-elimination during ligand preparation, forming a five-membered ring with a long 2-oxobutanoic acid side chain [Auzanneau, Charon and

Szabó, 1991; Clark *et al.*, 2016]. The flanking residue Asp325 was found to be essential in the Kdo recognition, making a direct hydrogen bond to the O6′ of the furanosyl ring however the 2-oxobutanoic acid was not defined and assumed to be free to move in the crystal structure. It was suggested by the authors that there may be a role for both of the flanking residues, Asp325 and Arg343 in recognising a complete six-membered Kdo ring as a simple overlay appears to suggest this is possible, however direct evidence for this is still required [Clark *et al.*, 2016]. Despite the second heptose being defined in the heptose disaccharide-bound complex, there was no evidence for the β-D-glucose which was substituted at O3′ position of the heptose ring [Wang *et al.*, 2008]. This is not surprising, however, when you consider the major interactions between the crystal contact and the second heptose are mainly mediated through the O6′ and O7′ hydroxyls in the heptose disaccharide-bound complex, suggesting there are no stabilising interactions between the β-D-glucose and the protein binding pocket.

Taking these studies together, it is clear from the high affinity for heptose and the ability of hSP-D to recognise the Hep-Kdo motif common to most lipopolysaccharides that recognition of rough mutants is primarily via the proximal heptose [Wang *et al.*, 2008; Clark *et al.*, 2016]. This may explain the propensity for hSP-D to recognise rough rather than smooth LPS mutants of Gram-negative bacteria as the long O-antigens of smooth LPS mutants may shield the inner core from protein recognition, as was suggested by *Clark* and colleagues [Clark *et al.*, 2016]. However, what these studies fail to do is to describe alternative epitopes that hSP-D may also be able to utilise in recognition of rough strains especially considering the flexibility of SP-Ds binding mechanism to recognise a plethora of different carbohydrates, demonstrated by the small ligand studies [Shrive *et al.*, 2003, 2009; Crouch, McDonald, *et al.*, 2006; Crouch *et al.*, 2009]. Further work is required to clarify this.

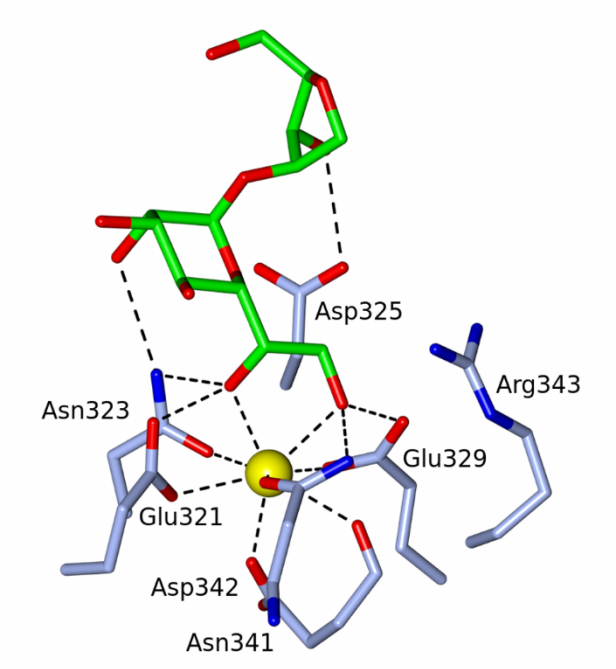


Figure 1.15: Calcium coordination of the hydrolysed inner core polysaccharide from *H. influenzae* Eagan 4A. Coordination of the calcium is completed through the O6′ and O7′ hydroxyls of the proximal heptose. Recognition is completed by Glu321, Asn323, Glu329 and Asn341 with addition hydrogen bonding between O6′ of the anhydro-Kdo and Asp325. Figures generated from the PDB accession code 4E52 [Clark *et al.*, 2016]

1.6.4 Recognition of gram-positive cell wall components

Gram-positive bacteria, including the clinically important species *Streptococcus pneumoniae* and *Staphlococcus aureus*, have been shown to be targets for surfactant protein D, leading to agglutination and activation of phagocytes [McNeely and Coonrod, 1993; Hartshorn *et al.*, 1998; Jounblat *et al.*, 2004; Leth-Larsen *et al.*, 2005; Zhang *et al.*, 2015]. How the recognition occurs has been investigated using a number of different components of the bacterial cell wall, identifying peptidoglycan and lipoteichoic acid as recognition site for gram-positive bacteria, however no structural investigations have been completed to date [van de Wetering *et al.*, 2001].

The study by *de Wetering* successfully demonstrated that unmodified peptidoglycan, which is representative of *in situ* peptidoglycan, could be bound by hSP-D in a semi-quantitative

assay that showed 7.4% of the total protein added successfully bound to the peptidoglycan. Inhibition assays of both peptidoglycan and lipoteichoic acid, using a range of monosaccharides, revealed that lipoteichoic acid has a higher affinity for hSP-D than peptidoglycan [van de Wetering *et al.*, 2001]. The lipoteichoic acid, from *B. subtilis*, was shown to be bound by hSP-D in a saturable manner suggesting that there are a limited number of binding sites. The recognition of peptidoglycan and lipoteichoic acid was calcium dependent and is thought to occur through the CRD as inhibition assays using inositol, maltose, glucose and mannose all successfully inhibited binding and are known to bind in the calcium binding pocket in the CRD. The mechanism for this remains to be elucidated.

1.7 Conclusions from work so far

The work so far has revealed a binding mechanism in SP-D that is dominated by the calcium coordination of two hydroxyls that adopt a conserved position in relation to the calcium ion and the surrounding side chains. Studies have revealed that this mechanism allows hSP-D to be versatile and recognise a range of mono- and oligosaccharides in the calcium binding pocket, completing protein-ligand interactions with Glu321, Asn323, Glu329 and Asn341 in all cases. The propensity to recognise a range of carbohydrates allows hSP-D to recognise a wide range of pathogens through their carbohydrate arrays on the surface, making it an essential part of the innate immune response and important regulator of the acquired immune response.

The crystal structures have also revealed important roles for the two binding site flanking residues, Asp325 and Arg343, in selecting and recognising natural ligands. These two side chains are in key positions to interact directly with bound oligosaccharides and orchestrate calcium coordination by orientating the hydroxyls in the binding site. Longer oligosaccharides have also highlighted the possibility of an extended binding surface

surrounding the main calcium coordinating pocket which may also play a vital role in recognising ligands and contribute to pathogen recognition by hSP-D.

1.8 Aims and objectives of this work

Studies presented here are a continuation and extension of work carried out by *Trevor Greenhough*, *Annette Shrive* and *Ian Burns*, and previous members of the research group *Amy Shaw, Carrie Smallcombe* and *Ruben da Silva* with some insight from *William Neale*, a current member of the group.

The characterisation of pathogen recognition by surfactant protein D is an important aspect of understanding the role of hSP-D in innate immune defence, particularly in establishing how SP-D is capable of recognising such a wide range of targets from all pathogenic phyla. Previous work in this group has established the native structure of a biologically and therapeutically active recombinant fragment of human SP-D and a number of ligand-bound complexes have also been solved, identifying the calcium-dependent mechanism of ligand recognition [Shrive *et al.*, 2003, 2009; Clark *et al.*, 2016]. The aims of this project are to extend this work to further characterise the binding determinants of hSP-D.

To do this, my project sets out to first determine how the recombinant fragment of SP-D recognises longer oligosaccharide ligands (trioses and above) that utilise the alternative anomer of D-glucose, β-D-glucose, common in pathogenic carbohydrate arrays and isomers of malt-N-oses. Ligand-protein complexes, using commercially available cellotriose (β(1→4)-linked), laminaritriose (β(1→3)-linked) and isomaltotriose (α(1→6)-linked), were investigated using crystallographic techniques to determine the structural determinants of ligand binding, compare them to maltotetraose (α(1→4)-linked

tetrasaccharide) and maltoheptaose (α(1→4)-linked heptasaccharide) and characterise recognition of non-terminal monosaccharide residues if any were present.

The second objective was to determine how SP-D recognises components of peptidoglycan, a structural component of the bacterial cell wall, using a muramyl disaccharide unit extracted from the Gram-positive Species *M. luteus*. Full length SP-D has been shown to recognise and bind to soluble and insoluble peptidoglycan however how the recognition occurred has not been elucidated at the molecular level. The study presented here describes for the first-time recognition of muramyl disaccharide and provides some insight into how longer peptidoglycan ligands may be accommodated by a recombinant fragment of hSP-D.

The final objective is to build on previous work in the group which identified a novel binding mechanism for the proximal Hep-Kdo motif of a lipopolysaccharide from the rough lipopolysaccharide mutant *H. influenzae* Eagan 4A [Clark *et al.*, 2016]. A structure of a delipidated oligosaccharide from *S. enterica* sv. Minnesota R7 mutant in complex with rfhSP-D has been investigated by a previous PhD student [Smallcombe, 2014]. However, my objective was to revisit the structure to resolve unanswered questions and provide a definitive description of the ligand-protein interactions, using the now fully characterised R5-bound complex structure as a comparison [Littlejohn *et al.*, 2018].

Chapter 2: Introduction to protein crystallography

In crystallography, there are a set of concepts that are essential to understanding how structural data can be derived from a crystal. This chapter is not an exhaustive review of protein crystallography. Instead, it intends to briefly explore the important concepts, to provide context to how the data was collected, analysed and how that data enabled the completion of the work throughout the remaining chapters.

2.1 A brief history of crystallography

2.1.1 The advent of the crystallographic method

X-ray diffraction was first described by *Laue* in 1912, following experiments with crystals of zinc sulphate [Friedrich, Knipping and Laue, 1913; Laue, 1913]. The diffraction pattern, thought to be evidence of a polychromatic source, was the first time that any structural insight into atomic positions was derived from a crystal. *Laue* proposed that the crystal was a three-dimensional array, with zinc and sulphur atoms at the corners of each of the cubes that made up the array.

In the same year, following *Laue's* discovery, the *Braggs* used the diffraction pattern from ZnS to elucidate *Bragg's Law* and, in doing so, provided the tool for solving all future structures from either chemical or protein sources (discussed in section 2.4.1, [discussed in section 2.4.1; Bragg, 1912]. *Lawrence* and *William Bragg* went on to use the structure of ZnS to prove *Bragg's Law*, determining that there are additional atoms in each face of the ZnS cube. Since then, *Bragg's Law* has been used to reveal a huge wealth of information about the atomic positions in not only some of the most basic materials, like sodium chloride, but also large protein complexes, viral capsids and, most famously, the structure of DNA [Bragg and Bragg, 1915; Watson and Crick, 1953].

2.1.2 Development of macromolecular crystallography

More complex structures of small biomolecules were soon solved using crystallographic techniques pioneered by Dorothy Hodgkin. In 1939, the first of the important biological molecule structures to be solved was cholesterol, an important component of the cell membrane, followed by penicillin in 1949 and vitamin B12 in 1956 [Carlisle and Crowfoot, 1945; Hodgkin, 1949; Hodgkin *et al.*, 1955]. This developed the crystallographic technique so far that the first structure of insulin was solved in 1969 – a project that was started 30 years before when the first diffraction patterns from a crystal of insulin were captured [Adams *et al.*, 1969].

In the intervening years, the first protein structures of myoglobin (1958) and haemoglobin (1960) were solved, followed by the first high-resolution structure of myoglobin in 1960, demonstrating that protein structures could be resolved to near-atomic resolution [Kendrew *et al.*, 1958, 1960; Perutz *et al.*, 1960]. Since then, the number of crystal structures of large macromolecules, including large protein complexes like the ribosome, which is over four million kDa, has increased year on year to beyond 135 000 at the time of writing [Berman *et al.*, 2000]. In no small part, this is thanks to the early pioneers in the field and the continued work to push and improve the techniques into the future.

2.1.3 The advent of modern computational techniques

The increasing complexity of the protein structures being investigated and the realisation that Fourier transforms can be used to help determine those structures led to the realisation that the emerging power of computation could be utilised. From the early days of crystallography, the first computers were used to accurately calculate the complex Fourier transforms required to reproduce electron density maps from the diffraction images (discussed in section 2.4.3). However, most of the process was still manually completed by painstakingly measuring each reflections position and intensity and then interpreting the

data output from the computational machines to elucidate the electron density and the molecular structure.

As the capacity of computers continued to increase through the 80s and 90s, so did the ability of the crystallographic tools to successfully process digital diffraction images and downstream processing, many of which form the basis of the programs still used today (discussed in chapter 3). A major improvement came with the ability to display electron density computationally and visualise it in three dimensions using programs like *O* written by *Alwyn Jones* [Jones *et al.*, 1991]. Since then, many of the earlier algorithms have been refined and optimised to produce programs that are able to deal with even difficult data, allowing for more and more protein structures to be successfully solved and deposited [PDB]. This has culminated in a concerted effort to continue progressing modern computational techniques as part of the collaborative computational number 4 (CCP4) and a similar suite of programs, Python-based Hierarchical Environment for Integrated Xtallography (PHENIX) [Adams *et al.*, 2010; Winn *et al.*, 2011]. These two program suites have revolutionised macromolecular crystallography.

2.2 Protein crystallisation

2.2.1 Crystallisation background

For X-ray crystallography to be possible, a crystal of the protein must be grown because a regular and repeating three dimensional array is needed to provide an interpretable diffraction pattern that allows for the nature and location of the atoms to be discerned. The 3D array is a result of all the atoms adopting equivalent positions in a regular lattice, related by symmetry operations, discussed in section 2.3. However, the array must be ordered and regular throughout the crystal for the diffraction pattern to useful and, therefore, the formation of the crystal must be carefully controlled.

The main obstacle to protein crystal formation is the free energy required for the critical nucleation of an ordered crystal that can then go on to grow into a crystal; using the initial nucleation as a template for the remaining protein molecules to bind with the surface of the crystal, in a lower energy process. To overcome the limiting energy, and for an ordered nucleus to propagate, the protein must enter a meta-stable, supersaturated condition that is beyond the solubility of the protein but, importantly, is not sufficient for the protein to precipitate out of the solution. Once the nucleation event occurs, the concentration of the protein in solution begins to fall as the crystal grows, with the concentration gradient driving crystal growth. However, when the concentration of the protein reaches equilibrium between the solution and the crystal, the concentration gradient stops, and the crystal growth stops as a result.

Control of the crystallisation must also extend to the production of crystals and not protein precipitate. This is achieved through careful increases in the concentration of a precipitating solution which drives the protein molecules into a super saturated state by gradually evaporating the water from the protein-precipitant mixture. This is usually achieved by a process of vapour diffusion. Two methods of vapour diffusion have been developed: Hanging drop method; and sitting drop method. Both methods are fundamentally the same, with a small volume (or drop) of protein solution and precipitant buffer spatially separated from a larger well of precipitant buffer in an air-tight environment. This allows for water to slowly diffuse out of the smaller volume and equilibrate with the reservoir which has the effect of gradually increasing the concentration of the protein and the precipitant in the drop, until the energy required for nucleation is overcome and crystal growth is initiated. The crystals are then able to grow until the concentration of free protein in the solution is equal to the protein concentration in the

crystal. The method used during the set of experiments described here was the sitting drop method [Figure 2.1].

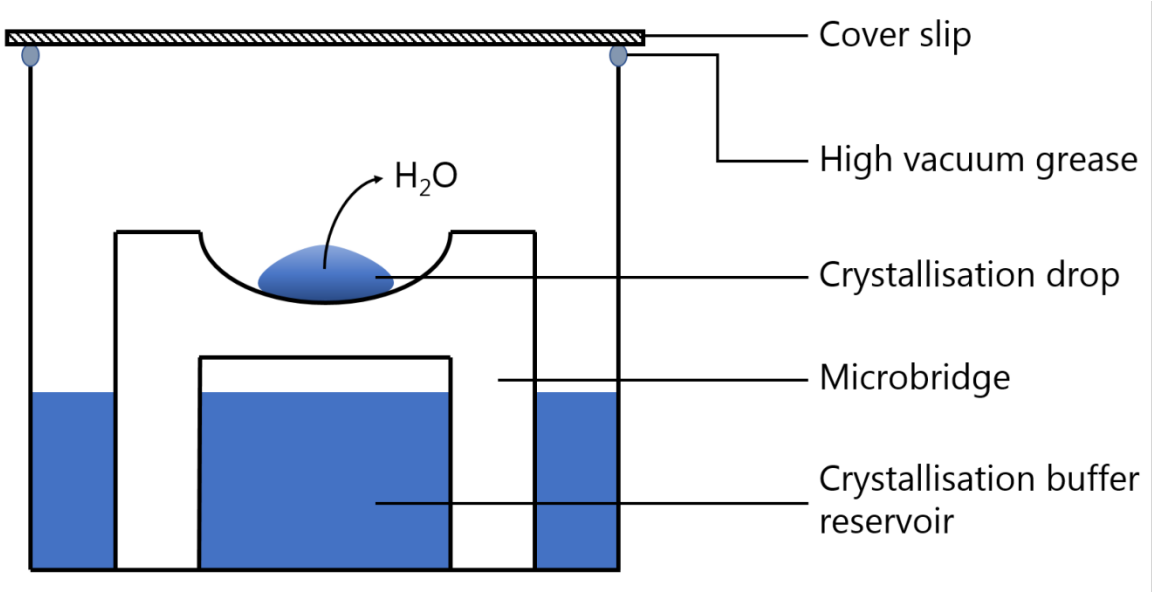


Figure 2.1: Sitting drop method. The figure shows the sitting drop method for vapour diffusion which allows for the gradual increase of protein concentration in the crystallisation drop until nucleation occurs and the protein crystal forms. In all of the experiments, 24-well linbro plates were used with approximately 1ml of reservoir solution and a 1:1 ratio of protein solution to precipitant buffer to a final volume of 4µl in the drop

2.2.2 Improving crystallisation

In each crystallisation experiment, there is some degree of improvement required to optimise the crystal quality, size and singularity because the specific conditions within the crystallisation well will differ between wells, even if the intended conditions are the same as a previously successful condition. The concentration of precipitant is commonly varied to optimise the crystallisation, but pH and type of salt are also regularly varied to improve both singularity of the crystals, reducing the size of crystal showers, and increasing the size of the crystals. These changes improve the crystal quality by changing either the stability of

the protein at supersaturation, encouraging better nucleation, or affecting the critical nucleation energy by either increasing it, to reduce the number of crystals nucleating, or reducing it to allow for better nucleation. Nucleation can also be aided by seeding microcrystals, which overcomes the critical nucleation energy and allows for crystals to grow in a more energetically favourable condition.

2.3 Symmetry and the crystal lattice

A crystal is, when broken down to its component parts, an array of a single repeating unit of 3D space, these units being related to each other by a number of translational and rotational symmetry elements that allow for the given 3D space to be completely filled. This single repeating unit is the unit cell. By characterising the relationships between the unit cells, the spacegroup can be derived using the symmetry elements. This short section will introduce the main concepts required to determine the spacegroup of protein crystals which are required for successful interpretation of the diffraction pattern and subsequent solution of the crystal structure.

2.3.1 Basic lattice principles

For diffraction to be useful in solving a molecular structure, an ordered three dimensional crystal comprising a repeated single unit, the unit cell, is required. The unit cell dimensions are described by three edges, *a*, *b*, and *c*, and the three angles between those edges, α, β, and γ [Figure 2.2]. Geometrical parameters of planes within the crystal, in terms of the unit cell parameters, can then be calculated, completing the three-dimensional description of the unit cell. The fixed size of the unit cell also allows the coordinates in the *x*, *y*, and *z* directions to be expressed as fractions of the unit cell and, by extension, in terms of the crystal planes which intersect these coordinates. These planes are described by the Miller indices *h*, *k*, *l* and intersect with the unit cell axes at *a/h, b/k, and c/l*. This property of Miller indices can then be used to determine the unit cell experimentally because the dissecting

planes are also the Bragg planes which, when aligned throughout the crystal according to the lattice of repeated unit cells, generate strong reflections in accordance with Bragg's law.

For the crystal to form correctly, the unit cell must also be able to tessellate (by translation) consistently throughout three-dimensional space so that there is no space between the unit cells. Within the unit cell itself the molecules are related by a number of symmetrical rotations, translations and mirror planes (the latter not being appropriate for proteins due to the enantiomeric nature of the amino acid). The repeated translational tessellation of unit cells throughout the crystal means that the symmetry elements within the unit cell also extend throughout the crystal.

There is also usually an asymmetric unit within the unit cell that does not have any crystal symmetry but may have local, or non-crystallographic, symmetry; such as the approximate 3-fold axis which relates the individual protein subunits in the recombinant fragment of human surfactant protein D (rfhSP-D) trimer [Håkansson *et al.*, 1999; Shrive *et al.*, 2003] Importantly, the symmetry elements mean that each unit cell shares exactly the same environmental conditions as every other unit cell in the crystal, making it possible to successfully diffract an incident beam of X-rays in a reproducible and constructive way resulting in a diffraction pattern that is interpretable in terms of the nature and position of each and every atom in the crystal and hence in the unit cell and the constituent molecules themselves.

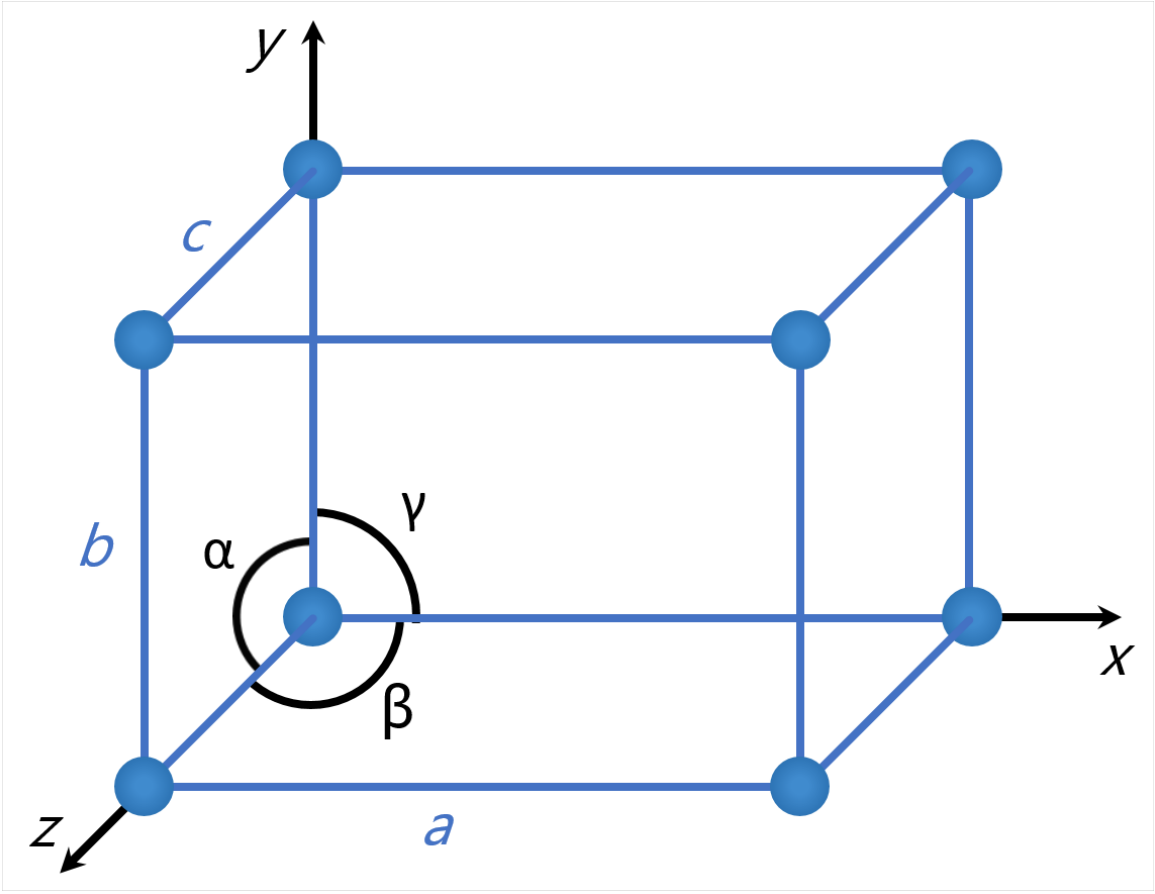


Figure 2.2: Unit cell parameters superimposed on a unit cell. The length of each axis (a, b, c) is determined by the distance between the lattice points (blue spheres) and describe the smallest basic repeating unit of a larger crystal lattice. The angles (α, β, γ) are measured between the three unit cell axes; α between b and c; β between a and c; and γ between a and b.

2.3.2 Bravais lattices and crystal systems

The geometry of the unit cell, and therefore the overall crystal, is described by the Bravais lattice which can then, more broadly, be described by the crystal system. There is a total of seven crystal systems describing the overall shape of the unit cell with the unit cell parameters (*a*, *b*, *c*, α, β, γ) measured from the diffraction pattern during indexing [table 2.1]. Within these seven systems, there are 14 Bravais lattices. If the lattice points are only found at the corners of the unit cell, they are identified as a *Primitive* lattice (P) but in the cases where there are equivalent lattice points not at the vertices, they are identified as *body-centred* (I), *face-centred* (all faces, F) and *face-centred* (one face, C). In the special case of a trigonal crystal lattice, there is also a *Rhombohedral-centred* (R) lattice with lattice points additional to those at the vertices of a primitive hexagonal lattice. In each case, there is a minimum amount of symmetry required for a successful lattice construct, however there are further possible symmetry operations in each case which are described by the point group and additional translation operations in the spacegroup.

Table 2.1: Summary of crystal systems and associated Bravias lattices. The table summarises the Bravais lattices, minimum symmetry requirements and unit cell constraints for each of the seven crystal systems

Crystal system	Bravais lattices	Minimum symmetry	Unit cell constraints
Triclinic	P	None	None
Monoclinic	P, C	One 2-fold axis	$\alpha=\gamma=90°$ or $\alpha=\beta=90°$
Orthorhombic	P, C, I, F	Three 2-fold axes[a]	$\alpha=\beta=\gamma=90°$
Tetragonal	P, I	One 4-fold axis	$\alpha=\beta=\gamma=90°$ $a=b$
Hexagonal	P	One 6-fold axis[b]	$\alpha=\gamma=90°$, $\beta=120°$ $a=b$
Trigonal	R[b]	One 3-fold axis	$\alpha=\beta=\gamma=60°$ $a=b=c$
Cubic	P, I, F	Four 3-fold axes	$\alpha=\beta=\gamma=90°$ $a=b=c$

[a]Parallel to *a*, *b* and *c*
[b]Special case of a primitive hexagonal system can produce a primitive trigonal lattice with a 3-fold axis where $\alpha=\gamma=90°$, $\beta=120°$ and $a=b$

2.3.3 Point groups

Each Bravais lattice has a set of minimum rotational symmetry operations required for an ordered, three-dimensional array of unit cells to "assemble" into a diffracting crystal. In many cases however, there are additional, but complimentary, rotation axes that enable better packing of the protein molecules [table 2.1]. In all cases, the rotational symmetry can be described using a "point group", which describes the *n*-fold axes in the lattice and, importantly, which axis the rotation occurs around. For example, the point group 222 would describe a lattice (and unit cell) with 2-fold rotational axes about x, y and z. There is a total of 32 unique point groups that can describe all rotation axes of all the crystal lattices. However, because the point group is determined from the overall lattice topology, any translational symmetry relationships are lost. As a result, screw axes (a rotation plus translation symmetry operator) are not identified within the point group.

2.3.4 Additional symmetry elements

Beyond unit cell translations and rotation axes, there are a second set of symmetry elements that combine a translation with another symmetry operation. In protein crystallography, these secondary elements are limited to screw axes since mirror planes and inversion centres are not allowed by the enantiomeric nature of amino acids and other biological molecules such as carbohydrates and DNA. A screw axis describes the translation and rotation of an asymmetric unit (and hence the molecules within that unit), along a crystallographic *n*-fold rotation axis.

The amount of translation and rotation between two molecules are described by the order of the rotation axis and the fractional translational shift between them in terms of the unit cell.

$$translation = \frac{1}{n}\ where\ n = 2,3,4\ or\ 6$$

$$rotation = \frac{360°}{n}\ where\ n = 2,3,4\ or\ 6$$

Therefore, for a screw axis associated with a 2-fold axis, *n* = 2, the asymmetric unit is translated by ½ a unit cell in a positive direction along the axis and rotated 180° about the axis.

The value of *n* is limited to 2, 3, 4 or 6 according to the seven crystal systems and associated Bravais lattices [see table 2.1]. However, both a 5-fold and 7-fold can exist as local, non-crystallographic symmetry, within the asymmetric unit.

2.3.5 Spacegroup

The spacegroup of any crystal is defined, primarily, by the Bravais lattice, the point group and any additional translations within the crystal, such as screw axes or glide planes. Thus, a crystal with a primitive (P) lattice and point group 222 where each 2-fold axis is a screw axis has spacegroup $P2_12_12_1$. In protein crystallography, the number of spacegroups is reduced from 230 to just 65 because of the enantiomeric properties of amino acids, preventing protein crystals from forming with mirror or inversion planes that would require the rotameric state of the L-amino acids to change to D-amino acids; which do not naturally exist.

A large proportion of proteins have been solved in the orthorhombic spacegroup, $P2_12_12_1$, or monoclinic spacegroup, $P12_11$ (written as $P2_1$). This is because these spacegroups are more flexible in terms of protein packing, allowing for a larger number of possible spatial arrangements of the protein molecules. Understanding the nomenclature of each spacegroup can be broken-down into three sections: (1) the Bravais symbol; (2) the crystallographic rotational symmetry operations; and (3) additional translations of the asymmetric unit within the unit cell. All of this information makes it is also possible to then derive the crystal system from the spacegroup. Taking the recombinant fragments of hSP-D, previously solved in spacegroup $P2_1$, as an example: P is representative of the Bravais lattice of the crystal, in this case primitive; 1 2 1 represents the crystallographic symmetry operations, in this case a 2-fold axis in the *b* direction; and 2_1 representing a screw axis with a translation of ½ a unit cell along *b* [Håkansson *et al.*, 1999; Shrive *et al.*, 2003; Crouch, McDonald, *et al.*, 2006].

In lower symmetry spacegroups, such as triclinic, monoclinic and orthorhombic, the symmetry operations describe the three axes, *a*, *b* and *c*, whereas in higher symmetry spacegroups this is not necessarily the case; for example, P422 indicates a 4-fold axis

parallel to c with 2-foldaxes along a and b and half way between. In all cases, where the axes of the unit cell have no associated symmetry operator, it is common for the axes to be removed from the spacegroup name and for the name to abbreviated to the axes where there is a symmetry operation, $P2_1$ in the rfhSP-D example.

2.3.6 Disorder and twinning

So far, the ideal crystalline array has been introduced, where all of the unit cells align along symmetry elements and form a consistent crystal lattice [Figure 2.3a]. However, protein crystals have a range of problems with the lattice that directly affect data collection. For simplicity, two of the difficulties experienced during this body of work are described here; disorder and twinning. Disorder is caused by the proteins not packing together in a repeating, symmetry related lattice, preventing the unit cell from fulfilling the basic requirement of a repeating environment [Figure 2.3c]. This results in poor diffraction as the requirement that the Bragg planes *h*, *k*, *l* within different unit cells throughout the crystal are aligned with each other is not met, which will limit the overall resolution of a dataset or, in the worst cases, prevents a crystal from diffracting.

Twinning refers to a crystal anomaly where two or more parts of the crystal have grown in two distinct orientations often related by a symmetry operator that is not part of the crystal point group. True twinning can only form in spacegroups where the two unit cell orientations can occupy the same three-dimensional space [Figure 2.3b]. In protein crystals, twinning is common because the protein molecules in the asymmetric unit have the potential to make more than one set of crystal contacts as they are arranged in the crystal, which can then nucleate a new domain based on the new crystal contacts. This is called pseudo-merohedral twinning. A very similar effect is also observed in merohedral twinning however the crystal domains are smaller and more dispersed through the crystal. In pseudo-merohedral and merohedral twins, the diffraction intensities from the two

domains (which diffract independently, producing their own diffraction patterns) form an average of the two diffraction patterns with no strong or weak intensities. As a result, the diffraction pattern is difficult to interpret.

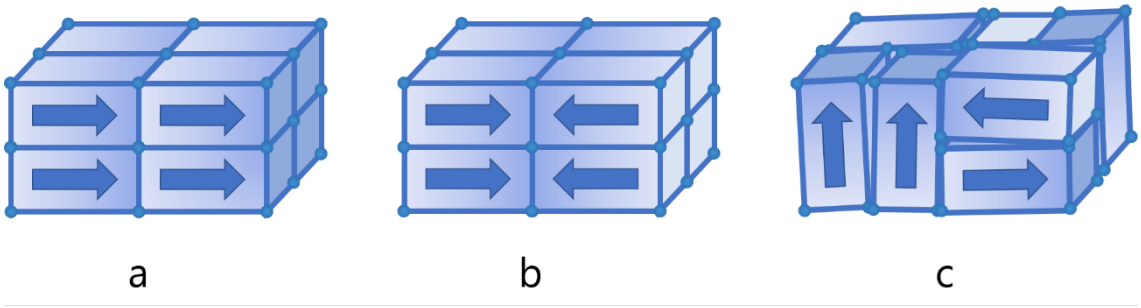


Figure 2.3: Examples of crystal lattice pathologies. An overall approximation of crystal lattice pathologies common in protein crystallography. (a) shows a well-ordered protein crystal with all unit cells sharing an identical environment; (b) shows an example of a twinned lattice, where two differently oriented unit cells can occupy the same 3D space; and (c) shows a disordered crystal, with each unit cell unaligned and orientated differently. In all cases, the abc axes are approximately in the same orientation with the ac face in plane with the paper.

2.4 X-ray diffraction methods

2.4.1 Electromagnetic waves and the X-ray

Electromagnetic waves are a spectrum of sinusoidal waves that interact and disrupt both electronic and magnetic environments and carry a fixed amount of energy. Sinusoidal waves oscillate through 360° (phase, ϕ) to complete a whole wavelength (λ), extending to a specific amplitude at the peak of the oscillation. X-rays are a short wavelength electromagnetic waves with a wavelength (λ) between $1x10^{-11}$m and $1x10^{-8}$m, or 0.1 to 100Å. This makes them useful for structure determination because the distance between two atoms ranges from 1.2Å to 1.5Å, allowing a wavelength of approximately 1.0Å to provide resolution of the atoms and therefore their positions in three dimensional space.

The amount of energy a wave carries depends on the amplitude of the wave, with larger amplitudes delivering more energy to any material the waves are incident on. The amount of energy delivered is the intensity of a beam. The incidence of an electromagnetic wave with electrons surrounding the atoms of a protein causes the electrons to absorb the energy from the electromagnetic wave, promoting the electrons to a higher energy orbital around the atom. However, the electrons are unable to maintain the excited state and rapidly return to their original energy state, emitting a photon in a random direction that has an equal energy and wavelength to the incident beam. The emitted photon is the scattered X-ray.

As with other electromagnetic waves, the emitted X-rays are able to have a summative effect on the overall emitted wave, exclusively affecting the amplitude of the waves without changing the overall wavelength of the outgoing wave. Any summation of waves is interference and can have three effects: (1) Constructive interference, whereby the waves are in phase and have an additive effect on the amplitude [Figure 2.4a]; (2) Incomplete

interference, where the waves are less than 180° out of phase causing the overall wave to shift along the wave-path (change of phase) as a function of the two (or more) waves and the amplitudes of the waves to be higher than either original wave but not increase in an additive fashion [Figure 2.4b] and; (3) Destructive interferences, where the two waves are 180° out of phase and cancel each other out, losing all amplitude [Figure 2.4c]. In crystallography, the effects of summation are evident from diffraction as each reflection is a sum of all the diffracted waves in that direction, leading to spots of different intensities.

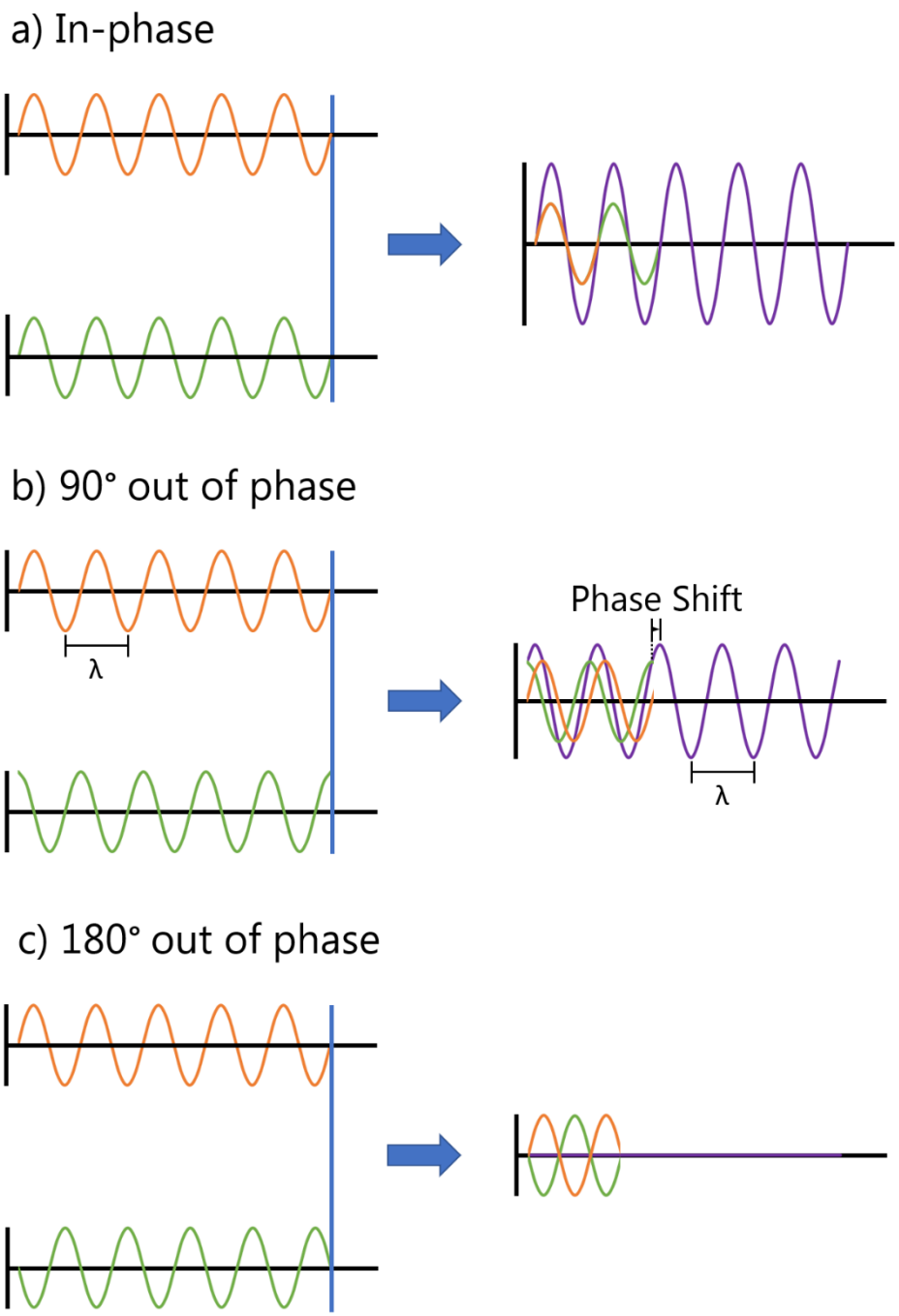


Figure 2.4: Summation of sinuosoidal waves and the resulting interference. Each figure shows two incoming waves (orange and green) and the resulting sum of those two waves (purple). Figure (a) shows constructive interference; (b) shows two sinusoidal waves 90° out of phase resulting in incomplete interference; and (c) shows two sinusoidal waves 180° out of phase resulting in destructive interference.

2.4.2 X-ray diffraction and Bragg's Law

Bragg's law (1) describes the conditions under which a crystalline material produces a reflected X-ray, called a Bragg's reflection. It ascertains that a plane of atoms is able to act as a mirror, reflecting an X-ray at the same angle as the incident beam meets the atomic plane. Therefore, if two monochromatic X-rays that are in the same phase interact with two atomic planes, with the correct distance (d) between them, the difference in path length ($2d$ sinθ) will be a whole number of wavelengths and n will an integer. If this is true, then the outgoing X-ray beams remain in phase and crystal will successfully diffract [Figure 2.5].

$$n\lambda = 2d\ sin\theta \quad (1)$$

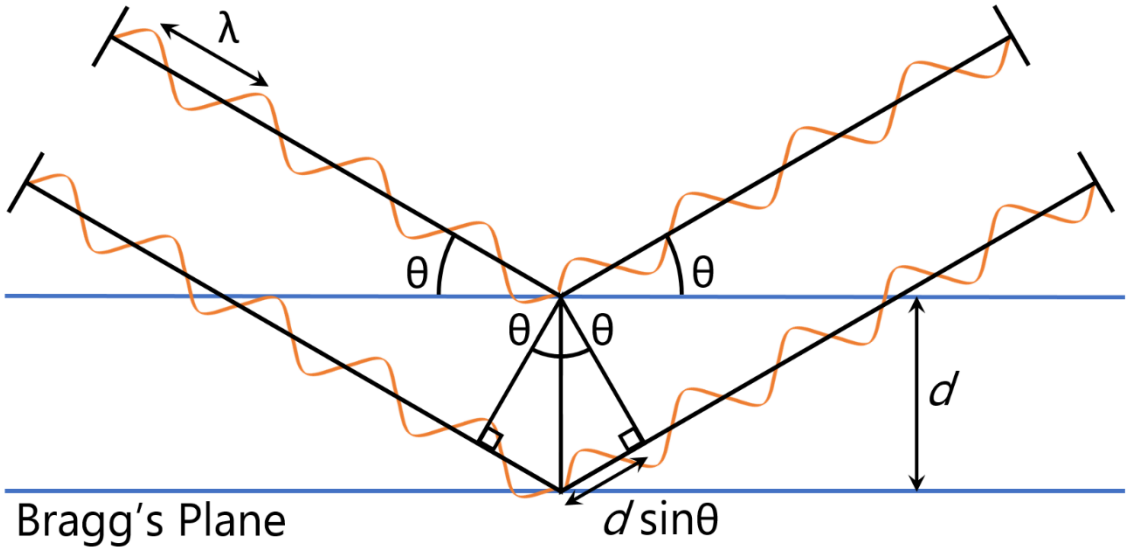


Figure 2.5: Proof of Bragg's Law. The figure demonstrates that for the two outgoing X-ray beams to be in phase, following reflection, the distance travelled by the second beam (d sinθ + d sinθ) must be a whole number (n) of wavelengths (λ), where θ is the angle of incidence.

A complete set of diffraction data comprises the diffraction maxima from each and every set of planes h, k, l in the crystal which, therefore, requires rotating so that all the planes are illuminated and allowed to diffract. Rotation changes the angle of incidence with the planes, resulting in a change in *d* and the set of miller indices *h, k, l* diffracting, allowing for a complete set of data to be collected. However, because the X-rays illuminate every atom in the crystal, the outgoing X-rays are a summation of all the diffracting atoms, both in the Bragg planes and between them, resulting in the information about the individual intensities and, most importantly, the phase of the outgoing X-ray being lost. Instead, the intensity provides some information about the average position of every atom associated with that reflection but without the phase, the individual positions cannot be determined. The lack of phase information is described as The Phase Problem and is biggest obstacle in solving a crystal structure (discussed further in section 2.4.3).

2.4.3 Importance of the crystal lattice at high resolution

The intensity of the reflections is also directly affected by the crystal lattice and the robustness of that lattice, particularly for high resolution data. For perfect diffraction, all of the atoms in the crystal would be completely stationary and in identical positions throughout the crystal to maximise the summative intensity of the diffracted X-ray. However, in a real crystal this can never be achieved because there is a fundamental amount of thermal energy in the highly conserved atomic positions and flexible regions in the protein that will not be in exactly the same position throughout the crystal, which will affect each and every diffraction spot and hence the information that can be extracted from them. In practice, the inherent flexibility of macromolecules limits the resolution of the data by compounding the natural drop off in scattering factor that results from increasing the scattering angle at high resolution. This is observed as a sharp drop, or edge, in a diffraction pattern beyond which there are no detectable spots.

The lattice also becomes increasing important at resolutions better than 2.0Å (where there are larger diffraction angles, θ) because there are more observations being made as a result of the large numbers of closely spaced Bragg planes in the crystal that meet Bragg's law and successfully diffract. As the interplanar spacing, *d*, falls, the angle of incidence must increase proportionally for successful reflection of an in-phase X-ray according to Bragg's law. The effect of this is to generate more microenvironments within each unit cell, as the miller indices increase, causing the position of equivalent atoms to have a larger impact on the overall crystal diffraction as very small differences in atomic position represent a large fraction of *d* at high diffraction angles.

2.4.4 Structure factors, Fourier transforms and The Phase Problem

The crystal lattice allows each reflection collected from a single crystal to be treated as an average of a near-infinite number of atoms arranged on and between a set of Bragg planes, with information about the sum of the individual scattered intensities. For the intensity of each diffracted spot, there is also an associated phase, which is lost during data collection, that is taken in combination with the intensity to describe the electron density around each atom in a unit cell. These complex numbers are the structure factors and can be expressed as a Fourier transform. To translate the structure factors back into real space (3D space), a second Fourier transform, an inverse Fourier transform, can be used to calculate the electron density map using the amplitude and phase for each atom.

The problem is, without knowing the phase for each reflection, the inverse Fourier Transform which requires intensity and phase, cannot be calculated; this is The Phase Problem. For small molecule crystals, the phase can be solved by direct methods based on atomic properties and probability, however for larger structures like protein crystals alternative methods must be used using either known properties of heavy atoms that can be soaked into the crystals, providing a reference phase, or molecular replacement using a

previously solved structure that is comparable to the protein in question, using the phase calculated from the known structure to solve the new protein structure. Molecular replacement by rigid-body refinement was used throughout this set of experiments.

2.4.5 Data collection techniques

The phase problem is created by the data collection methods necessary in crystallography. Unlike other microscopy techniques that use visible light or electrons, X-rays cannot be refocussed to generate a coherent image of the object and, as a result, the data must be collected from the diffracted X-rays and the image calculated computationally, using Fourier Transformations [Figure 2.6]. To make this possible, a range of methods and detectors have been developed to gather the diffraction data. Originally, data collection was collected using photo-reactive films that would permanently change colour when exposed to X-rays. However, this technique has been superseded by photon counting detectors such as charge-coupled detectors (CCDs) and silicon pixel detectors (DECTRIS Pilatus series) that measure the electrical changes associated with an incident X-ray. This has led to a huge reduction in overloads from the high intensity diffraction and maximised the amplification of the weak diffraction, so they can be measured above background. The improvements in detection hardware have also massively reduced the time taken to collect a dataset; from many hours for a single dataset to a few minutes with a live read out of the data as it is collected.

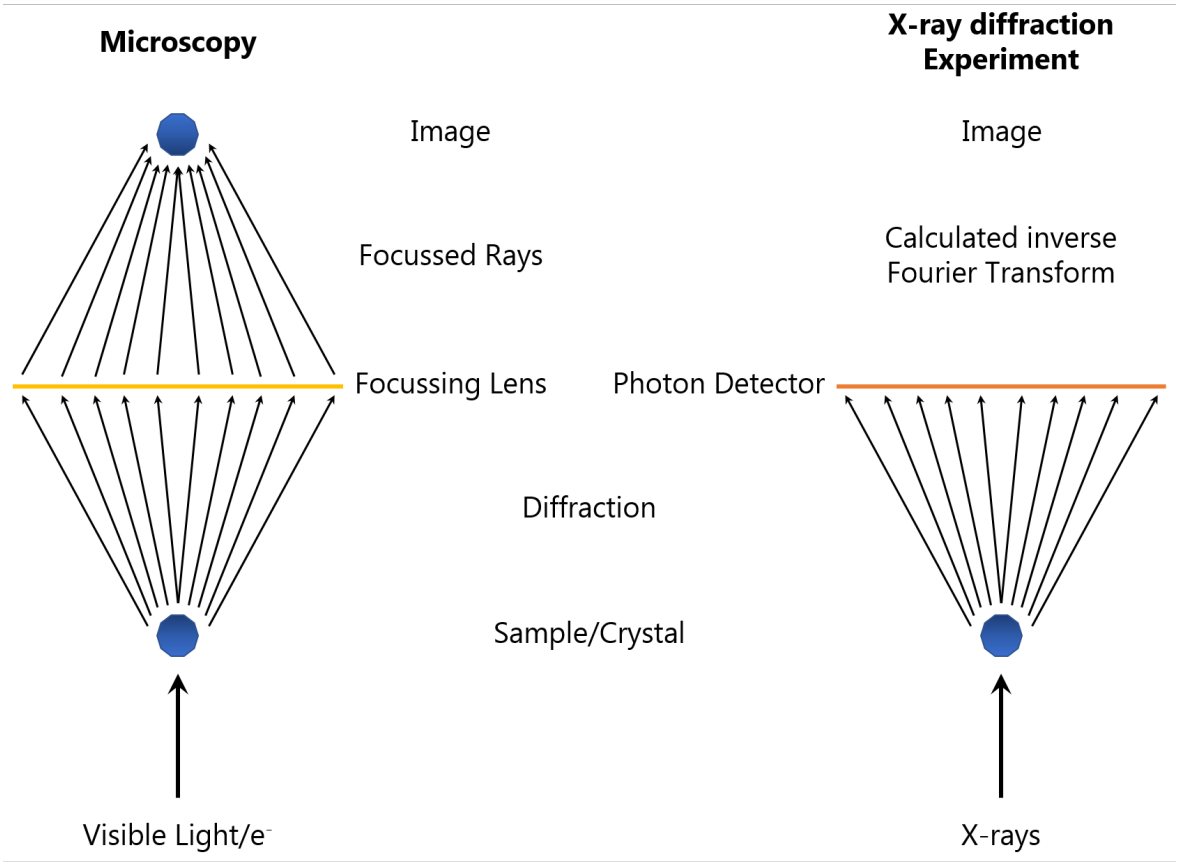


Figure 2.6: Comparison of microscopy and X-ray diffraction experiments. The figure compares the X-ray diffraction experimental set-ups to traditional and electron microscopy set ups. In microscopy, the diffracted rays/electrons (e-) are collected by a refractive lens that refocuses the rays into an image of the sample. In X-ray diffraction, it is currently impossible to refocus X-rays sufficiently to produce the image, so the inverse Fourier Transform is calculated to find the image from the diffraction pattern.

All of the data collected during the experimented presented in this work were collected using DECTRIS Pilatus 6M detectors at the synchrotron radiation source, *Diamond Light Source*, in Oxfordshire. Synchrotron radiation is beneficial to protein crystallography because of the high brilliance of the X-rays it produces and the consistent wavelengths that they are able to produce, by controlling the high-energy electrons very accurately; allowing for variables in the incident X-rays to be minimised. This, in turn means the data collected from the synchrotron beamlines are reliable at a wavelength of approximately 1Å, which is shorter than the interatomic distance of two carbon atoms (approximately 1.54Å). This means that at wavelengths of 1Å, data can be collected at subatomic resolution as a shorter wavelength means the distance between Bragg planes can be as small as 1Å, creating the conditions for the highest possible resolution. It is important for the detector to be at the correct distance from the crystal to maximise the separation between diffraction spots but also maximising the amount of data collected from the high-resolution diffraction shells.

In all of the experiments here, the crystals were also rotated by 0.1 – 0.2°, on a goniometer, between each data capture to ensure all of the Bragg planes were illuminated throughout the crystal, to ensure that a full set of reflections *hkl* were collected so that a full three-dimensional reconstruction of the asymmetric unit can be successfully made. The total oscillation of each crystal is determined by the symmetry of the spacegroup, with higher order symmetry requiring a smaller amount of rotation. For all of the data collections completed, a minimum of 180° were collected, enough for all spacegroups to be fully collected, but in many cases additional degrees were collected to ensure high completeness in the final dataset.

2.4.6 Radiation damage and the benefits of cryocooling

The strength of the high-resolution reflections can be improved by longer exposure times, effectively increasing the amount of energy delivered to the electron shells, increasing the chance of photon emission, however larger radiation doses increase the risk of radiation damage. Primary radiation damage occurs in a dose-dependent manner because some of the X-ray energy, carried by the photons, is absorbed by the atoms in the crystal, causing the atoms to lose electrons from their higher energy orbitals, often breaking covalent bonds and producing ionised radicals [Murray, Garman and Ravelli, 2004; Ravelli and Garman, 2006].

The ionised radical species then become mobile, causing secondary radiation damage to charged or highly energetic side chains in a time and thermal energy dependent manner. This is because the reactive species are able to move through the solvent channels, affecting other protein molecules in the crystal and exponentially increasing the number of reactive species in the crystal. Eventually this leads to degradation of the crystal and eventual collapse of the crystal lattice as every protein molecule is damaged beyond recognition, described as the Garman limit [Owen, Rudino-Pinera and Garman, 2006].

Primary radiation damage can be minimised by reducing the dose of radiation to the crystal, however this then limits the potential resolution of the data because a smaller dose prevents the high-resolution diffraction events from summing to an appropriate level, making this impractical for protein structure determination; particularly in ligand recognition studies where high resolution definition of bound ligands and their interaction with the protein is the goal. Secondary radiation damage, on the other hand, can be much more easily limited. As the damage is both time and temperature dependent, if the time taken for the dose to be delivered to the crystal can be reduced, and that dose can be delivered at low temperatures, the effects of secondary radiation damage can be

minimised. Synchrotrons have now made it possible to reduce data collection to a number of minutes by increasing the brightness of the X-ray beam, reducing the time that secondary radiation can have its effects. This leaves temperature as the final factor.

At room temperature, the reactive species are able to rapidly penetrate the crystal because of the thermal energy in the species itself and the surrounding solvent [Figure 2.7]. However, if the temperature can be reduced to absolute zero, there is no thermal energy present in the crystal system and the reactive species are unable to move from their initial positions. To achieve absolute zero, though, is impossible so the crystals are, instead, cooled to 100K in liquid nitrogen [Owen, Rudino-Pinera and Garman, 2006]. Cooling in liquid nitrogen substantially reduces the progression of the reactive species, localising it to the worst affected outer part of the crystal, whilst maintaining an unaffected core in the crystal. With modern synchrotron radiation, it is common to still see evidence of radiation damage in the electron density because of the intensity of the beam, however cryocooling has proven successful in limiting the amount of secondary radiation damage.

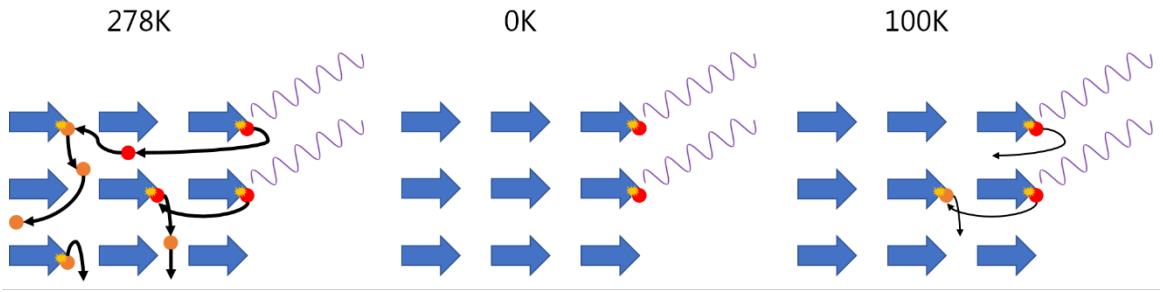


Figure 2.7: Radiation damage schematic. Figure shows three identical crystal systems (blue arrows) exposed to X-ray radiation (purple) causing primary radiation damage (yellow) and releasing reactive ionised species (red) that go onto create secondary events (orange). The speed of the diffusion is represented by the thickness of the directional arrows (black arrows), with thicker arrows corresponding to higher thermal energy.

Chapter 3: General materials and methods

3.1 Introduction

Much of the work completed used the same set of techniques and, therefore, those techniques can be generalised to ease understanding. This chapter describes the general methods used throughout the experimental work presented in this thesis. The fine details of each experiment are presented separately in their respective chapters which provide full details of the work carried out during the individual experiments.

3.2 Protein preparation

The recombinant fragment of human surfactant protein D was expressed, extracted and purified by Howard W. Clark, Rose-Marie Mackay and Jens Madsen and kindly provided for use in the structural studies. The methods used for preparing the protein are briefly described below.

3.2.1 Protein expression, extraction and purification

The recombinant fragment of human surfactant protein D, used throughout the ligand studies, contains a homotrimer consisting of 177 residues at the C-terminal of the native human SP-D, beginning at Gly179 [Shrive *et al.*, 2003]. This fragment, rfhSP-D, includes eight Gly-Xaa-Yaa repeats of the collagen domain (residues 179 – 202) at the N-terminal of the fragment, the α-coiled coil neck region (residues 203 – 235) and the carbohydrate recognition domain (residues 236 – 355). The protein sequence for the fragment is consistent with the native amino acid sequence of SP-D, with the exception of a substitution at residue 180 where proline is replaced with serine.

The fragment was overexpressed in *Escherichia coli* using a bacteriophage T7 expression system, expressed in the inclusion bodies [Madan *et al.*, 2001]. Denaturation and

renaturation allowed for the extraction of the correctly folded rfhSP-D. The protein was subsequently purified using ion-exchange, affinity and gel-filtration chromatography. The purity of the protein was confirmed by SDS-PAGE, immunoblotting and amino-terminal sequencing. Once the purified protein was received, it was buffer exchanged to remove the EDTA where necessary, then concentrated to a suitable concentration for crystallisation. Buffer exchange and concentration was completed by Ian Burns.

3.3 Crystallisation and cryoprotection

Following receipt and preparation of the recombinant fragment, crystallisation trials were carried out around the conditions that have previously grown diffracting crystals at high resolution [Shrive *et al.*, 2003]. The individual well conditions are given in the relevant chapters.

3.3.1 Sitting drop vapour diffusion method

In all of the studies in this thesis, the crystals were grown in 24-well linbro crystallisation trays (Jena Biosciences) using the sitting drop method (discussed in section 2.2.1). The reservoirs were prepared such that the required well conditions were achieved in a final volume of 1.0ml using 0.2µM filtered solutions. 2µl of the reservoir was then taken and added to the microbridge in a 1:1 ratio with the protein. The crystals were then stored at ambient temperature and monitored for crystal growth using light microscopy.

The sitting drop method was used because the materials are readily available in the laboratory where the work was completed, it allows for easy manual manipulation of the conditions as the wells are accessible and is historically the method preferred in the research group. Importantly, it allows for effective cryoprotection of the crystals *in situ*, removing the need to unnecessarily move the crystals between wells to facilitate effective

protection for cryocooling. This ensures the crystals are in the best condition for data collection as only one mounting is required.

3.3.2 Cryoprotection of crystals

The formation of ice during the cooling of protein crystals can be detrimental to both data quality and, in the worst case, can lead to the crystals fracturing or shattering as they are exposed to the cryo-temperatures of liquid nitrogen. However, cryocooling to around 100K is very beneficial for data collection as it reduces the degradation of the crystals by radiation damage and allows for easier transportation to the beamline. As a result, it is necessary to replace a proportion of the water in the solvent channels with a higher molecular weight solvent that does not form crystals at low temperatures.

Common cryoprotectants are short, hydrophilic molecules that can access the solvent channels in the crystal, replacing some of the water in the crystal and making hydrogen bonds to the remaining water molecules and protein in the crystal. This is important to avoid disrupting the matrix of interactions between the protein molecules that allow the protein to crystallise. Disruption of these interactions causes the crystals to dissolve. As a result, choosing the correct cryoprotectant is essential to maintain crystallinity and effectively protect the crystal.

In all the datasets presented in the following chapters, the cryoprotectant 2-methyl-2,4-pentandiol (MPD) was used. MPD was selected because it has previously been shown to effectively cryoprotect crystals of the rfhSP-D fragment used in these studies and has been shown on several occasions to allow the collection of high quality, high resolution data without affecting the crystal.

The cryoprotectant buffer was added to the well in 2µl aliquots, beginning with a 5% v/v MPD diluted in the conditions that the crystals were grown in to ensure there was no

chemical shock to the crystal system. Additional aliquots of 2µl were added in increasing increments of 5% v/v until a final concentration of 20% v/v MPD was achieved in the well. A final exchange was completed prior to mounting and cooling, where a volume equal to that added (usually 8µl) is removed from the well and replaced with 20% v/v MPD to ensure a final concentration approaching 20% v/v MPD. Throughout the additions, the crystals were monitored for any changes to the edges and faces.

3.3.3 Ligand soaking (general case)

To investigate the interactions of surfactant protein D with a range of ligands, crystals of the native recombinant fragment of human SP-D, described previously, were grown and the ligands of interest were subsequently soaked into the best crystals that formed. For all of the ligands used in these studies, the ligand additions were completed in the following way.

Soaking of the native protein crystals was completed during cryoprotection through additions of the ligand of interest to the cryobuffers used to protect the crystals, described in section 3.3.2. The ligand was added to the cryobuffer, at the concentrations described in each chapter, so that there was an excess of the relevant ligand compared to the number of rfhSP-D ligand-binding sites present in the well. The excess in all of the studies was more than 10-fold, based on the estimated concentration of the ligand and the approximate number of ligand-binding sites, following cryoprotectant exchange. This excess ensured that the majority of binding sites where binding was possible would be occupied in the crystal, thus ensuring that the electron density for the ligand would be strong enough for ligand fitting and determination of the binding mechanism.

The ligand was added to each aliquot of cryobuffer from the lowest concentration (5% v/v) to the highest (20% v/v) to achieve the desired ligand excess in each cryobuffer. This

allows for a gradual addition of the ligand to the well, ensuring the ligand is in a higher concentration in the well than the crystal, driving the diffusion gradient into the crystals. Gradually adding the ligand also minimises the risk of shocking the crystal with a high concentration of ligand, maintaining crystal quality. The overall time for the soak (and cryoprotection) is recorded in each chapter.

3.4 Data collection, processing and refinement

3.4.1 Data collection

Diffraction data was collected at *Diamond Light Source, Harwell Science & Innovation Campus, Didcot, Oxfordshire* on the macromolecular beamlines: I03, I04 and I04-1. All of the data was collected using the Dectris Pilatus 6M-F detectors, installed at the beamlines, which use hybrid photon counting technology to record the diffraction data [Henrich *et al.*, 2009]. This allowed for the rapid collection of a large number of diffraction images, ensuring complete datasets were collected. Specific details of the wavelength used to collect data and the number of images collected in each case are described in each chapter.

3.4.2 Data processing

Processing of the diffraction data was completed using a number of programs in the CCP4 suite [Winn *et al.*, 2011] to generate a set of structure factors that can be used to solve the structure of the ligand-bound protein structures The programs, and how they are used, are briefly described in this section.

MOSFLM, accessed and controlled using the *iMOSFLM* GUI, was used to initially index the unit cell of the protein crystal from two separate diffraction images, separated by 90°, to ensure a good estimate of the cell parameters can be determined [Battye *et al.*, 2011]. The Laue group was also determined at this stage, allowing for the data collection strategy to be optimised to maximise the completeness of the dataset when it is finally integrated. The

data presented in this thesis was all optimised to deliver a completeness of more than 90% where possible, with most datasets being processed to a completeness of close to 100% in strategy optimisation. Before integration and measurement of the individual intensities of the images, a final refinement of the cell parameters was completed using partial integration of two segments of data that are 90° apart [Battye *et al.*, 2011]. This allows for refinement of the cell parameters, but also improves the estimation of mosaicity and crystal orientation which is used to inform the predictions of the spot positions during integration. It was important at this stage to ensure that the spot predictions generated following cell refinement were correctly overlaying the diffraction spots in the images, to ensure that integration would succeed.

The final stage of *MOSFLM* is integration of the diffraction images [Leslie, 2006; Battye *et al.*, 2011]. In each image, the pixel density of each spot was measured and compared to the automatically calculated background scattering on the image. The pixel density was then used to integrate the images and derive the intensities for each diffraction spot. *MOSFLM* completes the integration in batches of ten images at a time, allowing for an average spot profile to be determined, so that quality of the data can monitored throughout the integration. Importantly, as part of the monitoring, each image undergoes refinement of the crystal position and mosaicity to ensure that there are not problems with the data collection but also to allow for a better summation of the spots which are partially recorded on multiple images.

Following integration of the images, the resulting intensity files were then analysed using the *POINTLESS-AIMLESS* pipeline [Evans, 2011]. *POINTLESS* uses the integrated intensity data to test the Laue group and spacegroup produced from *MOSFLM*, analysing the correlation coefficient between all related intensity pairs and potential symmetry elements [Evans, 2006, 2011]. The pairs are then grouped based on unrelated symmetry elements

and compared to each other, generating a *Z*-score. The scores can then be ranked, favouring the lowest symmetry, to determine the Laue and space group. For all of the crystals presented here, the space group was confirmed as $P2_1$ by *POINTLESS* and the data were automatically ran through *AIMLESS* for scaling and merging. *AIMLESS* utilises the scaling and merging algorithms first written for *SCALA*, but has additional parameters for assessing the data quality [Evans, 2006, 2011; Evans and Murshudov, 2013]. Briefly, *AIMLESS* scales each intensity to minimise error introduced by data collection, increasing internal consistency by minimising differences in intensities. This is completed by generating a common scale for all of the intensities based on a scaling model. The merging algorithms, as the name suggests, merges the partial intensities generated by the rapid data collection, into a single intensity.

In most cases, a number of *AIMLESS* runs were completed to ascertain what resolution the data was limited to, to maximise the data quality without losing too much high-resolution data. In all cases, the data were assessed for quality based on values for R_{merge}; I/σ(I); completeness; and the correlation coefficient (CC½). R_{merge} rationalises the convergence of all of the observations for a single unique reflection into a single *hkl*, with a lower R_{merge} representing better agreement between the observations. However, R_{merge} is dependent on the multiplicity, the number of repeated observations of the same reflection, of the data. I/σ(I) is an estimation of signal to noise in the data, with a higher I/σ(I) indicating that the data has a good signal and, therefore, is more accurate and more likely to be a genuine reflection. For the purposes of using I/σ(I) as a measure of data quality, a cut off of between 2.0 and 3.0 was used in selecting the resolution cut off. Completeness of the data is calculated by comparing the observed data to the calculated data and determining how many of the calculated reflections are present in the observed data. The final data quality test used to assess the data was the correlation coefficient (CC½). CC½ is calculated by

dividing the unmerged data for each unique reflection into two random halves, averaging the intensities from each subset and comparing the two values, where a unique reflection is the product of averaging multiple observations of the same reflection and its symmetry related reflections. This removes any bias introduced by multiplicity. CC½ is expressed as a fraction and, therefore, a value closer to one represents better agreement between the two subsets for each unique reflection and, therefore, across the whole dataset.

The data was divided into resolution ranges (bins) and statistics based on the four quality measures were calculated for the overall datasets and the highest resolution bin. The statistics from the highest resolution bin were used to determine the resolution of the data so that the structure can be reliably solved at high resolution. The data was interrogated and resolution bins where the R_{merge} was higher than 10% of the resolution were removed during the scaling and merging to improve the statistics. However, bins were not removed if they were detrimental to the overall completeness of the data.

Once merged, the intensities were converted to structure factors (amplitudes) using *CTRUNCATE* [French and Wilson, 1978]. The algorithm allows for the calculation of all the amplitudes from both positive and negative intensities, which can then be used to assess the data for any problems that can only be identified in merged data. Using the cumulative frequency distribution, Wilson plot and L-test, the data can be tested for any crystallographic twinning. If any twinning was detected, the dataset was not taken any further as untwinned crystals are readily available.

Following *CTRUNCATE*, the structure factors are ready to be used for structure solution. However, a 5% test set, comprising a random 5% of the data which do not contribute to subsequent refinements, needs to be selected to prevent the structure model from becoming biased during refinement. *Uniqueify* was used to randomly select the 5% test set

from the structure factors and record them in a separate column in the structure factor file, using *FreeR_Flag* to select and assign the flag [Winn *et al.*, 2011]. This test set will be refined separately the main dataset to ensure internal consistency.

The final stage of data processing is sorting the structure factor file into columns that can be read by the downstream programs used for structure refinement and validation. *SortMTZ* was used for all of the data presented here [Winn *et al.*, 2011]. It is worth noting that sorting the file is not strictly necessary, but it may prevent problems in the future.

3.4.3 Structure solution, refinement and validation

All the structures were solved using the high-resolution, native structure of the same recombinant fragment of human surfactant protein D, solved at 1.60Å [Shriv*e et al.*, 2003]. This structure had all of the calcium ions and waters removed to avoid biasing the occupancy of the calcium binding sites (Ca1, Ca2, Ca3 and Ca4) and any ligand density that may be present in the electron density. Using a 100% homogenous protein structure for the replacement allows for a rigid-body refinement, a constrained coordinate refinement of a group of atoms that rotate and translate as a single "rigid" body, to be used to solve the structure, avoiding the time-consuming process of a more traditional molecular replacement method. The phases derived from the native structure were sufficient to solve the structures for the cellotriose-, isomaltotriose-, laminaritriose-, muramyl disaccharide-, and the *Salmonella enterica* Minnesota R7 oligosaccharide proteins.

Each of the ligands were fitted using single carbohydrate monomers from the monomer library, a library of known small molecules commonly associated with macromolecular crystallography, using the molecular model building program, *Coot* [Emsley and Cowtan, 2004]. In general, the glycosidic linkages were described in the monomer library, mon_lib.cif, that is built into *Coot* and the structure refinement program *REFMAC5*.

However, if the linkage was not defined in the monomer library, a restraint library was generated by *REFMAC5* during the first round of restrained refinement [Murshudov *et al.*, 2011]. The restraint library, also called a ligand dictionary, provides a comprehensive list of bond lengths, bond angles, torsion angles and linkages between monomers based on the optimal topology from small molecule structures. For all of the ligands, the ligand dictionary was manually updated to include a full description of the glycosidic linkages using data from the monomer library (mon_lib.cif). For the *S. enterica* sv. Minnesota R7 polysaccharide complex, one of the carbohydrates was not described by the monomer library and had to be generated using a combination of *ChemDraw®* (PerkinElmer Informatics) and *JLigand* to generate a library [Lebedev *et al.*, 2012].

Restrained refinement of each structure was completed using *REFMAC5*, a robust refinement tool provided as part of *CCP*4 [Murshudov *et al.*, 2011; Winn *et al.*, 2011]. Refinement aims to match the atomic model to the experimental data by making small adjustments to the model, whilst maintaining the correct geometry of the protein backbone, side chains and ligands. This is completed based on two components: (1) prior knowledge of protein, ligand and other monomer geometry (f_{geom}) and; (2) the experimental X-ray data (f_{xray}) [Murshudov *et al.*, 2011].

$$f_{total} = f_{geom} + w f_{xray} \quad (1)$$

The initial phases, provided by the starting model used in the rigid-body solution, are refined using maximum likelihood methods to improve the electron density for manual model building and minimise f_{total}. The phases are then combined with the experimental and calculated structure factors using the methods described by *Murshudov* and colleagues (2011).

f_{total} also requires the geometry component to be calculated from the chemical dictionary, available in the program files. The geometry for the protein and ligand must be maintained for the model to be correct. As a result, equation 1 includes a weighting factor, *w*, that can be set to ensure good geometry is maintained in the final stages of model building by reducing the weight of the X-ray component.

To monitor each refinement cycle, a set of calculated structure factors, based on the current model of the structure, are compared to the experimentally observed structure factors to generate reliability factors (R_{factor} and R_{free}). Both of the reliability factors are calculated using equation 2:

$$R = \frac{\Sigma|F_{obs} - F_{calc}|}{\Sigma|F_{obs}|} \quad (2)$$

By calculating *R* for the working set (R_{factor}), excluding the 5% test set added by *Uniqueify*, and *R* for the 5% test set (R_{free}) independently, the reliability of the model, and how well the model predicts the experimental data, can be monitored [Brunger, 1992]. Due to the independent nature of the two $R_{factors}$, they should both converge during refinement as the model should be supported by the experimental data however, due to the small set of data used for R_{free}, the R_{free} value tends to be higher than the R_{factor}.

Refinement was completed in rounds of ten cycles, followed by manual rebuilding of the model. In the final refinement rounds, the number of cycles was reduced to prevent over refinement, leading to the divergence of the two $R_{factors}$. In some cases, a round of refinement was completed using *phenix.refine* for comparison of the refinement statistics, with approximate agreement being confirmed in all the cases it was used [Afonine *et al.*, 2012].

Model building was completed using the macromolecular model building software, *Coot*, which is distributed with *CCP*4 [Emsley and Cowtan, 2004; Emsley *et al.*, 2010; Winn *et al.*, 2011]. Two weighted electron density maps equivalent to the observed map ($2F_o$-F_c where F_o are the observed structure factors and F_c are the calculated structure factors based on the current model), and the difference map (F_o-F_c), traditionally generated using a fast Fourier transform in the program *FFT*, are generated from the MTZ file directly. The real-space refinement function, where the position of atoms is refined based on electron density peaks, and manual manipulation of the protein backbone were then used to fit the atoms into the electron density, ensuring that any side chains that were not well defined by the density were modelled into one of the commonly observed conformations (rotamers). Where it was necessary, atoms were moved out of negative density in the difference map and moved into the available density, while ensuring that the geometry was correct. Similarly, the ligands were fitted by manual manipulation of the atomic positions and torsion angles followed by rigid-body fitting into the density. Real space refinement was not used for the ligands as the carbohydrate rings would regularly be distorted.

Structure validation was completed throughout refinement of the structure to minimise any erroneous model building from the outset. To do this, the *MolProbity* server, hosted by Duke University, was used to carry out geometric and clash analysis following the addition of riding hydrogens (hydrogens that complete atomic valency) to all of the atoms [Chen *et al.*, 2010]. The Ramachandran plot, which plots the backbone dihedral angles against energetically favourable regions, was also checked for outliers and any outliers were corrected before the next round of refinement [Hooft, Sander and Vriend, 1997]. When the final structure with good $R_{factors}$ and geometry was completed, a final validation using *MolProbity* was carried out.

Additional validation of the polysaccharide ligands was completed using the new validation tool, *Privateer*, distributed as part of *CCP4i2* [Agirre and Cowtan, 2015]. *Privateer* uses the Cremer-Pople algorithm to analyse the ϕ and θ, to check they are correct, and generates a total puckering amplitude value (*Q* value) for each monosaccharide, based on the energetic favourability of the ring pucker [Cremer and Pople, 1975]. In ligand binding studies like these, the ligands should be in the lowest energy state; $^{4}C_{1}$ chair conformation for the pyranoses and ^{O}E envelope conformation for the furanoses. A second stage of validation is also carried out using the published stereochemistry of monosaccharides in the wwPDB database and a real space correlation is calculated to ensure the electron density supports the anomeric form in the model. There is also additional validation of any glycosidic linkages found in the model structure which checks the geometry of the linkages to ensure they are accurate.

Chapter 4: Structural insights into recognition of small glucosyl ligands by a recombinant fragment of human surfactant protein D

4.1 Introduction to alternative glucose oligosaccharides

Since the discovery of surfactant protein D, there has been a wide variety of ligands identified that SP-D can recognise through its carbohydrate recognition domains. The mechanism for this recognition has previously been investigated for a range of small mono-, di- and trisaccharides which have identified a calcium ion as being the main mediator of binding [discussed in sections 1.5 and 1.6; Shrive et al. 2003; Crouch et al. 2006; Wang et al. 2008; Shrive et al. 2009]. Many of these ligands have focussed on either lipopolysaccharide-derived polysaccharides or α(1→4)-linked glucose chains such as maltose and maltotriose [Shrive *et al.*, 2003; Crouch, McDonald, *et al.*, 2006; Clark *et al.*, 2016]. However, there is a catalogue of natural ligands for SP-D that utilise different glycosidic linkages and the anomeric form of D-glucose, β-D-glucose.

4.1.1 Anomeric glucoses

Glucose has two major anomeric forms, α-D-glucose and β-D-glucose, related through the chiral position of the hydroxyl on carbon one [C1'OH, Figure 4.1; Rao *et al.*, 1998]. These result in two distinct, non-superimposable glucose molecules that share functionality but do not share the same geometry; giving rise to different glucose chains. Many of the naturally occurring β-D-glucose chains have been identified in plants, fungi and bacteria and form part of the recognition motifs that the immune system can utilise to differentiate between self and non-self. Importantly, the arrangement of the anomeric glucoses and their associated chains allow molecules such as SP-D to recognise and agglutinate potentially pathogenic species and prevent an infection taking hold [Fakih *et al.*, 2015].

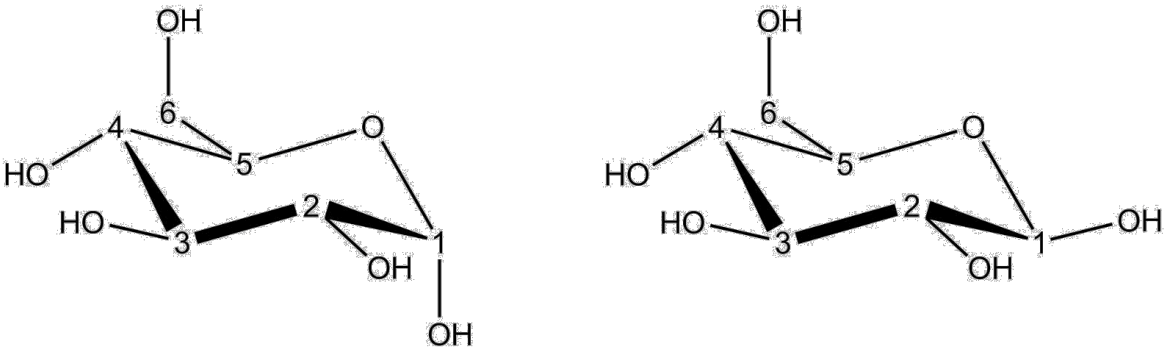


Figure 4.1: Anomeric forms of D-glucose with each carbon numbered. The alpha-anomer (left) has the hydroxyl group in an axial position and the beta-anomer (right) has the hydroxyl in the equatorial position. Carbon 5 is the anomeric reference in glucose

4.1.2 Alternative glycosidic linkages

The alternative anomers of glucose give rise to a variety of glucose chains that contain familiar glycosidic linkages, such as (1→4)-linkages, that form through the hydroxyl of the chiral C1 of glucose [Figure 4.1]. An α-linkage extends from the hydroxyl in an axial position, causing the second glucose to drop below the ring of the first glucose, and a β-linkage extends from the hydroxyl when it is in an equatorial position, creating a more planar bond to the next glucose [Rao *et al.*, 1998]. These alternate linkages can result in a considerable change in the overall topology of the oligosaccharide chain and the way that: (a) the organism interacts with its surroundings and; (b) how the immune system is able to interact with the carbohydrate array.

The α- and β-linkages are able to form between any of the hydroxyls on the adjacent glucose, allowing for further variations on the possible chain conformations. Most commonly found linkages include (1→4), (1→3) and (1→6) glycosidic linkages, usually as part of large, enzyme-mediated scaffolds such as starch or cellulose [Figure 4.2; Hehre, 1956; O'sullivan, 1997; Fujimoto *et al.*, 2009; Fesel and Zuccaro, 2016]. In many cases, the large glucose scaffolds form important structural components that are integral to the

survival of the organism, making them an ideal target for immune molecules during pathogen recognition and clearance.

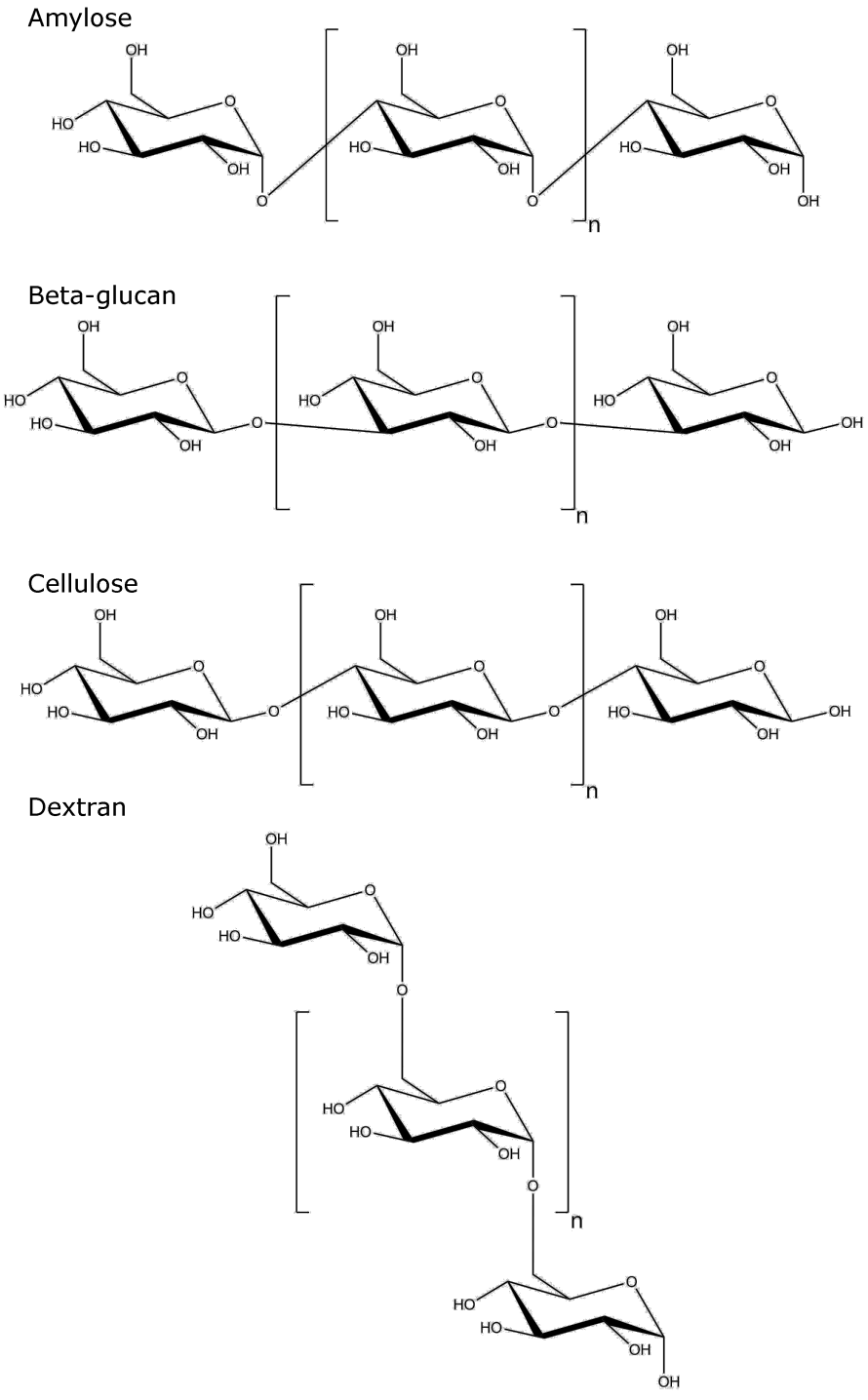


Figure 4.2: Structures of common carbohydrate structures on pathogens and other allergenic particles. Amylose is an α(1→4)-linked glucose polymer; cellulose is a β(1→4)-linked glucose polymer; dextran is an α(1→6)-linked glucose polymer; β-glucan is a β(1→3)-linked glucose polymer. Based on data from [Hehre, 1956; O'Sullivan, 1997; Rao *et al.*, 1998; Remaud-Simeon *et al.*, 2000; Fujimoto *et al.*, 2009]

4.1.3 Cellulose and β(1→4)-linked cellotriose

Cellulose is a large polymer of β-D-glucoses found in the cell walls of most plants, some algae and in the connective material of a number of bacterial biofilms, including species of *Salmonella* [Solomon *et al.*, 2005]. It has also been identified in the wall of pollen grains. The β-D-glucoses are linked together by β(1→4) glycosidic linkages in mostly linear chains of up to 15 000 β-D-glucoses, extending away from the surface of the membrane, forming large parallel chains arranged into microfibrils [Brown Jr and Montezinos, 1976; O'Sullivan, 1997; Brown Jr, 2004]. The microfibrils are then able to stack together, forming large hydrogen bond networks between the microfibrils, creating strong, insoluble complexes that are able to exclude water and other solutes.

Cellotriose is a small, β(1→4)-linked tri-glucose unit of cellulose that is produced during acid hydrolysis of naturally occurring cellulose I [Figure 4.3; Poletto, Ornaghi Júnior and Zattera, 2014]. This makes cellotriose a useful analogue for investigating the immune recognition of cellulose as it is more soluble than intact cellulose and is readily available from commercial sources.

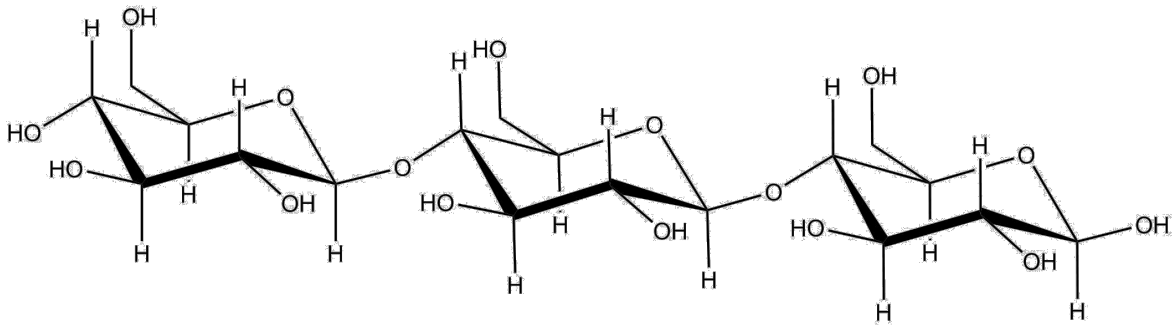


Figure 4.3: Structure of cellotriose. Three β-D-glucoses linked by β(1→4) glycosidic bonds.

4.1.4 Dextran and α(1→6)-linked isomaltotriose

Dextran is a large polymer of α-D-glucoses linked through a number of linkages depending on their biological source [Sloan *et al.*, 1954]. The main backbone is formed from a series of α(1→6) glycosidic linkages of variable lengths, with branches formed by either α(1→2), α(1→3) or α(1→4) linkages [Naessens *et al.*, 2005; Kasaai, 2012]. Branching occurs at different points depending on the species and strain of the bacteria producing a range of topologies. In most linear dextrans there are a small number of α(1→3)-linked branches to long α(1→6)-linked chains however there are examples of α(1→3) branches at regular intervals along the main backbone [Naessens *et al.*, 2005]. There are also a subset of dextrans, the alternans, that have an alternating backbone of α(1→6), α(1→ 3) linkages that are less common [Misaki *et al.*, 1980].

Isomaltotriose is an α(1→6)-linked tri-glucose unit analogous to the mostly linear dextran backbone [Figure 4.4]. A family of exocellular bacterial enzymes have been identified that hydrolyse dextrans into smaller isomaltose and isomaltotriose units [Jeanes *et al.*, 1953; Chludzinski, Germaine and Schachtele, 1974; Naessens *et al.*, 2005]. Isomaltotriose provides a useful insight into how the immune system is able to recognise these large, extracellular carbohydrate structures.

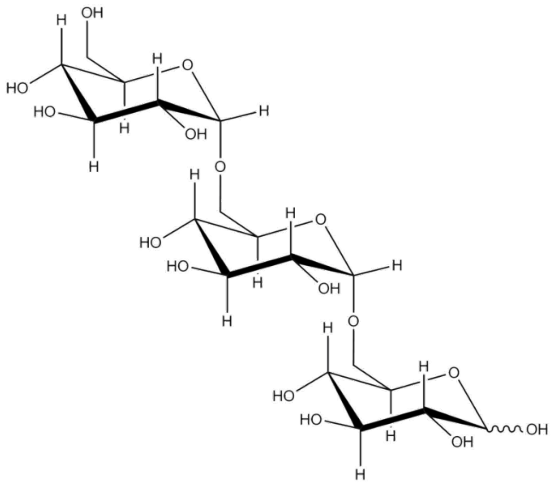


Figure 4.4: Structure of isomaltotriose. Three α-D-glucoses linked by α(1→6) glycosidic bonds.

4.1.5 β-glucans and β(1→3)-linked laminaritriose

β-glucans can be found in a huge range of organisms, including: euglenoid protozoa, certain species of algae, yeasts and fungi, higher order plants and, most importantly, potentially pathogenic bacteria that affect humans [McIntosh et al., 2004; Stone and Clarke, 1992]. The roles of these β-glucans range from storage polysaccharides in protozoa and algae to cell wall components in fungi and yeast. In higher order plants, a linear form of β-glucan, callose, forms an important part of the growth plates in the cell wall and is intrinsic to the cell wall of pollen grains [Ferguson *et al.*, 1998; Edlund, Swanson and Preuss, 2004; Schneider *et al.*, 2016]. A similar, largely linear form of β-glucan, curdlan, has been identified in the extracellular capsule of a number of bacteria that are medically important and has been implicated in the virulence of the bacteria in other animals [Nakanishi, 1976; Sutherland, 2001].

There are three main forms of β-glucan: a linear β(1→3)-linked polymer with minimal β(1→6) branches; a cyclic β(1→3)-, β(1→6)-linked β-D-glucose polymer; and a highly branched β(1→3)-, β(1→2)-linked form [Rolin et al., 1992; Kasai and Harada, 1980; Knecht et al., 1970]. Linear β-glucans form large, insoluble structures of up to 12 000 glucose units that crosslink via β(1→6) linkages between the chains to form an extracellular capsule and contribute to the formation of biofilms [Sutherland, 2001; Saito et al., 1968; Futatsuyama et al. 1999]. Cyclic β-glucans occur more commonly in symbiotic soil bacteria, forming large rings of β-D-glucoses of up to 15Å in diameter [Rolin et al., 1992; Miller et al., 1990]. In a number of species, the O6′ position of some of the β-D-glucoses are substituted with either phosphocholine or a laminaribiose molecules [Rolin et al., 1992; Bhagwat et al., 1999]. The highly branched forms of β-glucan have been identified in the capsules of a number of streptococcal and pneumococcal species and in some cases have a β(1→2)-linked β-D-

glucose substitution on the majority of β-D-glucoses in the β(1→3) backbone [Knecht et al., 1970; Henrichsen, 1995; Adeyeye et al., 1988].

Laminaritriose is a β(1→3)-linked tri-glucose unit of the linear backbone of β-glucan that can be produced by laminarinase, an endo-(1→3)-beta-glucanase originating from brown algae [Figure 4.5; Salyers et al., 1977]. This small, soluble unit of the β-glucans provides an opportunity to investigate the recognition of potentially pathogenic bacteria and fungi by surfactant protein D in the interstitial space of the alveoli. It also allows the effects of blocking the most well documented binding of carbohydrates by the O3′ and O4′ hydroxyls to be explored, as the β(1→3)-linkage prevents O3′ being recognised.

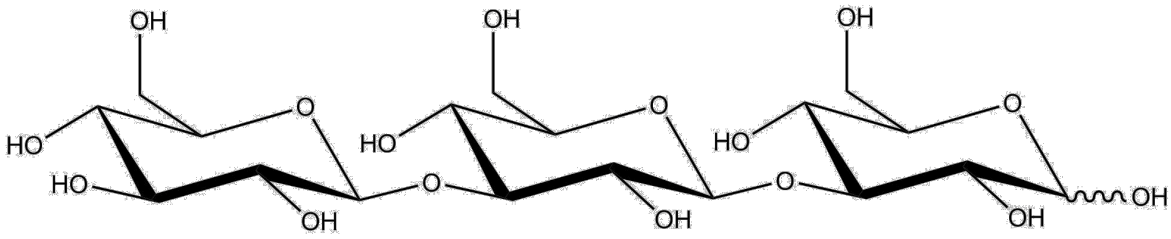


Figure 4.5: Structure of laminaritriose. Three β-D-glucoses linked by β(1 →3) glycosidic bonds.

4.1.6 Reducing and non-reducing ends of polysaccharide chains

The two terminal glucoses of a polysaccharide chain can be described as reducing and non-reducing based on their biochemical properties. The reducing terminal is characterised by the O1′ hydroxyl being available for further linkages or additions by other substituents that undergo nucleophilic attack. The non-reducing terminal does not have the O1′ hydroxyl available and, therefore, gives the polysaccharide directionality; as the chain can only be assembled in one direction [Figure 4.6].

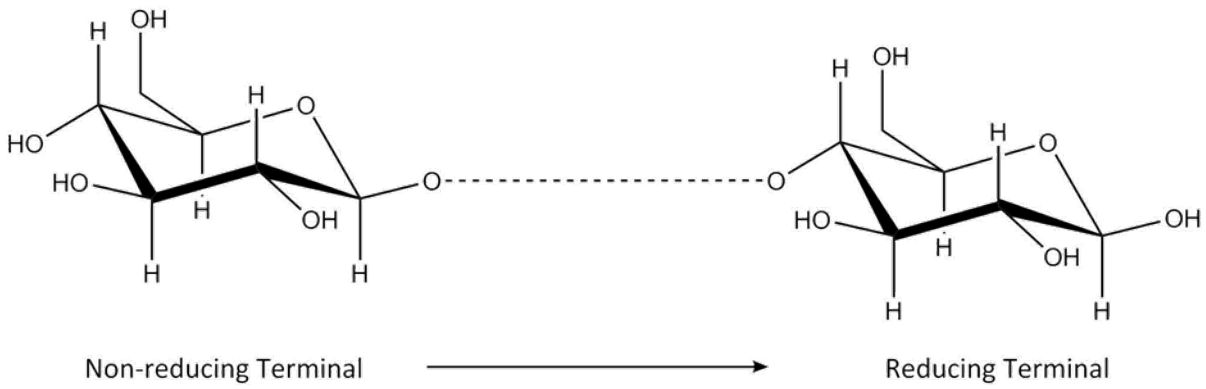


Figure 4.6: Identification of the non-reducing and reducing termini of oligosaccharides

Previous rfhSP-D structural studies using the glucosyl oligosaccharides maltose and maltotriose, two α(1→4)-linked oligosaccharides, revealed recognition by their non-reducing terminal glucose, allowing the reducing end to either project away from the binding site (maltose) or form a tentative π-interaction with Phe335 [Shrive *et al.*, 2003; Crouch, McDonald, *et al.*, 2006]. Using cellotriose, isomaltotriose and laminaritriose, the importance of the termini can be explored in more detail as the different linkages and β-D-glucoses affect the availability of O3′ and O4′ in the non-reducing residue.

4.1.7 Non-terminal recognition of glycosyl ligands by rfhSP-D

The potential for non-terminal recognition of small glucosyl ligand by surfactant protein D was extensively investigated using computational docking studies by Allen (2001), who identified a number of ligands that could be recognised through a non-terminal glucose [Allen *et al.*, 2001]. Of these ligands, isomaltotriose and the β-isoform, gentiotriose, are of particular interest as their longer polymers have been shown to effectively inhibit binding to *A. fumigatus* (IC_{50} values of 4.8±2.4μM and 1.0±0.3μM, respectively) [Allen, Voelker and Mason, 2001]. As a result, the experiments presented here can test the computational results crystallographically and provide direct evidence of any non-terminal binding in these triose ligands.

4.2 Materials and methods

4.2.1 Cellotriose soaking of native rfhSP-D

Cellotriose was soaked into native rfhSP-D crystals grown in a sitting drop as described in section 3.3.1. A final concentration of 20mM was achieved in the drop following an exchange of the drop volume (10µl) with 2µL of the cryoprotecting solution. Cryocooling to 100K was completed following cryoprotection with consecutive additions of 5%-20% 2-methyl-2,4-pentanediol (MPD) solutions, diluted with the well conditions to maintain the equilibrium that permitted crystal formation. The well conditions are detailed in table 4.1. Crystal cryoprotection, at the highest concentration of MPD, was completed for between 19 and 20 minutes, allowing sufficient time for the cellotriose and MPD to soak through the solvent channels and into the calcium binding sites.

Table 4.1: Table of well conditions for IB264C32 and IB264C33

IB264C32 and IB264C33	16% Polyethylene Glycol 6000 0.1M Tris pH 6.0 10mM $CaCl_2$

Data collection was carried out at Diamond Light Source, Oxfordshire on macromolecular crystallography beamline I02. A total of two high resolution datasets were collected between 1.4Å and 1.6Å from a total of two crystals. Processing of the diffraction data was completed according to section 3.4.2.

4.2.2 Isomaltotriose soaking of native rfhSP-D

Isomaltotriose was also soaked into native crystals of rfhSP-D, achieving a final concentration of 20mM in the drop by the same methods used in the cellotriose soaking. Cryoprotection was completed using 5%-20% MPD solutions, added in four aliquots of 2µL of increasing MPD concentration. The well conditions are detailed in table 4.2. Soaking of the ligand and MPD into the native crystals took between 12 and 19 minutes, allowing for

complete cryoprotection and time for the ligand to penetrate the solvent channels of the crystals.

Table 4.2: Table of well conditions for crystals IB254C12 and IB254C14

IB254C12 and IB254C14	16% Polyethylene Glycol 6000
	0.1M Tris pH 6.0
	10mM $CaCl_2$

Data collection was carried out at Diamond Light Source, Oxfordshire on macromolecular crystallography beamline I03 and I04-1. A total of two high resolution datasets were collected between 1.5Å and 1.7Å from two crystals. Processing of the diffraction data was completed according to section 3.4.2.

4.2.3 Laminaritriose soaking of native rfhSP-D

Laminaritriose was initially soaked into native rfhSP-D crystals in the same way as the previous ligands, as part of the cryoprotection to achieve a final concentration approaching 20mM, however the original crystals degraded and re-dissolved following the addition of the 15% MPD solution. Following monitoring of the cryoprotected well over the next two weeks, new crystals were identified in the drop and successfully cryocooled to 100K prior to data collection. The details of the well conditions prior to the initial laminaritriose soak, and the conditions that the new crystals later grew in, are detailed in table 4.3. As a result, the laminaritriose-bound crystals can be considered co-crystals of the rfhSP-D-laminaritriose complex.

Table 4.3: Table of well conditions for crystals IB319A53-A55

IB319A53 – A55	16% Polyethylene Glycol 6000
	0.1M Tris pH 7.0
	10mM $CaCl_2$
Following initial soak	15% 2-methyl-2,4-pentanediol
	20mM laminaritriose

Data collection was carried out at Diamond Light Source, Oxfordshire on macromolecular crystallography beamline I04-1. A total of three high resolution datasets were collected between 1.5Å and 1.66Å from three crystals. Processing of the diffraction data was completed according to section 3.4.2.

4.3 Results of small ligand binding

4.3.1 Data processing and data selection

Two high resolution datasets were collected from two separate cellotriose-soaked crystals, IB264C32 and IB264C33. The diffraction images from IB264C32 were indexed into the monoclinic primitive spacegroup $P2_1$, the spacegroup previously determined for the high resolution native structure of the same recombinant fragment of hSP-D used in these studies, with the final unit cell [Shrive *et al.*, 2003]:

$$a = 55.39\ b = 108.14\ c = 55.67 \qquad \alpha = \gamma = 90.0°\ \beta = 90.86°$$

The spacegroup was confirmed by Pointless as part of the Pointless-Aimless pipeline that was used to scale the integrated images generated by MOSFLM [Leslie, 2006; Battye *et al.*, 2011; Evans and Murshudov, 2013]. Scaling and merging of the data at a resolution of 1.59Å generated good overall R_{merge} statistics of 0.051 and 0.359 in the outershell. The overall I/σ and CC½ were 10.8 and 0.996, respectively, with the outershell being 2.9 and 0.846. Completeness was 99.5% overall and 99.3% in the outershell. Structure factors were generated by Truncate and twinning was assessed by cumulative frequency distribution and L-test. The data was not twinned.

IB264C33 was also indexed into monoclinic primitive $P2_1$, as was expected from previous structures. The final unit cell was refined in MOSFLM to:

$$a = 55.40\ b = 107.92\ c = 55.58 \qquad \alpha = \gamma = 90.0°\ \beta = 91.11°$$

Pointless confirmed the spacegroup and aimless scaled and merged the data. R_{merge} statistics were good at 0.061 overall and 0.255 in the outershell, at a resolution of 1.64Å. Overall I/σ was 8.4, outershell 2.6, and CC½ was 0.989, outershell 0.896. Overall completeness was 98.9% and completeness in the outershell was 98.8%. Structure factors were generated by Truncate, as before, and the twinning analysis revealed there was no twinning in the data.

Both structures were successfully solved by rigid-body refinement using the native structure of the same hSP-D fragment solved to 1.6Å with the calcium ions and solvent molecules removed [1PW9, Shrive *et al.*, 2003]. Assessment of the two maps showed comparable levels of information about the ligand binding, position of the ligand density compared to the coordinating calcium density and the number of residues present in the N-terminus. As a result, the highest resolution dataset collected from IB264C32 was carried forward for completion at a resolution of 1.59Å.

Two high resolution datasets were collected from two native crystals of rfhSP-D soaked with 20mM isomaltotriose during cryoprotection. The diffraction images from the two crystals, IB254C12 and IB254C14, were integrated and scaled using MOSFLM and Aimless as described section 3.4.2. In both cases the crystals were automatically indexed into the previously determined monoclinic spacegroup, $P2_1$ [Shrive *et al.*, 2003, 2009].

The unit cell of IB254C12 was subsequently refined to:

$$a = 55.26\ b = 107.81\ c = 55.47 \quad \alpha = \gamma = 90.0° \ \beta = 91.32°$$

The spacegroup was confirmed by Pointless, as part of the Pointless-Aimless pipeline in CCP4, as $P2_1$ and the intensities generated by MOSFLM were scaled and merged at a number of resolution cut offs. Unfortunately, the statistics for R_{merge}, I/σ(I) and CC½ did not

become acceptable until the data was limited to over 2.3Å. As a result, the dataset was discarded because the data from IB254C14 had good statistics at a higher resolution.

IB254C14 was indexed by MOSFLM into $P2_1$ with the following unit cell parameters:

$$a = 55.21\ b = 107.14\ c = 55.44 \quad \alpha = \gamma = 90.0° \ \beta = 91.29°$$

As before, the spacegroup was confirmed by Pointless before the integrated intensities were scaled and merged by Aimless at a resolution of 1.96Å. The merging statistics were good with the overall R_{merge} and R_{merge} in the outershell calculated to be 0.073 and 0.291, respectively. Data quality was confirmed by I/σ(I), 8.6 and 3.0 respectively, and CC½ was calculated to be 0.977 overall and 0.869 in the highest resolution shell, showing there is a high level of internal agreement in the data. Completeness for the dataset was found to be 89.4% overall and 84.2% for the outer shell. Structure factors were generated by Truncate and the analysis of twinning, using the tests described in section 3.4.2, revealed that there was no twinning in the data.

Due to the data from IB254C12 not merging and the resolution for IB254C14 being sub-2.0Å with good merging statistics, IB254C14 was taken forward for solution by rigid-body refinement. The structure for IB254C14 was successfully solved using the high resolution native structure of the same recombinant fragment (residues 179 – 355) of human SP-D with the calcium ions and solvent molecules removed to leave only the protein backbone and side chains. There was clearly density for the ligand in the density.

A total of six datasets for laminaritriose were collected from two separate wells and at different times. The first three datasets, IB254A51, IB254A52 and IB254A54, were collected but later found to be twinned following twinning analysis of the fast-processed data (fast_dp) generated on the beamline at Diamond Light Source. As a result, these three datasets were discarded as the recombinant fragment readily crystallises. The remaining

three datasets were produced by crystals that initially dissolved during cryoprotection but reformed in the following two weeks. IB319A53, IB319A54 and IB319A55 all diffracted well, and the images were indexed in a monoclinic spacegroup, $P2_1$, using MOSFLM.

The unit cell for IB319A53 was refined to:

$$a = 56.32\ b = 109.89\ c = 56.35 \quad \alpha = \gamma = 90.0° \ \beta = 90.11°$$

Diffraction images were then integrated using a strategy to maximise completeness and the intensities scaled and merged using Aimless. Pointless confirmed the spacegroup as $P2_1$ and the images were merged at a resolution of 1.80Å. The merged statistics showed a good R_{merge} was achieved overall, 0.052, and the highest resolution shell merged at 0.207. I/σ(I) for the overall dataset was 9.0 and 2.7 in the highest resolution shell while the CC½ was 0.990 and 0.638, with respect to the overall and outer shell. Completeness was 96.4% overall and 96.1% in the outer shell.

IB319A54 unit cell parameters were refined similarly to:

$$a = 56.18\ b = 110.02\ c = 56.36 \quad \alpha = \gamma = 90.0° \ \beta = 90.15$$

The diffraction images were integrated using MOSFLM, scaled and merged using Aimless using a resolution limit of 1.75Å. Pointless confirmed the spacegroup was $P2_1$ so no re-indexing was necessary. Data was successfully merged with R_{merge} calculated to be 0.052 overall and 0.328 in the outer shell. The overall I/σ(I) was 10.1 (outer shell 2.9) suggesting signal to noise was good while the correlation coefficient (CC½) was 0.989 overall and 0.841 in the outer shell. Completeness was high throughout the data, at 97.6% in the overall data and 97.3% in the outer shell.

The images collected from IB319A55 were indexed and refined with the following cell parameters:

$$a = 56.12\ b = 109.57\ c = 56.26 \quad \alpha = \gamma = 90.0° \ \beta = 90.05°$$

During scaling and merging it became apparent that there was a problem with the low resolution data that caused the R_{merge} and CC½ statistics to be beyond the acceptable limits at resolutions any lower than 1.6Å. As a result, this dataset was discarded as the data could not be relied upon and the data from IB319A54 was high resolution.

Due to the problems with data in IB319A55, no rigid-body solution was derived. IB319A53 and IB319A54 were both solved by rigid-body refinement using the native structure of the recombinant fragment of hSP-D solved at 1.6Å [Shrive *et al.*, 2003]. There was no difference in the information provided by the electron density, using Coot, so the highest resolution dataset was selected for refinement. As a result, IB319A54 was used to refine the structure and model laminaritriose into the density.

4.3.2 Initial map generation and density assessment

Initial maps for the cellotriose-, isomaltotriose- and laminaritriose-bound structures were generated using the rigid-body refinement function of REFMAC5 using the native structure of the same recombinant fragment of hSP-D used in this study with any calcium ions or water molecules removed, leaving only the protein backbone and sidechains to prevent any bias in the final map [1PW9, Shrive *et al.*, 2003]. The resulting weighted maps (equivalent to the traditional $2F_{obs}$-F_{calc} and F_{obs}-F_{calc} maps) were then displayed, with the solved protein structure, in Coot for density assessment and ligand fitting [Emsley and Cowtan, 2004; Emsley *et al.*, 2010].

In cellotriose, the rigid-body map for each carbohydrate binding site, before any rounds of refinement, is shown in Figure 4.7 with the rigid-body template displayed to allow for

orientation. The calcium ions in all three binding sites are clearly visible in the density from both the $2F_o$-F_c map (blue) and the F_o-F_c map (green). The F_o-F_c map, or difference map, shows there is strong occupancy of the calcium in all three sites as you would expect in the calcium-dependent binding site. There is also good resolution of the second and third calcium ions bound into each subunit of the trimer. However there is no density for the calcium, previously described as Ca4, in the neck domain of the trimer that is seen in the native structure of rfhSP-D [Shrive *et al.*, 2003].

In all three subunits, there is evidence for a terminal β-D-glucose of cellotriose binding the calcium, Ca1, in both of the maps [Figure 4.7]. In subunits B and C, the density covers the whole ring with good definition of the glycosidic linkage extending from the top of the bound glucose residue. In subunit A the whole ring is not defined, however the density clearly describes a glucose ring. Density for the remaining rings is not present in the rigid body maps in subunits A and C, beyond a small amount of density for where the glycosidic linkage to the second glucose would be. In subunit B there is some density for the second glucose and the glycosidic linkage however how the second ring will fit is unclear. There was no density for the third glucose ring in any of the subunits.

Definition at the N terminal of the recombinant hSP-D fragment extends to Ala205 in subunits A and B, with good density for the backbone throughout. The density does not extend as far in subunit C where definition does not extend beyond Ser206. There is no density for the remaining amino acids in the fragment, residues 179 to 204/5, and they were not modelled.

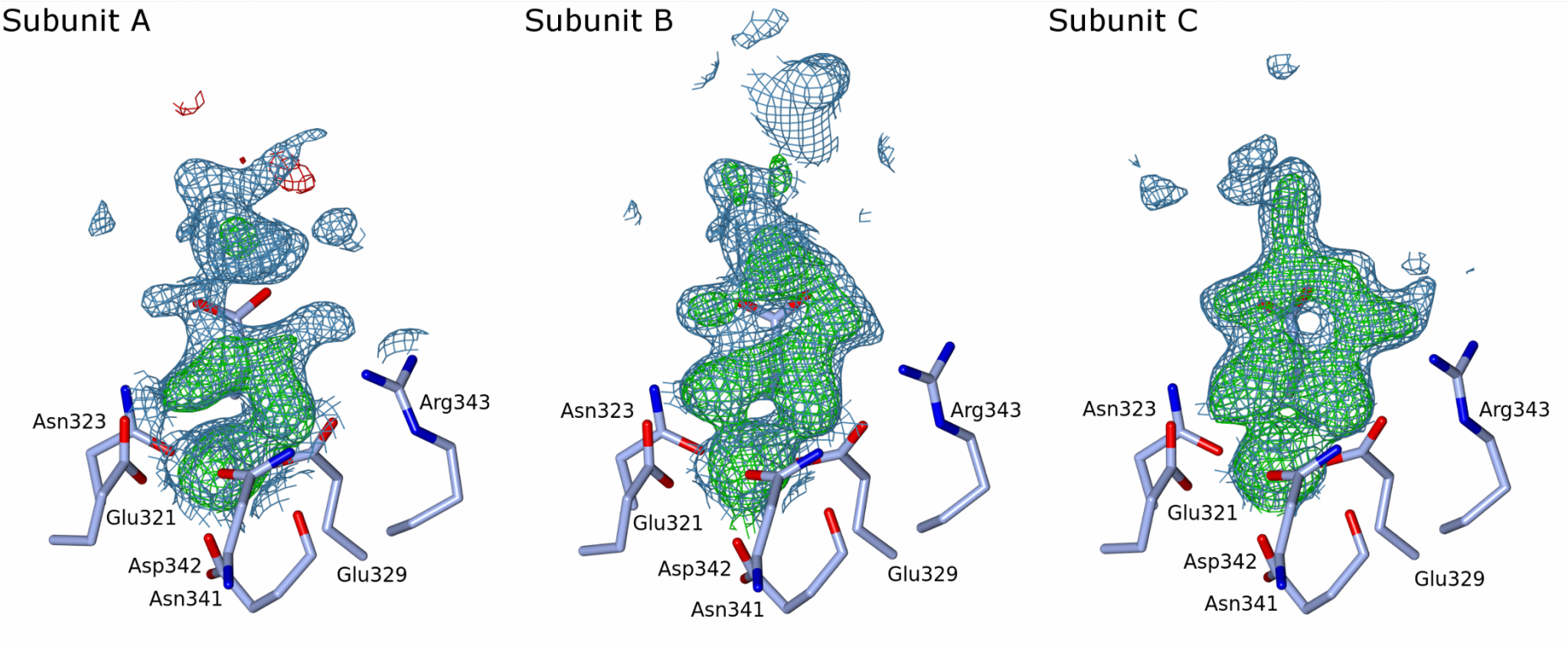


Figure 4.7: Rigid-body electron density maps for the cellotriose in complex with rfhSP-D. The observed electron density map displayed at 1.0σ ($2F_{obs}$-F_{calc}, blue) and difference map at ±3.0σ (F_{obs}-F_{calc}, green (+) and red (-)) from the rigid-body solution. The template structure is displayed to allow for orientation of the binding site. Electron density shows the presence of cellotriose in all three chains, coordinated by the clearly defined calcium ion.

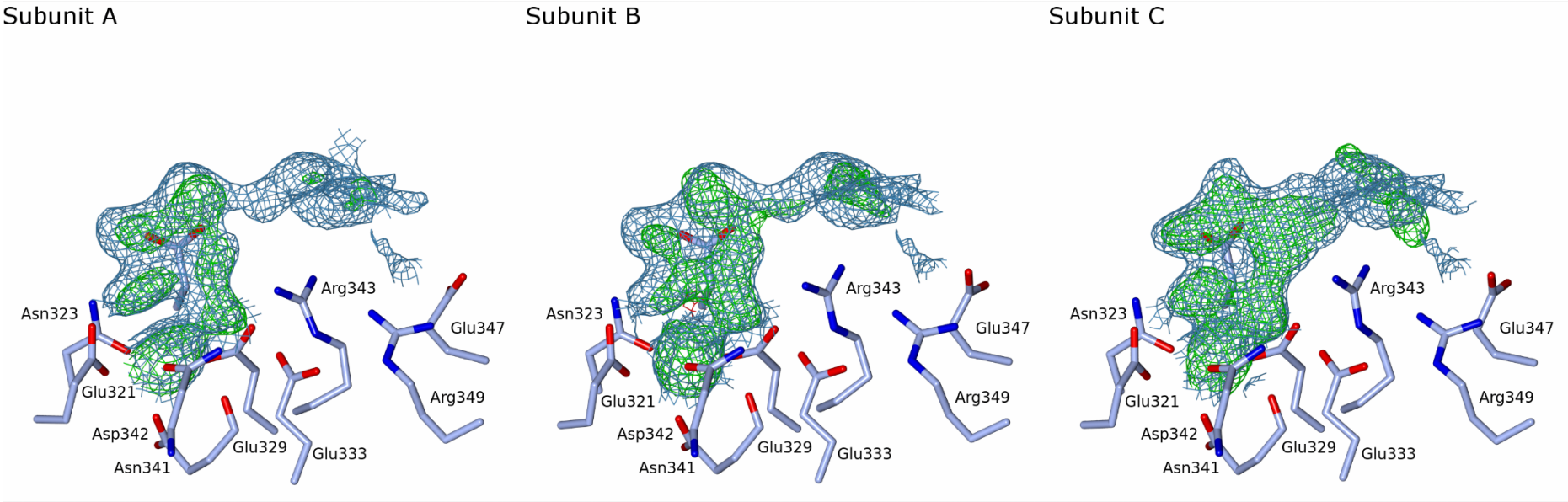


Figure 4.8: Rigid-body electron density maps for the isomaltotriose in complex with rfhSP-D. The observed electron density map at 1.0σ ($2F_{obs}$-F_{calc}, blue) and difference map at ±3.0σ (F_{obs}-F_{calc}, green (+) and red (-)) from rigid-body solution. The template structure is displayed to allow for orientation of the binding site. Electron density shows the presence of isomaltotriose in all three chains, with additional density extending towards Glu347 and Arg349. Coordinating calcium (Ca1) is clearly present in the electron density from both maps.

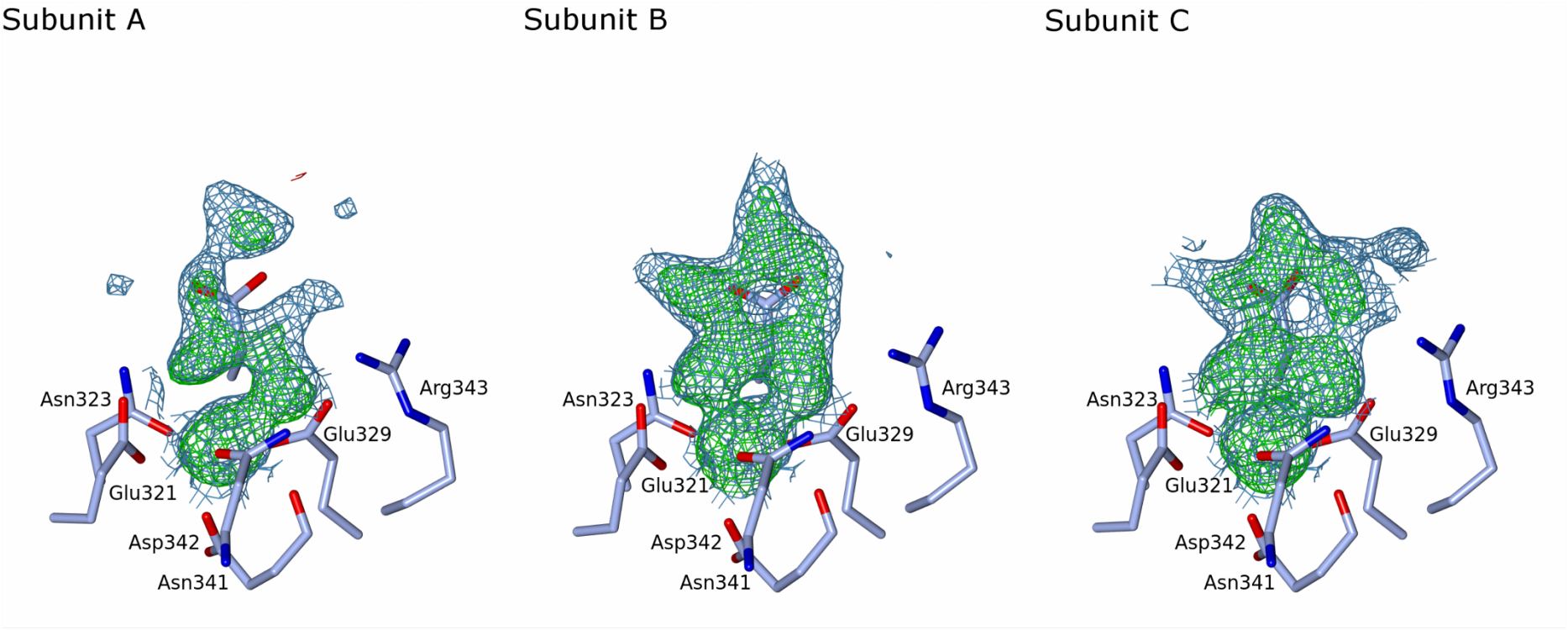


Figure 4.9: Rigid-body electron density maps for the laminaritriose in complex with rfhSP-D. The observed electron density map at 1.0σ ($2F_{obs}$-F_{calc}, blue) and difference map at ±3.0σ (F_{obs}-F_{calc}, green (+) and red (-)) from rigid-body solution. The template structure is displayed to allow for orientation of the binding site. Electron density shows the presence of laminaritriose in all three chains and the coordinating calcium (Ca1) in both maps.

Both rigid-body maps ($2F_o$-F_c in blue, F_o-F_c in green) from the isomaltotriose-bound structure, for all three carbohydrate recognition sites and the Ca4 binding site, are shown in Figure 4.8. As with the cellotriose-bound structure, there is no density in the neck of the trimer for the fourth calcium. The remaining calcium ions in each subunit are clearly defined in the electron density, confirmed by the strong positive signal in the difference weighted map, supported by the positions of the coordinating side chains described by Shrive (2003) [Shrive *et al.*, 2003].

There is clear density for a terminal glucose residue, at coordinating distances to the density for Ca1, in all three subunits and in both electron density maps [Figure 4.8]. In subunits A and C, the density continues across the second glucose, covering most of the ring and, importantly, the glycosidic linkage. The third glucose has some density from the second glycosidic linkage, extending across Arg343 and back towards the protein surface at Arg349. The positive density in the difference map shows that the occupancy is best for the primary coordinating glucose in all three chains whereas the density is less good for the second and third glucoses. In subunit B, there is clear density for a glucose in a similar position to the first and second glucose residues in subunits A and C however the density extending beyond the second glucose is more extensive.

Electron density extends across the backbone to residue 206 in the N-terminal direction in subunit A, 204 in subunit B and 205 in subunit C. The side chains are less well defined in the N-terminal of the visible fragment compared to the CRD. The remaining N-terminal of the recombinant fragment (residues 179 – 204/5/6) is not defined by the density and therefore cannot be determined in the structure.

The laminaritriose-bound maps for each carbohydrate binding site and the Ca4-binding site, from the rigid-body refinement, are presented in Figure 4.9. Density for all three

calcium ions in subunits A, B and C is present in positions comparable to the native rfhSP-D structure, suggesting that the three main calcium ions are present in the structure. Positive density confirms that the calcium ions are present in the difference map. There does not appear to be any density in the position where Ca4 is present in the native structure, with limited evidence for the associated changes in the side chains of Tyr228 and Lys232 [Shrive *et al.*, 2003, 2009].

There is density for a glucose in the binding site of subunits B and C in the observed map (blue) with good density for the glucose in the difference map (green), confirming that there is ligand bound into B and C subunits [Figure 4.9]. There is some evidence in the density that there are extensions from the glucose in subunit B, extending away from ring above Glu321 and, potentially from the top of ring, away from the protein. Interestingly in subunit C, the limited extensions appear to be from both sides of the binding glucose. It is not clear, however, if there are any further glucoses visible in the density as the observed density is fragmented. The density in subunit A is less conclusive in either map, as a glucose ring is not completely covered by the density, however the density present is of a sufficient size to fit a β-D-glucose [Figure 4.9]. The difference map in subunit A appears to confirm the position of the coordinating hydroxyls as there is positive density supporting the observed map. There is no density beyond the first ring in subunit A.

Density for the neck domain in the laminaritriose-bound structure is less well defined than in the cellotriose- and isomaltotriose-bound structures. The observed map extends across the amino acid backbone to Ala205 in subunit A, Ala205 in subunit B and Ser206 in subunit C however there is large amount of negative density around subunit A, suggesting that there may not be complete definition to Ala205 in the final model. As with the other two structures, there is no density for the remaining N-terminal amino acids up to and including Gly179 that does not readily stabilise in the crystal.

4.3.3 Final structures and binding mechanisms

The overall structure for all three ligand bound structures is a trimeric unit containing the carbohydrate recognition domain and neck domain from residue 205 onwards [Figure 4.10 and 4.11] except for laminaritriose where the N-terminal is disordered to residue 212 in subunit A [Figure 4.12]. The calcium ion found in the neck of the trimer in other published structures, is not present in any of the structures presented here. Rounds of refinement were completed after the addition of the first β-D-glucose, correction of the protein backbone and side chains, and addition of the solvent molecules to the structure, with 10 cycles of maximum-likelihood phase refinement for each round. Any additional changes made to the structure were then re-refined to check the fit and calculate the final R-factors.

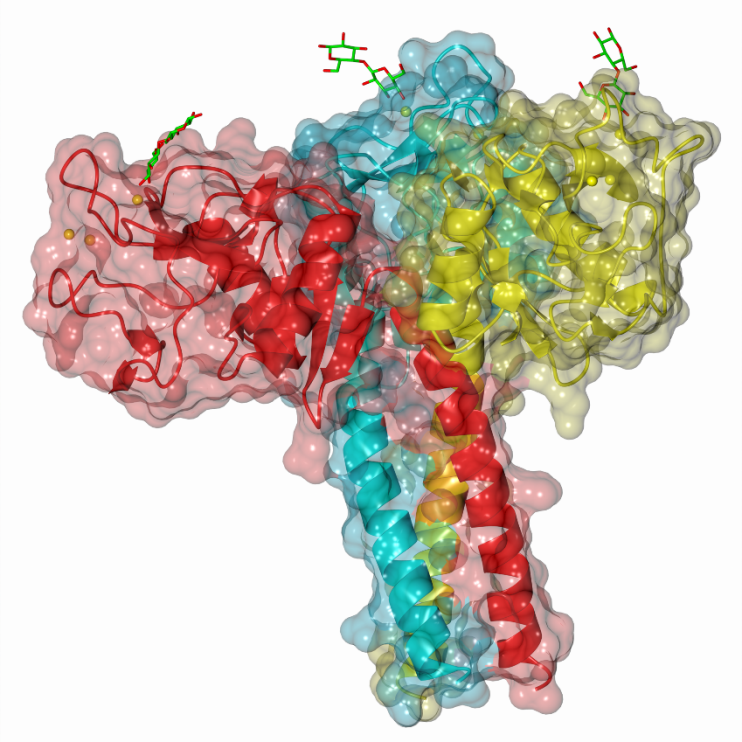


Figure 4.10: Overall structure of rfhSP-D in complex with cellotriose. The overall orientation of cellotriose (green) in all three ligand binding sites shows the ligand extending away from the surface of the CRD of each subunit (red, A; yellow, B; cyan, C). Calcium ions are represented in yellow.

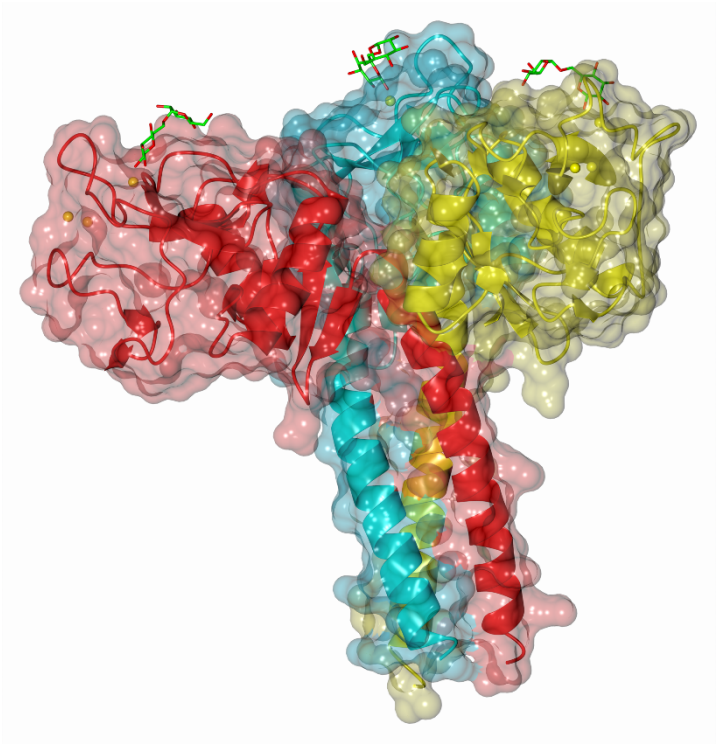


Figure 4.11: Overall structure of rfhSP-D in complex with isomaltotriose. The overall orientation of isomaltotriose (green) in all three ligand binding sites shows the ligand extending across the surface of the CRD of each subunit (red, A; yellow, B; cyan, C). Calcium ions are represented in yellow.

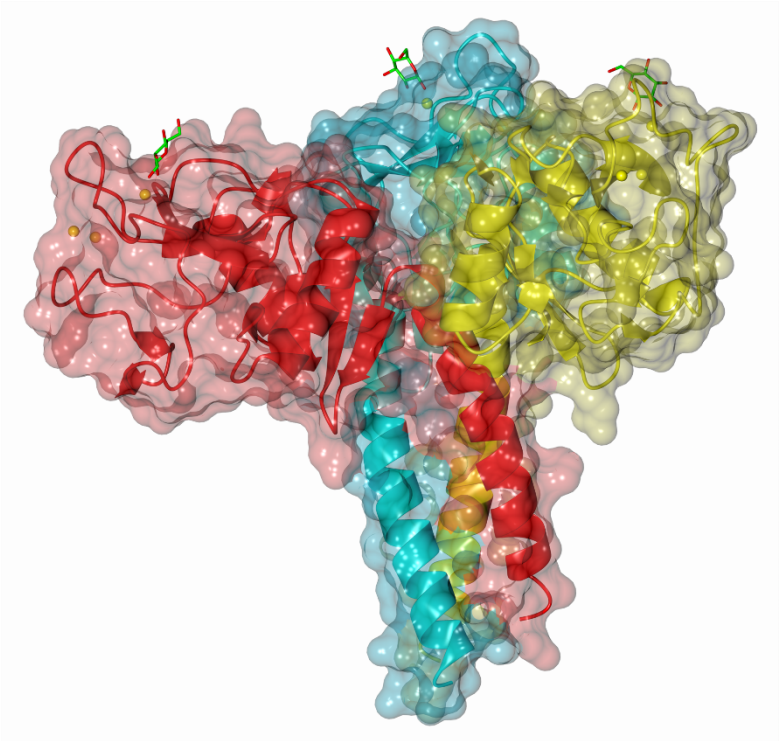


Figure 4.12: Overall structure of rfhSP-D in complex with laminaritriose. The overall orientation of laminaritriose (green) in all three ligand binding sites shows a single reducing terminal glucose bound in the CRD of each subunit (red, A; yellow, B; cyan, C). Calcium ions are represented in yellow.

Table 4.4: Data collection and refinement statistics from the small ligand bound structures. Data collected on beamlines I02, I03 and I04-1.

	Cellotriose	**Isomaltotriose**	**Laminaritriose**
Data collection			
Wavelength (Å)	0.9795	0.9200	0.9795
Temperature (K)	100	100	100
Space group	$P2_1$	$P2_1$	$P2_1$
Cell Dimensions			
a (Å)	55.39	55.21	56.18
b (Å)	108.18	107.14	110.02
c (Å)	55.61	55.44	56.36
β (°)	90.86	91.29	90.15
Maximal resolution (Å)	1.59	1.96	1.75
Observations	290 585 (14 509)	134 991 (8 047)	184 804 (10 054)
Unique reflections	87 424 (4 288)	41 033 (2 722)	67 179 (3 663)
Completeness (%)	99.5 (99.3)	89.4 (84.2)	97.6 (97.3)
R_{merge}[a]	0.051 (0.359)	0.073 (0.291)	0.052 (0.328)
$I/\sigma(I)$	10.8 (2.9)	8.6 (3.0)	10.1 (2.9)
Refinement			
Protein atoms[b]	3 478	3 471	3 365
Residues, chain A	205 - 355	205 - 355	212 - 355
Residues, chain B	205 - 355	204 - 355	209 - 355
Residues, chain C	206 - 355	206 - 355	208 - 355
Other atoms			
Calcium ions	9	9	9
Ligand	A = 2, B = 2, C = 2	A = 2, B = 2, C = 2	A =1, B = 1, C = 1
Water	405	350	430
Resolution range (Å)	55.6 – 1.59	55.43 – 1.96	50.16 – 1.75
R_{work}[c]	0.1879	0.1664	0.1808
R_{free}[d]	0.2106	0.2084	0.2193
Average *B*-values (Å^2)			
Protein main chain	25.77	24.46	28.21
Ligand	40.68	32.34	32.41
Water	36.54	35.50	39.02
Ramachandron plot values[e] (%)			
Favoured	97.98	97.76	97.69
Allowed	2.02	2.24	2.31
Disallowed	0.00	0.00	0.00

[a]$R_{merge} = \Sigma_h\Sigma_j|I_{h,j}-I_h| \,/\, \Sigma_b\Sigma_j|I_{h,j}|$, where $I_{h,j}$ is the *j*th observation of reflection *h* and I_h is the mean of *j* for reflection *h*

[b]Total number of protein atoms used in refinement

[c]$R_{work} = \Sigma_h||F_{oh}|-|F_{ch}|| \,/\, \Sigma_h|F_{oh}|$, where F_{oh} and F_{ch} are the observed and calculated structure factor amplitudes, respectively, for reflection *h*.

[d]R_{free} = R_{work} but for a random 5% subset of reflections

[e]defined by MolProbity analysis

Table 4.5: Calcium coordination distances in the cellotriose-bound, isomaltotriose-bound and laminaritriose-bound complexes with recombinant fragment of human surfactant protein D

		Cellotriose			Isomaltotriose			Laminaritriose		
Atom 1	**Atom 2**	**A**	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**
Ca1	Glu321 OE1	2.66	2.59	2.59	2.52	2.67	2.64	2.67	2.61	2.60
	Asn323 OD1	2.41	2.38	2.33	2.46	2.43	2.48	2.47	2.44	2.43
	Glu329 OE1	2.42	2.44	2.42	2.38	2.35	2.30	2.42	2.38	2.41
	Asn341 OD1	2.40	2.39	2.39	2.47	2.45	2.37	2.46	2.42	2.45
	Asp342 OD1	2.41	2.34	2.38	2.38	2.37	2.30	2.47	2.30	2.33
	Asp342 O	2.53	2.52	2.54	2.58	2.57	2.57	2.57	2.52	2.58
	Glc1 O3′/ Glc3 O2′	2.35	2.35	2.39	2.36	2.36	2.37	2.42	2.39	2.44
	Glc1 O4′/ Glc3 O1′	2.36	2.42	2.41	2.37	2.38	2.35	2.43	2.49	2.44
Glc1 O3′/ Glc3 O2′	Glu321 OE1	3.02	3.04	3.08	3.06	3.09	2.97	3.12	3.04	3.04
	Glu321 OE2	2.56	2.60	2.57	2.66	2.59	2.81	2.69	2.47	2.56
	Asn323 OD1	2.97	2.97	2.95	2.95	2.95	2.92	3.13	2.99	3.05
	Asn323 ND2	2.76	2.92	2.86	2.97	2.81	2.87	2.95	2.96	2.95
	Asn341 OD1	3.17	3.15	3.20	-	-	-	3.20	3.19	-
Glc1 O4′/ Glc3 O1′	Glu329 OE1	3.01	3.00	3.04	2.88	3.01	2.98	3.02	3.06	2.97
	Glu329 OE2	2.55	2.53	2.57	2.59	2.53	2.58	2.57	2.56	2.61
	Asn341 ND2	2.97	3.01	2.99	3.08	3.05	2.94	3.09	3.08	3.03
	Asn341 OD1	-	-	3.16	-	3.18	3.11	-	-	-
	Asp342 O	2.84	2.94	2.90	2.90	2.94	2.84	2.92	2.98	2.92
Glc1 O6′	Arg343 NH1	2.64	-	2.99	-	-	-	-	-	-
Glc3 O6′	Arg343 NH1	-	-	-	-	-	-	-	3.06	3.09

Table 4.5 continued

		Cellotriose			Isomaltotriose			Laminaritriose		
Atom 1	**Atom 2**	**A**	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**
Glc3 O5′	Arg343 NH1	-	-	-	3.07	3.18	3.05	-	-	-
	Arg343 NH2	-	-	-	3.17	-	3.15	-	-	-
Glc2 O3′	Glc1 O5′	2.83	3.10	3.18	-	-	-	-	-	-
	Glu347 OE1	-	-	-	3.03	-	2.57	-	-	-
Glc2 O3′	Glu347 OE2	-	-	-	2.80	2.64	-	-	-	-
Glc2 O4′	Glu347 OE1	-	-	-	2.51	2.86	-	-	-	-
	Arg349 NH2	-	-	-	2.72	2.99	2.97	-	-	-
Glc1 O4′	Arg349 NH1	-	-	-	-	2.93		-	-	-
Ca2	Asp297 OD1	2.60	2.68	2.61	2.65	2.62	2.69	2.64	2.64	2.68
	Asp297 OD2	2.52	2.47	2.47	2.37	2.38	2.44	2.54	2.47	2.54
	Glu301 OE1	2.52	2.52	2.46	2.54	2.51	2.52	2.49	2.53	2.52
	Glu301 OE2	2.41	2.51	2.45	2.51	2.51	2.52	2.47	2.50	2.51
	Asp324 OD1	2.51	2.52	2.63	2.54	2.52	2.65	2.55	2.58	2.57
	Glu329 O	2.43	2.39	2.41	2.54	2.43	2.44	2.49	2.49	2.40
	Asp330 OD1	2.41	2.38	2.41	2.32	2.30	2.43	2.38	2.40	2.49
	W	2.43	2.34	2.37	2.30	2.30	2.32	2.40	2.38	2.30
Ca3	Glu301 OE1	2.32	2.37	2.38		2.31		2.42	2.36	2.36
	Asp330 OD1	2.55	2.50	2.53		2.52		2.56	2.49	2.54
	Asp330 OD2	2.44	2.50	2.46		2.45		2.42	2.53	2.56
	W	2.30	2.39	2.30		2.30		2.42	2.36	2.30
	W	2.30	2.42	2.30		2.33		2.45	2.38	2.36
	W	2.34	2.30	2.35		2.35		2.30	2.41	2.39
	W	2.35	2.34	2.43		2.44			2.43	2.30

All numbers quoted are in Å. Ca1 refers to the non-reducing glucose terminal. Ca3 refers to the reducing glucose terminal. W represents water molecules in important binding positions

Final refinement statistics for the cellotriose-bound structure were 0.1879 and 0.2106 for R_{factor} and R_{free} respectively. Initially, the structure only showed density for the coordinating β-D-glucose with some definition of the glycosidic linkage. Following refinement, additional density for the second β-D-glucose became apparent, enabling fitting of the second sugar in all three subunits. Coordination to Ca1 is through the O3′ and O4′ hydroxyls, of the non-reducing glucose, at distances of 2.35Å and 2.36Å in subunit A; 2.35Å and 2.42Å in subunit B; and 2.39Å and 2.41Å in subunit C [Figure 4.13]. As with previously solved ligand-bound structures, interactions with the first glucose are completed by protein-ligand interactions between the O3′ and O4′ hydroxyls and side chains of Glu321 (2.57Å, 3.08Å), Asn323 (2.86Å, 2.95Å), Glu329 (2.57Å, 3.04Å), Asn341 (2.99Å, 3.16Å, 3.20Å) and the carbonyl of Asp342 (2.90Å). Additional protein-ligand interactions occur between O6′ hydroxyl and Arg343 in subunits A (2.64Å) and C (2.99Å). This interaction is not present in subunit B.

Density for the second glucose is best defined in subunit A where a water bridge between the O6′ hydroxyl and the main ring oxygen (O5′) stabilises the ring and minimises its movement, allowing it to be seen in the density. One further hydrogen bond is formed between the O3′ hydroxyl of the second glucose and the O5′ oxygen in the binding glucose (2.83Å). Density for the second glucose in subunits B and C is less well defined, particularly in subunit C where only half of the ring is defined. In subunit B, the only interaction for the second ring is a 3.10Å hydrogen bond to the O5′ oxygen in the first glucose whereas in subunit C this interaction is 3.18Å. At 1.0rmsd, there is no density for the glycosidic linkage in any of the subunits in the cellotriose-bound structures suggesting that there is mobility in the reducing residue as it projects into the space between the molecules in all three subunits.

The overall orientations of the rings in subunit A are comparable to the maltose-bound structure however there is a general shift in the binding glucose to accommodate the β-linkage. The second glucose is rotated 180° with respect to the glycosidic linkage and approximately 24.5° towards GlnB282 of the symmetry related molecule. The orientation of Arg343, an important binding-site flanking residue, is rotated towards the ligand in all three subunits whereas Asp325 is in a more comparable position to that previously seen and makes no direct interactions with the ligand.

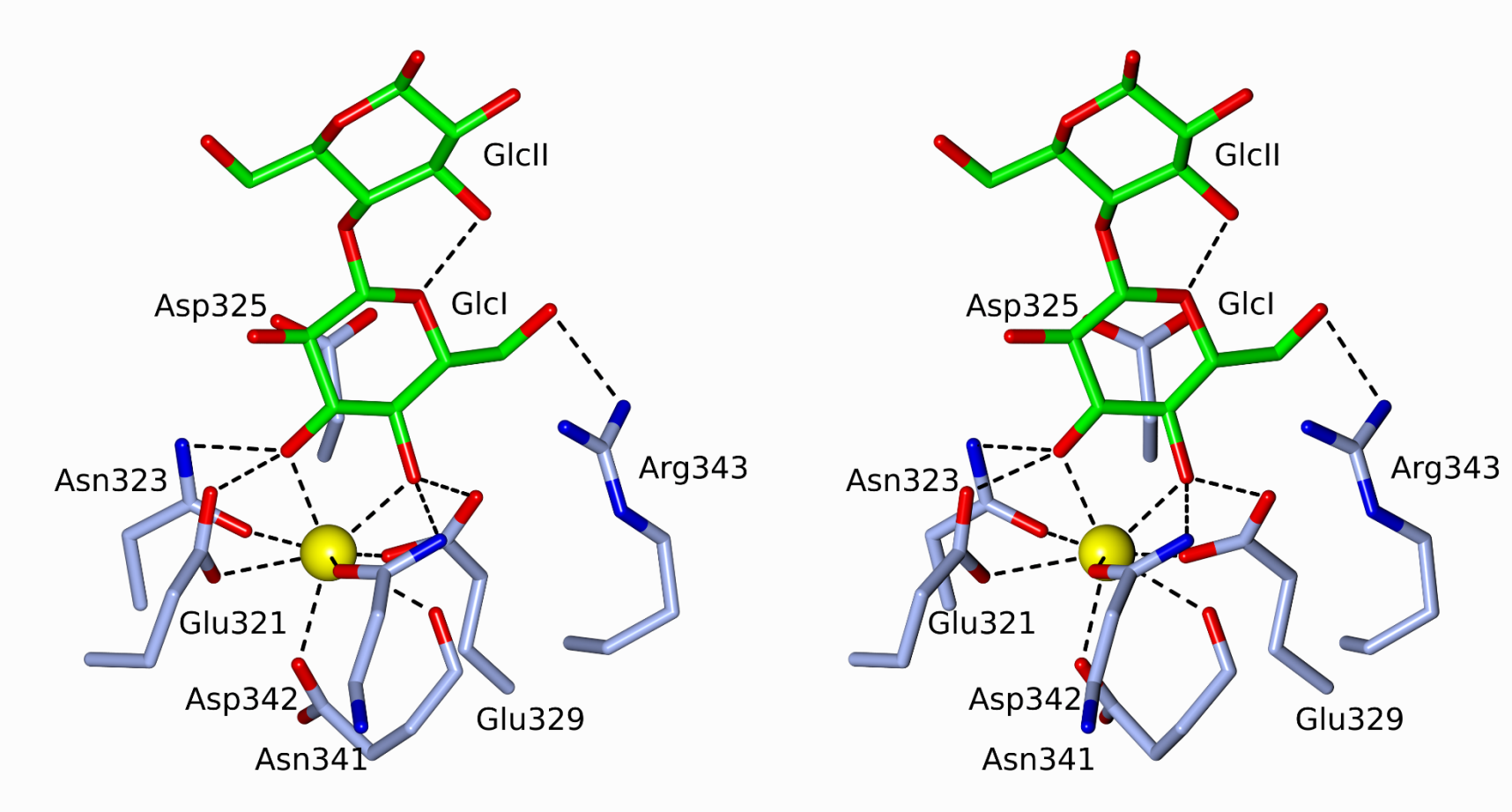


Figure 4.13: Stereo view of calcium coordination in the cellotriose-bound rfhSP-D complex. The binding calcium (Ca1, yellow) is coordinated by the O3' and O4' hydroxyl groups of the non-reducing terminal β-D-glucose (GlcI) with additional interactions with Glu321, Asn323, Glu329, Asn341 and Arg343. Representative of all three chains in the cellotriose-bound complex.

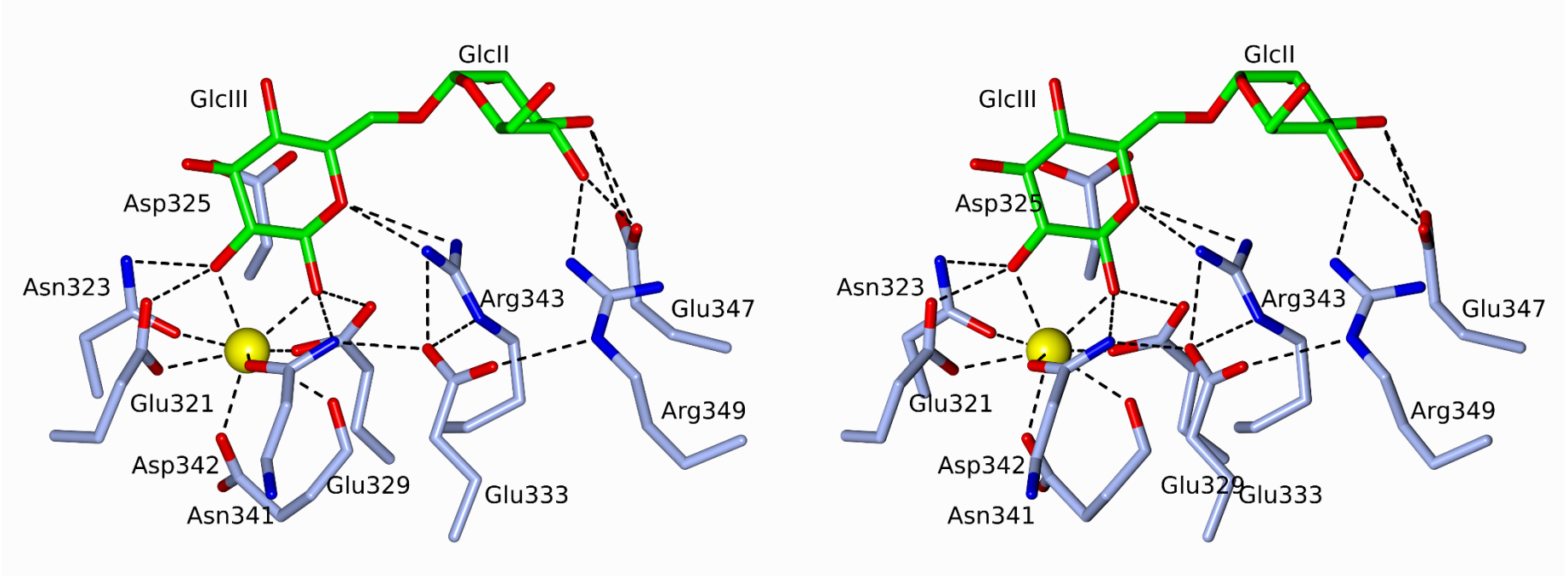


Figure 4.14: Stereo view of the calcium coordination in the isomaltotriose-bound rfhSP-D complex. The binding calcium (Ca1, yellow) is coordinated by the O1′ and O2′ hydroxyls of the reducing terminal β-D-glucose (GlcIII) with additional interactions with Glu321, Asn323, Glu329 and Asn341. The interactions with the protein are completed by Arg343 and O5′ ring oxygen and a hydrogen bond network between the non-terminal glucose (GlcII), Glu347 and Arg349. Representative of all three subunits in the isomaltotriose-bound rfhSP-D complex.

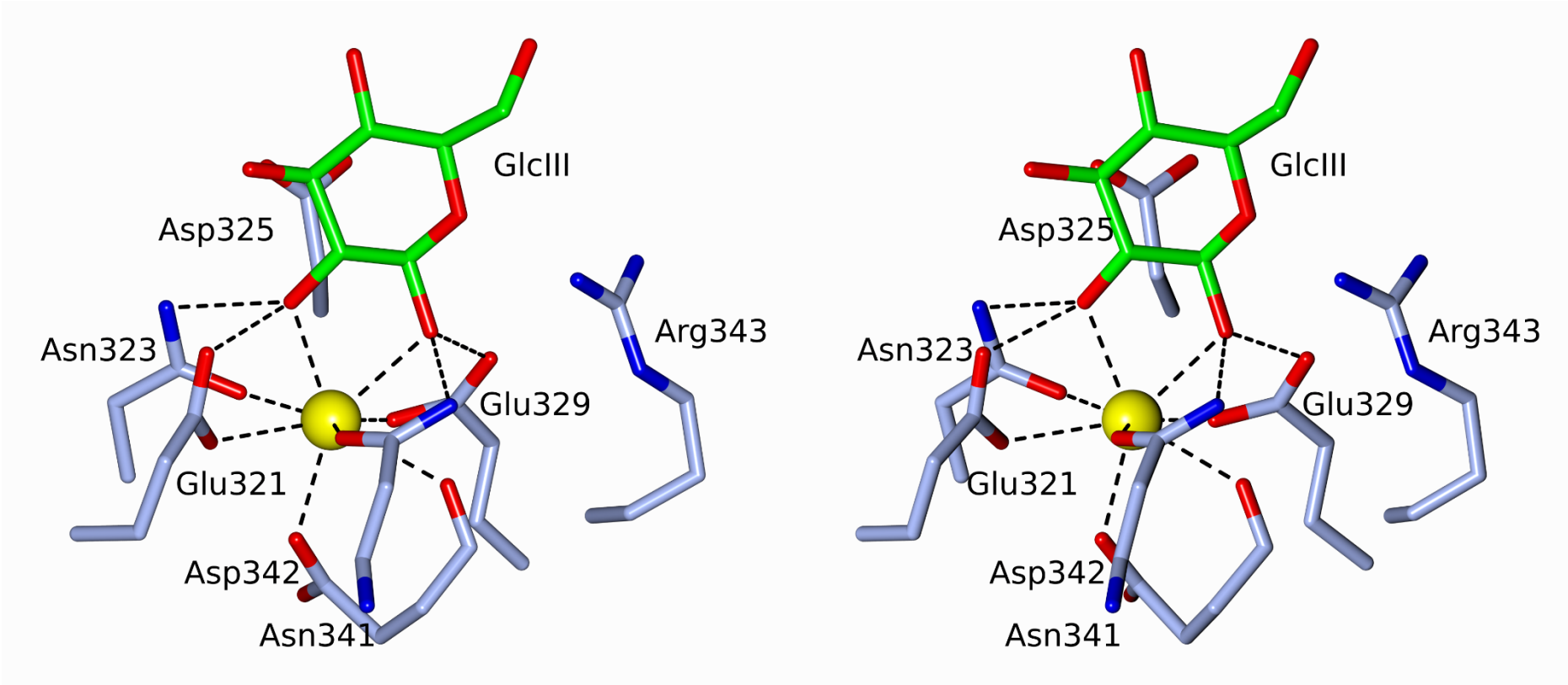


Figure 4.15: Stereo view of calcium coordination in the laminaritriose-bound rfhSP-D complex. The binding calcium (Ca1, yellow) is coordinated by the O1′ and O2′ hydroxyls of the reducing terminal glucose (GlcIII) with additional interactions with Glu321, Asn323, Glu329 and Asn341.

Restrained refinement for the final model of isomaltotriose-bound rfhSP-D refined with statistics of 0.1664 for the R_{factor} and 0.2084 for the R_{free}. These two statistics support the structure being an accurate representation of the data as both of the *R* statistics converged throughout refinement. The final structure contains two reducing terminal α(1→6)-linked glucose residues (GlcII and GlcIII) in each subunit, coordinating Ca1 in a novel O1′ and O2′ hydroxyl mediated mechanism. Interestingly, the coordinating glucose (GlcIII) is not only the reducing terminal of the glucose chain, it is also in the anomeric beta form. This allows the two coordinating hydroxyls to acquire the same position as the O3′ and O4′ hydroxyls in the cellotriose-bound structure presented here and in a variety of previously published ligand-bound structures [Shrive *et al.*, 2003, 2009; Crouch, McDonald, *et al.*, 2006]. O1′ coordinates Ca1 at 2.37Å and O2′ is coordinated at 2.36Å (table 4.5, figure 4.14). The remaining coordination to the binding glucose (GlcIII) is completed by a series of protein-ligand interactions of between 2.59Å and 3.17Å, mostly between calcium coordinating residues (Glu321, Asn323, Glu329, Asn341 and Asp342) and the O1′ and O2′ hydroxyls, with additional hydrogen bonds between the O5′ oxygen and Arg343 (3.07Å, 3.17Å). Interestingly, the alternative hydroxyl coordination results in the α(1→6) linkage extending away from the binding site in an alternative direction to other malto-N-oses, completing the protein interactions at a secondary binding site with Glu347 and Arg349.

The second glucose (GlcII) is positioned above Arg343, forcing it into a more compressed conformation, forming a hydrogen bond network between the O3′ and O4′ hydroxyls, NH2 of Arg349 and both oxygen groups of Glu347 [Figure 4.14]. The O2′ hydroxyl completes the interactions with a water bridge to the main chain of the symmetry related molecule in subunit A. In subunits B and C, O2′ completes this water bridge with Asp325 and Asp329 due to a small rotation in the plane of the coordinating glucose that brings the second glucose (GlcII) closer to the protein surface. The density for the non-reducing terminal

glucose (GlcI) was not sufficient to confidently model the glucose residue however it is suggested that there are transient interactions with the protein because of a concerted rearrangement of Glu289 towards a pocket where GlcI would tentatively recognise the protein; potentially explaining the electron density seen in the crystal structure.

The laminariotriose-bound structure was refined with only a single glucose, achieving an R_{factor} of 0.1808 and R_{free} of 0.2193 which converged during refinement. The calcium coordination occurred via the O1′ and O2′ hydroxyls, of the reducing terminal, in the same orientation as isomaltotriose, with coordinating distances between 2.39Å and 2.49Å respectively [Figure 4.1]. The O1′ interactions with protein are completed by Glu329 (2.56Å – 3.06Å), Asn341 (3.03Å – 3.09Å) and the carbonyl of Asp342 (2.92Å – 2.98Å); for O2′ the interactions are completed by Glu321 (2.47Å - 3.12Å) and Asn323 (2.95Å - 3.13Å). In addition, there is an interaction between Asn341 and O2′ in subunits A and B which is not present in subunit C. However, there is a long hydrogen bond between the O6′ hydroxyl and side chain of Arg343.

Beyond the density for the β-D-glucose binding to Ca1, there is very little density for the remaining two glucosyl residues in subunits A and C, apart from a small amount of positive density, supported in the $2F_o$-F_c equivalent map, where you would expect the next glucose to be; however, there is insufficient density for confident fitting. In subunit B, there is better definition of the glycosidic bond with density coming off the bound glucose at C3′ and extending into a patch positive density. The density does not extend across enough of the second glucose ring, however, so no fitting could be completed.

4.4 Discussion

4.4.1 Novel interactions with β-D-glucose and β-linked oligosaccharides

This is the first time that interactions between human surfactant protein D and β-glucosyl ligands have been investigated. Two alternative binding mechanisms have been identified in β-D-glucose ligands: (1) by the well documented O3′ and O4′ mannose-type interaction first seen in maltose and; (2) by the novel O1′ and O2′ hydroxyls. Cellotriose coordinates the calcium in the binding site via the O3′ and O4′ hydroxyls, of the non-reducing glucose, in a similar manner to maltose and maltotriose, whereas both laminaritriose and isomaltotriose both reveal recognition by O1′ and O2′ of the reducing terminal glucose [Shrive *et al.*, 2003; Crouch, McDonald, *et al.*, 2006].

The orientation of the O1′ and O2′ hydroxyls in laminaritriose and isomaltotriose are comparable to the mannose-type, equatorial hydroxyls that coordinate the calcium in other α-D-glucoses, allowing them to coordinate to Ca1. In both of these ligands bound structures, the binding glucose is rotated, compared to the terminal glucose of maltose, by 180° about the plane of the ring, presenting the binding hydroxyls to Ca1. There is a further rotation compared to the terminal glucose in maltose that moves the oxygen in the glucosyl ring into the proximity of Arg343, in a similar position to that seen in the galactose-bound structure [Shrive *et al.*, 2009]. As expected, the hydroxyls in all three structures superimpose with each other and adopt a position similar to the equatorial hydroxyls, however their overall position is an average of 0.18Å closer to the calcium in comparison to maltose.

Initially it was surprising that a β-D-glucose was the coordinating glucose in the isomaltotriose-bound structure as dextran is an α-D-glucose polysaccharide. However, during acid hydrolysis of the glycosidic linkage to produce the triose, carried out by the

supplier (dextra laboratories), the O1′ hydroxyl can undergo an unpredictable rearrangement resulting in a mixture of reducing terminal glucose types. More interestingly, the O3′ and O4′ hydroxyls of the non-reducing terminal glucose were available for binding in the crystal but the O1′ and O2′ hydroxyls were selected. This may be a result of the crystal artificially orientating the ligand into the binding site in subunit A, however there is a sufficient space around subunits B and C which may suggest the selection is genuine and that SP-D has a higher affinity for β-D-glucose ligands compared to its α-D-glucose equivalent. Further work is necessary to determine whether this is the case or not using binding affinity studies.

It was suggested by Allen and colleagues (2001) that surfactant protein D is able to recognise non-terminal glucoses through calcium coordination following molecular docking simulations [Allen *et al.*, 2001]. One of the objectives of this study was to investigate these interactions crystallographically using one of the identified ligands, isomaltotriose. The results here did not provide any insight into the non-terminal interactions however they do not rule out the possibility that they could occur and simply highlight the affinity for the terminal glucoses. To investigate this further, substitutions of the terminal glucose hydroxyls with groups that cannot donate coordinate bonds, promoting coordination of the non-terminal glucose.

4.4.2 Extended binding site on surfactant protein D

An extended binding site was first suggested by Crouch and colleagues who identified an interaction between the terminal residue of maltotriose and Phe335 which is positioned at the bottom of a shallow cleft on the surface of surfactant protein D [Crouch, McDonald, *et al.*, 2006]. The isomaltotriose structure also presented a secondary binding site but extending from the other side of the binding site, across the flanking residue, Arg343, and extending to Glu347 and Arg349 where the second glucose in the ligand forms at least two

interactions with the side chains. This appears to be, in part, because of the alternative coordination of Ca1 that forces isomaltotriose to exit the binding site in the direction of Arg343, potentially aided by the increased height and flexibility in the α(1→6)-linkage. This appears to allow isomaltotriose to form a 'bridge' across Arg343, which is in a retracted position compared to the native and other known ligand-bound structures [Shrive *et al.*, 2003, 2009; Crouch, McDonald, *et al.*, 2006; Wang *et al.*, 2008; Clark *et al.*, 2016].

Sequence analysis reveals that Glu347 and Glu349 are both highly conserved across the collectins and a range of mammalian and non-mammalian species. This suggests that the extended binding site, identified in the isomaltotriose-bound structure, is important in the recognition of longer polysaccharide ligands found on the surface of pathogenic and allergenic species. However further investigation with longer ligands is necessary to completely characterise the mid to long distance binding surface of SP-D.

Neither cellotriose or laminaritriose completed any interactions beyond the calcium binding site, particularly because cellotriose only differs from maltotriose by the topology of the glycosidic bond [Pendrill, Säwen and Widmalm, 2013]. However, the binding affinity of rfhSP-D for maltotriose is higher than for maltose (0.94mM compared to 2.3mM) whilst full length cellulose was found not to inhibit SP-D binding to its natural ligand from *A. fumigatus* [Allen, Voelker and Mason, 2001; Crouch, McDonald, *et al.*, 2006]. The difference between maltotriose and cellulose may suggest that the more planar chain in cellotriose, as a result of the equatorial β-linkage, prevents the Phe335 interaction which is present in the maltotriose-bound structure [Crouch *et al.*, 2006], however data for cellotriose affinity is necessary to confirm this. The small amount of density for the two undefined β-D-glucoses in the laminariotriose-bound structure projects above Glu321 in the direction of Pro319, where there is little to coordinate the undefined glucoses.

Gentiotriose, a β(1→6)-linked unit of pustulan, was also soaked into native crystals and successfully diffracted, however there was no ligand present in the binding site of any subunits (data not shown). It was later determined by the supplier of the ligand that the product did not contain the expected ligand and lacked the small chain gentiooligosaccharides. If the soak had been successful, the extending binding site, and influence of the flanking residues, could have been further characterised.

This adds to evolving understanding of how SP-D recognises longer ligands and merited revisiting previous data using longer malto-N-oses that contain α(1→4) glycosidic linkages [Shaw, 2009]. Two rfhSP-D complexes with maltotetraose (four α-D-glucose) and maltoheptaose (seven α-D-glucoses) were reinterpreted following the identification of a β-D-glucose in the calcium binding pocket and revealed a β-D-glucose in the binding site of subunits B and C in the heptose-bound complex [Figure 4.16, to be published]. The coordination of Ca1 occurs by a similar mechanism to the isomaltotriose-bound and laminaritriose-bound complexes, mediated through coordination bonds from the O1′ and O2′ hydroxyls, from the reducing terminal glucose, in the equatorial positions permitted by the β-anomeric form of glucose. In subunit A, recognition occurs through the non-reducing terminal glucose by the well-established O3′ and O4′ hydroxyl mechanism first observed in the maltose-bound and seen here in the cellotriose-bound complexes [Figure 4.17].

The new analysis of the maltotetraose-bound complex also revealed an extended binding site in subunit A. Calcium recognition occurs by the same mechanism as maltose, by the O3′ and O4′ equatorial hydroxyls, but the remaining three glucose residues extend across the surface of the CRD in a similar orientation to the maltotriose-bound complex and make additional hydrogen bonds to Thr336 and the main chain of Asn337 through non-reducing glucoses [GlcIV, Figure 4.18; Crouch *et al.*, 2006]. The additional glucose in maltotetraose

causes the oligosaccharide to form a higher "arch" over the surface, stabilised by hydrogen bonds between the glucose residues, and prevents Phe335 from interacting with the ligand. Interestingly, the fourth glucose was also in its β-anomeric form, suggesting that the method for producing the smaller oligosaccharides from longer polysaccharides causes a rearrangement of the reducing terminal monosaccharide as it appears in all of the ligands α-oligosaccharides in this study.

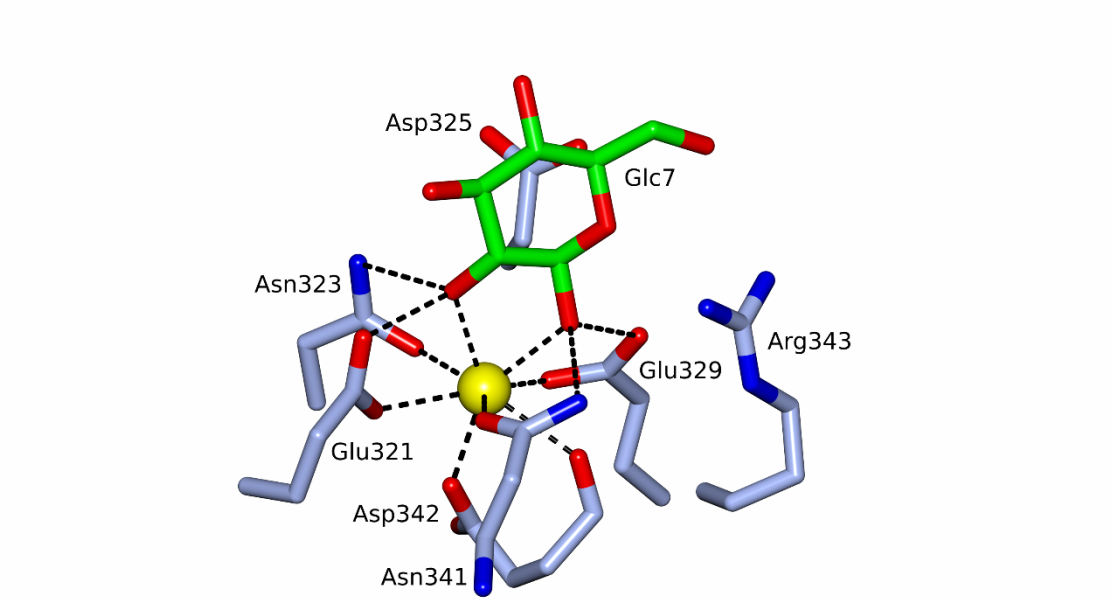


Figure 4.16: Calcium coordination in subunit B of the maltoheptaose-bound rfhSP-D complex. Coordination of the calcium (Ca1, yellow) is by O1' and O2' hydroxyls of the reducing terminal glucose (Glc7). Recognition is completed by interactions with Glu321, Asn323, Glu329 and Asn341. Representative of subunit C.

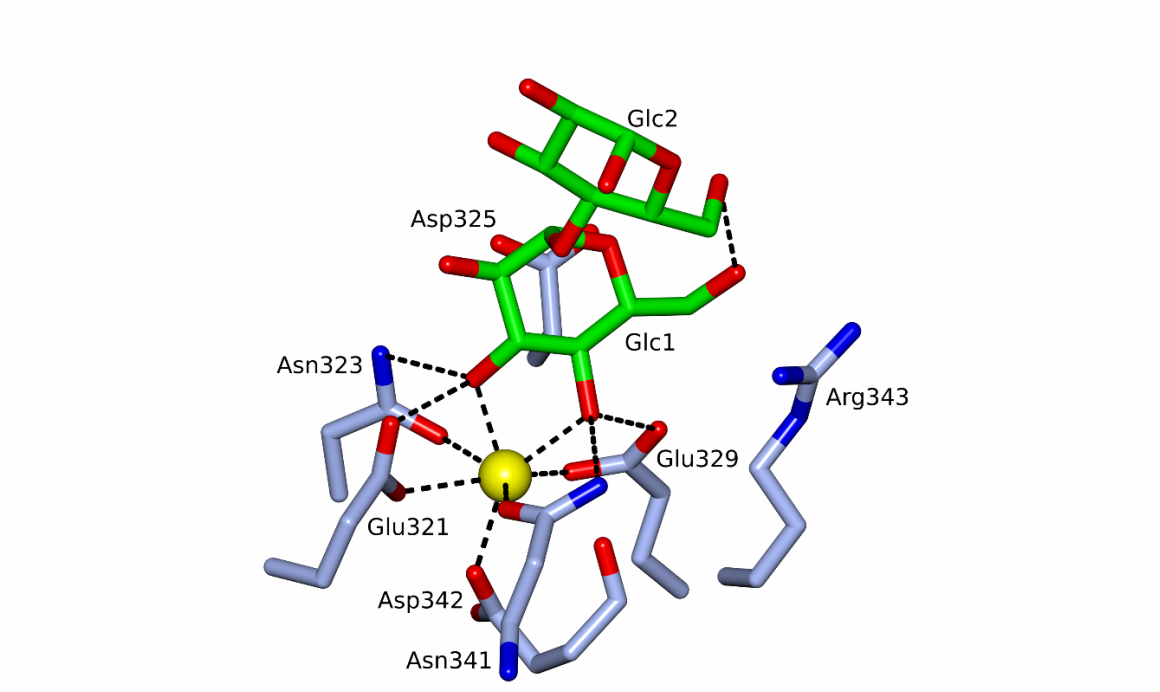


Figure 4.17: Calcium coordination in subunit A of the maltoheptaose-bound rfhSP-D complex. Coordination of the calcium (Ca1, yellow) is by O3′ and O4′ hydroxyls of the non-reducing glucose (Glc1). Recognition is completed by interactions with Glu321, Asn323, Glu329 and Asn341 with a single interaction to the second α-D-glucose (Glc2) between the two O6′ hydroxyls.

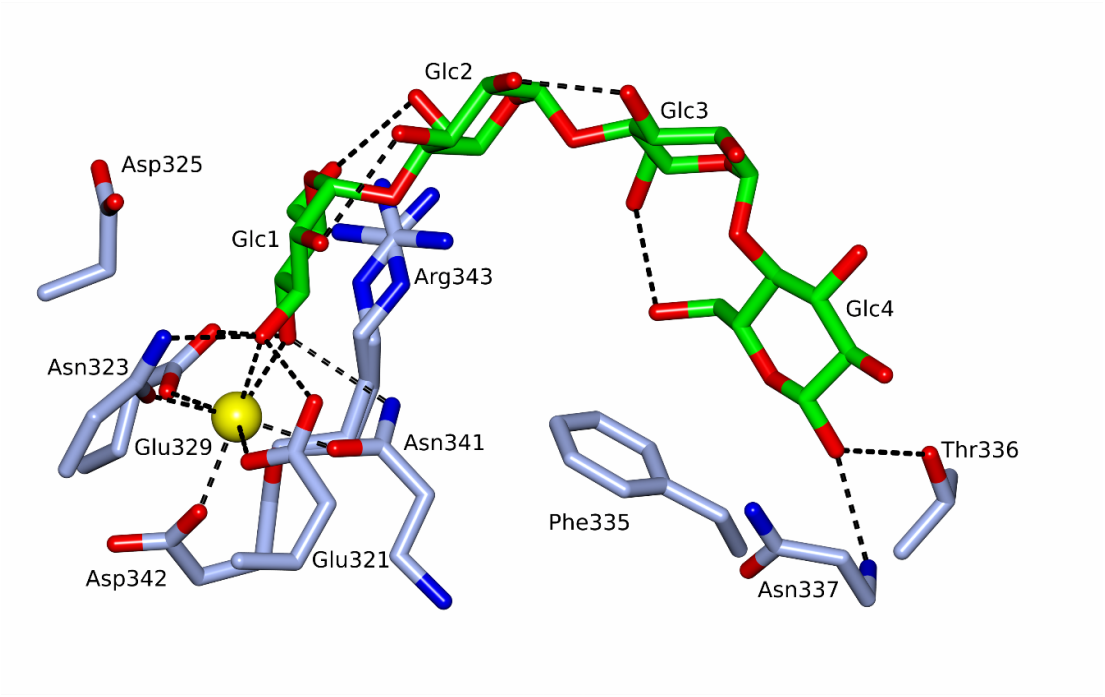


Figure 4.18: Calcium coordination in subunit A of the maltotetraose-bound rfhSP-D complex. Coordination of the calcium (Ca1, yellow) is by the O3′ and O4′ hydroxyls, completed by interactions with Glu321, Asn323, Glu329 and Asn341. Glc4 completes the interactions with the protein through hydrogen bonds to Thr336 and the backbone of Asn337

4.4.3 Importance of the flanking residues

Since the earliest ligand binding studies, the importance of the, widely conserved, Asp325 and Arg343 side chains as selection factors for ligands has been investigated through numerous ligand binding studies and mutation studies [Shrive *et al.*, 2003, 2009, Crouch *et al.*, 2007, 2009; Wang *et al.*, 2008; Clark *et al.*, 2016; Reinhardt *et al.*, 2016]. These studies have revealed that Asp325 and Arg343 are able to effectively select and direct the ligands into the calcium binding site due to their key positions either side of the binding pocket, forming two oppositely charged islands on the surface that orientates the ligand, presumably to allow the coordinating hydroxyls to interact with Ca1 in the correct positions. The significance of this selectivity was demonstrated using D325A + R343V and R343K mutants, which showed increased affinity for a range of ligands, including manno-N-oses that are typical of viral glycoproteins, that usually have significantly lower affinity in the native CRD of SP-D [Crouch *et al.*, 2009; Goh *et al.*, 2013].

The structures presented, and in particular the isomaltotriose-bound structure, demonstrate a more direct role for Arg343 in the recognition of β-D-glucoses, where it interacts with both isomaltotriose and laminaritriose. This data also suggests that Arg343 may be more dynamic and flexible, allowing simple glucosyl ligands access to the extended binding surface. Interestingly, altered positions of Arg343 still maintain the interaction with Glu333 (first reported in the maltose-bound structure), suggesting that the Glu347/Arg349 binding surface is semi-rigid and that SP-D is predisposed to exposing the extended recognition surface to natural polysaccharide ligands [Shrive *et al.*, 2003].

Chapter 5: Recognition of cell wall muramyl disaccharide from gram-positive bacteria

5.1 Introduction to gram-positive cell wall and peptidoglycan

Surfactant protein D is best known for recognising gram-negative bacteria and viral particles, including rough variants of *Haemophilus influenzae* and human influenza A virus [Hillaire *et al.*, 2013; Clark *et al.*, 2016]. However, the recognition role of SP-D has been shown to extend to gram-positive bacteria, a structurally distinct group of bacteria that are characterised by a cell wall that contains high levels of peptidoglycan and lacks the lipopolysaccharides found in gram-negative bacteria [Hartshorn *et al.*, 1998; Palaniyar, Nadesalingam and Reid, 2002; Kishore *et al.*, 2006; Vollmer, Blanot and De Pedro, 2008]. For the first time, the molecular basis of the recognition of muramyl disaccharide, a primary unit of peptidoglycan, has been determined and is presented here.

5.1.1 The gram-positive cell wall

The gram-positive cell wall is primarily composed of peptidoglycan, a long chain polysaccharide crosslinked by short peptide chains, forming a strong, semi-rigid structure that protects the cell from changes in turgor pressure and controls the size and shape of the cell; expanding and contracting in response to the environment to maxima and minima, preventing the cells from bursting [Schleifer and Kandler, 1972]. There are also other, highly glycosylated structures that cross the peptidoglycan layer and interact with the host, including lipoteichoic acids (LTAs), which are membrane-anchored, and peptidoglycan-anchored teichoic acids (TAs). The cell wall is completed by a range of transport and synthetic proteins that are involved in nutrient transport and peptidoglycan maintenance [Schleifer and Kandler, 1972; Park and Uehara, 2008].

An important part of cell wall maintenance is shedding of peptidoglycan from the cell wall surface, releasing between 5% and 50% of the total peptidoglycan as the cell wall matures [Park and Uehara, 2008]. In gram-negative bacteria, the outer membrane prevents the release of the peptidoglycan, allowing it to be recycled into new peptidoglycan, but the lack of the outer membrane in gram-positive bacteria allows the peptidoglycan to be released into its surroundings in a soluble form. It is the soluble form of peptidoglycan that has been shown to be immunogenic, stimulating a strong response from phagocytes and macrophages and stimulating a release of chemokines and cytokines leading to inflammation and further recruitment of immunocytes [De Kimpe *et al.*, 1995].

The number of exposed carbohydrate structures on the surface of gram-positive bacteria provide possible recognition sites for surfactant protein D as peptidoglycan and lipoteichoic acid are both known to be pathogen-associated molecular patterns used by numerous innate immune molecules to recognise gram-positive bacteria [van de Wetering *et al.*, 2001; Bertsche *et al.*, 2015]. It was suggested by Palaniyar in 2002 that both soluble and insoluble, cell wall-bound peptidoglycan that are exposed to the environment are targets for surfactant protein D via interactions between the core polysaccharide and carbohydrate recognition domain [Palaniyar, Nadesalingam and Reid, 2002].

5.1.2 Recognition of gram-positive cell wall components

A number of gram-positive bacteria have been identified as targets for surfactant protein D; many of which are known to be medically important in respiratory tract infections, including *Staphylococcus aureus* and *Streptococcus pneumoniae* [Table 1.3; Hartshorn *et al.*, 1998]. It has been shown that recognition of *S. pneumoniae* by SP-D causes extensive agglutination and, in a strain dependent manner, increases uptake by neutrophils [Hartshorn *et al.*, 1998; Jounblat *et al.*, 2004]. A similar effect was also seen in *S. aureus* but

to a lesser extent, suggesting that neutrophilic uptake varies between species as well as strains.

How these important gram-positive species are recognised by SP-D and SP-A was further characterised by *van de Wetering* and colleagues in 2001, who successfully showed that both surfactant proteins were able to recognise and adhere to LTAs and an insoluble, unmodified fragment of peptidoglycan [van de Wetering *et al.*, 2001]. For SP-D, the recognition of LTA extracted from *Bascillus subtilis* occurred via the carbohydrate recognition domain, in a calcium dependent mechanism, by a limited number of recognition motifs on the LTA. Similarly, the unmodified peptidoglycan from *S. aureus* was also recognised by the CRD of hSP-D but with lower affinity compared to LTA. Interestingly, a soluble fraction of peptidoglycan has also been recognised by SP-D however the details of this recognition are not clear [Palaniyar, Nadesalingam and Reid, 2002].

Peptidoglycan was shown to be recognised via the N-acetylglucosamine moiety by the related collectin mannose binding protein (MBP), through the CRD interaction that is suggested to be responsible for recognising peptidoglycan in SP-D [Nadesalingam *et al.*, 2005]. Interestingly, this study also suggested that MBP has a higher affinity for the soluble form of peptidoglycan compared to the insoluble peptidoglycan and, in fact, competed closely with the high affinity MBP ligand, mannan. As MBP shares the characteristic CRD with SP-D, it is proposed that SP-D may also interact with the two forms of peptidoglycan through the N-acetylglucosamine and to test this, a muramyl disaccharide unit extracted from *Micrococcus luteus*, another gram-positive species, has been soaked into native crystals of a biologically and therapeutically active recombinant fragment of human SP-D (rfhSP-D) to identify which of the two available saccharide moieties SP-D recognises.

5.1.3 Muramyl disaccharide

Muramyl disaccharide is the repeating unit of the polysaccharide at the core of peptidoglycan and is formed from alternating units of N-acetyl-β-D-glucosamine (GlcNAc) and N-acetyl-β-D-muramic acid (MurNAc), linked by a β(1→4) glycosidic linkage. In peptidoglycan, this disaccharide unit then repeats up to 60 times, linked by further β(1→4) glycosidic linkages, to form long, structurally strong polymers that are then cross-linked by the O-linked tetra- or penta-peptides that extend from the lactyl group attached to position three of the main saccharide ring [Figure 5.1; Schleifer and Kandler, 1972; Yang *et al.*, 2017]. The crosslinked polysaccharides then form the characteristic cell wall observed in gram-positive bacteria.

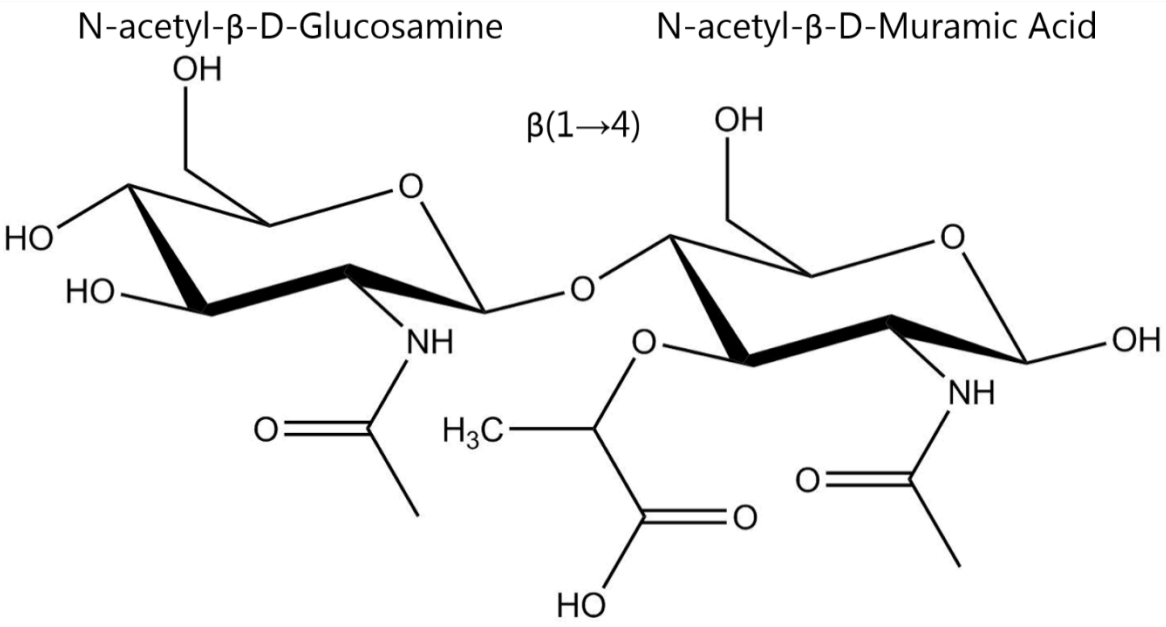


Figure 5.1: Representation of Muramyl Disaccharide from Gram-Positive Cell Wall. Figure shows a representation of muramyl disaccharide extracted from the gram-positive bacteria, *Micrococcus luteus*. It comprises an N-acetyl-β-D-glucosamine linked to an N-acetyl-β-D-muramic acid by a β(1→4) glycosidic linkage.

The aim of this study is to characterise how human surfactant protein D recognises the natural muramyl disaccharide from *Micrococcus luteus* at the atomic level and infer how a longer fragment of peptidoglycan, soluble or insoluble, may be recognised during an immune challenge by gram-positive bacteria such as *Streptococcus pneumoniae*. This will also provide further insight into which carbohydrate moiety of the muramyl disaccharide is the primary recognition target of rfhSP-D.

5.2 Materials and methods

5.2.1 Protein preparation

The rfhSP-D was received from the Southampton group at a concentration of 8.53mgml^{-1}, following purification using affinity chromatography. The protein was subsequently buffer exchanged into phosphate-buffered saline (PBS) before crystallisation. See section 3.2.1 for full details.

5.2.2 Muramyl disaccharide soaking of native rfhSP-D

Bacterial cell wall muramyl disaccharide, from *Micrococcus luteus* (Dextra Laboratories), was soaked into native rfhSP-D crystals grown in a sitting drop (see section 3.3.1). Cryocooling to 100K was completed following cryoprotection with consecutive additions of 5%-20% 2-methyl-2,4-pentandiol (MPD), diluted with the well conditions (table 5.1) to maintain the equilibrium that permitted crystal formation. A final concentration of 20mM was achieved in the drop following an exchange of the drop volume, 12µl, with 5µl of the cryoprotection buffer. The well conditions are detailed in table 5.1. Crystals were exposed to the highest concentration cryoprotectant for a total of nine minutes before the crystal was mounted, allowing sufficient time for the muramyl disaccharide and MPD to soak through the solvent channels and into the calcium binding sites of the protein.

Table 5.1: Table of well conditions for crystal IB264A53

IB264A53	16% Polyethylene Glycol 6000
	0.1M Tris pH 6.0
	10mM $CaCl_2$

Data collection was carried out at Diamond Light Source, Oxfordshire on macromolecular crystallography beamline I03. One high resolution dataset was collected at a resolution of 1.69Å. Processing of the diffraction data was completed according to section 3.4.2.

5.3 Results of muramyl disaccharide binding

5.3.1 Data processing and data selection

High resolution data was collected from one muramyl disaccharide-soaked crystal, IB264A53. The diffraction images from this crystal were indexed into monoclinic primitive spacegroup $P2_1$, the spacegroup previously determined for the high resolution native structure of the same recombinant fragment of human SP-D used in these studies, with a final unit cell of [Shrive *et al.*, 2003]:

$$a = 55.28\text{Å} \quad b = 108.18\text{Å} \quad c = 55.74\text{Å}$$

$$\alpha = \gamma = 90.0° \quad \beta = 91.14°$$

The spacegroup was confirmed using Pointless during the Pointless-Aimless pipeline which was used to scale the integrated data, generated from the raw diffraction data by MOSFLM; all distributed as part of the CCP4 suite of programs for macromolecular structure solution [Leslie, 2006; Winn *et al.*, 2011; Evans and Murshudov, 2013]. Scaling and merging the data at a maximum resolution of 1.95Å generated good overall R_{merge} statistics of 6.6% and 21.1% in the outer shell. The overall I/σ(I) and CC½ were 8.0 and 0.990, respectively, with the outershell values calculated to be 3.3 and 0.889. Completeness for the dataset was 92.5% overall and 93.5% in the outershell. Structure factors were generated using Truncate and twinning was assessed using cumulative frequency distribution and L-test. The data was not twinned.

The structure was successfully solved by rigid-body refinement (REFMAC5) using the native structure of the same hSP-D fragment solved to 1.6Å with the calcium ions and solvent molecules removed [1PW9, Shrive *et al.*, 2003; Murshudov *et al.*, 2011]. As there was only one dataset, the resulting electron density maps were assessed to identify if ligand was

present in the binding site before continuing restrained refinement. As there was density for the ligand in the calcium-binding pocket, the data was selected for further restrained refinement and model building [Håkansson *et al.*, 1999; Shrive *et al.*, 2003].

5.3.2 Initial map generation and density assessment

The initial maps for the muramyl disaccharide-bound structure were generated using the rigid-body refinement function of REFMAC5 using the native structure of the same recombinant fragment of hSP-D as used in this study as the starting model. All of the calcium ions and solvent molecules were removed from the model, leaving the protein backbone and side chains of residues 205-355 to prevent any bias in the final electron density maps [1PW9, Shrive *et al.*, 2003]. This was sufficient to solve the protein structure and generate two weighted maps (equivalent to the $2F_{obs}$-F_{calc} and F_{obs}-F_{calc}) that were displayed in Coot for density assessment and ligand modelling [Emsley and Cowtan, 2004; Emsley *et al.*, 2010].

The resulting electron density maps of the carbohydrate binding pocket for each subunit are shown in Figure 5.2 with the template rfhSP-D structure displayed to allow for orientation of the binding site. There is good electron density, in both the $2F_o$-F_c and F_o-F_c maps, for the three primary calcium ions in the CRD in all three subunits, suggesting there is good occupation in these three calcium binding sites. The fourth calcium binding site, identified in the neck domain of the native structure, does not appear to have any density in the observed map ($2F_o$) or the difference map (F_o).

In the carbohydrate binding site of subunit A, there is some evidence for calcium coordinating atoms in the observed map [Figure 5.2]. However, the density does not extend beyond the coordinating atoms and suggests that there is no ligand bound in subunit A, allowing for water molecules to coordinate the calcium as in the native structure

[Shrive *et al.*, 2003]. In subunits B and C, both maps describe the primary binding N-acetyl-β-D-glucosamine, with strong density for all of the atoms of the ring and the N-acetyl extension at position 2 of the ring [Figure 5.2]. There is also good resolution for the glycosidic linkage and the main ring N-acetyl-α-D-muramic acid in the observed map of both subunits and some density defining the orientation of the ring in the difference map in subunit B. The O-linked carboxymethyl side chain, in subunit B, has some definition in the initial observed map however the other side chain, the N-linked acetyl group, is not clearly defined. The definition of the two side chains in subunit C is considerably better and the positions of the two side chains is clearly defined, despite the density for the ring being less complete than in subunit B.

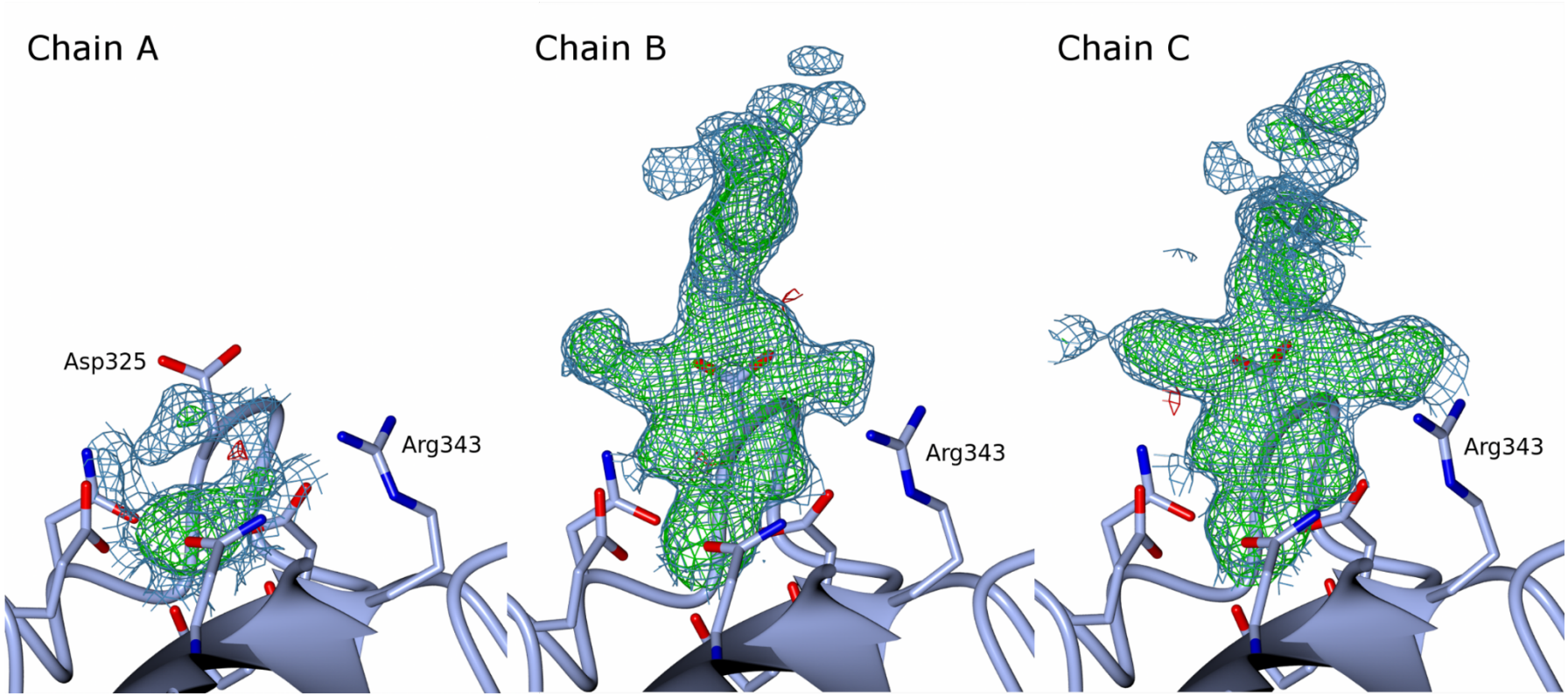


Figure 5.2: Electron density maps from initial rigid body refinement to solve the muramyl disaccharide structure. The $2F_{obs}$-F_{calc} map (1.0 sigma; blue) from subunits B and C show that the ligand is successfully bound, whereas in subunit A, there is no evidence of ligand. This is also the case for the F_{obs}-F_{calc} map (3.0 sigma; green, red), or difference maps, which show strong peaks for the ligand in subunits B and C.

5.3.3 Final structures and binding mechanisms

The overall structure is a trimeric unit containing the alpha-helical coiled coil neck domain and the carbohydrate recognition domain from residue 205 onwards in subunits A and B and 206 in subunit C [Figure 5.3]. The remaining the N-terminal residues (179 - 204) of the recombinant fragment used are not defined in the electron density. The three main calcium ions (Ca1, Ca2 and Ca3) are present in all three subunits however Ca4, identified in the native and other ligand-bound structures, is not present in the structure presented here [Shrive *et al.*, 2003, 2009]. Rounds of refinement were completed following manual corrections of the protein backbone and side chains, addition of the N-acetyl-β-D-glucosamine (GlcNAc) and N-acetyl-α-D-muramic acid (MurNAc), and addition of solvent molecules, with up to 10 cycles of maximum-likelihood phase refinement in each round. Additional rounds of refinement were completed following any further changes to the structure, and the final R-factors were calculated.

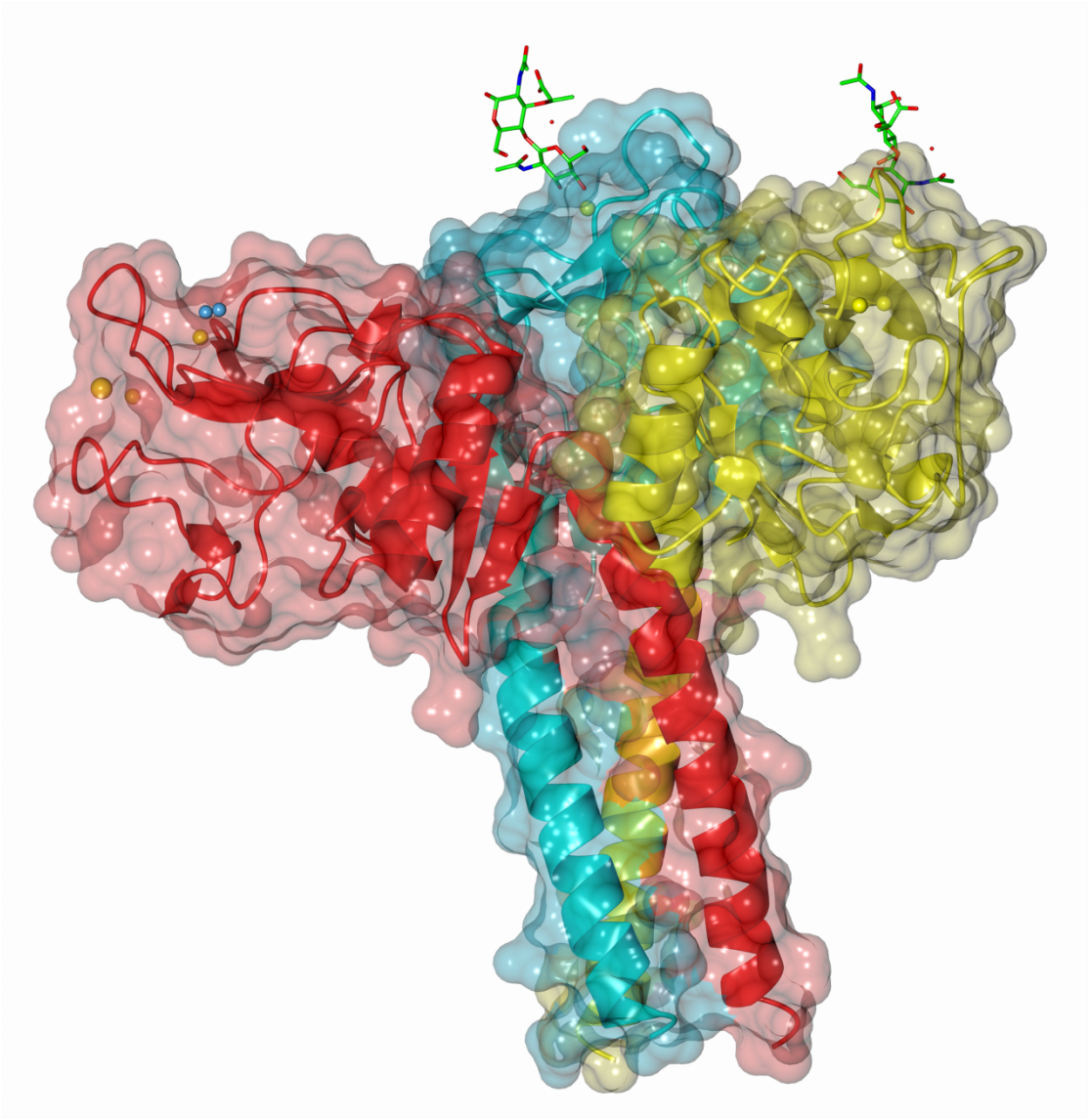


Figure 5.3: Overall structure of rfhSP-D in complex with muramyl disaccharide. The overall orientation of muramyl disaccharide (green) in the ligand binding sites of subunits B (yellow) and C (cyan). Figure shows the ligand extending away from the surface of the CRD of each subunit where it is bound. Calcium ions are represented in yellow and Ca1 coordinating water molecules in blue. Subunit A is represented in red.

Table 5.2: Data collection (DLS I04-1) and refinement statistics from the muramyl disaccharide-bound structures of a recombinant fragment of human surfactant protein D.

	Muramyl Disaccharide
Data collection	
Wavelength (Å)	0.9763
Temperature (K)	100
Space group	$P2_1$
Cell Dimensions	
a (Å)	55.28
b (Å)	108.18
c (Å)	55.74
β (°)	91.14
Maximal resolution (Å)	1.95
Resolution Range (Å)	49.54 – 1.95 (8.94 – 1.95)
Observations	106 081 (6 906)
Unique reflections	44 108 (3 156)
Completeness (%)	92.5 (93.5)
R_{merge}[a]	0.066 (0.211)
I/σ(I)	8.0 (3.3)
CC½	0.990 (0.889)
Refinement	
Protein atoms[b]	3 464
Residues, subunit A	205 – 355
Residues, subunit B	205 – 355
Residues, subunit C	206 - 355
Other atoms	
Calcium ions	9
Ligand	A = 0, B = 2, C=2
Water	336
Resolution range (Å)	55.73 – 1.95
R_{work}[c]	0.1708
R_{free}[d]	0.2169
Average *B*-values (Å^2)	
Protein main chain	28.41
Heteroatoms	32.11
Water	36.39
Ramachandron plot values[e] (%)	
Favoured	97.98
Allowed	2.02
Disallowed	0

[a] $R_{merge} = \Sigma_h\Sigma_j|I_{h,j}-I_h| \,/\, \Sigma_b\Sigma_j|I_{h,j}|$, where $I_{h,j}$ is the *j*th observation of reflection *h* and I_h is the mean of *j* for reflection *h*

[b] Total number of protein atoms used in refinement

[c] $R_{work} = \Sigma_h||F_{oh}|-|F_{ch}|| \,/\, \Sigma_h|F_{oh}|$, where F_{oh} and F_{ch} are the observed and calculated structure factor amplitudes, respectively, for reflection *h*.

[d] $R_{free} = R_{work}$ but for a random 5% subset of reflections

[e] defined by MolProbity analysis

Table 5.3: Calcium coordination distances and protein-ligand interactions in the muramyl disaccharide bound recombinant fragment of human surfactant protein D.

		Muramyl Disaccharide		
Atom 1	**Atom 2**	**A**	**B**	**C**
Ca1	Glu321 OE1	2.67	2.55	2.71
	Asn323 OD1	2.38	2.46	2.44
	Glu329 OE1	2.42	2.46	2.33
	Asn341 OD1	2.49	2.35	2.38
	Asp342 OD1	2.28	2.28	2.29
	Asp342 O	2.61	2.53	2.55
	GlcNAc1 O3′ / W	2.53	2.57	2.51
	GlcNAc1 O4′ / W	2.51	2.53	2.56
GlcNAc1 O3′ / W	Glu321 OE2	-	2.49	2.60
	Asn323 ND2	-	3.08	2.86
GlcNAc1 O4′ / W	Glu329 OE2	-	2.63	2.47
	Asn341 ND2	-	3.06	3.01
GlcNAc1 O6′	Arg343 NH1	-	3.04	2.74
Ca2	Asp297 OD1	2.54	2.79	2.65
	Asp297 OD2	2.45	2.49	2.29
	Glu301 OE1	2.57	2.73	2.62
	Glu301 OE2	2.61	2.89	2.64
	Asp324 OD1	2.46	2.78	2.66
	Glu329 O	2.46	2.46	2.47
	Asp330 OD1	2.39	2.47	2.47
	W	2.28	2.22	2.25

		Muramyl Disaccharide		
Atom 1	**Atom 2**	**A**	**B**	**C**
Ca3	Glu301 OE1	2.35	2.38	2.53
	Asp330 OD1	2.57	2.64	2.50
	Asp330 OD2	2.38	2.77	2.49
	W	2.26	2.38	2.29
	W	2.27	2.40	2.46
	W	2.29	-	2.46
	W	2.50	-	-

All numbers quoted in Å. GlcNAc represents the coordinating N-acetyl-β-D-glucosamine. W represents waters in key binding positions.

The final refinement statistics following completion of the model were 0.1708 and 0.2169 for R_{work} and R_{free} respectively, showing good overall agreement between the model and data (table 5.2). The difference in $R_{factors}$ (0.0461) is larger than expected at 1.95Å however, as that difference was constant throughout the refinement and both $R_{factors}$ fell equally throughout refinement cycles, it can be accepted. The density for both monosaccharides and the glycosidic β-linkage was well defined from the initial electron density maps in subunits B and C, allowing for the whole ligand to be fitted, including the N-linked acetyl groups and the O-linked lactyl group of the muramic acid. There was no density for the ligand in subunit A.

In subunits B and C, calcium-dependent recognition occurs via the N-acetyl-β-D-glucosamine (GlcNAc1) through the O3′ and O4′ equatorial hydroxyl groups at 2.57Å and 2.53Å, respectively, in subunit B and 2.51Å and 2.56Å in subunit C (table 5.3, Figures 5.4 and 5.5). The recognition is completed by a number of protein-ligand interactions that are common to other ligand-bound structures of rfhSP-D, where O3′ is bound via hydrogen bonds by Glu321 (2.49Å in B, 2.60Å in C) and Asn323 (3.08Å in B, 2.86Å in C) and O4′ is coordinated by Glu329 (2.63Å in B, 2.47Å in C) and Asn341 (3.06Å in B, 3.01Å in C). The protein interactions are completed by Arg343 and O6′ hydroxyl of GlcNAc1, through a single hydrogen bond of 3.04Å and 2.74Å (B and C, respectively); an interaction not always present in other glucosyl ligands. In addition, there is also a conserved water bridge between O7′ of GlcNAc1 and Asp325.

Interestingly, the water involved in forming the water bridge between Asp325 and O7′ also interacts with the O10′ on the carboxylethyl side chain of the N-acetyl-α-D-muramic acid (MurNAc1), forming the only major interaction between the protein and the second monosaccharide residue. It was not expected that the second monosaccharide would be in its α-anomeric form, as naturally it is in the β-form, however the density clearly showed the

hydroxyl on carbon 1, the anomeric carbon, was in an axial position in relation to the ring. The remaining coordination between the protein and the MurNAc occurs through an extensive water network that connects the protein, O6′ of MurNAc and N2 of GlcNAc. The N-linked acetyl group of the MurNAc residue does not appear to interact with the protein in the asymmetric unit, however it appears to be stabilised in the crystal through weak interactions with a polar pocket formed by the main chain oxygens of Thr262 and Gln263 (3.53Å to 4.80Å).

The overall orientation of the GlcNAc in the binding site is conserved across both subunits with the exception of two small rotations about the N-acetyl linkage and the C5-C6 bond as a result of an additional water bridge between O6′ and the O-linked lactyl group of the MurNAc residue in subunit B. Interestingly, the rotation of the O6′ position in the C subunit also results in an additional water bridge, not present in subunit B, to the side chain of Glu347, bringing O6′ closer to Arg343 (2.74Å compared to 3.04Å) which may suggest variability in the position of O6′ during recognition [Figures 5.4 and 5.5]. This is compounded by an apparent shift in the side chain of Glu347 in subunit B, which, when compared to the native structure or subunit C of the muramyl disaccharide-bound structure, moves closer to the binding site and interacts with Arg343 but with no direct or indirect interaction with the ligand. The MurNAc residue is rotated approximately 100° with respect to the plane of the GlcNAc residue, around the glycosidic bond, in both subunits where the ligand is present. Importantly, the O1′ hydroxyl, where the next repeating unit of GlcNAc-MurNAc would be in a longer ligand, is completely accessible and positioned so that a longer peptidoglycan unit would be able to extend from away from the protein surface [Figure 5.6]. Furthermore, the lactyl group of the MurNAc is also available and positioned away from the protein surface where, in peptidoglycan, the tetra- or pentapeptide would be linked [Vollmer and Seligman, 2010].

There is very little movement in the side chain of the major binding-site flanking residue Arg343 to facilitate recognition of the O6′ compared to the native structure of rfhSP-D, however, in the opposing flanking residue, Asp325, there is a substantial rotation of 90° in both subunits where the ligand is present [Shrive *et al.*, 2003]. This is surprising since the interactions with the ligand are water mediated in both cases, especially as in other bound structures this is not the case. Furthermore, the proximity of the symmetry related molecule to Asp325, where it forms a crystal contact with Lys230, appears to prevent the O-linked lactyl side chain of MurNAc from rotating (67°) and directly interacting with the Asp325; as is the case in other natural ligand-bound structures [Clark *et al.*, 2016].

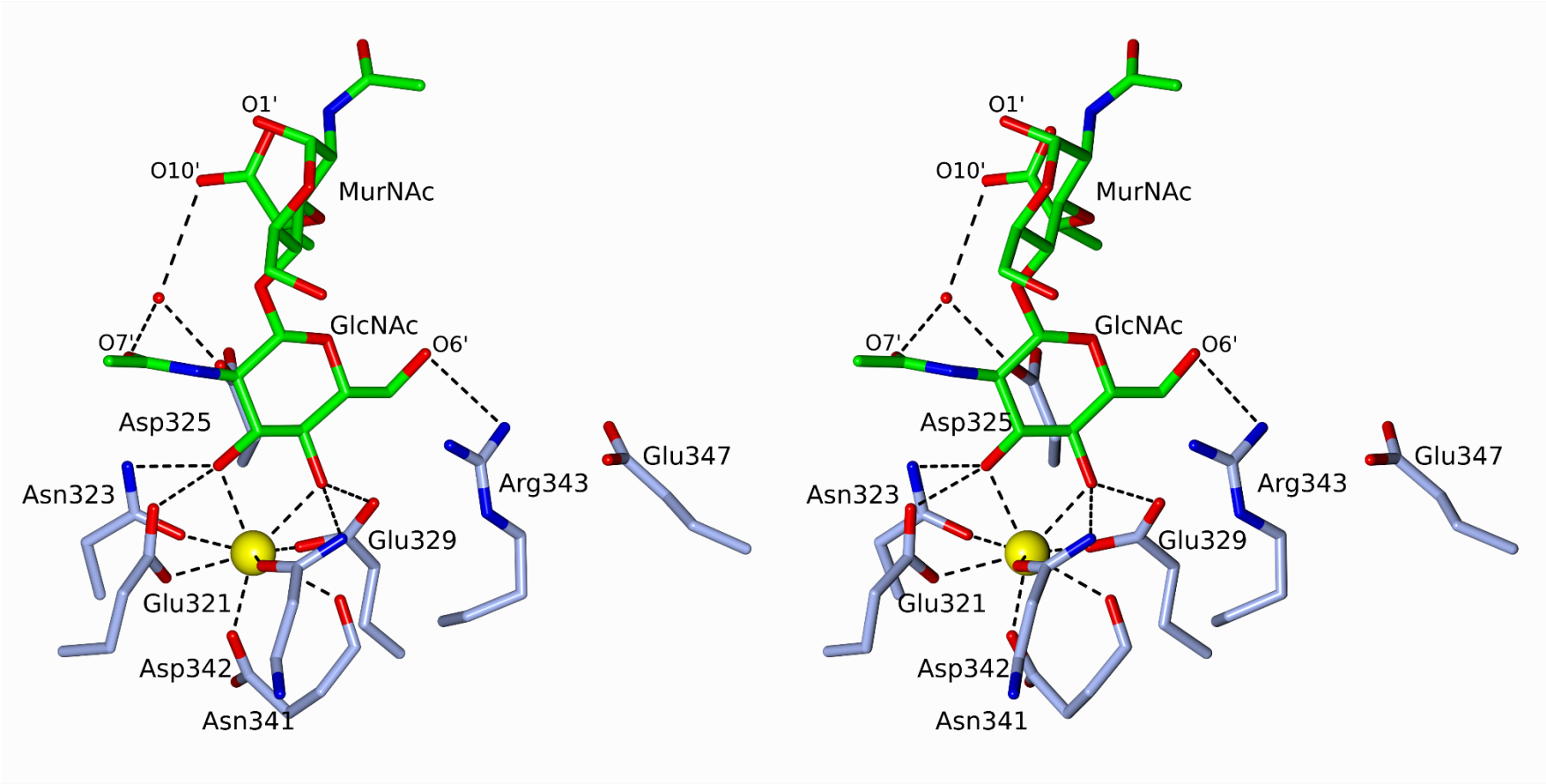


Figure 5.4: Stereo view of the binding mechanism by muramyl disaccharide in subunit B. Binding is through coordination of the calcium ion in the main binding pocket (Ca1, yellow) in muramyl disaccharide complex with rfhSP-D. The main coordination is via the O3′ and O4′ hydroxyls of GlcNAc residue. Recognition is completed by Glu321, Asn323, Glu329, Asn341 and Arg343 (which interacts with the O6′ hydroxyl. The interactions with the MurNAc residue are completed by a water bridge (red sphere) between Asp325, GlcNAc and the lactyl group of MurNAc.

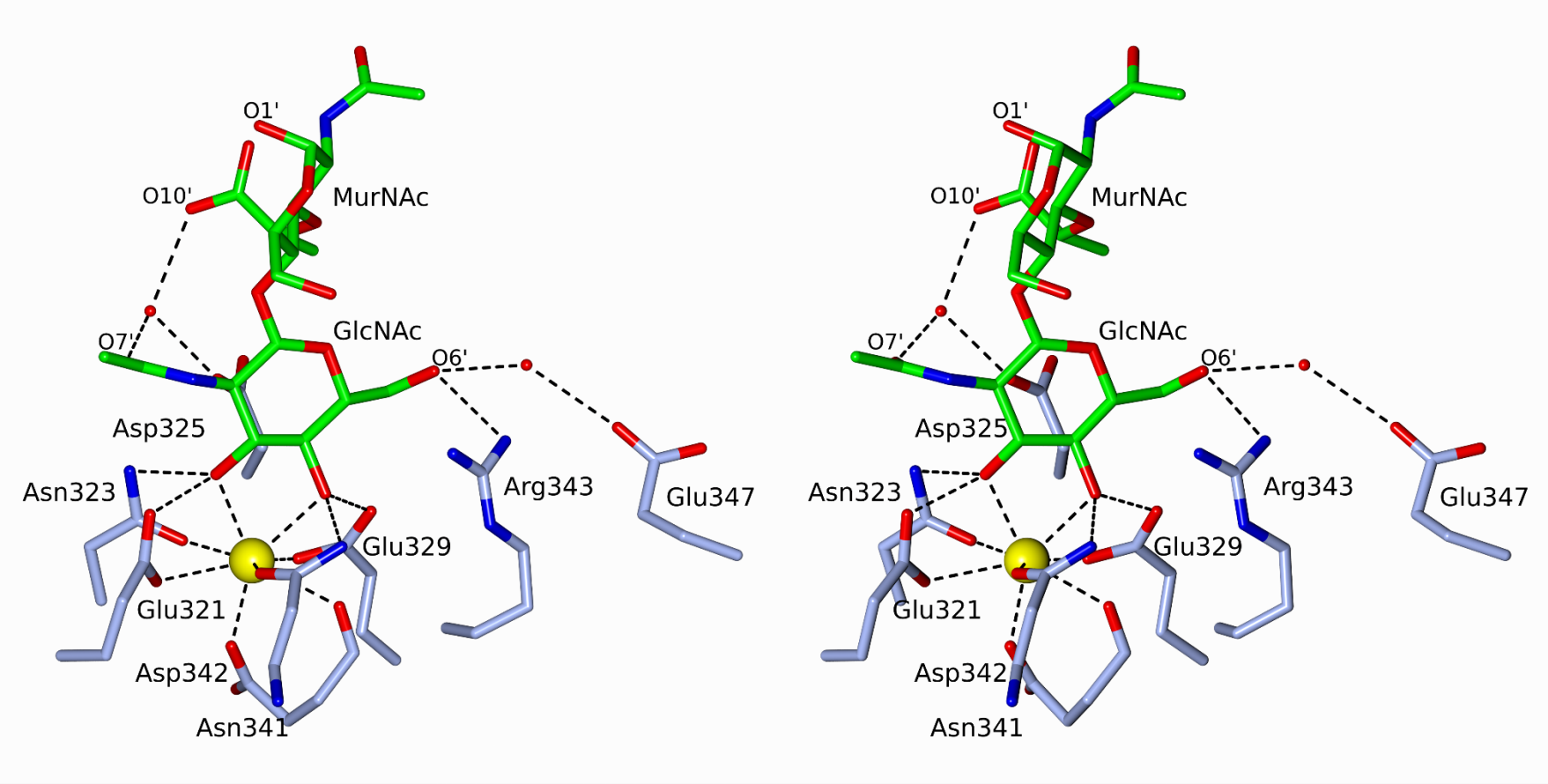


Figure 5.5: Stereo view of the binding mechanism by muramyl disaccharide in subunit C. Binding is through coordination of the calcium ion in the main binding pocket (Ca1, yellow) in muramyl disaccharide complex with rfhSP-D. The main coordination is via the O3′ and O4′ hydroxyls of the GlcNAc residue. Recognition is completed by Glu321, Asn323, Glu329 and Asn341, with an additional hydrogen bond with Arg343. Distance to Arg343 is shorter in subunit C compared to subunit B. The water bridge (red sphere) between Asp325, GlcNAc and the lactyl group of the MurNAc residue.

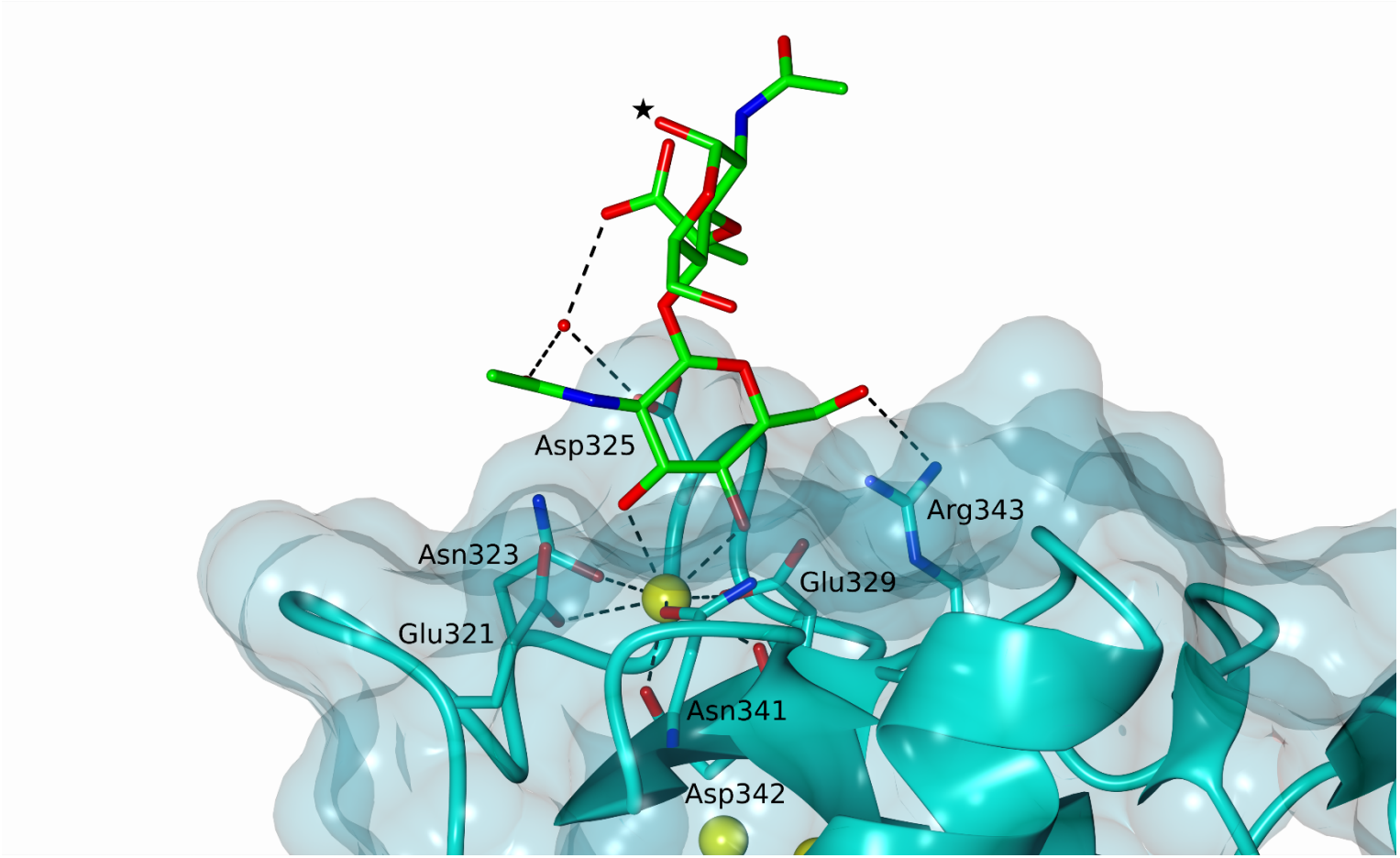


Figure 5.6: Subunit C of the muramyl disaccharide complex showing the muramyl disaccharide extending away from the surface of the protein. The figure shows that the ligand extends away from the protein surface (cyan, transparent), showing a longer ligand can be accommodated. (star) position of the MurNAc where further disaccharides extend from in longer ligands.

5.4 Discussion

5.4.1 Recognition of muramyl disaccharide

Recognition of gram-negative bacteria has previously been investigated using both naturally derived inner core polysaccharides and other synthetic analogues such as α(1→3)-diheptose [Wang *et al.*, 2008; Clark *et al.*, 2016; Reinhardt *et al.*, 2016]. For the first time, the structure presented here provides a much-needed insight into how hSP-D recognises the structurally distinct group of gram-positive bacteria that have a peptidoglycan-rich cell wall which is exposed to the surrounding environment [van de Wetering *et al.*, 2001; Palaniyar, Nadesalingam and Reid, 2002]. Muramyl disaccharide is the repeating unit of the structural polysaccharide in cell wall peptidoglycan, containing a β(1→4)linked N-acetyl-β-D-glucosamine and N-acetyl-β-D-muramic acid (GlcNAc-β(1→4)-MurNAc), that, in this case, was extracted and hydrolysed from the gram-positive bacteria, *Micrococcus luteus*; making it a suitable and natural analogue for a longer peptidoglycan molecule [Dextra Laboratories].

The muramyl disaccharide was primarily recognised through the O3′ and O4′ hydroxyls of the GlcNAc residue. This conforms to the well-established recognition pattern of the calcium ion in the primary binding site, Ca1, coordinating with two equatorial hydroxyls on adjacent carbon atoms of the main carbohydrate ring or, in the case of heptose, the dihydroxyethyl side chain [Shrive *et al.*, 2003, 2009; Wang *et al.*, 2008; Crouch *et al.*, 2009]. Coordination by these two hydroxyls of GlcNAc results in the possible ring orientations being restricted by the N-linked acetyl group at carbon 2 because of the proximity of Arg343 to the binding pocket. This confirms the findings of Allen and his colleagues in 2004, who described an R343V mutant with increased affinity for GlcNAc [Allen *et al.*, 2004]. A similar effect is also alluded to in the N-acetyl-D-mannose-bound (ManNAc-bound) structure, where the ManNAc is recognised in a comparable way to GlcNAc, with O3′ and

O4' functioning as the calcium coordinating atoms, and the flanking Arg343 forcing the N-acetyl group into a small pocket above Glu321 and Asn323 [Figure 5.4; Crouch *et al.*, 2007]. However, there is no direct interaction between the N-acetyl side and any of the side chains in the GlcNAc pocket here, other than a weak interaction between the carbonyl group of the acetate and ND2 of Asn323 (up to 3.59Å) that is not present in the ManNAc-bound structure.

Interestingly, the ManNAc structure does have an extra interaction between the nitrogen of the mannosamine moiety and Asp325, in all three chains as a result of the axial position of the nitrogen [Crouch *et al.*, 2007], compared to the equatorial position in GlcNAc. This is despite a much larger rotation of Asp325 in the GlcNAc-bound structure. The rotation moves the interacting oxygen (OD2) closer, overall, to the GlcNAc, possibly suggesting that the carbonyl of the N-linked acetyl group, which interacts weakly with Asn323, is also interacting weakly with Asp325 through a long hydrogen bond (~4.0Å) and brings about this change in orientation, reinforcing the suggestion that Asp325 is directly involved in selecting and orientating ligands in the binding site of hSP-D.

Beyond the well characterised interactions present in many other ligand-bound structures of rfhSP-D, GlcNAc also interacts with Glu347 through a water bridge in subunit B. The importance of this water bridge is unclear as it is not present in subunit C, however it is sufficient to cause the side chain of Glu347 to move towards the binding site, suggesting that Glu347 may have larger role in ligand recognition than was originally thought, as this is the first structure to demonstrate this shift in position. It has been previously suggested that Glu347 may play a role in ligand recognition due to its close proximity to the main binding site, potentially working in tandem with Arg349, as both groups are conserved in surfactant protein A and D, however this is the first evidence of any involvement and something that requires further investigation [Håkansson *et al.*, 1999].

The N-acetyl-muramic acid residue does not appear to make any direct interactions with the protein which is perhaps surprising considering the good definition of the ring in the electron density. It appears that the extensive, water-mediated hydrogen bond network is sufficient to stabilise the MurNAc. The strongest of these interactions (2.82 Å) is a single water bridge between Asp325, the lactyl side chain of MurNAc and the acetyl group of the GlcNAc residue which is conserved in subunits B and C. It is interesting, however, that the proximity of the symmetry related molecule to the MurNAc potentially exerts a stabilising effect on the ring through long interactions between the N-acetyl group and the main chain carbonyls. This has not been identified in previous disaccharide-bound structures, where definition of the second ring is usually present in subunit A, suggesting that the size of the muramyl disaccharide unit and its side chains probe this polar pocket of the symmetry related molecule where smaller, glucosyl ligands, are unable to [Shrive *et al.*, 2003, 2009]. The size of muramyl disaccharide may also explain why the ligand was unable to bind in subunit A as the symmetry related molecule restricts the binding site.

There is also a rotation of the MurNAc in relation to the plane of the GlcNAc ring, in subunits B and C, which may be an artefact from the natural peptidoglycan backbone source of the disaccharide unit used in this study. A 3-fold and 4-fold screw axis around the oligosaccharide core has been observed in complete sulci, peptidoglycan fragments and synthetic peptidoglycan fragments from various gram-positive and gram-negative species, with pentapeptides extending away from the core every 120° and 90°, respectively [Vollmer, Blanot and De Pedro, 2008; Ślusarz, Szulc and Madaj, 2014; de Pedro and Cava, 2015; Kim, Chang and Singh, 2015]. This is further supported by the positions of the N-linked acetyl groups that are rotated approximately 120° in relation to each other, which would produce the 3-fold screw axis in a longer chain. However, the proximity of the symmetry related

molecule will also be exerting an effect on this orientation and any rotation may be more directly associated with the accessibility of the binding site.

5.4.2 Possible rotation of MurNAc and an extended role for Asp325

The role of Asp325 in ligand recognition has been well studied using both native and D325A/R343V mutants and has been shown to have a selective role for ligands through either blocking the main calcium binding site, in combination with Arg343, or directly interacting with the ligand to improve binding affinity [Crouch *et al.*, 2005, 2009; Clark *et al.*, 2016]. It is suggested here that Asp325 may also be involved in recognising muramyl disaccharide and, by extension, peptidoglycan from bacterial cell wall. It is proposed that the proximity of the symmetry related molecule, whilst stabilising the MurNAc residue and making it visible in the electron density map, may also be preventing the carboxylic acid of the O-linked side chain from interacting with Asp325; with which it would otherwise interact naturally. In more general terms, Asp235 is an important recognition residue for longer, naturally derived ligands and may contribute to the selection of peptidoglycan as a ligand through the positioning of the terminal N-acetyl-glucosamine in the correct orientation for the calcium to coordinate with the O3' and O4' hydroxyls.

The role of the Asp325 rotation in the muramyl disaccharide-bound structure is not clear because in other ligand-bound structures, including the *H. influenzae* Eagan 4A polysaccharide-bound structure, the rotation is not observed [Crouch *et al.*, 2006; Wang *et al.*, 2008; Shrive *et al.*, 2009; Goh *et al.*, 2013; Clark *et al.*, 2016]. This is particularly of interest because it is conserved in both subunits where muramyl disaccharide is bound, with the Asp325 accommodating the native position in subunit A where no ligand is bound. This suggests an active role in the recognition of the disaccharide despite not making any direct interactions with the GlcNAc or MurNAc beyond a water bridge to the acetyl side chain of GlcNAc. It is worthy of note, however, that out of the many ligand-bound rfhSP-D

structures checked by model superposition, the water involved in the water bridge to GlcNAc only appears to be present in the muramyl disaccharide structure, suggesting that Asp325 may be rotating to fulfil this water interaction specifically for GlcNAc [Shrive *et al.*, 2003, 2009, Crouch *et al.*, 2007, 2009; Wang *et al.*, 2008; Clark *et al.*, 2016].

5.4.3 Longer peptidoglycan chain recognition

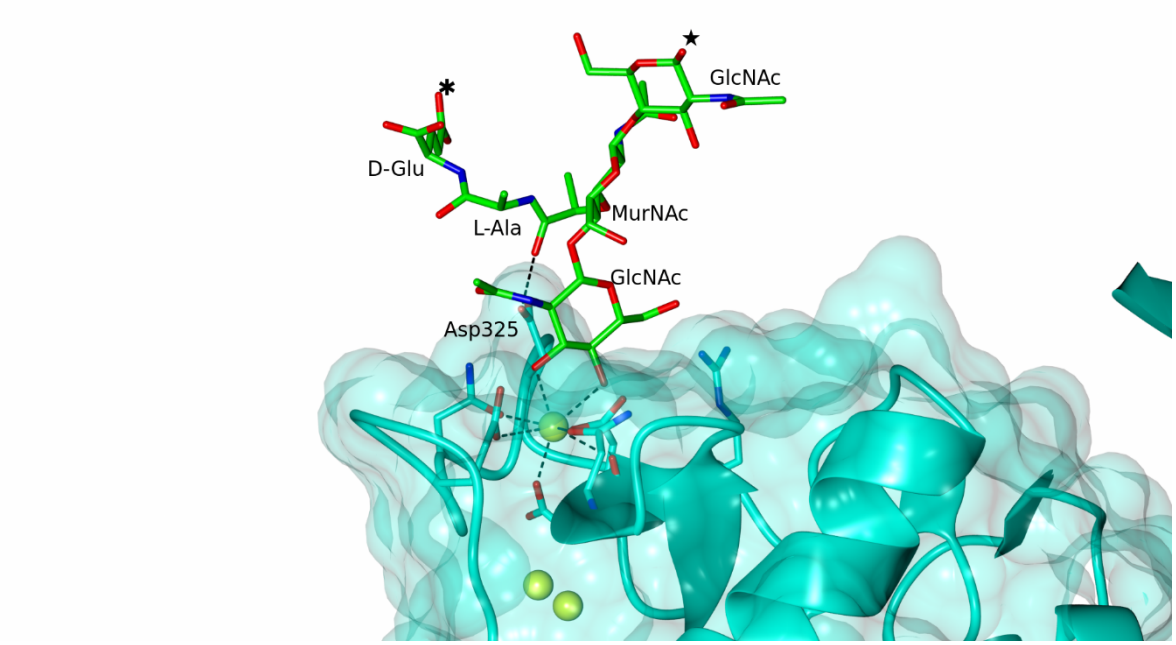


Figure 5.7: Simple model of a longer fragment of peptidoglycan. Model of a longer fragment of peptidoglycan, produced by superposition of the fragment onto the crystal structure, demonstrates a longer peptidoglycan can be accommodated by the surface of rfhSP-D (cyan). The lactyl group of MurNAc is able of interacting directly with Asp325 whilst still allowing the oligopeptide to extend above the surface (two amino acids shown). (star) oligosaccharide core extension; (*) position where remaining peptide extends.

Peptidoglycan is naturally a long chain polymer of muramyl disaccharide linked by β(1→4) glycosidic bonds with tetra- or pentapeptides O-linked to the lactyl moiety of the N-acetyl-muramic acid [Schleifer and Kandler, 1972]. Therefore, it is important to consider whether longer peptidoglycan chains could be accommodated by surfactant protein D. Simple modelling by superposing a longer peptidoglycan fragment onto the muramyl disaccharide structure reveals that the polysaccharide backbone of the peptidoglycan extends away

from the surface of the carbohydrate recognition domain without making any further interactions with the protein [Figure 5.7]. Similarly, the oligopeptide extensions of the peptidoglycan can also be accommodated by the surface of hSP-D, as the peptide would extend above Asp325 and the rest of the protein based on the lactyl group position in the disaccharide-bound structure. Moreover, if the lactyl group is able to directly interact physiologically with Asp325, as is proposed here, the position of the carboxyl group of the lactyl side chain would also permit the peptide extension without clashing with the CRD [Figure 5.7].

This model suggests that the shorter, soluble peptidoglycan and the longer, insoluble peptidoglycan could both be recognised by hSP-D if a GlcNAc residue is available at the non-reducing termini of the polymer. The GlcNAc residue is clearly available in both types of peptidoglycan as MBP was shown to recognise GlcNAc in the soluble and insoluble fragments, with high affinity [Nadesalingam *et al.*, 2005]. It is not clear from this structural study if the affinity for muramyl disaccharide reflects the high affinity binding observed for MBP, however it has previously been shown that the affinity between wildtype hSP-D and an insoluble peptidoglycan from *S. aureus* is lower than the affinity for lipoteichoic acid and can be completely blocked by 10 mM maltose, mannose or glucose, suggesting that the interaction may be quite weak [van de Wetering *et al.*, 2001]. This is not surprising as hSP-D has been shown to have a relatively low affinity for GlcNAc, compared to its preferred ligand maltose, with an IC_{50} of between 9.8 mM and 14 mM being reported [Crouch *et al.*, 2005; Crouch *et al.*, 2006]. The model of the longer peptidoglycan fragment may also go some way to explaining the lower affinity, as the peptide extensions may cause some steric clashes that cannot be directly observed since they were not included in the actual soaked ligand used here. Conversely, the peptide extensions may increase the affinity of a single muramyl disaccharide unit by interacting with other parts of the protein. However, for any

conclusion to be drawn, there is a considerable amount of further work required using longer, intact peptidoglycan from soluble and insoluble fragments.

5.4.4 Implications for GlcNAc recognition

Recognition of a terminal GlcNAc residue also extends the number of possible ligands and their binding modes that hSP-D can recognise, particularly gram-negative lipopolysaccharides and other glycosylated proteins. It is well established that gram-negative lipopolysaccharides (LPS) contain a range of conserved, LPS-specific monosaccharides such as the inner core heptose and 3-deoxy-D-*manno*-oct-2-ulosonic acid residues, but beyond this core motif there are a number of other monosaccharides that complete the inner and outer core of LPS, including N-acetylglucosamine [Rund *et al.*, 1999; Kabanov and Prokhorenko, 2010; Holst, 2016]. These include the LPS of most important pathogens such as *Escherichia coli*, which in some strains contain terminal GlcNAc residues in the outer core, and *Neisseria meningitidis*, where a GlcNAc is bound to HepI of the inner core polysaccharide [Michael *et al.*, 2009; Holst, 2011]. In both of these examples, the two O3' and O4' hydroxyls are thought to be available and unsubstituted, making them good targets for hSP-D based on the data presented here.

Chapter 6: Recognition of R7 rough mutant lipopolysaccharide from *Salmonella enterica* sv. Minnesota

As part of the ongoing investigation of LPS/SP-D interactions the structures of rfhSP-D complexed with *Salmonella enterica* sv. Minnesota R5 and R7 polysaccharide (Sm-R5 and Sm-R7 respectively) have been revisited in order to resolve several unanswered questions and provide a definitive description of both the bound ligand and the interaction with hSP-D [Smallcombe, 2014; da Silva, 2016]. The fully refined structure of the Sm-R5 complex has now been determined by other members of the Research Group and has been provided for comparison with the R5 structure whose re-analysis and full refinement is described here and in Appendix II which contains Littlejohn *et al.*, 2018, " Structural definition of hSP-D recognition of *Salmonella enterica* LPS inner core oligosaccharides reveals alternative binding modes for the same LPS."

6.1 Introduction to rough mutant *Salmonella enterica*

6.1.1 Gram-negative rough mutants

Gram-negative bacteria are adorned with a range of glycosylated structures on their outer membrane, one of which is the highly immunogenic, lipopolysaccharide (LPS) [Kabanov and Prokhorenko, 2010]. There are two major forms of LPS, smooth LPS and rough LPS, which vary between species and strains. Smooth LPS contain the inner core polysaccharide, outer core polysaccharide and the O-antigen, a long, repeating, antigen-specific oligosaccharide that extends from the surface of the bacteria [Erridge, Bennett-Guerrero and Poxton, 2002; Holst, 2011]. Rough LPS strains, on the other hand, lack the repeating O-antigen, leaving

the inner and outer core polysaccharides exposed to the immune system [Holst, 2016]. In many species, including *S. enterica*, the core polysaccharides are a conserved motif, providing a possible recognition site for the innate immune system and human surfactant protein D (hSP-D).

6.1.2 Rough mutants of *Salmonella enterica* sv. Minnesota

While the core polysaccharide is largely conserved in *S. enetrica* Minnesota, the extent of the core varies due to mutations in the synthetic enzymes responsible for assembling the LPS, resulting in truncated polysaccharide cores [Lüderitz, Staub and Westphal, 1966; Mansfield and Forsythe, 2001]. Synthesis of the LPS is sequential and, therefore, requires the correct arrangement of saccharide residues, including any phosphate substitutions, for the next monosaccharide to be added. If there are any interruptions to this, the remaining polysaccharide is not attached and the LPS is expressed in its truncated form [Figure 6.1].

The R7 mutant of *S. enterica* Minnestoa is a deep rough mutant strain that lacks the outer core oligosaccharides of the Ra mutant and GlcI, GalI and HepIII of the Rb3 mutant, equivalent to the Rd1 mutant in Figure 6.1. This leaves a short inner core of an α(1→3)-linked diheptose linked to the α(2→4)-linked tri-Kdo through a single α(1→5) between HepI and KdoI. Human SP-D has been shown to recognise both R7 and R5 mutants, Rd1 and Rc phenotypes respectively, however how this recognition occurs, and by which monosaccharide residue, has not been elucidated.

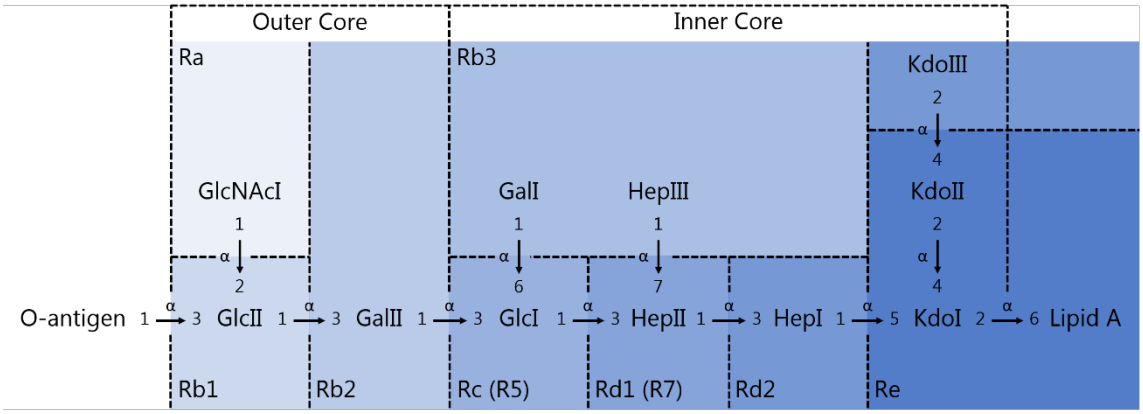


Figure 6.1: Graphical Schematic of the Core Polysaccharide from a Lipopolysaccharide of *Salmonella enterica*. The figure summarises the core polysaccharide from a lipopolysaccharide of *Salmonella enterica*, including all glycosidic bond donors and acceptors (numbered according to the associated carbon atom) and the geometry of the bond (i.e. α). The lines and colours (shades of blue) represent where different mutants of S. enterica extend to in the core polysaccharide (i.e. the Ra mutant includes the whole core polysaccharide and is represented by the palest blue). Abbreviations: Kdo, 3-deoxy-D-manno-oct-2-ulosonic acid; Hep, heptose, Glc, glucose; Gal, galactose; GlcNAc, N-acetyl-glucosamine. Figure based on data from: [Mansfield and Forsythe, 2001]

6.1.3 Recognition of inner core polysaccharides from lipopolysaccharides

Recently, the structure of a rough mutant *H. influenzae* Eagan 4A was successfully crystallised in complex with a biologically and therapeutically active C-terminal fragment of hSP-D [Clark *et al.*, 2016]. The LPS in this study contained three core monosaccharide residues: Kdo, heptose and a β-D-glucose linked by α(1→5) and β(1→3) glycosidic linkages, respectively. Recognition occurred through calcium-dependent binding to the dihydroxyethyl side chain of the proximal heptose and identified an anhydrous form of Kdo which made a further hydrogen bond to the Asp325 side chain, completing the recognition of the core. It is thought that the dominant anhydrous Kdo was a product of mild hydrolysis, which was required to cleave the lipid A portion of the LPS, forming an enantiomeric 5-membered furanoid derivative of the Kdo as a result of a β-elimination of a phosphate group.

Recognition of the inner core heptose by hSP-D has also been explored using synthetic LPS cores, which showed that the binding affinity for the cores containing a heptose was higher than those where a heptose was not available for recognition [Reinhardt *et al.*, 2016]. This suggests that the heptose is a key residue for LPS recognition, supporting the findings of *Clark* and colleagues in the Eagan-bound structure [Clark *et al.*, 2016]. The R7 rough strain of LPS allows for the fidelity of this proximal heptose recognition where more complex inner core LPS oligosaccharides are present that contain more than one heptose residue.

To complete the characterisation of the R7 polysaccharide and the binding mechanism that hSP-D uses to recognise the inner core of rough mutants of *S. enterica*, the three dimensional structure of a biologically and therapeutically active fragment of the C-terminal of hSP-D (rfhSP-D) in complex with a delipidated LPS from *S. enterica* Minnesota R7 mutant strain has been determined by x-ray crystallography. The data presented here was originally collected by Dr. Carrie Smallcombe and Dr. Ruben da Silva, however the data was reprocessed from the raw diffraction images as significant questions remained about the interaction between hSP-D and the polysaccharides from Sm-R7 and Sm-R5 [Smallcombe, 2014; da Silva, 2016].

6.2 Materials and methods

6.2.1 Preparation of *S. enterica* Minnesota R7 oligosaccharide

The delipidated R7 oligosaccharide portion of *Salmonella enterica* sv. Minnesota (Enzo Lifesceinces-ALX-581-018-L002) was prepared by a previous PhD student, under mild hydrolysis conditions with 2% acetic acid, following the method published by *Masoud* and co-workers and *Clark et al.* [Masoud *et al.*, 1994; Smallcombe, 2014; Clark *et al.*, 2016]. The molecular weight of the whole R7 LPS was calculated to be 2863 Da by the addition of all glycoconjugates and Lipid A. The lipid A portion accounted for approximately 1800 Da, making up ≈63% of the total LPS [Que *et al.*, 2000]. 2 mg of R7 LPS were hydrolysed, yielding approximately 0.57 mg of the oligosaccharide which represents a hydrolysis efficiency of 77%. The resulting R7 oligosaccharide was dissolved in deionised water to give a final KdoI-HepI-HepII (accounting for the β-elimination of the KdoII-KdoIII extension, see section 6.3.3) concentration of ≈18 mM stock that was used for crystal soaking and cryoprotection.

6.2.2 Protein preparation

The rfhSP-D was received from the Southampton group at a concentration of 8.8 mg ml^{-1} in phosphate buffered saline (PBS), following purification by affinity chromatography. An additional 10 mM $CaCl_2$ was added to the protein before crystallisation to ensure the protein was calcium-bound and able to bind the R7 ligand.

6.2.3 R7 Oligosaccharide soaking of native hSP-D crystals

The delipidated R7 oligosaccharide was soaked into native rfhSP-D crystals grown in a sitting drop (see section 3.3.2). A final concentration of ≈13 mM of the KdoI-HepI-HepII portion of the oligosaccharide was achieved in the drop following an exchange of the drop volume, 6 µl, with 6 µl of cryoprotection buffer. Flash cryocooling to 100 K was completed

following cryoprotection with consecutive additions of 5%-15% 2-methyl-2,4-pentandiol (MPD), diluted with the well conditions, in 2 µl aliquots. The well conditions are detailed in table 6.1. Crystals were soaked with the R7 oligosaccharide for six minutes before cryocooling to allow the ligand to saturate the crystal.

Table 6.1: Table of Well Conditions for Crystal CCS16A11

CCS16A11	16% Polyethylene Glycol 10 000 0.1 M Tris pH 7.0 10 mM $CaCl_2$

6.2.4 Data collection

Data for the rfhSP-D in complex with R7 oligosaccharide were collected at 100K at Diamond Light Source, Oxfordshire, on macromolecular crystallography beamline I03, using the Dectris Pilatus 6M hybrid photon counting detector. A single, high resolution dataset was collected from crystal CCS16A11 which was processed following the protocol outlined in section 3.4.2.

6.3 Results

6.3.1 Data processing

The high resolution dataset collected from the crystal CCS16A11, soaked with the delipidated R7 oligosaccharide, has been reprocessed following recent results in the research group [Smallcombe, 2014]. The diffraction images were indexed in monoclinic primitive spacegroup $P2_1$ using MOSFLM [Leslie, 2006; Battye *et al.*, 2011]. This is in agreement with the previously solved native structure of the same recombinant fragment of hSP-D used in these studies [Shrive *et al.*, 2003]. The final unit cell was refined to:

$$a = 55.33\text{Å} \quad b = 108.14\text{Å} \quad c = 55.67\text{Å}$$

$$\alpha = \gamma = 90° \quad \beta = 91.82°$$

Pointless confirmed the spacegroup during the Pointless-Aimless pipeline and was used to scale the integrated data, distributed as part of the CCP4 suite of programs for macromolecular structure solution [Evans, 2006; Winn *et al.*, 2011; Evans and Murshudov, 2013]. Scaling and merging the data at a maximum resolution of 1.75Å generated an R_{merge} of 35.9% in the highest resolution bin (6.6% overall); an I/σ(I) of 7.3 overall and 2.2 in the outer bin; CC½ of 0.991 overall and 0.774 in the outer bin. Structure factors were generated using Truncate and the twinning was assessed using cumulative frequency distribution and L-test statistics. The data was not twinned and was taken forward for further refinement.

The structure was successfully solved using the rigid-body refinement function in REFMAC5 using the native structure of hSP-D, solved to 1.6Å with the calcium ions and solvent molecules removed [1PW9, Shrive *et al.*, 2003; Murshudov *et al.*, 2011]. The resulting electron density maps were assessed, using the native structure of hSP-D for orientation, to identify if there was density for the ligand in the binding pocket. The density for the ligand

was present in subunits B and C so the dataset was selected for further restrained refinement and model building.

6.3.2 Initial map generation and density assessment

The initial maps for the R7 oligosaccharide-bound complex were generated using the rigid-body refinement function of REFMAC5 using the native structure of the same recombinant fragment of hSP-D used in this study with all calcium ions and water molecules removed, leaving only the protein backbone and sidechains to prevent any bias in the final map [1PW9, Shrive *et al.*, 2003]. The resulting weighted maps (equivalent to the $2F_{obs}$-F_{calc} and F_{obs}-F_{calc} maps) were displayed, with the solved protein structure as a template, in Coot for density assessment and ligand fitting [Emsley and Cowtan, 2004; Emsley *et al.*, 2010].

The rigid-body map for each carbohydrate binding site, prior to any refinement, is shown in Figure 6.2 with the rigid-body template displayed to allow for orientation. The calcium ions in all three binding sites are clearly visible in the electron density from both maps, $2F_o$-F_c (blue) and F_o-F_c (green), in a similar position to the native rfhSP-D structure. There is also good resolution of the remaining two calcium ions in the CRD which also appear to occupy the same positions as the native. There was no density for the calcium, Ca4, in the neck domain that has previously been present in the native and other bound structures of rfhSP-D [Shrive *et al.*, 2003, 2009].

In subunits B and C, there is evidence for a non-terminal heptose coordinating the calcium in the main binding pocket (Ca1) with good electron density for the heptose in the difference map of both subunits (green, Figure 6.2). There is further density in the observed map (blue) for the two additional monosaccharides that make up the R7-oligosaccharide. The density for the second heptose (HepII) is less well defined than the calcium coordinating heptose but the main ring has good definition in the electron density

with some density extending in the direction of the hydroxylethyl side chain (C6′ and C7′). There is also evidence for the Kdo in the density of both subunits, extending between the two flanking residues, Asp325 and Arg343. From the density, the Kdo appears to be in an anhydrous furanose form, with some density extending in the direct of the 2-oxobutanoic acid. In both subunits, the difference map supports the occupancy of all three monosaccharides that make up the R7 oligosaccharide. There was no clear density for any ligand bound into subunit A.

Electron density extends across the protein backbone from residue 205 in subunits A and C, with extra density for residue 204 being present in subunit B. At the N-terminal, the side chains are less well defined than the highly ordered side chains associated with the lectin fold of the CRD. The remaining N-terminal of the recombinant fragment (residues 179 – 204/3/4) is not defined in the electron density and cannot be determined in the structure.

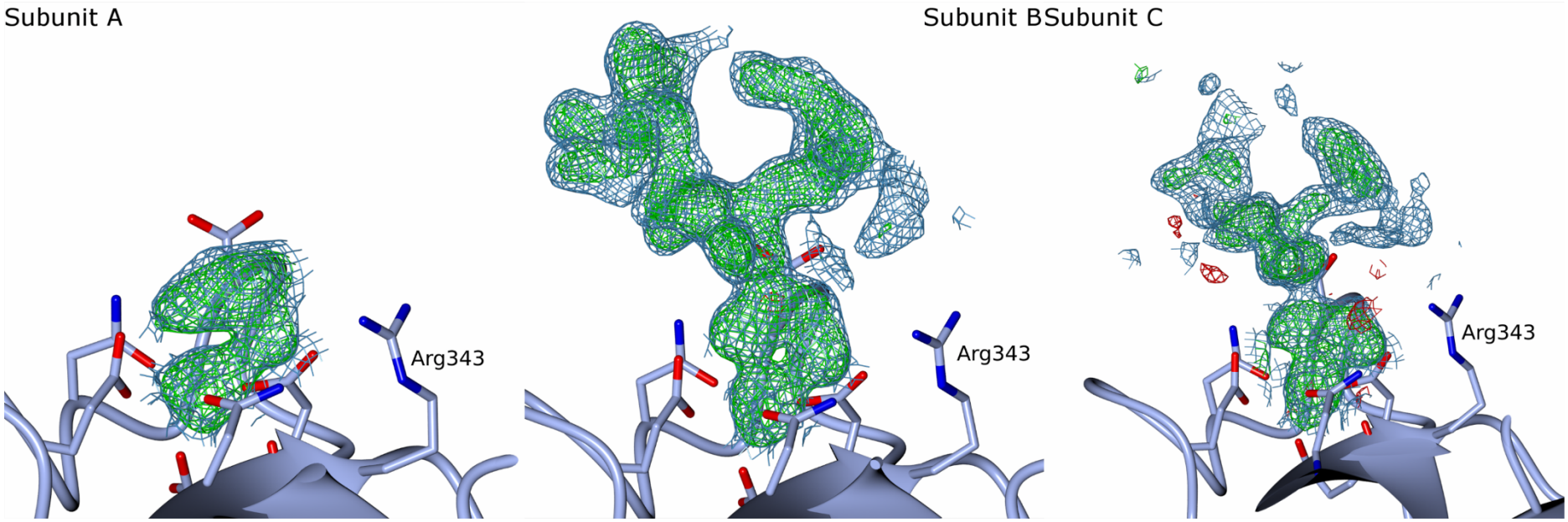


Figure 6.2: Rigid-body electron density maps for the *Salmonella enterica* R7-oligosaccharide in complex with rfhSP-D. The observed electron density map is displayed at 1.0σ ($2F_{obs}$-F_{calc}, blue) and difference map at ±3.0σ (F_{obs}-F_{calc}, green (+) and red (-)) from the rigid body solution. The template structure is displayed to allow for orientation of the binding site. The electron density shows that subunit A does not have any ligand in the binding site whereas there is clearly electron density for the *Salmonella enterica* R7 oligosaccharide in the calcium (Ca1) binding site.

6.3.3 Ligand characterisation

The electron density for the R7 oligosaccharide clearly defines the KdoI-HepI-HepII [Figure 6.3] and the Kdo is in the anhydrous form first reported by *Clark* and coworkers in the *H. influenzae* Eagan 4A bound rfhSP-D structure [Clark *et al.*, 2016]. However, more detail of the 5-membered anhydro-Kdo ring substituents are present in the R7 oligosaccharide, with definition of the 2-oxobutanoic acid sidechain at C4 in a beta conformation, with respect to the glycosidic bond [Figure 6.4]. The anhydro-Kdo is indicative of a β-elimination of the KdoII-KdoIII which, as would be expected, is not present in the electron density at position 4 of the Kdo as the hydroxyl group is now involved in completing the furanose ring.

Both heptose residues (HepI and HepII) are in their α-anomer, as you would expect for the R7 oligosaccharide core. The anhydro-Kdo is in its β-form, where the 2-oxobutanoic acid side chain extends from the main 5-membered ring equatorially, following the plane of the ring. This suggests the β-form predominates the α-form following β-elimination of the KdoII-KdoIII fragment.

Figure 6.3: Structure of the R7 Oligosaccharide Product of Mild Acid Hydrolysis. Mild hydrolysis of the R7 LPS removes the lipid A and KdoI-KdoII, leaving the product with an *anhydro-Kdo. Hep, L-D-heptose; KdoI*, 4,7-anhydro-3-deoxy-D-manno-oct-2-ulosonic acid furanose.

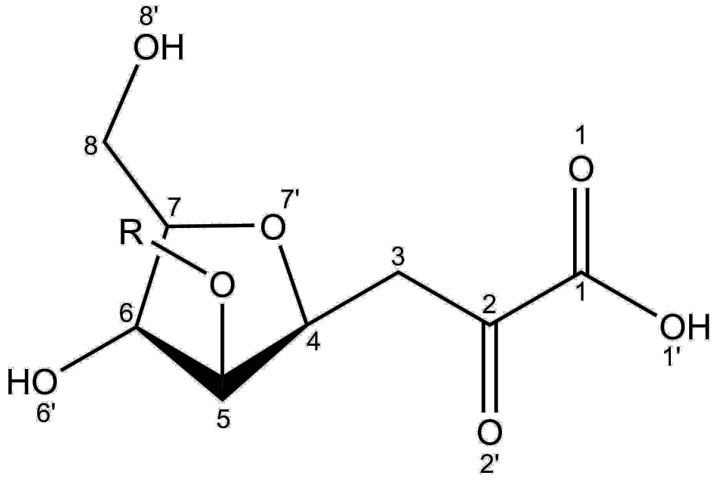


Figure 6.4: Structure and Numbering of the 4,7-anhydro-3-deoxy-D-manno-oct-2-ulosonic acid Furanoid Derivative. Numbering is retained from the KD5 structure associated with PDB entry 4E52 [Clark et al., 2016].

6.3.4 Final structures and binding mechanisms

Restrained refinement was completed in rounds following protein backbone and side chain corrections, addition of HepI, HepII and anhydro-Kdo, and addition of solvent molecules; with up to 10 cycles of maximum-likelihood phase refinement in each round. Additional rounds were completed following any further changes to the structure, and the final R factors were calculated [table 6.2].

The overall structure of the rfhSP-D and R7 oligosaccharide complex is a homotrimer of three identical subunits, A, B and C, which contain residues 205-355, 204-355 and 205-355, respectively, of the α-coiled coil neck domain and the complete carbohydrate recognition domain at the C-terminus [Figure 6.5]. The remaining N-terminal residues (179-204 in subunits A and C, 179-203 in subunit B) are disordered. All three subunits of the rfhSP-D contain three calcium ions (Ca1, Ca2 and Ca3) in the CRD but there is no calcium (Ca4) present in the neck domain [Shrive *et al.*, 2003]. The structure reveals that the oligosaccharide ligand is bound in subunit B and C but is not present in subunit A due to the constraints of crystal contacts [Figure 6.5].

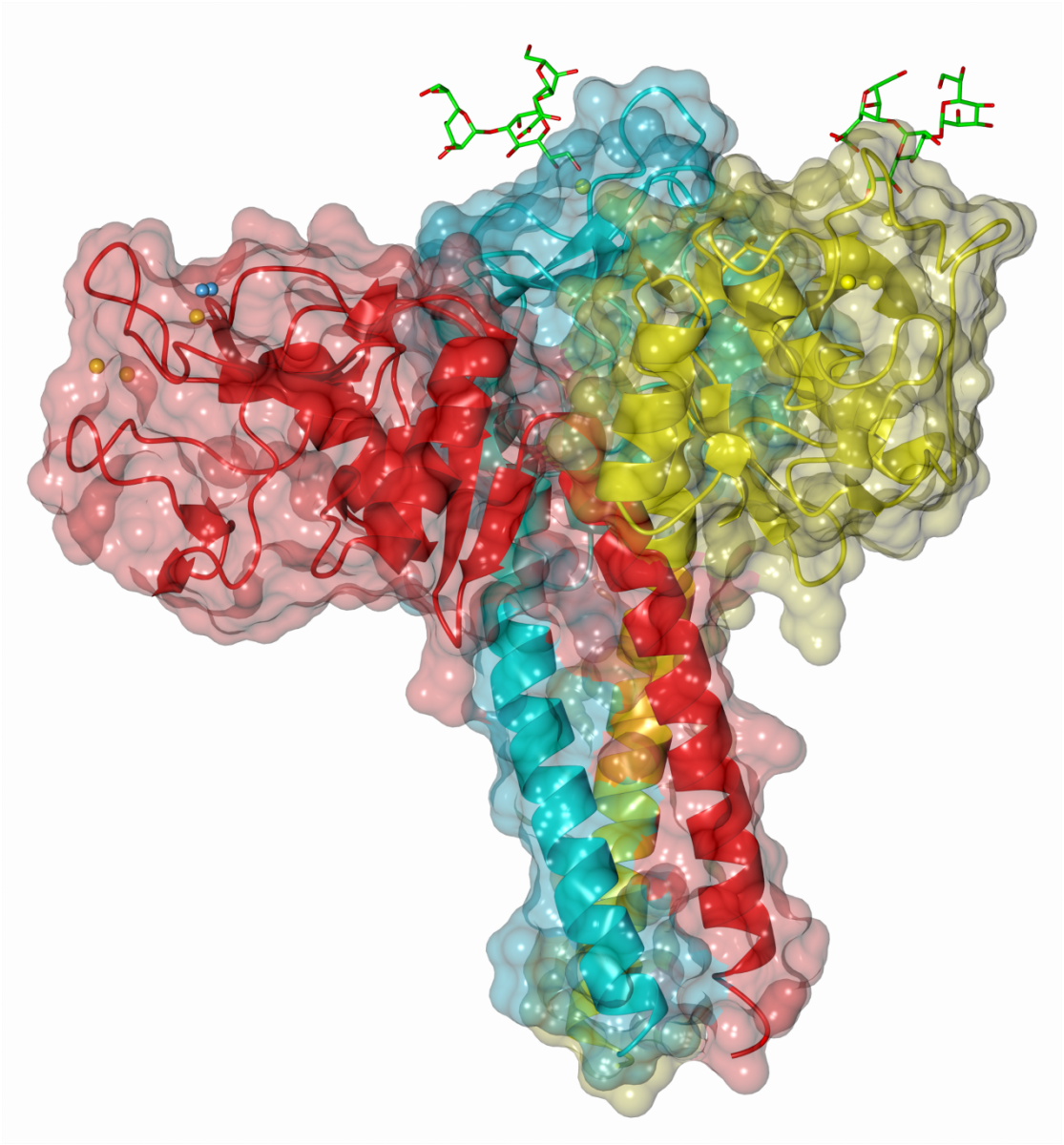


Figure 6.5: Overall structure of rfhSP-D in complex with *Salmonella enterica* Minnesota R7 polysaccharide. The overall orientation of *Salmonella enterica* Minnesota R7 polysaccharide (green) in the ligand binding sites of subunits B (yellow) and C (cyan). Figure shows the ligand extending away from the surface of the CRD of each subunit where it is bound. Calcium ions are represented in yellow and Ca1 coordinating water molecules in blue. Subunit A is represented in red.

Table 6.2: Data Collection and Processing Statistics for CCS16A11

	R7 Oligosaccharide
Data collection	
Wavelength (Å)	0.97625
Temperature (K)	100
Space group	$P2_1$
Cell Dimensions	
a (Å)	55.33
b (Å)	108.14
c (Å)	55.67
β (°)	91.82
Maximal resolution (Å)	1.75
Resolution Range (Å)	55.3 – 1.75 (1.78 - 1.75)
Observations	143 142 (7 653)
Unique reflections	61 250 (3 299)
Completeness (%)	93.2 (91.2)
R_{merge}[a]	0.066 (0.359)
I/σ(I)	7.3 (2.2)
CC½	0.991 (0.774)
Refinement	
Protein atoms[b]	3483
Residues, subunit A	205 – 355
Residues, subunit B	204 – 355
Residues, subunit C	205 - 355
Other atoms	
Calcium ions	9
Ligand	A = 0, B = 3, C=3
Water	479
Resolution range (Å)	55.3 – 1.75
R_{work}[c]	17.3
R_{free}[d]	19.6
Average *B*-values (Å^2)	
Protein main chain	23.1
Heteroatoms	31.1
Water	34.2
Ramachandron plot values[e] (%)	
Favoured	97.8
Allowed	2.2
Disallowed	0

[a] $R_{merge} = \Sigma_h\Sigma_j|I_{h,j}-I_h| \,/\, \Sigma_b\Sigma_j|I_{h,j}|$, where $I_{h,j}$ is the *j*th observation of reflection *h* and I_h is the mean of *j* for reflection *h*

[b] Total number of protein atoms used in refinement

[c] $R_{work} = \Sigma_h||F_{oh}|-|F_{ch}|| \,/\, \Sigma_h|F_{oh}|$, where F_{oh} and F_{ch} are the observed and calculated structure factor amplitudes, respectively, for reflection *h*.

[d] R_{free} = R_{work} but for a random 5% subset of reflections

[e] defined by MolProbity analysis

Table 6.3: Coordination distances to calcium and R7 Oligosaccharide from *S. enterica* Minnesota in the ligand-bound recombinant fragment of human surfactant protein D

		R7 Oligosaccharide					R7 Oligosaccharide		
Atom 1	**Atom 2**	**A**	**B**	**C**	**Atom 1**	**Atom 2**	**A**	**B**	**C**
Ca1	Glu321 OE1	2.65	2.60	2.64	Ca2	Asp297 OD1	2.67	2.62	2.70
	Asn323 OD1	2.49	2.44	2.38		Asp297 OD2	2.44	2.40	2.45
	Glu329 OE1	2.51	2.41	2.34		Glu301 OE1	2.51	2.57	2.53
	Asn341 OD1	2.37	2.43	2.44		Glu301 OE2	2.53	2.58	2.52
	Asp342 OD1	2.35	2.34	2.28		Asp324 OD1	2.43	2.60	2.64
	Asp342 O	2.54	2.52	2.51		Glu329 O	2.50	2.32	2.37
	HepI O6′ / W	-	2.36	2.37		Asp330 OD1	2.34	2.39	2.39
	HepI O7′ / W	-	2.35	2.34		W	2.36	2.31	2.32
HepI O6′ / W	Glu321 OE2	-	2.64	2.52	Ca3	Glu301 OE1	2.30	2.43	2.32
	Asn323 ND2	-	3.03	2.93		Asp330 OD1	2.50	2.50	2.55
HepI O7′ / W	Glu329 OE2	-	2.64	2.57		Asp330 OD2	2.43	2.45	2.45
	Asn341 ND2	-	2.94	3.07		W	2.23	2.26	2.17
KdoI O1′	Arg343 NH2	-	3.46	3.51		W	2.35	2.33	2.31
KdoI O2′	Arg343 NH1	-	3.10	3.07		W	2.34	2.38	2.33
KdoI O6′	Asp325 OD2	-	2.76	2.69		W	2.38	2.42	2.34

All numbers quoted in Å. Abbreviations: Hep, L-D-manno-heptose; Kdo, furanoid derivative 4,7-anhydro-3-deoxy-D-manno-oct-2-ulosonic acid (anhydro-Kdo). W represents waters in key binding positions.

Recognition of the R7 oligosaccharide occurred in a calcium dependent interaction between Ca1 and the hydroxyethyl side chain of HepI, coordinated by O6′ and O7′ that are positioned in a similar position to other carbohydrate-bound structures of rfhSP-D fragments. In subunits B and C, the calcium coordination distances to O6′ are 2.36Å and 2.37Å, respectively, and 2.35Å and 2.34Å to O7′ (table 6.3). Recognition in the primary binding pocket is completed by the usual amino acid side chains first identified in the rfhSP-D-maltose complex: Glu321 and Asn323 coordinate with O6′ through hydrogen bonds (2.64Å and 3.03Å, respectively in subunit B; 2.52Å and 2.93Å in subunit C); and Glu329 and Asn341 (2.64Å and 2.94Å in subunit B; 2.57Å and 3.07Å in subunit C) [Shrive *et al.*, 2003]. These four side chains also coordinate the calcium, Ca1, that coordinates the ligand through interactions that range between 2.34Å and 2.64Å in each subunit, with Asp342 completing the calcium coordination through two coordinating atoms, one in the side chain (2.34Å in B, 2.28Å in C) and the other in the main chain carbonyl (2.52Å in B, 2.51Å in C). Ca1 coordination is completed in a similar manner in subunit A (table 6.3).

Recognition of the R7 oligosaccharide is completed by a series of hydrogen bond mediated interactions with the flanking residues of the Ca1 binding pocket, mainly with the anhydro-Kdo residue [Figure 6.6, 6.7]. The main ring hydroxyl, O6′, directly interacts with the side chain of Asp325 through a single hydrogen bond at a distance of 2.76Å in subunit B and 2.69Å in subunit C. This does not disrupt the critical crystal contact between Asp325 and the symmetry related molecule. The final direct contacts with the protein are between the 2-oxobutanoic acid side chain at position 4 of the anhydro-Kdo, which interacts with the flexible side chain of Arg343 in both side chains, through hydrogen bonds to the O2′ α-keto group [Figure 6.6, 6.7] at a distance of 3.10Å in subunit B and 3.07Å in subunit C. There are also additional weak interactions with O1 of the Kdo side chain which are 3.46Å

and 3.51Å in length, in subunits B and C respectively. These stabilise the side chain in the crystal, making them observable in the electron density.

The second heptose residue (HepII) does not make any direct contacts with the protein, allowing to take up two different orientations in the two subunits where it is bound. In subunit B, a single water bridge mediates an interaction between O7′ of HepII and O2′ of HepI, which goes on to hydrogen bond to the ND2 group of Asn323 (3.15Å). The only other major interaction of HepII in the crystal is a water mediated interaction between O6′ and the amide group of Ser239C. In subunit C, the position of the main heptose ring is rotated approximately 45° around the α(1→3) linkage in relation to HepII in subunit B, moving HepII away from the surface of the protein and allowing it to interact directly with the symmetry related molecule of rfhSP-D [Figure 6.7]. O6′ and O7′ make direct interactions with the main chain carbonyl (3.01Å) and the amide linkage (2.99Å) of Gly241A, respectively, in the crystal contact. The interaction of HepII with the symmetry related molecule is completed by a water molecule network to Lys229A.

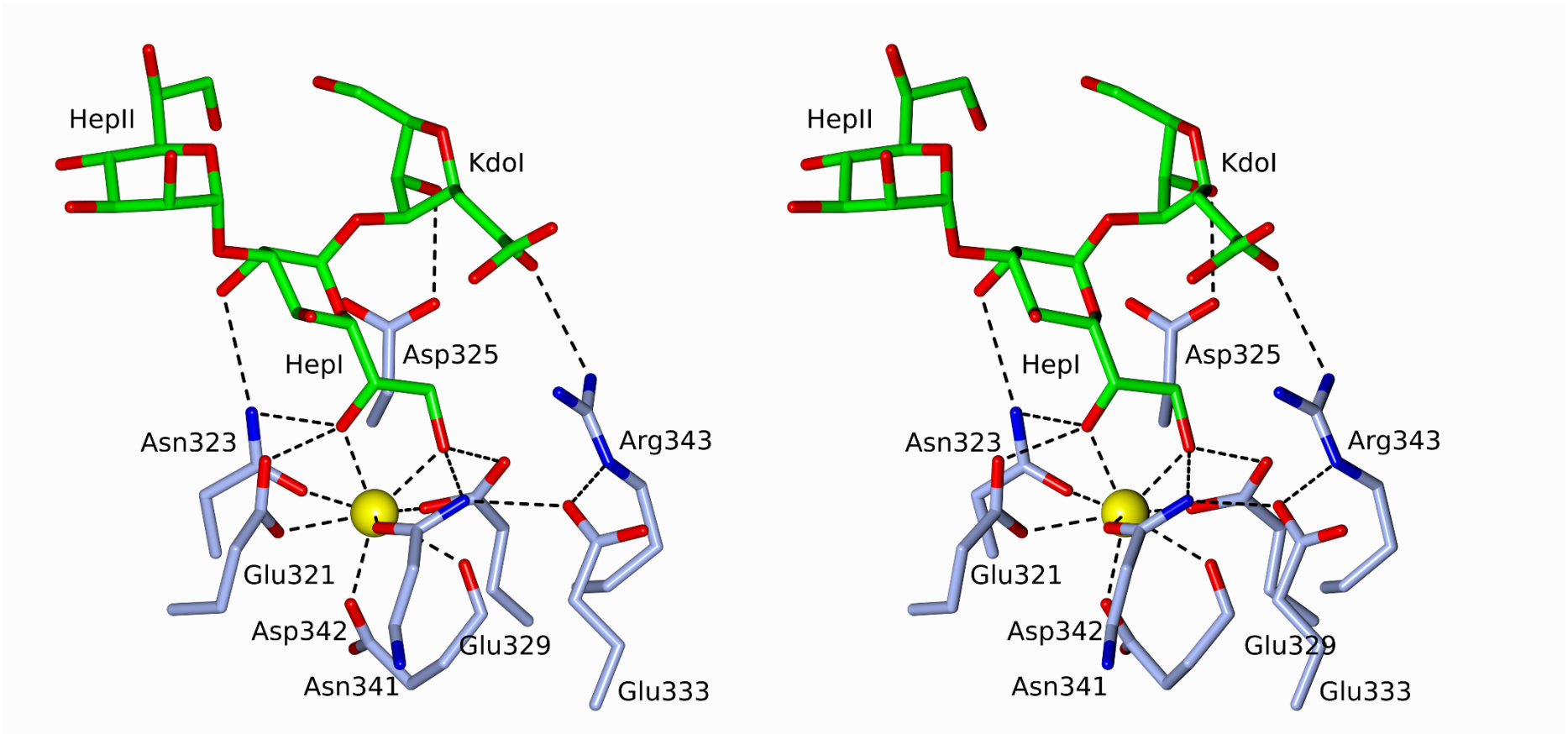


Figure 6.6: Stereo view of calcium coordination in subunit B of the *Salmonella enterica* R7 oligosaccharide-bound rfhSP-D. Ligand recognition occurs through O6' and O7' hydroxyls of the proximal heptose (HepI) coordinating the calcium ion (Ca1, yellow). Recognition of HepI is completed by interactions with Glu321, Asn323, Glu329 and Asn341. The anhydro-kdo (KdoI) interacts with both flanking residues, Asp325 and Arg343. HepII interacts with the symmetry related molecule.

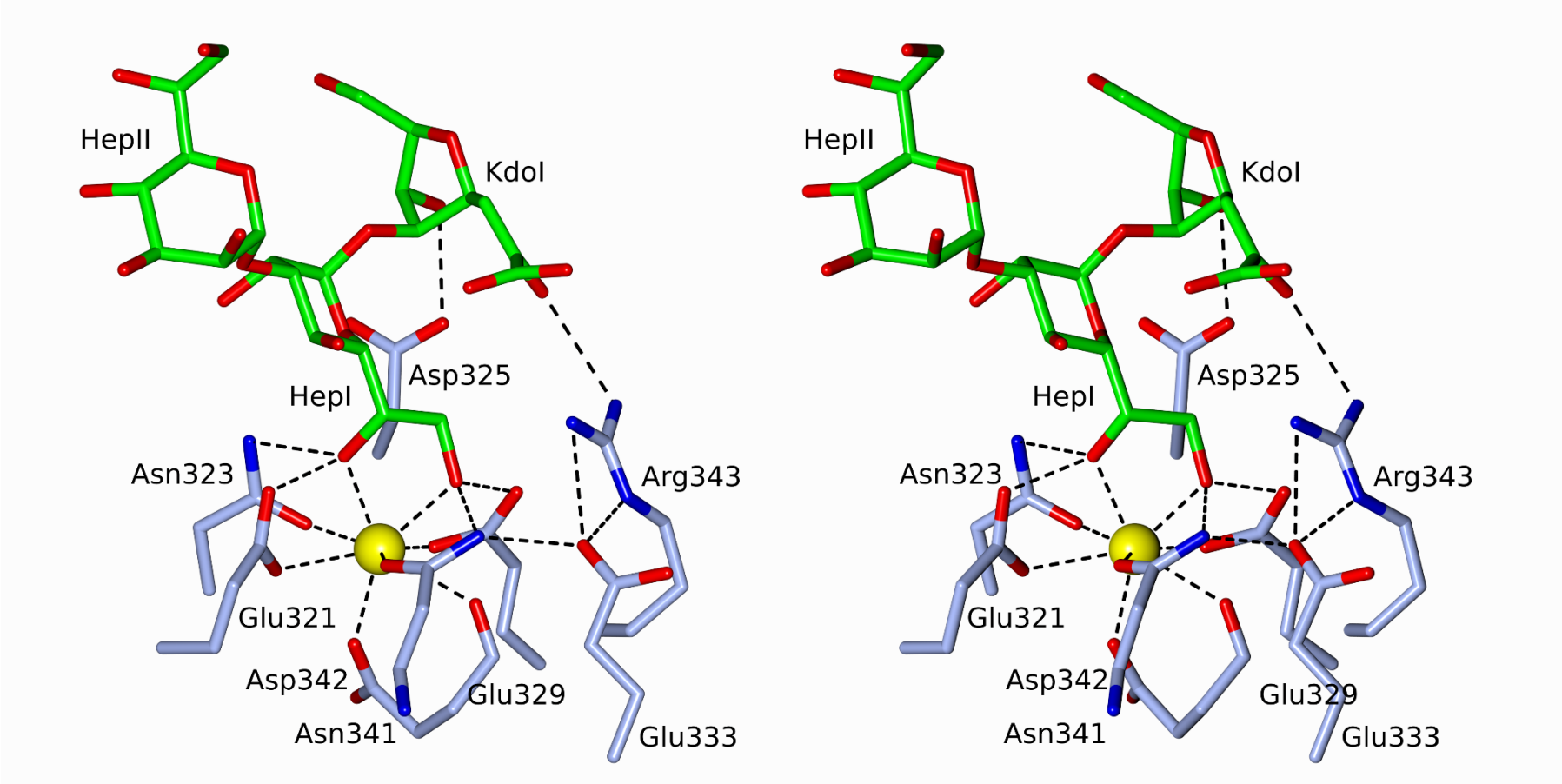


Figure 6.7: Stereo view of calcium coordination in subunit C of the *Salmonella enterica* R7 oligosaccharide-bound rfhSP-D complex. Ligand recognition occurs through O6' and O7' hydroxyls of the proximal heptose (HepI) coordinating the calcium ion (Ca1, yellow). Recognition of HepI is completed by interactions with Glu321, Asn323, Glu329 and Asn341. The anhydro-kdo (KdoI) interacts with both flanking residues, Asp325 and Arg343. HepII interacts with the symmetry related molecule. Note the change in rotation of HepII compared to HepII in Figure 6.6.

6.4 Discussion

6.4.1 Comparison of the R5 and R7 mutants of *S. enterica* Minnesota

As a way of comparing more complex inner core polysaccharides from LPS, the delipidated inner core polysaccharide of the R5 mutant of *S. enterica* sv. Minnesota, equivalent to the Rc mutant, has also been successfully crystallised in complex with the same biologically active recombinant fragment of hSP-D by a former PhD student in the research group at a resolution of 1.65Å [da Silva, 2016; Littlejohn *et al.*, 2018]. This was used for comparison and elucidation of the wider recognition mechanism of *S. enterica* sv. Minnesota LPS. The R5 mutant contains an additional α(1→3)-linked α-D-glucose at position 3 of HepII compared to the R7 mutant but is otherwise identical, including the formation of the furanoid anhydro-Kdo seen in the R7 complex and the previously published *H. influenzae* Eagan 4A complex [Figure 6.8; Clark *et al.*, 2016].

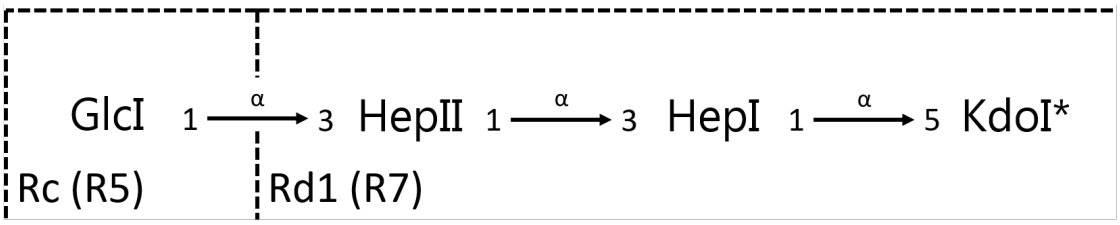


Figure 6.8: Structure of the R5 and R7 Oligosaccharide Products from Mild Hydrolysis. Mild hydrolysis of the R5 LPS removes the lipid A and KdoI-KdoII, leaving the product with an *anhydro-Kdo. Hep, L-D-heptose; KdoI*, furanoid 4,7-anhydro-3-deoxy-D-manno-oct-2-ulosonic acid.

6.4.2 Recognition of the Hep-Kdo motif

The high resolution structures of the oligosaccharide portions of LPS from deep rough mutants of *S. enterica* Minnesota R7 and R5 in complex with a recombinant fragment of hSP-D reveal that hSP-D targets the proximal inner core Hep-Kdo motif, through calcium-dependent recognition of the non-terminal heptose residue. This initial binding via Ca1 is supplemented by interactions between heptose and Kdo with the two flanking residues, Arg343 and Asp325, revealing for the first time the 2-oxobutanoic acid side chain of an anhydro-form of Kdo is able to interact extensively with the side chain of Arg343; providing further detail of the anhydro-Kdo itself, which was not fully defined in the *H. influenzae* Eagan 4A complex [Clark *et al.*, 2016].

The structure presented here confirms the 5-membered furanoid derived structure of the previously reported anhydro-Kdo with more details of the long 2-oxobutanoic side chain extending from carbon 4 [Figure 6.6, 6.7; Clark *et al.*, 2016]. This provides the first direct confirmation that removal of the lipid A from intact LPS, using mild hydrolysis, is accompanied by an elimination of the α(2→4)-linked KdoII-KdoIII moiety from carbon 4 of KdoI, in a manner equivalent to the β-elimination of the phosphate in the Eagan 4A-bound structure [Clark *et al.*, 2016]. The 2-oxobutanoic acid side chain is also present in the R5-bound structure, superposing very closely with the anhydro-Kdo in the R7-bound structure, confirming that the anhydro-Kdo is present in the β-anomer in both structures. It is unclear whether this is due to rfhSP-D having selectivity for the β-anomer, or whether it is a product of the elimination of the KdoII-KdoIII pair that restricts the newly formed anhydro-Kdo to the β-conformation rather than producing a racemic mixture of α and β anomers [Auzanneau, Charon and Szabó, 1991].

In subunit B and C of the R7- and the R5-bound structures, recognition of the proximal heptose (HepI) is calcium dependent, coordinating the O6′ and O7′ hydroxyls of the

hydroxyethyl side chain, with the remaining recognition occurring through further interactions of these two hydroxyls with the side chains of Glu321, Asn323, Glu329 and Asn341 (2.34Å to 3.38Å). Recognition of the inner core is completed through the anhydro-Kdo, which extends between the two flanking residues, Asp325 and Arg343, which directly bind to the furanose ring hydroxyl, O6', and the α-keto/carboxylic acid group of the C4 side chain (O2' and O1', Figure 6.6, 6.7, 6.9).

Superposition of all the HepI coordinating subunits, including those from the *H. influenzae* Eagan 4A structures reveal an essentially conserved position for both the furanose Kdo ring and the pyranose ring of HepI, including the acidic side chain present in the two *Salmonella*-bound structures. The conserved topology of the Hep-Kdo (anhydro) results in conservation of the Asp325 – O6' (anhydro-Kdo) interaction across the three structures, with only a small variation between the three oligosaccharides (2.56 – 2.76Å). However, this is not the case for the Arg343 – Kdo (anhydro) interaction. There are large variations in the length of the interactions between the carboxylic acid O1' and the NH2 of Arg343, ranging from strong hydrogen bonds (2.87Å) in the subunit C of the R5-bound structure, to long range weak interactions (>3.5Å) in subunit C of R7 [Appendix II; Clark *et al.*, 2016; Littlejohn *et al.*, 2017]. This suggests that there is an inherent flexibility in the Arg343 – Kdo interaction. It is well documented that Arg343 is able to change the conformation of its side chain in response to different ligands, allowing it to influence ligand selection and recognition [Allen *et al.*, 2004; Shrive *et al.*, 2009]. There also appears to be some conformational variability in the acidic side chain of the anhydro-Kdo which may explain why there is very little electron density for the long, acidic side chain in the *H. influenzae* Eagan 4A complex, with the exception of a small amount of density at the terminus of the side chain that was tentatively identified as an oxygen [Clark *et al.*, 2016]. In combination,

the structural evidence does suggest that the Arg343 – ligand interaction is a flexible determinant of specificity.

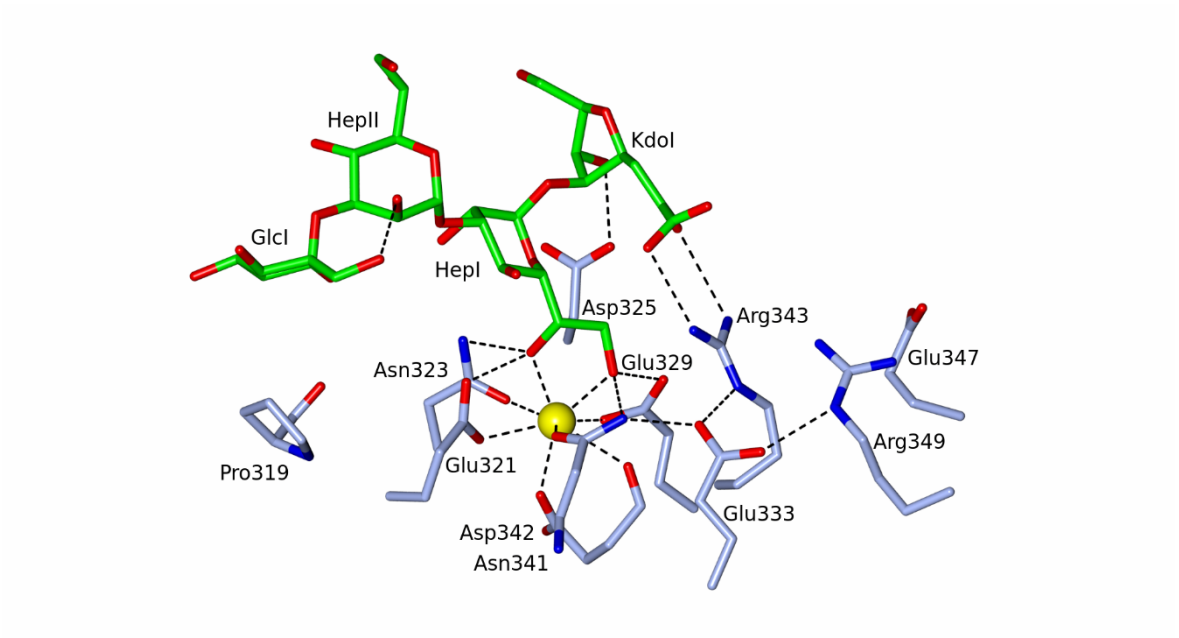


Figure 6.9: Calcium coordination in subunit C of the *Salmonella enterica* sv. Minnesota R5 oligosaccharide in complex with rfhSP-D. Calcium (Ca1, yellow) us coordinated by the O6′ and O7′ hydroxyls of the proximal heptose (HepI). Recognition is completed by Glu321, Asn323, Glu329 and Asn341. Additional recognition occurs through interactions with the anhydro-Kdo (KdoI) from the two flanking residues, Asp325 and Arg343. Figure adapted from [Littlejohn *et al.*, 2017]

6.4.3 Flexibility and versatility of ligand recognition

The major difference between the rfhSP-D complexes with R7 and R5 oligosaccharides is the presence of the ligand in subunit A. In the R7-bound complex, there is no ligand in subunit A. Instead, the calcium ion in the main binding pocket (Ca1) is coordinated by two water molecules which reflects the high resolution native structure [Shrive *et al.*, 2003]. However, in subunit A of the R5-bound structure, the extended R5 oligosaccharide (GlcI-HepII-HepI-Kdo(anhydro)) allows Ca1 to recognise and bind to the terminal glucose residue in a calcium dependent recognition of the O3′ and O4′ hydroxyl groups, allowing the remaining oligosaccharide to extend along the intermolecular solvent channel, over

Arg343 and towards the side chain of Arg349 in a direction approximately orthogonal to the HepI-bound subunits. Interestingly, the role of Arg343 remains fundamental to recognising the R5 oligosaccharide [Figure 6.10], despite the completely different position of HepI, with respect to the protein structure. The side chain of Arg343 forms the only direct contact with oligosaccharide outside of the main binding pocket, through a single hydrogen bond between NH2 and O4' of HepI which may be evidence of Arg343 directly contributing to directing the glucose in to the calcium binding site [Crouch *et al.*, 2009; Shrive *et al.*, 2009].

It appears that the crystal packing is largely responsible for precluding the recognition of the inner core proximal heptose, providing an explanation for the lack of R7-oligosaccharide in the binding site of subunit A, but it is fortuitous as it clearly demonstrates that the same LPS can be recognised by hSP-D using alternative recognition strategies and, importantly, shows that hSP-D is able to recognise more than just the inner core heptose and Kdo residues in natural LPS ligands. This potentially provides some explanation for the extraordinary ability of hSP-D to recognise such a variety of Gram-negative bacteria by utilising the Ca1-dependent binding of more than just the inner core heptose and extends the recognition to multiple saccharide residues in the LPS [Wang *et al.*, 2008; Clark *et al.*, 2016]. This also demonstrates a versatility of recognition in the carbohydrate recognition domain, providing evidence that hSP-D is able to target more than one saccharide *in vivo*, allowing it to target the variety of different carbohydrate residues on the surface of bacteria.

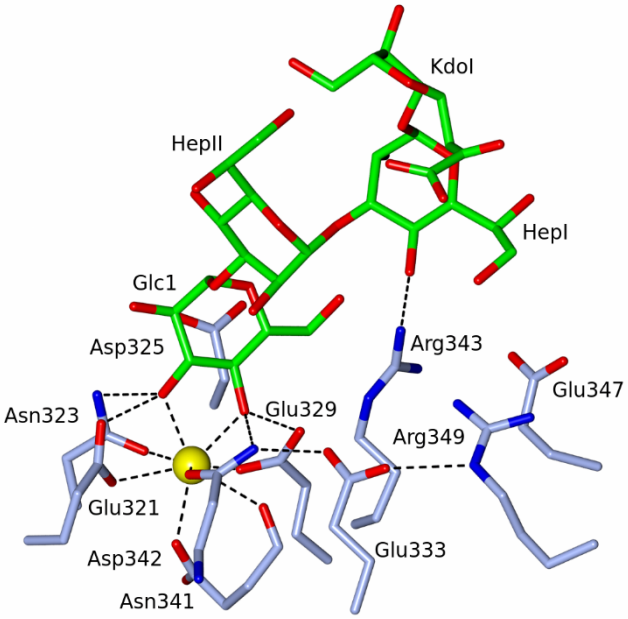


Figure 6.10: Calcium coordination in subunit A of the *Salmonella enterica* sv Minnesota R5 oligosaccharide in complex with rfhSP-D. Calcium (Ca1, yellow) is coordinated by the O3' and O4' hydroxyls of the terminal glucose (GlcI). Recognition of GlcI is completed by Glu321, Asn323, Glu329 and Asn341. Arg343 plays a pivotal role in completing the recognition by interacting with HepI. Figure adapted from [Littlejohn *et al.*, 2017]

6.4.4 Accommodation of the full R5 inner core polysaccharide

Human SP-D is known to recognise and bind to the LPS molecules from *S. enterica* Minnesota deep rough mutants R5 and R7 [Kuan, Rust and Crouch, 1992]. To explore how the whole inner core may be accommodated by hSP-D, a simple overlay of the full R5 oligosaccharide, including the KdoI-KdoII pair and an intact KdoI, onto the Hep-Kdo (anhydro) unit of the R5 oligosaccharide in complex with the rfhSP-D was carried out. The intact 6-membered KdoI ring was easily positioned between the two flanking residues Asp325 and Arg343 revealing that the ring may be orientated to interact with either flanking residue. A rotation through the α(1→5) glycosidic bond allows the dihydroxyethyl side chain of KdoI (O7' and O8') to rotate between the two side chains and interact with both side chains; based on the position of the α(1→5) glycosidic bond in the R7- and R5-bound crystal structures [Figure 6.11]. In the position in this simple model, O7' interacts

with the side chain of Asp325, allowing the remaining KdoII-KdoIII chain to be easily accommodated by the flexible Arg343 side chain. The KdoII-KdoIII pair extends across the surface of the CRD and can be readily placed in the proximity of Glu349 and the conserved side chains of Arg349 and Glu333, which are present in both hSP-A and SP-D [Figure 6.11]. Importantly, 2'OH of KdoI is positioned to allow the large lipid A moiety to exit the CRD, extending out of the CRD and into space.

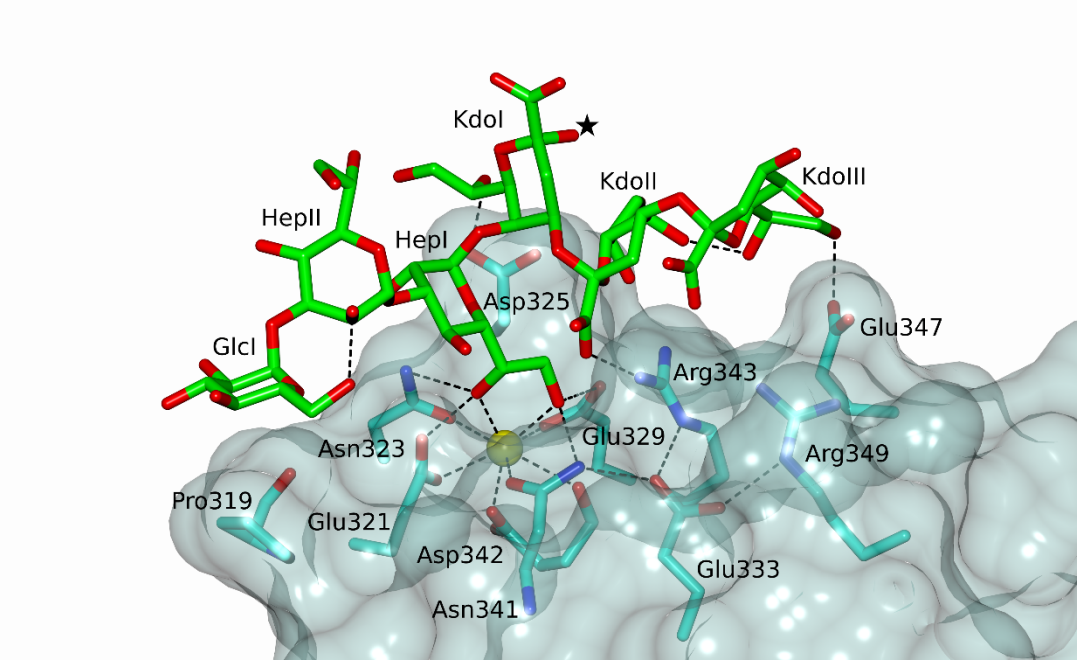


Figure 6.11: Model of the complete *Salmonella enterica* R5 oligosaccharide bound to rfhSP-D with the inner core HepI in the calcium binding site. KdoI (intact) is positioned to interact with Arg325, with KdoII and KdoIII interacting with Arg343 and Glu347 respectively. The star indicates the position of O2' where the lipid A moiety would be attached.

The model also suggests that longer *S. enterica* mutants with more extensive polysaccharide cores, including those from the Rb mutant which is known to be recognised by hSP-D, could be accommodated by the protein surface [Kuan, Rust and Crouch, 1992]. In these longer inner cores, positions 7 and 4 of HepII are phosphorylated and GlcI has galactose substitutions [GalI and GalII, Figure 6.1] at positions 1 and 3, which in this model would all extend away from the CRD [Mansfield and Forsythe, 2001]. This strongly suggests

that the Hep-Kdo motif present in the R7- and R5-bound structures can be recognised in other longer LPS ligands from other species that hSP-D, and the recombinant fragment used in this study, are able to recognise. This includes important pathogens such as *Pseudomonas aeruginosa*, *Escherichia coli* and *Klebsiella pneumoniae*, with the potential for increased recognition of the multiple Kdo structures known to be present in their LPS cores [Lim *et al.*, 1994; Olsthoorn, Haverkamp and Thomas-Oates, 1999; Holst, 2011].

Chapter 7: General discussion and future work

7.1 General discussion

The work presented in this thesis provides new insights into how human surfactant protein D is capable of recognising a variety of carbohydrate ligands commonly found on the surfaces of a number of pathogens. Surfactant protein D is a member of the highly conserved family of C-type lectins, the collectins, that have been identified in phylogenetically ancient species including tunicates that share a highly conserved carbohydrate recognition domain, usually at their C-terminus [Veldhuizen, Van Eijk and Haagsman, 2011; Drickamer and Taylor, 2015]. The carbohydrate recognition domain is linked at the N-terminus to an α-coiled coil neck domain which leads to a long, mostly flexible collagen-like domain and the N-terminal cysteine-rich domain [Epstein *et al.*, 1996; Lu *et al.*, 2002]. The CRDs of the collectins have been shown to have an up to 70% conserved sequence identity with differences in the side chains surrounding the calcium binding pocket almost completely conserved. changing the carbohydrate specificity of the collectin.

7.1.1 Identification of a new recognition mechanism for β-D-glucoses

Human SP-D has a high affinity for α(1→4)-linked maltose whereas MBP and hSP-A prefer acetylated amino sugars, N-acetylglucosamine and N-acetylmannosamine, respectively [Haagsman *et al.*, 1987; Haurum *et al.*, 1993; Crouch *et al.*, 2005]. Interestingly though, *Allen* and colleagues (2001) demonstrated that hSP-D is also able to recognise β-D-glucose ligands with relatively high affinity compared to α-D-glucose ligands of a similar size [Allen *et al.*, 2001]. β-D-glucose chains are common on the cell surface of a broad range of pathogens that hSP-D is known to recognise, including the pathogenic fungus *Aspergillus fumigatus* which is the causative agent in the pulmonary infection aspergillosis [Allen,

Voelker and Mason, 2001]. They also form an important part of the cell wall of allergenic particles, such as pollen and dust mite allergen, which hSP-D has been suggested to recognise and play a functional role in controlling the response [Wang *et al.*, 1996, 1998; Ferguson *et al.*, 1998; Schneider *et al.*, 2016]. Alternative glycosidic linkages, for example α(1→6) glycosidic linkages, are also common features of the microbial extracellular space in biofilms that are exposed to the immune system and, therefore, a potential target for the innate immune system. For the first time, small oligosaccharides (cellotriose, β(1→4)-linked; isomaltotriose, α(1→6)-linked; laminaritriose, β(1→3)-linked) containing these common alternative anomers of glucose and isomeric chains have been complexed with a recombinant fragment of hSP-D and the recognition mechanism fully described at the atomic level. Maltotetraose and maltoheptaose (α(1→4)) were also solved for comparison to longer α-glucooligosaccharides [Shaw, 2009].

In the cellotriose/rfhSP-D complex and the maltoheptaose/rfhSP-D complex, where the non-reducing terminal glucose coordinates to Ca1 (all three subunits in cellotriose-bound and subunit A of the maltoheptaose complex), recognition occurs through the equatorial hydroxyls O3′ and O4′ in a comparable position to the well-documented recognition mechanism for a wide range of mono-, di- and trisaccharides [Shrive *et al.*, 2003, 2009; Crouch, McDonald, *et al.*, 2006; Shaw, 2009; Goh *et al.*, 2013]. However, where the reducing terminal glucose is coordinated by Ca1, or O3′ is unavailable for binding, coordination occurs through O1′ and O2′ in a similar mechanism to galactose recognition [Shrive *et al.*, 2009]. This is possible because the β-anomeric conformation of glucose presents the O1′ and O2′ hydroxyls in position analogous to the O3′ and O4′ hydroxyls that are usually recognised in maltooligosaccharides. Interestingly, one of the small α-glucooligosaccharides, isomaltotriose, was recognised by the novel O1′/O2′ mechanism providing the first direct evidence of the isomerism of glucose polysaccharides following

chemical degradation. However, why the β-anomer is selected for over the O3′ and O4′ hydroxyls, available on both of the other two glucose residues in isomaltotriose, is unclear. This is especially intriguing considering that cellotriose is recognised by the O3′/O4′ hydroxyls despite O1′ and O2′ being available in the reducing terminal; unless the same isomerism seen in isomaltotriose produces an α-D-glucose following degradation during cellotriose production, removing the availability of the O1′ and O2′ hydroxyls.

7.1.2 Recognition mechanism of muramyl disaccharide is via N-acetylglucosamine

The major component of the Gram-positive cell wall, peptidoglycan, has also been shown to be bound by surfactant protein D in both its insoluble form, found in the cell wall, and its soluble form which is shed from the surface of gram-positive, and to a lesser extent gram-negative, bacteria [Schleifer and Kandler, 1972; Bertsche *et al.*, 2015]. Peptidoglycan is composed of repeating units of muramyl disaccharide, an N-acetylglucosamine β(1→4)-linked to an N-acetylmuramic acid unit, and tetra- or pentapeptides extending in an orthogonal direction to the polysaccharide core. The oligopeptides crosslink with each other through a mixture of direct interactions between polar groups, such as lysine, and polyalanine bridging peptides, forming an elastic but strong layer that can resist changes to turgor pressure in the cell with either 3-fold or 4-fold screw-axes around the polysaccharide cores [Vollmer, Blanot and De Pedro, 2008; de Pedro and Cava, 2015].

The length of the polysaccharide core is dependent on the number of muramyl disaccharide repeats and varies across species depending on the direction of the growth in the cell wall. Some of the peptidoglycan potentially presents itself to the immune system inside the cell wall, whilst the soluble form of peptidoglycan is exposed to the immune system when it is released from the cell wall, stimulating an immune response. How hSP-D recognises either the insoluble form or the soluble forms of peptidoglycan has not previously been characterised at the atomic level, but the work presented here provides the

first insight into how the muramyl disaccharide repeating unit is recognised by the CRD of a recombinant fragment of hSP-D that is therapeutically and biologically active.

Recognition occurs via the N-acetylglucosamine O3' and O4' hydroxyls coordinating with the Ca1 in the main binding pocket, with recognition being completed by the usual amino acid side chains (Glu321, Asn323, Glu329 and Asn341) in a similar mechanism to maltose and ManNAc recognition [Shrive *et al.*, 2003; Crouch *et al.*, 2007]. This allows the N-linked acetyl group to extend above the side chains of Glu321 and Asn323, forming a water bridge network, with the β(1→4)-linked N-acetylmuramic acid extending away from the surface of the protein and distantly interacting with a symmetry related molecule. A simple model of a longer peptidoglycan, including the first three amino acids of the peptide crosslinkers, demonstrates that the protein surface could easily accommodate the peptide as it extends into space with the remaining polysaccharide chain also extending away from the surface of the CRD in a roughly orthogonal direction. This suggests the possibility that hSP-D may be able to recognise peptidoglycan embedded in the insoluble fragments of peptidoglycan if the GlcNAc residues are presented at the exposed face of the cell wall. Similarly, it also provides much needed insight into how hSP-D recognises soluble peptidoglycan, by the terminal GlcNAc, and how this might allow the potential clearance of gram-positive bacteria and their breakdown products.

7.1.3 Flexibility and versatility of rfhSP-D in *S. enterica* sv. Minnesota recognition

Surfactant protein D is also able to recognise the core polysaccharide of the gram negative lipopolysaccharides found in the outer membrane [Sahly *et al.*, 2002; Wu *et al.*, 2003]. There has been extensive work showing that hSP-D is capable of recognising a range of gram-negative species with highest affinity for rough LPS mutants that lack the long repeating O-antigen, which can be highly variable, leaving the more conserved core polysaccharide accessible to the immune system and hSP-D. In particular, heptose-rich

inner core oligosaccharides have been shown to be particularly susceptible to recognition by hSP-D, elaborated on by a number of structural investigations into heptose recognition by rfhSP-D [Wang *et al.*, 2008; Reinhardt *et al.*, 2016]. This is typified by the successful high-resolution structure of rfhSP-D in complex with the delipidated oligosaccharide core from *H. influenzae* Eagan 4A LPS. The structure reveals recognition of the inner core oligosaccharide via the O6′-O7′ hydroxyl groups of surface-proximal heptose [Clark *et al.*, 2016]. The work done by *Clark* and colleagues (2016) also revealed an anhydrous form of Kdo, 4,7-anhydro-3-deoxy-D-*manno*-oct-2-ulosonic acid, produced from the mild acid hydrolysis used to cleave the lipid A moiety, however the complete anhydro-Kdo was not visible in the electron density. Building on this work, previous Keele structural biology PhD students [Smallcombe, 2014; da Silva, 2016] successfully soaked crystals of rfhSP-D with delipidated oligosaccharides from the Rc (R5) and Rd1 (R7) rough mutants of *S. enterica* sv. Minnesota. More recent advances in our understanding of the recognition mechanism has led to the reinvestigation of the structures reported here.

The analysis of the structures reveals two binding mechanisms for the longer R5 oligosaccharide, produced as a result of the crystal system restricting the binding site of subunit A. In subunit B and C, the proximal heptose, HepI, is coordinated by the calcium through the O6′ and O7′ of the dihydroxyethyl side chain and the main mannosyl ring extends between Asn323 and Glu329 in a comparable conformation to the coordinating heptose in the *H. influenzae* Eagan 4A oligosaccharide complex and the L-D-diheptose complex [Wang *et al.*, 2008; Clark *et al.*, 2016]. This is mirrored by the R7 oligosaccharide complex where the proximal heptose is the main recognition residue, however, because R7 lacks the terminal glucose, no ligand is present in subunit A and instead two waters complete the coordination of Ca1. While in all of the subunits where HepI is the primary site of recognition, the α(1 →5)-linked Kdo is in the anhydrous form previously identified in

the Eagan-bound structure, the re-investigation provides the first direct evidence of β-elimination of larger Kdo substituents and the 2-oxobutanoic acid side chain has, for the first time, been completely characterised in its β-anomeric form [Auzanneau, Charon and Szabó, 1991]. Subunit A of the R5 oligosaccharide complex contains the delipidated R5 oligosaccharide bound via the terminal α-D-glucose, coordinated by the O3′ and O4′ hydroxyls. The difference in binding may be an artefact of crystallisation but it does provide a unique insight into rfhSP-Ds propensity to recognise multiple epitopes in the LPS and the versatility to employ multiple binding mechanisms beyond the inner core Hep-Kdo mechanism when that epitope is not accessible.

7.1.4 Asp325, Arg343 and the extended binding surface of hSP-D

All of the structures presented here confirm the importance of the two flanking residues Asp325 and Arg343 in recognising oligosaccharide ligands. In the small ligand-bound structures, Arg343 was involved in recognising the terminal glucose coordinated by Ca1 in all three structures and demonstrates the flexibility of this side chain to alter its conformation to interact with the O6′ or O5′, in the case of isomaltotriose. The isomaltotriose complex emphasises the inherent flexibility in Arg343, which accommodates a much less extended conformation to allow the ligand to extend over the side chain and interact with Glu347 and Arg349; all of which is stabilised by a hydrogen bond between Arg343 and O4′ of the second glucose. The importance of Arg343 in ligand recognition has been well studied and confirmed using crystallography, *in silico* modelling and biochemical analysis, but it has also been shown that both Arg343 and Asp325 can act as a selection factor for ligands by restricting the access to one side of the binding site [Shrive *et al.*, 2003, 2009; Allen *et al.*, 2004; Crouch *et al.*, 2009]. This selective role is apparent in the muramyl disaccharide complex where the orientation of the terminal GlcNAc is restricted so that the N-linked acetyl branch extends above the calcium coordinating side

chains of Glu321 and Asn323 whilst maintaining the O3'-O4' coordination to Ca1. Furthermore, subunit A of the R5-bound structure [da Silva, 2016; Littlejohn *et al.*, 2018] reveals that Arg343 provides an important anchoring position for HepI in the LPS oligosaccharide, potentially directing the terminal glucose into the calcium binding pocket and permitting coordination to O3' and O4'. In subunits B and C of the two *S. enterica* Minnesota oligosaccharide complexes, where HepI is coordinated by Ca1, Arg343 is fundamental in recognising the proximal Hep-Kdo in combination with Asp325, which directly interact with the anhydro-Kdo and its 2-oxobutanoic acid side chain by adopting a conformation unique to the Hep-Kdo recognition mechanism [Littlejohn *et al.*, 2018].

The role of Asp325 in selecting and coordinating ligands has been extensively explored using small ligand studies and site directed mutations, demonstrating a clear importance in ligand recognition [Crouch *et al.*, 2009; Shrive *et al.*, 2009]. Results presented here support the importance of Asp325 as it plays a critical role in recognising the anhydro-Kdo and is hypothesised to interact with a complete 6-membered Kdo based on a simple overlaid model. Asp325 may also provide a single contact to the N-acetylmuramic acid of muramyl disaccharide, through the lactic acid chain [Clark *et al.*, 2016; da Silva, 2016; Littlejohn *et al.*, 2018]. The crystal contacts prevent the interaction of MurNAc with Asp325 because of steric clashes in the crystal system, however, *in vivo*, the lactic acid side chain would be free to rotate around the O3' ester bond and form a direct hydrogen bond with Asp325 OE1, while maintaining sufficient space for the peptide crosslinker.

The small ligand complexes and the simple model of the complete R5 oligosaccharide complex indicate that the extended binding surface, first suggested by *Crouch* and colleagues in 2006, is more extensive than only Phe335 and extend in multiple directions away from the binding site, partly enabled by Arg343 [Crouch *et al.*, 2006]. Maltotetraose extends towards Phe335 but is further from the surface of rfhSP-D than maltotriose,

forming an extended arch, which allows the reducing terminal glucose to interact with Thr336 and Asn337 rather than stacking with the Phe335 phenyl ring as observed for maltotriose [Crouch, McDonald, *et al.*, 2006; Shaw, 2009]. In contrast, isomaltotriose, due to the flexibility afforded by the α(1→6) glycosidic linkage, extends above Arg343, causing the compressed conformation previously mentioned, towards a pocket created by Glu347 and Arg349. These two side chains coordinate the nonterminal glucose with the non-reducing terminal glucose thought to make transient interactions with the protein. Similarly, the KdoII-KdoIII, removed during delipidation in the R5 oligosaccharide bound structure, could easily be accommodated by the Glu347/Arg349 pocket that coordinates isomaltotriose.

The R5 oligosaccharide is particularly interesting as a recent study suggested that synthetic LPS oligosaccharides have secondary recognition sites to the main Ca1 binding pocket, suggesting that surrounding residues may be important for the recognition of long, naturally occurring polysaccharides associated with LPS [Reinhardt *et al.*, 2016]. However, this is the first experimental evidence for this extended binding surface in surfactant protein D, beyond Phe335, providing valuable insight into how SP-D is able to recognise longer polysaccharide ligands from a variety of pathogens and allergens, using a flexible and versatile mechanism across its surface.

7.2 Future work

This study has identified a new binding mechanism for β-D-glucoses and an extended binding surface for longer ligands where Glu347 and Arg349 appear to form an important secondary binding pocket. To explore this more fully, mutation studies of both Glu347 and Arg349 would provide insight into their role in the binding affinity for longer polysaccharide ligands such as mannan and pustulan. A starting point may be a charge switch between the two side chains (i.e. E347R/R349E) which would maintain the hydrogen bond network seen in the studies in this thesis, progressing to alanine mutations to explore the importance of the secondary binding pocket. Similarly, the structures reported here appear to highlight the importance of Thr336 and Asn337 in recognising tetrasaccharides, but this interaction requires further investigation. Future work should continue to characterise the binding surface beyond the central calcium binding pocket, Ca1, as this surface may have an additive effect to the selectivity already described for Asp325 and Arg343.

Further work is required into the recognition mechanism for gram-positive bacterial cell wall components and their associated soluble markers. The muramyl disaccharide complex demonstrates that hSP-D is able to recognise a small fragment of the polysaccharide core of peptidoglycan but provides no information about how, for example, the pentapeptides may affect the recognition or whether they may, in fact, interact with a further extended binding surface that has not yet been identified. Larger fragments of intact peptidoglycan will make this possible, with an emphasis on including the oligopeptides. It is unclear how successful a soaking protocol would be for these larger fragments as the solvent channels and crystal system may prevent the ligand from binding Ca1, in which case a co-crystallisation protocol may be more suitable, allowing the ligand to bind prior to crystallisation. Surfactant protein D also interacts with lipoteichoic acids on the surface of

gram-positive bacteria however this interaction has not been structurally characterised and will provide further insight into how gram-positive bacteria are recognised [van de Wetering *et al.*, 2001]. It is not immediately clear from the structure of lipoteichoic acid what epitope hSP-D recognises as there are only two β(1→6)-linked glucoses in the whole chain, but a successful structural study would reveal, at the atomic level, how SP-D is able to bind this immunologically important gram-positive cell surface component.

The work presented here in combination with the work of previous members of the research group provides an important foundation in characterising how human surfactant protein D recognises the inner core polysaccharide of gram-negative bacterial lipopolysaccharide, identifying two recognition mechanisms and an anhydro-Kdo [Smallcombe, 2014; Clark *et al.*, 2016; da Silva, 2016; Littlejohn *et al.*, 2018]. In the future, however, it is essential to gain a better understanding of how hSP-D recognises lipopolysaccharides that represent *in vivo* LPS more closely in both its intact, membrane-bound form, and the polysaccharide released from the bacterial surface. This will require: (1) a method for maintaining the KdoI as a six-membered ring following delipidation or solubilising LPS; (2) characterisation of how polysaccharide release affects the structural components of the polysaccharide.

A method for delipidating LPS using hydrogen fluoride has been reported that preserves the Kdo in its natural, six-membered form which may enable characterisation of the interactions suggested by the model of the R5 oligosaccharide [Phillips *et al.*, 1992]. It remains unclear if the anhydro-Kdo exists naturally following polysaccharide release from LPS. As previous studies have shown, mass spectrometry is a useful tool in characterising the structures in the LPS, however further techniques could be useful, including nuclear magnetic resonance (NMR) and X-ray based techniques, to determine if the Kdo is intact in the polysaccharide fragments or if it is truly a product only of mild acid hydrolysis.

Structural characterisation of the fine detail of the interaction between hSP-D and cell surface receptors is an essential step towards completing the picture of hSP-D in the immune system, with the potential to provide structural insight into how hSP-D is able to regulate the immune system so effectively. This would be an ambitious project requiring production of recombinant hSP-D, both full length trimers and the fragment used in these studies, and, more importantly, recombinant forms of putative cell surface receptors. As hSP-D is a C-type lectin, it may be necessary to express the receptors in expression systems that permit glycosylation, which SP-D may recognise, presenting its own problems. Furthermore, membrane-bound proteins present additional problems to protein crystallography in that solubilised protein is required. However, with the increasing accessibility of techniques such as crystallography using lipidic cubic phase, which provides a matrix for the insoluble domains of receptors, and increasingly powerful and high resolution cryo-electron microscopy (cryo-EM) techniques, the investigation of protein-protein complexes is rapidly entering the mainstream and may the applied to hSP-D in the future.

The ability of surfactant protein D to recognise, opsonise and clear gram-positive and gram-negative bacteria, viruses, fungi and other important pathogenic material makes it a tantalising therapeutic target, especially as cases of antibiotic resistance in mainly gram-negative bacteria continue to increase and traditional treatments become less effective. The recombinant fragments of the neck and CRD have already shown promise in animal models using minimal invasive administration, however further structural work to determine how hSP-D, in particular, recognises both pathogens and host cells is required to maximise the safety and efficacy of potential hSP-D based therapeutics.

Reference list

Adams, M. J., Blundell, T. L., Dodson, E. J., Dodson, G. G., Vijayan, M., Baker, E. N., Harding, M. M., Hodgkin, D. C., Rimmer, B. and Sheat, S. (1969) 'Structure of Rhombohedral 2 Zinc Insulin Crystals', *Nature*, **224**(5218), pp. 491–495.

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. and Zwart, P. H. (2010) 'PHENIX: A Comprehensive Python-based System for Macromolecular Structure Solution', *Acta Crystallographica*, **D66**, pp. 213–221.

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. and Adams, P. D. (2012) 'Towards automated crystallographic structure refinement with phenix.refine', *Acta Crystallographica*. International Union of Crystallography, **D68**(4), pp. 352–367.

Agirre, J. and Cowtan, K. D. (2015) 'Validation of carbohydrate structures in CCP4 6.5', *Computational Crystallography Newsletter*, **6**(1), pp. 10–12.

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', *Biochemistry*, **40**(26), pp. 7789–7798.

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

Allen, M. J., Voelker, D. R. and Mason, R. J. (2001) 'Interactions of surfactant proteins A and D with Saccharomyces cerevisiae and Aspergillus fumigatus.', *Infection and immunity*, **69**(4), pp. 2037–44.

Auzanneau, F., Charon, D. and Szabó, 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', *Journal of the Chemical Society*, **Perkin Transactions 1**(3), pp. 509–517.

Axelgaard, E., Jensen, L., Dyrlund, T. F., Nielsen, H. J., Enghild, J. J., Thiel, S. and Jensenius, J. C. (2013) 'Investigations on collectin liver 1', *Journal of Biological Chemistry*, **288**(32), pp. 23407–23420.

Barrow, A. D., Palarasah, Y., Bugatti, M., Holehouse, A. S., Byers, D. E., Holtzman, M. J., Vermi, W., Skjødt, K., Crouch, E. C., Colonna, M. and Skjoedt, K. (2015) 'OSCAR Is a Receptor for Surfactant Protein D That Activates TNF-alpha Release from Human CCR2+ Inflammatory Monocytes', *The Journal of Immunology*, **194**(7), pp. 3317–3326.

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. and Leslie, A. G. W. (2011) 'iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM', *Acta Crystallographica*. International Union of Crystallography, **D67**(4), pp. 271–281.

Benne, C. A., Kraaijeveld, C. A., van Strijp, J. A. G., Brouwer, E., Harmsen, M., Verhoef, J., van Golde, L. M. G. and Van Iwaarden, J. F. (1995) 'Interactions of surfactant protein a with influenza a viruses: Binding and neutralization', *Journal of Infectious Diseases*, **171**(2), pp. 335–341.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) 'The protein data bank.', *Nucleic acids research*, **28**(1), pp. 235–242.

Bertsche, U., Mayer, C., Götz, F. and Gust, A. A. (2015) 'Peptidoglycan perception-Sensing bacteria by their common envelope structure', *International Journal of Medical Microbiology*. Elsevier GmbH., **305**(2), pp. 217–223.

Blok, J., Air, G. M., Laver, W. G., Ward, C. W., Lilley, G. G., Frank Woods, E., Roxburgh, C. M. and Inglis, A. S. (1982) 'Studies on the size, chemical composition, and partial sequence of the neuraminidase (NA) from type A influenza viruses show that the N-terminal region of the NA is not processed and serves to anchor the NA in the viral membrane', *Virology*, **119**(1), pp. 109–121.

Bowman, S. M. and Free, S. J. (2006) 'The structure and synthesis of the fungal cell wall', *BioEssays*, **28**(8), pp. 799–808.

Bragg, W. H. (1912) 'X-rays and Crystals', *Nature*, **90**(2243), pp. 219.

Bragg, W. H. and Bragg, W. L. (1915) *X Rays and Crystal Structure*. 2nd edn. London: G. Bell and Sons Ltd.

Brennan, P. J. and Nikaido, H. (1995) 'The Envelope of Mycobacteria', *Annual Review of Biochemistry*, **64**(1), pp. 29–63.

Briken, V., Porcelli, S. A., Besra, G. S. and Kremer, L. (2004) 'Mycobacterial lipoarabinomannan and related lipoglycans: From biogenesis to modulation of the immune response', *Molecular Microbiology*, **53**(2), pp. 391–403.

Brown-Augsburger, P., Hartshorn, K. L., Chang, D., Rust, K., Fliszar, C., Welgus, H. G. and Crouch, E. C. (1996) 'Site-directed Mutagenesis of Cys-15 and Cys-20 of Pulmonary Surfactant Protein D: Expression of a Trimeric Protein with Altered Anti-viral Properties', *Journal of Biological Chemistry*, **271**(23), pp. 13724–13730.

Brown Jr, R. M. (2004) 'Cellulose Structure and Biosynthesis: What is in Store for the 21st Century?', *Journal of Polymer Science, Part A: Polymer Chemistry*, **42**(3), pp. 487–495.

Brown Jr, R. M. and Montezinos, D. (1976) 'Cellulose microfibrils: Visualization of biosynthetic and orienting complexes in association with the plasma membrane (freeze-etching/Oocystis/enzyme complex/granule bands/terminal synthesis)', *Proceedings of the National Academy of Sciences*, **73**(1), pp. 143–147.

Carlisle, C. H. and Crowfoot, D. (1945) 'The Crystal Structure of Cholesteryl Iodide', *Proceedings of the Royal Society*, **A184**(996), pp. 64–83.

Caroff, M. and Karibian, D. (2003) 'Structure of bacterial lipopolysaccharides', *Carbohydrate Research*, **338**(23), pp. 2431–2447.

Chan, J., Fan, X., Hunter, S. W., Brennan, P. J. and Bloom, B. R. (1991) 'Lipoarabinomannan, a Possible Virulence Factor Involved in Persistence of Mycobacterium-Tuberculosis Within Macrophages', *Infection and Immunity*, **59**(5), pp. 1755–1761.

Chaplin, D. D. (2006) '1. Overview of the human immune response', *Journal of Allergy and Clinical Immunology*, **117**(2 SUPPL. 2), pp. 430–435.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) 'MolProbity : all-atom structure validation for macromolecular crystallography', *Acta Crystallographica*. International Union of Crystallography, **D66**(1), pp. 12–21.

Chiba, H., Pattanajitvilai, S., Evans, A. J., Harbeck, R. J. and Voelker, D. R. (2002) 'Human surfactant protein D (SP-D) binds Mycoplasma pneumoniae by high affinity interactions with lipids', *Journal of Biological Chemistry*, **277**(23), pp. 20379–20385.

Chludzinski, A. M., Germaine, G. R. and Schachtele, C. F. (1974) 'Purification and Properties of Dextransucrase from Streptococcus mutans', *Journal of Bacteriology*, **118**(1), pp. 1–7.

Clark, H. W., Mackay, R.-M. A., Deadman, M. E., Hood, D. W., Madsen, J., Moxon, E. R., Townsend, J. P., Reid, K. B. M., Ahmed, A. D., Shaw, A. J., Greenhough, T. J. and Shrive, A. K. (2016) 'Crystal structure of a complex of Surfactant Protein D and Haemophilus influenzae lipopolysaccharide reveals shielding of core structures in SP-D resistant strains', *Infection and Immunity*, **84**(5), pp. 1585–1592.

Coombs, P. J., Graham, S. A., Drickamer, K. and Taylor, M. E. (2005) 'Selective binding of the scavenger receptor C-type lectin to Lewis x trisaccharide and related glycan ligands', *Journal of Biological Chemistry*, **280**(24), pp. 22993–22999.

Cremer, D. and Pople, J. A. (1975) 'A General definition of ring puckering coordinates', *Journal of the American Chemical Society*, **97**(6), pp. 1354–1358.

Crouch, E. C., Hartshorn, K. L., Horlacher, T., McDonald, B., Smith, K., Cafarella, T. R., Seaton, B. A., Seeberger, P. H. and Head, J. F. (2009) 'Recognition of mannosylated ligands and influenza a virus by human surfactant protein D: Contributions of an extended site and residue 343', *Biochemistry*, **48**(15), pp. 3335–3345.

Crouch, E. C., McDonald, B., Smith, K., Cafarella, T. R., Seaton, B. A. and Head, J. F. (2006) 'Contributions of phenylalanine 335 to ligand recognition by human surfactant protein D: Ring interactions with SP-D ligands', *Journal of Biological Chemistry*, **281**(26), pp. 18008–18014.

Crouch, E. C., McDonald, B., Smith, K., Roberts, M., Mealy, T. R., Seaton, B. A. and Head, J. F. (2007) 'Critical role of Arg/Lys343 in the species-dependent recognition of phosphatidylinositol by pulmonary surfactant protein D', *Biochemistry*, **46**(17), pp. 5160–5169.

Crouch, E. C., Persson, A., Chang, D. and Heuser, J. (1994) 'Molecular structure of pulmonary surfactant protein D (SP-D).', *Journal of Biological Chemistry*, **269**(25), pp. 17311–17319.

Crouch, E. C., Smith, K., McDonald, B., Briner, D., Linders, B., McDonald, J., Holmskov, U., Head, J. F. and Hartshorn, K. L. (2006) 'Species differences in the carbohydrate binding preferences of surfactant protein D', *American Journal of Respiratory Cell and Molecular Biology*, **35**(1), pp. 84–94.

Crouch, E. C., Tu, Y., Briner, D., McDonald, B., Smith, K., Holmskov, U. and Hartshorn, K. L. (2005) 'Ligand Specificity of Human Surfactant Protein D', *Journal of Biological Chemistry*, **280**(17), pp. 17046–17056.

DiAngelo, S., Lin, Z., Wang, G., Phillips, S., Ramet, M., Luo, J. and Floros, J. (1999) 'Novel, non-radioactive, simple and multiplex PCR-cRFLP methods for genotyping human SP-A and SP-D marker alleles.', *Disease Markers*, **15**(4), pp. 269–81.

Drickamer, K., Dordal, M. S. and Reynolds, L. (1986) 'Mannose-binding proteins isolated from rat liver contain carbohydrate-recognition domains linked to collagenous tails. Complete primary structures and homology with pulmonary surfactant apoprotein', *Journal of Biological Chemistry*, **261**(15), pp. 6878–6887.

Drickamer, K. and Taylor, M. E. (2015) 'Recent insights into structures and functions of C-type lectins in the immune system', *Current Opinion in Structural Biology*, **34**(1), pp. 26–34.

Edlund, A. F., Swanson, R. and Preuss, D. (2004) 'Pollen and Stigma Structure and Function: The Role of Diversity in Pollination', *The Plant Cell*, **16**(suppl_1), pp. S84–S97.

Van Eijk, M., Rynkiewicz, M. J., White, M. R., Hartshorn, K. L., Zou, X., Schulten, K., Luo, D., Crouch, E. C., Cafarella, T. R., Head, J. F., Haagsman, H. P. and Seaton, B. A. (2012) 'A unique sugar-binding site mediates the distinct anti-influenza activity of pig surfactant protein D', *Journal of Biological Chemistry*, **287**(32), pp. 26666–26677.

Emsley, P. and Cowtan, K. D. (2004) 'Coot: Model-building tools for molecular graphics', *Acta Crystallographica*. International Union of Crystallography, **D60**(12 I), pp. 2126–2132.

Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. D. (2010) 'Features and development of Coot', *Acta Crystallographica*. International Union of Crystallography, **D66**(4), pp. 486–501.

Epstein, J., Eichbaum, Q., Sheriff, S. and Ezekowitz, R. A. B. (1996) 'The collectins in innate immunity', *Current Opinion in Immunology*, **8**(1), pp. 29–35.

Erridge, C., Bennett-Guerrero, E. and Poxton, I. R. (2002) 'Structure and function of lipopolysaccharides', *Microbes and Infection*, **4**(8), pp. 837–851.

Esser, A. F. (1994) 'The membrane attack complex of complement. Assembly, structure and cytotoxic activity', *Toxicology*, **87**(1–3), pp. 229–247.

Evans, P. R. (2006) 'Scaling and assessment of data quality', *Acta Crystallographica*. International Union of Crystallography, **D62**(1), pp. 72–82.

Evans, P. R. (2011) 'An introduction to data reduction: Space-group determination, scaling and intensity statistics', *Acta Crystallographica*. International Union of Crystallography, **D67**(4), pp. 282–292.

Evans, P. R. and Murshudov, G. N. (2013) 'How good are my data and what is the resolution?', *Acta Crystallographica*. International Union of Crystallography, **D69**(7), pp. 1204–1214.

Fakih, D., Pilecki, B., Schlosser, A., Jepsen, C. S., Thomsen, L. K., Ormhøj, M., Watson, A., Madsen, J., Clark, H. W., Barfod, K. K., Hansen, S. W. K., Marcussen, N., Jounblat, R., Chamat, S., Holmskov, U. and Sorensen, G. L. (2015) 'Protective effects of surfactant protein D treatment in 1,3-β-glucan-modulated allergic inflammation.', *American Journal of Physiology: Lung Cellular and Molecular Physiology*, **309**(11), pp. L1333-43.

Feinberg, H., Taylor, M. E. and Weis, W. I. (2007) 'Scavenger receptor C-type lectin binds to the leukocyte cell surface glycan Lewisx by a novel mechanism', *Journal of Biological Chemistry*, **282**(23), pp. 17250–17258.

Feinberg, H., Torgersen, D., Drickamer, K. and Weis, W. I. (2000) 'Mechanism of pH-dependent N-acetylgalactosamine binding by a functional mimic of the hepatocyte asialoglycoprotein receptor', *Journal of Biological Chemistry*, **275**(45), pp. 35176–35184.

Ferguson, C., Teeri, T. T., Siika-Aho, M., Read, S. M. and Bacic, A. (1998) 'Location of cellulose and callose in pollen tubes and grains of Nicotiana tabacum', *Planta*, **206**(3), pp. 452–460.

Ferguson, J. S., Voelker, D. R., McCormack, F. X. and Schlesinger, L. S. (2010) 'Surfactant Protein D Binds to Mycobacterium tuberculosis Bacilli and Lipoarabinomannan via Carbohydrate-Lectin Interactions Resulting in Reduced Phagocytosis of the Bacteria by Macrophages', *The Journal of Immunology*, **163**, pp. 312–321.

Fesel, P. H. and Zuccaro, A. (2016) 'β-glucan: Crucial component of the fungal cell wall and elusive MAMP in plants', *Fungal Genetics and Biology*, **90**(1), pp. 53–60.

Forsberg, L. S., Bhat, U. R. and Carlson, R. W. (2000) 'Structural characterization of the O-antigenic polysaccharide of the lipopolysaccharide from Rhizobium etli strain CE3', *Journal of Biological Chemistry*, **275**(25), pp. 18851–18863.

Fournier, B., Andargachew, R., Robin, A. Z., Laur, O., Voelker, D. R., Lee, W. Y., Weber, D. and Parkos, C. A. (2012) 'Surfactant protein D (SP-D) binds to membrane-proximal domain (D3) of signal regulatory protein α (SIRPα), a site distant from binding domain of CD47, while also binding to analogous region on signal regulatory protein β (SIRPβ)', *Journal of Biological Chemistry*, **287**(23), pp. 19386–19398.

French, S. and Wilson, K. S. (1978) 'On the treatment of negative intensity observations', *Acta Crystallographica*, **A34**(4), pp. 517–525.

Friedrich, W., Knipping, P. and Laue, M. V. (1913) 'Interferenz-Erscheinungen bei Rontgenstrahlen', *Annalen der Physik*, **346**(10), pp. 971–988.

Fujimoto, Y., Hattori, T., Uno, S., Murata, T. and Usui, T. (2009) 'Enzymatic synthesis of gentiooligosaccharides by transglycosylation with β-glycosidases from Penicillium multicolor', *Carbohydrate Research*, **344**(8), pp. 972–978.

Gaiha, G. D., Dong, T., Palaniyar, N., Mitchell, D. A., Reid, K. B. M. and Clark, H. W. (2008) 'Surfactant protein A binds to HIV and inhibits direct infection of CD4+ cells, but enhances dendritic cell-mediated viral transfer.', *The Journal of Immunology*, **181**(1), pp. 601–609.

Gamblin, S. J. and Skehel, J. J. (2010) 'Influenza hemagglutinin and neuraminidase membrane glycoproteins', *Journal of Biological Chemistry*, **285**(37), pp. 28403–28409.

Gander, J. E. (1974) 'Fungal cell wall glycoproteins and peptido-polysaccharides', *Annual Review of Microbiology*, **28**(1), pp. 103–119.

Gardai, S. J., Xiao, Y.-Q., Dickinson, M., Nick, J. A., Voelker, D. R., Greene, K. E. and Henson, P. M. (2003) 'By Binding SIRP-alpha or Calreticulin/CD91, Lung Collectins Act as Dual Function Surveillance Molecules to Suppress or Enhance Inflammation', *Cell*, **115**(1), pp. 13–23.

Geijtenbeek, T. B. H., Kwon, D. S., Torensma, R., van Vliet, S. J., van Duijnhoven, G. C. ., Middel, J., Cornelissen, I. L. M. H. ., Nottet, H. S. L. ., KewalRamani, V. N., Littman, D. R., Figdor, C. G. and van Kooyk, Y. (2000) 'DC-SIGN, a Dendritic Cell–Specific HIV-1-Binding Protein that Enhances trans-Infection of T Cells', *Cell*, **100**(5), pp. 587–597.

Gingras, A. R., Girija, U. V., Keeble, A. H., Panchal, R., Mitchell, D. A., Moody, P. C. E. and Wallis, R. (2011) 'Structural basis of mannan-binding lectin recognition by its associated serine protease MASP-1: Implications for complement activation', *Structure*, **19**(11), pp. 1635–1643.

Girija, U. V., Furze, C. M., Gingras, A. R., Yoshizaki, T., Ohtani, K., Marshall, J. E., Wallis, A. K., Schwaeble, W. J., El-Mezgueldi, M., Mitchell, D. A., Moody, P. C. E., Wakamiya, N. and Wallis, R. (2015) 'Molecular basis of sugar recognition by collectin-K1 and the effects of mutations associated with 3MC syndrome.', *BMC Biology*, **13**(1), p. 27.

Goh, B. C., Rynkiewicz, M. J., Cafarella, T. R., White, M. R., Hartshorn, K. L., Allen, K., Crouch, E. C., Calin, O., Seeberger, P. H., Schulten, K. and Seaton, B. A. (2013) 'Molecular mechanisms of inhibition of influenza by surfactant protein D revealed by large-scale molecular dynamics simulation', *Biochemistry*, **52**(47), pp. 8527–8538.

Green, P. L., Luty, A., Nair, S. V., Radford, J. and Raftos, D. A. (2006) 'A second form of collagenous lectin from the tunicate, Styela plicata', *Comparative Biochemistry and Physiology Part B*, **144**(3), pp. 343–350.

Green, P. L., Nair, S. V. and Raftos, D. A. (2003) 'Secretion of a collectin-like protein in tunicates is enhanced during inflammatory responses', *Developmental and Comparative Immunology*, **27**(1), pp. 3–9.

Gringhuis, S. I., den Dunnen, J., Litjens, M., van der Vlist, M. and Geijtenbeek, T. B. H. (2009) 'Carbohydrate-specific signaling through the DC-SIGN signalosome tailors immunity to Mycobacterium tuberculosis, HIV-1 and Helicobacter pylori', *Nature Immunology*, **10**(10), pp. 1081–1088.

Gringhuis, S. I., Kaptein, T. M., Wevers, B. A., Mesman, A. W. and Geijtenbeek, T. B. H. (2014) 'Fucose-specific DC-SIGN signalling directs T helper cell type-2 responses via IKKε- and CYLD-dependent Bcl3 activation', *Nature Communications*, **5**, pp. 3898.

Haagsman, H. P., Hawgood, S., Sargeant, T., White, R. T., Drickamer, K. and Benson, B. J. (1987) 'The Major Lung Surfactant Protein, SP 28-36, Is a Calcium-dependent, Carbohydrate-binding Protein', *Journal of Biological Chemistry*, **262**(29), pp. 13877–13880.

Haishima, Y., Holst, O. and Brade, H. (1992) 'Structural investigation on the lipopolysaccharide of Escherichia coli rough mutant F653 representing the R3 core type', *European Journal of Biochemistry*, **203**, pp. 127–134.

Håkansson, K., Lim, N. K., Hoppe, H.-J. and Reid, K. B. M. (1999) 'Crystal structure of the trimeric alpha-helical coiled-coil and the three lectin domains of human lung surfactant protein D.', *Structure*, **7**(3), pp. 255–264.

Hansen, S. W. K., Holm, D., Moeller, V., Vitved, L., Bendixen, C., Reid, K. B. M., Skjoedt, K. and Holmskov, U. (2002) 'CL-46, a novel collectin highly expressed in bovine thymus and liver.', *The Journal of Immunology*, **169**(10), pp. 5726–34.

Hansen, S. W. K., Holm, D., Moeller, V., Vitved, L., Bendixen, C., Skjoedt, K. and Holmskov, U. (2003) 'Genomic and molecular characterization of CL-43 and its proximal promoter', *Biochimica et Biophysica Acta*, **1625**(1), pp. 1–10.

Hansen, S. W. K. and Holmskov, U. (2002) 'Lung surfactant protein D (SP-D) and the molecular diverted descendants: conglutinin, CL-43 and CL-46.', *Immunobiology*, **205**(4–5), pp. 498–517.

Hansen, S. W. K., Selman, L., Palaniyar, N., Ziegler, K., Brandt, J., Kliem, A., Jonasson, M., Skjoedt, M.-O., Nielsen, O., Hartshorn, K. L., Jørgensen, T. J. D., Skjoedt, K., Holmskov, U. and Skjødt, K. (2010) 'Collectin 11 (CL-11, CL-K1) Is a MASP-1/3-Associated Plasma Collectin with Microbial-Binding Activity', *The Journal of Immunology*, **185**(10), pp. 6096–6104.

Hartshorn, K. L., Crouch, E. C., White, M. R., Colamussi, M. L., Kakkanatt, A., Tauber, B., Shepherd, V. L. and Sastry, K. N. (1998) 'Pulmonary surfactant proteins A and D enhance neutrophil uptake of bacteria.', *The American Journal of Physiology*, **274**(6 Pt 1), pp. L958–L969.

Hartshorn, K. L., Crouch, E. C., White, M. R., Eggleton, P., Tauber, A. I., Chang, D. and Sastry, K. N. (1994) 'Evidence for a protective role of pulmonary surfactant protein D (SP-D) against influenza A viruses.', *Journal of Clinical Investigation*, **94**(1), pp. 311–9.

Hartshorn, K. L., Halmskov, U., Hansen, S. W. K., Zhang, P., Meschi, J., Mogues, T., White, M. R. and Crouch, E. C. (2002) 'Distinctive anti-influenza properties of recombinant collectin 43', *Biochemical Journal*, **366**(1), pp. 87–96.

Hartshorn, K. L., White, M. R., Jensenius, J. C., Morris, S. M., Tauber, A. I., Crouch, E. C. and Reid, K. B. M. (1996) 'Neutrophil deactivation by influenza A viruses: mechanisms of protection after viral opsonization with collectins and hemagglutination-inhibiting antibodies', *Blood*, **87**(8), pp. 3450–3461.

Haurum, J. S., Thiel, S., Haagsman, H. P., Laursen, S. B., Larsen, B. and Jensenius, J. C. (1993) 'Studies on the carbohydrate-binding characteristics of human pulmonary surfactant-associated protein A and comparison with two other collectins: mannan-binding protein and conglutinin', *Biochemical Journal*, **293**(3), pp. 873–878.

Hayhurst, E. J., Kailas, L., Hobbs, J. K. and Foster, S. J. (2008) 'Cell wall peptidoglycan architecture in Bacillus subtilis.', *Proceedings of the National Academy of Sciences*, **105**(38), pp. 14603–8.

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', *Journal of Biological Chemistry*, **278**(44), pp. 43254–43260.

Hehre, E. J. (1956) 'Natural Synthesis of Low Molecular Weight (Clinical Type) Dextran by a Streptococcus Strain', *Journal of Biological Chemistry*, **222**, pp. 739–750.

Henrich, B., Bergamaschi, A., Broennimann, C., Dinapoli, R., Eikenberry, E. F., Johnson, I., Kobas, M., Kraft, P., Mozzanica, A. and Schmitt, B. (2009) 'PILATUS: A single photon counting pixel detector for X-ray applications', *Nuclear Instruments and Methods in Physics Research A*, **607**(1), pp. 247–249.

Henriksen, M. L., Brandt, J., Andrieu, J.-P., Nielsen, C., Jensen, P. H., Holmskov, U., Jørgensen, T. J. D., Palarasah, Y., Thielens, N. M. and Hansen, S. W. K. (2013) 'Heteromeric complexes of native collectin kidney 1 and collectin liver 1 are found in the circulation with MASPs and activate the complement system.', *The Journal of Immunology*, **191**(12), pp. 6117–27.

Henriksen, M. L., Brandt, J., Iyer, S. S. C., Thielens, N. M. and Hansen, S. W. K. (2013) 'Characterization of the interaction between collectin 11 (CL-11, CL-K1) and nucleic acids', *Molecular Immunology*, **56**(4), pp. 757–767.

Hickling, T. P., Bright, H., Wing, K., Gower, D., Martin, S. L., Sim, R. B. and Malhotra, R. (1999) 'A recombinant trimeric surfactant protein D carbohydrate recognition domain inhibits respiratory syncytial virus infection in vitro and in vivo.', *European Journal of Immunology*, **29**(11), pp. 3478–84.

Hillaire, M. L. B., Haagsman, H. P., Osterhaus, A. D. M. E., Rimmelzwaan, G. F. and Van Eijk, M. (2013) 'Pulmonary surfactant protein D in first-line innate defence against influenza A virus infections', *Journal of Innate Immunity*, **5**(3), pp. 197–208.

Hodgkin, D. C. (1949) 'The X-ray analysis of the structure of penicillin', *Advances in Science*, **6**(22), pp. 85–89.

Hodgkin, D. C., Pickworth, J., Robertson, J. H., Trueblood, K. N., Prosen, R. J. and White, J. G. (1955) 'Structure of Vitamin B12 : The Crystal Structure of the Hexacarboxylic Acid derived from B12 and the Molecular Structure of the Vitamin', *Nature*, **176**(4477), pp. 325–328.

Holmskov, U., Laursen, S. B., Malhotra, R., Wiedemann, H., Timpl, R., Stuart, G. R., Tornøe, I., Madsen, P. S., Reid, K. B. M. and Jensenius, J. C. (1995) 'Comparative study of the structural and functional properties of a bovine plasma C-type lectin, collectin-43, with other collectins.', *Biochemical Journal*, **305**, pp. 889–896.

Holmskov, U., Lawson, P., Teisner, B., Tornøe, I., Willis, A. C., Morgan, C., Koch, C. and Reid, K. B. M. (1997) 'Isolation and characterization of a new member of the scavenger receptor superfamily, glycoprotein-340 (gp-340), as a lung surfactant protein-D binding molecule', *Journal of Biological Chemistry*, **272**(21), pp. 13743–13749.

Holst, O. (2007) 'The structures of core regions from enterobacterial lipopolysaccharides - an update', *FEMS Microbiology Letters*, **271**(1), pp. 3–11.

Holst, O. (2011) 'Structure of the Lipopolysaccharide Core Region', in Knirel, Y. A. and Valvano, M. A. (eds) *Bacterial Lipopolysaccharides: Structure, Chemical Synthesis, Biogenesis and Interaction with Host Cells*. 1st edn. Vienna: Springer-Verlag Wein, pp. 21–39.

Holst, O. (2016) 'Structures of Lipid A and Core Region of Bacterial Lipopolysaccharides - An Update', *Trends in Carbohydrate Research*, **8**(1), pp. 1–13.

Holst, O. and Brade, H. (1991) 'Structural studies of the core region of the lipopolysaccharide from Salmonella minnesota strain R7 (rough mutant chemotype Rd1)', *Carbohydrate Research*, **219**, pp. 247–251.

Holst, O. and Brade, H. (1992) 'Chemical structure of the core region of lipopolysaccharides', in Morrison, D. C. and Ryan, J. L. (eds) *Bacterial endotoxic lipopolysaccharides*. Boca Raton: CRC Press, pp. 135–154.

Hooft, R. W., Sander, C. and Vriend, G. (1997) 'Objectively judging the quality of a protein structure from a Ramachandran plot.', *Computer Applications in the Biosciences*, **13**(4), pp. 425–430.

Hoppe, H.-J., Barlow, P. N. and Reid, K. B. M. (1994) 'A parallel three stranded ??-helical bundle at the nucleation site of collagen triple-helix formation', *FEBS Letters*, **344**(2–3), pp. 191–195.

Hunter, S. W., Gaylor, H. and Brennan, P. J. (1986) 'Structure and antigenicity of the phosphorylated lipopolysaccharide antigens from the leprosy and tubercle bacilli', *Journal of Biological Chemistry*, **261**(26), pp. 12345–12351.

Iobst, S. T. and Drickamer, K. (1994) 'Binding of Sugar Ligands to Ca2+-dependent Animal Lectins: II. Generation of High-Affinity Galactose Binding by Site-Directed Mutagenesis', *Journal of Biological Chemistry*, **269**(22), pp. 15512–15519.

Iobst, S. T., Wormald, M. R., Weis, W. I., Dwek, R. A. and Drickamer, K. (1994) 'Binding of sugar ligands to Ca2+-dependent animal lectins: I. Analysis of mannose binding by site-directed mutagenesis and NMR', *Journal of Biological Chemistry*, **269**(22), pp. 15505–15511.

Ip, W. K. E., Takahashi, K., Moore, K. J., Stuart, L. M. and Ezekowitz, R. A. B. (2008) 'Mannose-binding lectin enhances Toll-like receptors 2 and 6 signaling from the phagosome', *The Journal of Experimental Medicine*, **205**(1), pp. 169–181.

Van Iwaarden, J. F., Pikaar, J. C., Storm, J., Brouwer, E., Verhoef, J., Oosting, R. S., Van Golde, L. M. G. and van Strijp, J. A. G. (1994) 'Binding of surfactant protein A to the lipid A moiety of bacterial lipopolysaccharides.', *Biochemical Journal*, **303**(2), pp. 407–411.

Van Iwaarden, J. F., van Strijp, J. A. G., Visser, H., Haagsman, H. P., Verhoef, J. and van Golde, L. M. G. (1992) 'Binding of surfactant protein A (SP-A) to herpes simplex virus type 1-infected cells is mediated by the carbohydrate moiety of SP-A', *Journal of Biological Chemistry*, **267**(35), pp. 25039–25043.

Jack, D. L., Read, R. C., Tenner, A. J., Frosch, M., Turner, M. W. and Klein, N. J. (2001) 'Mannose-binding lectin regulates the inflammatory response of human professional phagocytes to Neisseria meningitidis serogroup B.', *The Journal of infectious diseases*, **184**(9), pp. 1152–62.

Jakel, A., Qaseem, A. S., Kishore, U. and Sim, R. B. (2013) 'Ligands ans receptors of Lung Surfactant Proteins SP-A and SP-D', *Frontiers in Bioscience*, **18**(3), pp. 1129–1140.

Jakel, A. and Sim, R. B. (2012) 'The human lung surfactant proteins A (SP-A) and D (SP-D) share similar binding mechanisms and common ligands on macrophages and dendritic cells', *Journal of Unsolved Questions*, **2**(2), pp. 12–18.

Janeway, C. A. (1989) 'Approaching the asymptote? Evolution and Revolution in Immunology', *Cold Spring Harbor Symposia on Quantitative Biology*, **54**(1), pp. 1–13.

Jayaraman, A., Koh, X., Li, J., Raman, R., Viswanathan, K., Shriver, Z. and Sasisekharan, R. (2012) 'Glycosylation at Asn [91] of H1N1 haemagglutinin affects binding to glycan receptors', *Biochemical Journal*, **444**(3), pp. 429–435.

Jeanes, A., Wilham, C. A., Jones, R. W., Tsuchna, H. M. and Rist, C. E. (1953) 'Isomaltose and Isomaltotriose from Enzymic Hydrolyzates of Dextran', *Journal of the American Chemical Society*, **75**(23), pp. 5911–5915.

Jones, T. A., Zou, J.-Y., Cowan, S. W. and Kjeldgaard, M. (1991) 'Improved Methods for the Building of Protein Models in Electron Density Maps and the Location of Errors in these Models', *Acta Crystallographica*, **A47**, pp. 110–119.

de Jong, M. A. W. P., Vriend, L. E. M., Theelen, B., Taylor, M. E., Fluitsma, D., Boekhout, T. and Geijtenbeek, T. B. H. (2010) 'C-type lectin Langerin is a β-glucan receptor on human Langerhans cells that recognizes opportunistic and pathogenic fungi', *Molecular Immunology*, **47**(6), pp. 1216–1225.

Jounblat, R., Kadioglu, A., Iannelli, F., Pozzi, G., Eggleton, P. and Andrew, P. W. (2004) 'Binding and Agglutination of Streptococcus pneumoniae by Human Surfactant Protein D (SP-D) Vary between Strains, but SP-D Fails to Enhance Killing by Neutrophils', *Infection and Immunity*, **72**(2), pp. 709–716.

Kabanov, D. S. and Prokhorenko, I. R. (2010) 'Structural analysis of lipopolysaccharides from Gram-negative bacteria', *Biochemistry (Moscow)*, **75**(4), pp. 383–404.

Kang, P. B., Azad, A. K., Torrelles, J. B., Kaufman, T. M., Beharka, A., Tibesar, E., DesJardin, L. E. and Schlesinger, L. S. (2005) 'The human macrophage mannose receptor directs *Mycobacterium tuberculosis* lipoarabinomannan-mediated phagosome biogenesis', *The Journal of Experimental Medicine*, **202**(7), pp. 987–999.

Kasaai, M. R. (2012) 'Dilute solution properties and degree of chain branching for dextran', *Carbohydrate Polymers*, **88**(1), pp. 373–381.

Kawabata, S. ichiro and Iwanaga, S. (1999) 'Role of lectins in the innate immunity of horseshoe crab', *Developmental and Comparative Immunology*, **23**(4–5), pp. 391–400.

Kawai, T. and Akira, S. (2010) 'The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors.', *Nature immunology*, **11**(5), pp. 373–84.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. and Phillips, D. C. (1958) 'A three-dimensional model of the myoglobin molecule obtained by X-ray analysis', *Nature*, **181**, pp. 662–666.

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. and Shore, V. C. (1960) 'Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2A. Resolution', *Nature*, **185**, pp. 422–427.

Keshi, H., Sakamoto, T., Kawai, T., Ohtani, K., Katoh, T., Jang, S.-J., Motomura, W., Yoshizaki, T., Fukuda, M., Koyama, S., Fukuzawa, J., Fukuoh, A., Yoshida, I., Suzuki, Y. and Wakamiya, N. (2006) 'Identification and Characterization of a Novel Human Collectin CL-K1', *Microbiology and Immunology*, **50**(12), pp. 1001–1013.

Kim, S. J., Chang, J. and Singh, M. (2015) 'Peptidoglycan architecture of Gram-positive bacteria by solid-state NMR', *Biochimica et Biophysica Acta*, **1848**(1), pp. 350–362.

De Kimpe, S. J., Kengatharan, M., Thiemermann, C. and Vane, J. R. (1995) 'The cell wall components peptidoglycan and lipoteichoic acid from Staphylococcus aureus act in synergy to cause shock and multiple organ failure.', *Proceedings of the National Academy of Sciences*, **92**(22), pp. 10359–63.

King, R. J., Simon, D. and Horowitz, P. M. (1989) 'Aspects of secondary and quaternary structure of surfactant protein A from canine lung', *Biochimica et Biophysica Acta*, **1001**(3), pp. 294–301.

Kishore, U., Ghai, R., Greenhough, T. J., Shrive, A. K., Bonifati, D. M., Gadjeva, M. G., Waters, P., Kojouharova, M. S., Chakraborty, T. and Agrawal, A. (2004) 'Structural and functional anatomy of the globular domain of complement protein C1q', *Immunology Letters*, **95**(2), pp. 113–128.

Kishore, U., Greenhough, T. J., Waters, P., Shrive, A. K., Ghai, R., Kamran, M. F., Bernal, A. L., Reid, K. B. M., Madan, T. and Chakraborty, T. (2006) 'Surfactant proteins SP-A and SP-D: Structure, function and receptors', *Molecular Immunology*, **43**(9), pp. 1293–1315.

Kishore, U. and Reid, K. B. M. (2000) 'C1q: Structure, function, and receptors', *Immunopharmacology*, **49**(1–2), pp. 159–170.

Kjaer, T. R., Jensen, L., Hansen, A., Dani, R., Jensenius, J. C., Dobó, J., Gál, P. and Thiel, S. (2016) 'Oligomerization of Mannan-binding Lectin Dictates Binding Properties and Complement Activation', *Scandinavian Journal of Immunology*, **84**(1), pp. 12–19.

Van Kooyk, Y. and Geijtenbeek, T. B. H. (2003) 'DC-SIGN: Escape mechanism for pathogens', *Nature Reviews Immunology*, **3**(9), pp. 697–709.

Kuan, S. F., Rust, K. and Crouch, E. C. (1992) 'Interactions of Surfactant Protein D with Bacterial Lipopolysaccharides', *Journal of Clinical Investigation*, **90**(July), pp. 97–106.

Kudo, K., Sano, H., Takahashi, H., Kuronuma, K., Yokota, S., Fujii, N., Shimada, K., Yano, I., Kumazawa, Y., Voelker, D. R., Abe, S. and Kuroki, Y. (2004) 'Pulmonary collectins enhance phagocytosis of Mycobacterium avium through increased activity of mannose receptor.', *The Journal of Immunology*, **172**(12), pp. 7592–602.

Laue, M. V. (1913) 'Eine quantitative prufung der theorie fur die interferenz-erscheinungen bei Rontgenstrahlen', *Annalen der Physik*, **346**(10), pp. 989–1002.

Lebedev, A. A., Young, P., Isupov, M. N., Moroz, O. V., Vagin, A. A. and Murshudov, G. N. (2012) 'JLigand: A graphical tool for the CCP4 template-restraint library', *Acta Crystallographica*, **D68**(4), pp. 431–440.

Lee, Y. M., Leiby, K. R., Allar, J., Paris, K., Lerch, B. and Okarma, T. B. (1991) 'Primary structure of bovine conglutinin, a member of the C-type animal lectin family', *Journal of Biological Chemistry*, **266**(5), pp. 2715–2723.

Lenardon, M. D., Munro, C. A. and Gow, N. A. R. (2010) 'Chitin synthesis and fungal pathogenesis', *Current Opinion in Microbiology*, **13**(4), pp. 416–423.

Leslie, A. G. W. (2006) 'The integration of macromolecular diffraction data', *Acta Crystallographica*, **D62**(1), pp. 48–57.

Leth-Larsen, R., Floridon, C., Nielsen, O. and Holmskov, U. (2004) 'Surfactant protein D in the female genital tract', *Molecular Human Reproduction*, **10**(3), pp. 149–154.

Leth-Larsen, R., Garred, P., Jensenius, H., Meschi, J., Hartshorn, K. L., Madsen, J., Tornøe, I., Madsen, H. O., Sorensen, G. L., Crouch, E. C. and Holmskov, U. (2005) 'A Common Polymorphism in the SFTPD Gene Influences Assembly, Function, and Concentration of Surfactant Protein D', *The Journal of Immunology*, **174**(3), pp. 1532–1538.

LeVine, A. M., Elliott, J., Whitsett, J. A., Srikiatkhachorn, A., Crouch, E. C., DeSilva, N. and Korfhagen, T. R. (2004) 'Surfactant protein-D enhances phagocytosis and pulmonary clearance of respiratory syncytial virus', *American Journal of Respiratory Cell and Molecular Biology*, **31**(2I), pp. 193–199.

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 or D Deficiency During Bacterial Infection on the Lung', *The Journal of Immunology*, **165**(7), pp. 3934–3940.

LeVine, A. M., Whitsett, J. A., Hartshorn, K. L., Crouch, E. C. and Korfhagen, T. R. (2001) 'Surfactant protein D enhances clearance of influenza A virus from the lung in vivo.', *The Journal of Immunology*, **167**(10), pp. 5868–5873.

Li, H., Yang, T., Liao, T., Debowski, A. W., Nilsson, H.-O., Fulurija, A., Haslam, S. M., Mulloy, B., Dell, A., Stubbs, K. A., Marshall, B. J. and Benghezal, M. (2017) 'The redefinition of Helicobacter pylori lipopolysaccharide O-antigen and core- oligosaccharide domains', *PLOS Pathogens*, **13**(3), pp. 1-21.

Lim, B. L., Wang, J., Holmskov, U., Hoppe, H.-J. and Reid, K. B. M. (1994) 'Expression of the Carbohydrate Recognition Domain of Lung Surfactant Protein D and Demonstration of its Binding to Lipopolysaccharides from Gram-negative Bacteria', *Biochemical and Biophysical Research Communications*, **202**(3), pp. 1674–1680.

Littlejohn, J. R., da Silva, R., Neale, W. A., Smallcombe, C., Clark, H. W., Mackay, R.-M. A., Watson, A., Madsen, J., Hood, D. W., Burns, I., Greenhough, T. J. and Shrive, A. K. (2018) 'Structural definition of hSP-D recognition of Salmonella enterica LPS inner core oligosaccharides reveals alternative binding modes for the same LPS', *PLOS ONE* (under review).

Lobato-Pascual, A., Saether, P. C., Fossum, S., Dissen, E. and Daws, M. R. (2013) 'Mincle, the receptor for mycobacterial cord factor, forms a functional receptor complex with MCL and FcεRI-γ', *European Journal of Immunology*, **43**(12), pp. 3167–3174.

Lu, J., Teh, C., Kishore, U. and Reid, K. B. M. (2002) 'Collectins and ficolins: Sugar pattern recognition molecules of the mammalian innate immune system', *Biochimica et Biophysica Acta*, **1572**(2–3), pp. 387–400.

Lu, J., Wiedemann, H., Holmskov, U., Thiel, S., Timpl, R. and Reid, K. B. M. (1993) 'Structural similarity between lung surfactant protein D and conglutinin: Two distinct, C-type lectins containing collagen-like sequences', *European Journal of Biochemistry*, **215**, pp. 793–799.

Luderitz, O., Galanos, C., Lehmann, V., Nurminen, M., Rietschel, E. T., Rosenfelder, G., Simon, M. and Westphal, O. (1973) 'Lipid A: Chemical Structure and Biological Activity', *The Journal of Infectious Disease*, **128**, pp. S17–S29.

Lüderitz, O., Staub, a M. and Westphal, O. (1966) 'Immunochemistry of O and R antigens of Salmonella and related Enterobacteriaceae.', *Bacteriological Reviews*, **30**(1), pp. 192–255.

Ma, Y. J., Skjoedt, M.-O. and Garred, P. (2013) 'Collectin-11/MASP Complex Formation Triggers Activation of the Lectin Complement Pathway - The Fifth Lectin Pathway Initiation Complex', *Journal of Innate Immunity*, **5**(3), pp. 242–250.

MacLeod, M. K. L., Kappler, J. W. and Marrack, P. (2010) 'Memory CD4 T cells: Generation, reactivation and re-assignment', *Immunology*, **130**(1), pp. 10–15.

Madan, T., Eggleton, P., Kishore, U., Strong, P., Aggrawal, S. S., Sarma, P. U. and Reid, K. B. M. (1997) 'Binding of pulmonary surfactant proteins A and D to Aspergillus fumigatus conidia enhances phagocytosis and killing by human neutrophils and alveolar macrophages', *Infection and Immunity*, **65**(8), pp. 3171–3179.

Madan, T., Kishore, U., Singh, M., Strong, P., Clark, H. W., Hussain, E. M., Reid, K. B. M. and Sarma, P. U. (2001) 'Surfactant proteins A and D protect mice against pulmonary hypersensitivity induced by Aspergillus fumigatus antigens and allergens', *Journal of Clinical Investigation*, **107**(4), pp. 467–475.

Madsen, J., Gaiha, G. D., Palaniyar, N., Dong, T., Mitchell, D. A. and Clark, H. W. (2013) 'Surfactant Protein D Modulates HIV Infection of Both T-Cells and Dendritic Cells', *PLOS ONE*, **8**(3), pp. 1–12.

Madsen, J., Kliem, A., Tornøe, I., Skjoedt, K., Koch, C. and Holmskov, U. (2000) 'Localization of lung surfactant protein D on mucosal surfaces in human tissues.', *The Journal of Immunology*, **164**(11), pp. 5866–5870.

Maeda, N., Nigou, J., Herrmann, J. L., Jackson, M., Amara, A., Lagrange, P. H., Puzo, G., Gicquel, B. and Neyrolles, O. (2003) 'The cell surface receptor DC-SIGN discriminates between Mycobacterium species through selective recognition of the mannose caps on lipoarabinomannan', *Journal of Biological Chemistry*, **278**(8), pp. 5513–5516.

Malhotra, R. and Sim, R. B. (1995) 'Collectins and viral infection', *Trends in Microbiology*, **3**(6), pp. 240–244.

Mansfield, L. P. and Forsythe, S. J. (2001) 'Demonstration of the Rb1 lipopolysaccharide core structure in Salmonella strains with the monoclonal antibody M105', *Journal of Medical Microbiology*, **50**(4), pp. 339–344.

Marchalonis, J. J. and Edelman, G. M. (1968) 'Isolation and characterization of a hemagglutinin from Limulus polyphemus', *Journal of Molecular Biology*, **32**(2), pp. 453–465.

Masoud, H., Altman, E., Richards, J. C. and Lam, J. S. (1994) 'General strategy for structural analysis of the oligosaccharide region of lipooligosaccharides. Structure of the oligosaccharide component of Pseudomonas aeruginosa IATS serotype 06 mutant R5 rough-type lipopolysaccharide', *Biochemistry*, **33**(1991), pp. 10568–10578.

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', *Journal of Biological Chemistry*, **274**(5), pp. 3173–3181.

McCormack, F. X., Gibbons, R., Ward, S. R., Kuzmenko, A., Wu, H. and Deepe, G. S. (2003) 'Macrophage-independent fungicidal action of the pulmonary collectins', *Journal of Biological Chemistry*, **278**(38), pp. 36250–36256.

McNeely, T. B. and Coonrod, J. D. (1993) 'Comparison of the opsonic activity of human surfactant protein A for Staphylococcus aureus and Streptococcus pneumoniae with rabbit and human macrophages.', *The Journal of Infectious Disease*, **167**(1), pp. 91–7.

Medzhitov, R. (2007) 'Recognition of microorganisms and activation of the immune response', *Nature*, **449**(7164), pp. 819–826.

Meschi, J., Crouch, E. C., Skolnik, P., Yahya, K., Holmskov, U., Leth-Larsen, R., Tornøe, I., Tecle, T., White, M. R. and Hartshorn, K. L. (2005) 'Surfactant protein D binds to human immunodeficiency virus (HIV) envelope protein gp120 and inhibits HIV replication', *Journal of General Virology*, **86**(11), pp. 3097–3107.

Michael, F. S., Vinogradov, E., Wenzel, C. Q., Mcintosh, B., Li, J., Hoe, J. C., Richards, J. C. and Cox, A. D. (2009) 'Phosphoethanolamine is located at the 6-position and not at the 7-position of the distal heptose residue in the lipopolysaccharide from Neisseria meningitidis', *Glycobiology*, **19**(12), pp. 1436–1445.

Mir-shekari, S. Y., Ashford, D. A., Harvey, D. J., Dwek, R. A. and Schulze, I. T. (1997) 'The Glycosylation of the Influenza A Virus Hemagglutinin by Mammalian Cells', *Journal of Biological Chemistry*, **272**(7), pp. 4027–4036.

Misaki, A., Torii, M., Sawai, T. and Goldstein, I. J. (1980) 'Structure of the dextran of Leuconostoc mesenteroides B-1355', *Carbohydrate Research*, **84**(2), pp. 273–285.

Mitchell, D. A., Fadden, A. J. and Drickamer, K. (2001) 'A Novel Mechanism of Carbohydrate Recognition by the C-type Lectins DC-SIGN and DC-SIGNR', *Journal of Biological Chemistry*, **276**(31), pp. 28939–28945.

Murakami, S., Iwaki, D., Mitsuzawa, H., Sano, H., Takahashi, H., Voelker, D. R., Akino, T. and Kuroki, Y. (2002) 'Surfactant protein A inhibits peptidoglycan-induced tumor necrosis factor-α secretion in U937 cells and alveolar macrophages by direct interaction with toll-like receptor 2', *Journal of Biological Chemistry*, **277**(9), pp. 6830–6837.

Murray, E., Khamri, W., Walker, M. M. M., Eggleton, P., Moran, A. P., Ferris, J. A., Knapp, S., Karim, Q. N., Worku, M., Strong, P., Reid, K. B. M. and Thursz, M. R. (2002) 'Expression of surfactant protein D in the human gastric mucosa and during Helicobacter pylori infection.', *Innate Immunity*, **70**(3), pp. 1481–1487.

Murray, J. W., Garman, E. F. and Ravelli, R. B. G. (2004) 'X-ray absorption by macromolecular crystals: The effects of wavelength and crystal composition on absorbed dose', *Journal of Applied Crystallography*, **37**(4), pp. 513–522.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. and Vagin, A. A. (2011) 'REFMAC5 for the refinement of macromolecular crystal structures', *Acta Crystallographica*, **D67**(4), pp. 355–367.

Nadesalingam, J., Dodds, A. W., Reid, K. B. M. and Palaniyar, N. (2005) 'Mannose-binding lectin recognizes peptidoglycan via the N-acetyl glucosamine moiety, and inhibits ligand-induced proinflammatory effect and promotes chemokine production by macrophages.', *Journal of Immunology*, **175**(3), pp. 1785–94.

Nadesalingam, J., Reid, K. B. M. and Palaniyar, N. (2005) 'Collectin surfactant protein D binds antibodies and interlinks innate and adaptive immune systems', *FEBS Letters*, **579**(20), pp. 4449–4453.

Naessens, M., Cerdobbel, A., Soetaert, W. and Vandamme, E. J. (2005) 'Leuconostoc dextransucrase and dextran: Production, properties and applications', *Journal of Chemical Technology and Biotechnology*, **80**(8), pp. 845–860.

Nair, S. V., Pearce, S., Green, P. L., Mahajan, D., Newton, R. A. and Raftos, D. A. (2000) 'A collectin-like protein from tunicates', *Comparative Biochemistry and Physiology Part B*, **125**(2), pp. 279–289.

O'Riordan, D. M., Standing, J. E., Kwon, K. Y., Chang, D., Crouch, E. C. and Limper, A. H. (1995) 'Surfactant protein D interacts with *Pneumocystis carinii* and mediates organism adherence to alveolar macrophages', *Journal of Clinical Investigation*, **95**(6), pp. 2699–2710.

O'Sullivan, A. C. (1997) 'Cellulose: the structure slowly unravels', *Cellulose*, **4**(3), pp. 173–207.

Oberley, R. E., Ault, K. A., Neff, T. L., Khubchandani, K. R., Crouch, E. C. and Snyder, J. M. (2004) 'Surfactant proteins A and D enhance the phagocytosis of Chlamydia into THP-1 cells', *American Journal of Physiology: Lung Cellular and Molecular Physiology*, **287**(10), pp. 296–306.

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**(12), pp. 861–870.

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', *Infection and Immunity*, **69**(1), pp. 24–33.

Ogasawara, Y., Kuroki, Y. and Akino, T. (1992) 'Pulmonary surfactant protein D specifically binds to phosphatidylinositol', *Journal of Biological Chemistry*, **267**(29), pp. 21244–21249.

Ohtani, K., Suzuki, Y., Eda, S., Kawai, T., Kase, T., Keshi, H., Sakai, Y., Fukuoh, A., Sakamoto, T., Itabe, H., Suzutani, T., Ogasawara, M., Yoshida, I. and Wakamiya, N. (2001) 'The Membrane-type Collectin CL-P1 Is a Scavenger Receptor on Vascular Endothelial Cells', *Journal of Biological Chemistry*, **276**(47), pp. 44222–44228.

Ohtani, K., Suzuki, Y., Eda, S., Kawai, T., Kase, T., Yamazaki, H., Shimada, T., Keshi, H., Sakai, Y., Fukuoh, A., Sakamoto, T. and Wakamiya, N. (1999) 'Molecular Cloning of a Novel Human Collectin from Liver (CL-L1)', *Journal of Biological Chemistry*, **274**(19), pp. 13681–13689.

Ohtani, K., Suzuki, Y. and Wakamiya, N. (2012) 'Biological functions of the novel collectins CL-L1, CL-K1, and CL-P1', *Journal of Biomedicine and Biotechnology*, **2012**, pp. 1–8.

Ohya, M., Nishitani, C., Sano, H., Yamada, C., Mitsuzawa, H., Shimizu, T., Saito, T., Smith, K., Crouch, E. C. and Kuroki, Y. (2006) 'Human pulmonary surfactant protein D binds the extracellular domains of Toll-like receptors 2 and 4 through the carbohydrate recognition domain by a mechanism different from its binding to phosphatidylinositol and lipopolysaccharide', *Biochemistry*, **45**(28), pp. 8657–8664.

Olsthoorn, M. M. A., Haverkamp, J. and Thomas-Oates, J. E. (1999) 'Mass spectrometric analysis of Klebsiella pneumoniae ssp. Pneumoniae rough strain R20 (O1-:K20-) lipopolysaccharide preparations: Identification of novel core oligosaccharide components and three 3-deoxy-D-manno-oct-2-ulopyranosonic artifacts', *Journal of Mass Spectrometry*, **34**(6), pp. 622–636.

Olsthoorn, M. M. A., Petersen, B. O., Schlecht, S., Haverkamp, J., Bock, K., Thomas-Oates, J. E. and Holst, O. (1998) 'Identification of a Novel Core Type in Salmonella Lipopolysaccharide: Complete structural analysis of the core region of the lipopolysaccharide from Salmonella enterica sv. Arizonae O62', *Journal of Biological Chemistry*, **273**(7), pp. 3817–3829.

Ooi, E. H., Wormald, P. J., Carney, A. S., James, C. L. and Tan, L. W. (2007) 'Surfactant protein d expression in chronic rhinosinusitis patients and immune responses in vitro to Aspergillus and alternaria in a nasal explant model', *Laryngoscope*, **117**(1), pp. 51–57.

Owen, R. L., Rudino-Pinera, E. and Garman, E. F. (2006) 'Experimental determination of the radiation dose limit for cryocooled protein crystals', *Proceedings of the National Academy of Sciences*, **103**(13), pp. 4912–4917.

Palaniyar, N., Nadesalingam, J. and Reid, K. B. M. (2002) 'Pulmonary innate immune proteins and receptors that interact with gram-positive bacterial ligands.', *Immunobiology*, **205**(4–5), pp. 575–594.

Palaniyar, N., Zhang, L., Kuzmenko, A., Ikegami, M., Wan, S., Wu, H., Korfhagen, T. R., Whitsett, J. A. and McCormack, F. X. (2002) 'The role of pulmonary collectin N-terminal domains in surfactant structure, function, and homeostasis in Vivo', *Journal of Biological Chemistry*, **277**(30), pp. 26971–26979.

Pape, K. A., Taylor, J. J., Maul, R. W., Gearhart, P. J. and Jenkins, M. K. (2011) 'Different B Cell Populations Mediate Early and Late Memory During an Endogenous Immune Response', *Science*, **331**(6021), pp. 1203–1207.

Park, J. T. and Uehara, T. (2008) 'How Bacteria Consume Their Own Exoskeletons (Turnover and Recycling of Cell Wall Peptidoglycan)', *Microbiology and Molecular Biology Reviews*, **72**(2), pp. 211–227.

de Pedro, M. A. and Cava, F. (2015) 'Structural constraints and dynamics of bacterial cell wall architecture', *Frontiers in Microbiology*, **6**(1), pp. 1–10.

Peiser, L., Mukhopadhyay, S. and Gordon, S. (2002) 'Scavenger receptors in innate immunity', *Current Opinion in Immunology*, **14**(1), pp. 123–128.

Pendrill, R., Säwen, E. and Widmalm, G. (2013) 'Conformation and Dynamics at a Flexible Glycosidic Linkage Revealed by NMR Spectroscopy and Molecular Dynamics Simulations: Analysis of β-L-Fucp-(1→6)-α-D-Glcp-OMe in Water Solution', *The Journal of Physical Chemsitry B*, **117**, pp. 14709–14722.

Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G. and North, A. C. (1960) 'Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis', *Nature*, **185**(4711), pp. 416–422.

Phillips, N. J., Apicella, M. A., Mcleod Griffiss, J. and Gibson, B. W. (1992) 'Structural Characterization of the Cell Surface Lipooligosaccharides from a Nontypable Strain of Haemophilus influenzae', *Biochemistry*, **31**, pp. 451–5.

Pikaar, J. C., Voorhout, W. F., van Golde, L. M. G., Verhoef, J., van Strijp, J. A. G. and Van Iwaarden, J. F. (1995) 'Opsonic activities of Surfactant porteins A nd D in phagocytosis of Gram negative bacteria by alveolar macrophages', *Journal of Infectious Diseases*, **172**, pp. 481–489.

Poletto, M., Ornaghi Júnior, H. L. and Zattera, A. J. (2014) 'Native cellulose: Structure, characterization and thermal properties', *Materials*, **7**(9), pp. 6105–6119.

Pondman, K. M., Paudyal, B., Sim, R. B., Kaur, A., Kouser, L., Tsolaki, A. G., Jones, L. A., Salvador-Morales, C., Khan, H. A., Ten Haken, B., Stenbeck, G. and Kishore, U. (2017) 'Pulmonary surfactant protein SP-D opsonises carbon nanotubes and augments their phagocytosis and subsequent pro-inflammatory immune response', *Nanoscale*, **9**(3), pp. 1097-1109.

Que, N. L. S., Lin, S., Cotter, R. J. and Raetz, C. R. H. (2000) 'Purification and Mass Spectrometry of Six Lipid A Species from the Bacterial Endosymbiont Rhizobium etli', *Journal of Biological Chemistry*, **275**(36), pp. 28006–28016.

Rao, V. S. R., Qasba, P. K., Balaji, P. V. and Chandrasekaran, R. (1998) *Conformation of Carbohydrates*. Amsterdam: Harwood Academic Publishers.

Raska, M., Takahashi, K., Czernekova, L., Zachova, K., Hall, S., Moldoveanu, Z., Elliott, M. C., Wilson, L., Brown, R., Jancova, D., Barnes, S., Vrbkova, J., Tomana, M., Smith, P. D., Mestecky, J., Renfrow, M. B. and Novak, J. (2010) 'Glycosylation patterns of HIV-1 gp120 depend on the type of expressing cells and affect antibody recognition', *Journal of Biological Chemistry*, **285**(27), pp. 20860–20869.

Ravelli, R. B. G. and Garman, E. F. (2006) 'Radiation damage in macromolecular cryocrystallography', *Current Opinion in Structural Biology*, **16**(5), pp. 624–629.

Reading, P. C., Holmskov, U. and Anders, E. M. (1998) 'Antiviral activity of bovine collectins against rotaviruses', *Journal of General Virology*, **79**(9), pp. 2255–2263.

Reinhardt, A., Wehle, M., Geissner, A., Crouch, E. C., Kang, Y., Yang, Y., Anish, C., Santer, M. and Seeberger, P. H. (2016) 'Structure binding relationship of human surfactant protein D and various lipopolysaccharide inner core structures', *Journal of Structural Biology*, **195**, pp. 387–395.

Remaud-Simeon, M., Willemot, R. M., Sarçabal, P., Potocki De Montalk, G. and Monsan, P. (2000) 'Glucansucrases: Molecular engineering and oligosaccharide synthesis', *Journal of Molecular Catalysis - B Enzymatic*, **10**(1–3), pp. 117–128.

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', *American Journal of Respiratory Cell and Molecular Biology*, **21**(5), pp. 576–585.

Roche, A.-C. and Monsigny, M. (1974) 'Purification and properties of limulin: A lectin (agglutinin) from hemolymph of Limulus polyphemus', *Biochimica et Biophysica Acta*, **371**(1), pp. 242–254.

Rosseau, S., Guenther, A., Seeger, W. and Lohmeyer, J. (1997) 'Phagocytosis of viable Candida albicans by alveolar macrophages: lack of opsonin function of surfactant protein A', *Journal of Infectious Diseases*, **175**(2), pp. 421–428.

Rothmann, A. B., Mortensen, H. D., Holmskov, U. and Hojrup, P. (1997) 'Structural Characterization of Bovine Collectin-43', *European Journal of Biochemistry*, **243**(3), pp. 630–635.

van Rozendaal, B. A. W. M., van Spriel, A. B., van de Winkel, J. G. and Haagsman, H. P. (2000) 'Role of pulmonary surfactant protein D in innate defense against Candida albicans.', *The Journal of Infectious Disease*, **182**(3), pp. 917–22.

Rund, S., Lindner, B., Brade, H. and Holst, O. (1999) 'Structural Analysis of the Lipopolysaccharide from Chlamydia trachomatis Serotype L2', *Journal of Biological Chemistry*, **274**(24), pp. 16819–16824.

Rynkiewicz, M. J., Wu, H., Cafarella, T. R., Nikolaidis, N. M., Head, J. F., Seaton, B. A. and McCormack, F. X. (2017) 'Differential Ligand Binding Specificities of the Pulmonary Collectins Are Determined by the Conformational Freedom of a Surface Loop', *Biochemistry*, **56**(31), pp. 4095–4105.

Saar Dover, R., Bitler, A., Shimoni, E., Trieu-Cuot, P. and Shai, Y. (2015) 'Multiparametric AFM reveals turgor-responsive net-like peptidoglycan architecture in live streptococci', *Nature Communications*. Nature Publishing Group, **6**(May), pp. 7193.

Sahly, H., Ofek, I., Podschun, R., Brade, H., He, Y., Ullmann, U. and Crouch, E. C. (2002) 'Surfactant protein D binds selectively to Klebsiella pneumoniae lipopolysaccharides containing mannose-rich O-antigens.', *The Journal of Immunology*, **169**(6), pp. 3267–74.

Sano, H., Sohma, H., Muta, T., Nomura, S., Voelker, D. R. and Kuroki, Y. (1999) 'Pulmonary surfactant protein A modulates the cellular response to smooth and rough lipopolysaccharides by interaction with CD14.', *The Journal of Immunology*, **163**(1), pp. 387–95.

Sarashina-Kida, H., Negishi, H., Nishio, J., Suda, W., Nakajima, Y., Yasui-Kato, M., Iwaisako, K., Kang, S., Endo, N., Yanai, H., Asagiri, M., Kida, H., Hattori, M., Kumanogoh, A. and Taniguchi, T. (2017) 'Gallbladder-derived surfactant protein D regulates gut commensal bacteria for maintaining intestinal homeostasis', *Proceedings of the National Academy of Sciences*, **114**(38), pp. 10178-10183.

Sato, M., Sano, H., Iwaki, D., Kudo, K., Konishi, M., Takahashi, H., Takahashi, T., Imaizumi, H., Asai, Y. and Kuroki, Y. (2003) 'Direct Binding of Toll-Like Receptor 2 to Zymosan, and Zymosan-Induced NF- B Activation and TNF- Secretion Are Down-Regulated by Lung Collectin Surfactant Protein A', *The Journal of Immunology*, **171**(1), pp. 417–425.

Sato, Y., Morimoto, K., Kubo, T., Sakaguchi, T., Nishizono, A., Hirayama, M. and Hori, K. (2015) 'Entry inhibition of influenza viruses with high mannose binding lectin ESA-2 from the red alga Eucheuma serra through the recognition of viral hemagglutinin', *Marine Drugs*, **13**(6), pp. 3454–3465.

Schleifer, K. H. and Kandler, O. (1972) 'Peptidoglycan types of bacterial cell walls and their taxonomic implications.', *Bacteriological Reviews*, **36**(4), pp. 407–477.

Schneider, R., Hanak, T., Persson, S. and Voigt, C. A. (2016) 'Cellulose and callose synthesis and organization in focus, what's new?', *Current Opinion in Plant Biology*, **34**, pp. 9–16.

Schwaeble, W. J., Dahl, M. R., Thiel, S., Stover, C. and Jensenius, J. C. (2002) 'The mannan-binding lectin-associated serine proteases (MASPs) and MAp19: four components of the lectin pathway activation complex encoded by two genes.', *Immunobiology*, **205**, pp. 455–466.

Shaw, A. J. (2009) *Structural studies of ligand recognition by the collectins hSP-D, CL-46 and Conglutinin*. Keele University.

Sheriff, S., Chang, C. Y. and Ezekowitz, R. A. B. (1994) 'Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple α-helical coiled-coil', *Nature Structural Biology*, **1**(11), pp. 789–794.

Shirato, K., Miyoshi, H., Goto, A., Ako, Y., Ueki, T., Kariwa, H. and Takashima, I. (2004) 'Viral envelope protein glycosylation is a molecular determinant of the neuroinvasiveness of the New York strain of West Nile virus', *Journal of General Virology*, **85**(12), pp. 3637–3645.

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. and Greenhough, T. J. (2009) 'Structural Characterisation of Ligand-Binding Determinants in Human Lung Surfactant Protein D: Influence of Asp325', *Journal of Molecular Biology*. Elsevier Ltd, **394**(4), pp. 776–788.

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', *Journal of Molecular Biology*, **331**(2), pp. 509–523.

da Silva, R. (2016) *Structural Studies of Surfactant Protein D in Complex with Bacterial Lipopolysaccharide Ligands*. Keele University.

Sloan, J. W., Alexander, B. H., Lohmar, R. L., Wolff, I. A. and Rist, C. E. (1954) 'Determination of Dextran Structure by Periodate Oxidation Techniques', *Journal of the American Chemical Society*, **76**(17), pp. 4429–4434.

Ślusarz, R., Szulc, M. and Madaj, J. (2014) 'Molecular modeling of Gram-positive bacteria peptidoglycan layer, selected glycopeptide antibiotics and vancomycin derivatives modified with sugar moieties', *Carbohydrate Research*, **389**(1), pp. 154–164.

Smallcombe, C. (2014) *Structural studies of the recognition of bacterial lipopolysaccharides by human surfactant protein D*. Keele University.

Snyder, G. A., Ford, J., Torabi-parizi, P., Arthos, J. A., Schuck, P., Colonna, M. and Sun, P. D. (2005) 'Characterization of DC-SIGN / R Interaction with Human Immunodeficiency Virus Type 1 gp120 and ICAM Molecules Favors the Receptor ' s Role as an Antigen-Capturing Rather than an Adhesion Receptor', *Journal of Virology*, **79**(8), pp. 4589–4598.

Sodeik, B. and Krijnse-Locker, J. (2002) 'Assembly of vaccinia virus revisited: De novo membrane synthesis or acquisition from the host?', *Trends in Microbiology*, **10**(1), pp. 15–24.

Solomon, E. B., Niemira, B. A., Sapers, G. M. and Annous, B. A. (2005) 'Biofilm Formation, Cellulose Production, and Curli Biosynthesis by Salmonella Originating from Produce, Animal, and Clinical Sources', *Journal of Food Protection*, **68**(5), pp. 906–912.

Somers, W. S., Tang, J., Shaw, G. D. and Camphausen, R. T. (2000) 'Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLeX and PSGL-1 (Cell 103:3 (467-479))', *Cell*, **103**(3), pp. 467–479.

Stahl, P. D., Rodman, J. S., Miller, M. J. and Schlesinger, P. H. (1978) 'Evidence for receptor-mediated binding of glycoproteins, glycoconjugates, and lysosomal glycosidases by alveolar macrophages.', *Proceedings of the National Academy of Sciences of the United States of America*, **75**(3), pp. 1399–403.

Takizawa, H. and Manz, M. G. (2007) 'Macrophage tolerance: CD47-SIRP-alpha-mediated signals matter.', *Nature immunology*, **8**(12), pp. 1287–1289.

Taylor, P. R., Gordon, S. and Martinez-Pomares, L. (2005) 'The mannose receptor: Linking homeostasis and immunity through sugar recognition', *Trends in Immunology*, **26**(2), pp. 104–110.

Tenner, A. J., Robinson, S. L., Borchelt, J. and Wright, J. R. (1989) 'Human pulmonary surfactant protein (SP-A), a protein structurally homologous to C1q, can enhance FcR- and CR1-mediated phagocytosis', *Journal of Biological Chemistry*, **264**(23), pp. 13923–13928.

Thawer, S., Auret, J., Schnoeller, C., Chetty, A., Smith, K., Darby, M., Roberts, L., Mackay, R.-M. A., Whitwell, H. J., Timms, J. F., Madsen, J., Selkirk, M. E., Brombacher, F., Clark, H. W. and Horsnell, W. G. C. (2016) 'Surfactant Protein-D Is Essential for Immunity to Helminth Infection', *PLOS Pathogens*, **12**(2), pp. 1–18.

Ujma, S., Horsnell, W. G. C., Katz, A. A., Clark, H. W. and Schäfer, G. (2017) 'Non-Pulmonary Immune Functions of Surfactant Proteins A and D', *Journal of Innate Immunity*, **9**(1), pp. 3–11.

Varghese, J. N., Laver, W. G. and Colman, P. M. (1983) 'Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution', *Nature*, **303**(5912), pp. 35–40.

Veldhuizen, E. J. A., Van Eijk, M. and Haagsman, H. P. (2011) 'The carbohydrate recognition domain of collectins', *FEBS Journal*, **278**(20), pp. 3930–3941.

Vigerust, D. J. and Shepherd, V. L. (2007) 'Virus glycosylation: role in virulence and immune interactions', *Trends in Microbiology*, **15**(5), pp. 211–218.

Vollmer, W., Blanot, D. and De Pedro, M. A. (2008) 'Peptidoglycan structure and architecture', *FEMS Microbiology Reviews*, **32**(2), pp. 149–167.

Vollmer, W. and Seligman, S. J. (2010) 'Architecture of peptidoglycan: more data and more models', *Trends in Microbiology*, **18**(2), pp. 59–66.

Voorhout, W. F., Velavan, T. P., Kuroki, Y., Ogasawara, Y., van Golde, L. M. G. and Geuze, H. J. (1992) 'Immunocytochemical localization of surfactant protein D (SP-D) in type II cells, Clara cells, and alveolar macrophages of rat lung.', *The Journal of Histochemistry and Cytochemistry*, **40**(10), pp. 1589–1597.

Voss, T., Eistetter, H., Schafer, K. P. and Engel, J. (1988) 'Macromolecular organization of natural and recombinant lung surfactant protein SP 28-36. Structural homology with the complement factor C1q', *Journal of Molecular Biology*, **201**(1), pp. 219–227.

Wallis, R. (2002) 'The Lectin-Pathway of Complement Activation : MBL , other Collectins and Ficolins. Structural and Functional Aspects of Complement Activation by Mannose-binding Protein', *Immunobiology*, **205**, pp. 433–445.

Wang, H., Head, J. F., Kosma, P., Brade, H., Müller-Loennies, S., Sheikh, S., McDonald, B., Smith, K., Cafarella, T. R., Seaton, B. A. and Crouch, E. C. (2008) 'Recognition of Heptoses and the Inner Core of Bacterial Lipopolysaccharides by Surfactant Protein D', *Biochemistry*, **47**(2), pp. 710–720.

Wang, J., Kishore, U., Lim, B. L., Strong, P. and Reid, K. B. M. (1996) 'Interaction of human lung surfactant proteins A and D with mite (Dermatophagoides pteronyssinus) allergens', *Clinical & Experimental Immunology*, **106**(2), pp. 367–373.

Wang, J. Y., Shieh, C. C., You, P. F., Lei, H. Y. and Reid, K. B. M. (1998) 'Inhibitory effect of pulmonary surfactant proteins A and D on allergen-induced lymphocyte proliferation and histamine release in children with asthma.', *American Journal of Respiratory and Critical Care Medicine*, **158**(2), pp. 510–518.

Watford, W. T., Ghio, a J. and Wright, J. R. (2000) 'Complement-mediated host defense in the lung.', *American Journal of Physiology: Lung Cellular and Molecular Physiology*, **279**(5), pp. L790-8.

Watford, W. T., Wright, J. R., Hester, C. G., Jiang, H. and Frank, M. M. (2001) 'Surfactant protein A regulates complement activation.', *The Journal of Immunology*, **167**(11), pp. 6593–600.

Watson, J. . and Crick, F. H. C. (1953) 'Molecular Structure of Nucleic Acids', *Nature*, **171**, pp. 737–738.

Weis, W. I. and Drickamer, K. (1994) 'Trimeric structure of a C-type mannose-binding protein', *Structure*, **2**(12), pp. 1227–1240.

Weis, W. I., Taylor, M. E. and Drickamer, K. (1998) 'The C-type lectin superfamily in the immune system', *Immunological Reviews*, **163**, pp. 19–34.

Welsch, S., Keppler, O. T., Habermann, A., Allespach, I., Krijnse-Locker, J. and Kräusslich, H. G. (2007) 'HIV-1 buds predominantly at the plasma membrane of primary human macrophages', *PLOS Pathogens*, **3**(3), pp. 36-47.

Werninghaus, K., Babiak, A., Groß, O., Hölscher, C., Dietrich, H., Agger, E. M., Mages, J., Mocsai, A., Schoenen, H., Finger, K., Nimmerjahn, F., Brown, G. D., Kirschning, C., Heit, A., Andersen, P., Wagner, H., Ruland, J. and Lang, R. (2009) 'Adjuvanticity of a synthetic cord factor analogue for subunit *Mycobacterium tuberculosis* vaccination requires FcRγ–Syk–Card9–dependent innate immune activation', *The Journal of Experimental Medicine*, **206**(1), pp. 89–97.

van de Wetering, J. K., van Eijk, M., van Golde, L. M. G., Hartung, T., van Strijp, J. A. G. and Batenburg, J. J. (2001) 'Characteristics of Surfactant Protein A and D Binding to Lipoteichoic Acid and Peptidoglycan, 2 Major Cell Wall Components of Gram-Positive Bacteria', *The Journal of Infectious Disease*, **184**(9), pp. 1143–1151.

Van De Wetering, J. K., Van Remoortere, A., Vaandrager, A. B., Batenburg, J. J., van Golde, L. M. G., Hokke, C. H. and Van Hellemond, J. J. (2004) 'Surfactant protein D binding to terminal alpha-1-3-linked fucose residues and to Schistosoma mansoni', *American Journal of Respiratory Cell and Molecular Biology*, **31**(5), pp. 565–572.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. A. and Wilson, K. S. (2011) 'Overview of the CCP4 suite and current developments', *Acta Crystallographica*. International Union of Crystallography, **D67**(4), pp. 235–242.

Wright, J. R. (2005) 'Immunoregulatory functions of surfactant proteins.', *Nature Reviews Immunology*, **5**(1), pp. 58–68.

Wu, H., Kuzmenko, A., Wan, S., Schaffer, L., Weiss, A., Fisher, J. H., Kim, K. S. and McCormack, F. X. (2003) 'Surfactant proteins A and D inhibit the growth of Gram-negative bacteria by increasing membrane permeability', *Journal of Clinical Investigation*, **111**(10), pp. 1589–1602.

Yang, H., Singh, M., Kim, S. J. and Schaefer, J. (2017) 'Characterization of the Tertiary Structure of the Peptidoglycan of Enterococcus faecalis', *Biochimica et Biophysica Acta*. Elsevier B.V., **1859**(11), pp. 2171–2180.

Yoshida, T., Tsuruta, Y., Iwasaki, M., Yamane, S., Ochi, T. and Suzuki, R. (2003) 'SRCL/CL-P1 recognizes GalNAc and a carcinoma-associated antigen, Tn antigen', *Journal of Biochemistry*, **133**(3), pp. 271–277.

Zelensky, A. N. and Gready, J. E. (2005) 'The C-type lectin-like domain superfamily', *FEBS Journal*, **272**(24), pp. 6179–6217.

Zhang, Z., Abdel-Razek, O., Hawgood, S. and Wang, G. (2015) 'Protective Role of Surfactant Protein D in Ocular Staphylococcus aureus Infection', *PLOS ONE*. Edited by Y. P. Di, **10**(9), pp. 138597.

Zhou, L., Hinerman, J. M., Blaszczyk, M., Miller, J. L. C., Conrady, D. G., Barrow, A. D., Chirgadze, D. Y., Bihan, D., Farndale, R. W. and Herr, A. B. (2016) 'Structural basis for collagen recognition by the immune receptor OSCAR', *Blood*, **127**(5), pp. 529–537.

Appendices

Appendix I:

Refinement statistics for maltotetraose and maltoheptaose complexed with rfhSP-D

Refinement statistics from the maltotetraose and maltoheptaose bound complexes of recombinant fragment human surfactactant protein D. The data was reprocessed as part of the small ligand studies (chapter 4) following the revelation that β-D-glucose can be recognised by hSP-D.

	Maltotetraose	Maltoheptaose
Data collection		
Wavelength (Å)	1.117	1.488
Temperature (K)	100	100
Space group	$P2_1$	$P2_1$
Cell Dimensions		
a (Å)	55.53	55.41
b (Å)	108.45	108.14
c (Å)	55.76	55.67
β (°)	90.94	91.24
Maximal resolution (Å)	1.50	1.65
Resolution range (Å)	55.75–1.50 (1.54-1.50)	55.43-1.65 (1.74-1.65)
Observations	1 004 446 (59 772)	190 221 (22 059)
Unique reflections	104 894 (15 315)	60 189 (6 974)
Completeness (%)	99.7	95.4
R_{merge}[a]	0.382	
$I/\sigma(I)$	2.0	2.4
Refinement		
Protein atoms[b]	3 475	3 469
Residues, chain A	206 - 355	205 - 355
Residues, chain B	205 - 355	204 - 355
Residues, chain C	206 - 355	206 - 355
Other atoms		
Calcium ions	10	10
Ligand	A = 4, B = 1, C = 1	A = 2, B = 1, C = 1
Water	639	525
Resolution range (Å)	55.75 – 1.50	55.43 – 1.65
R_{work}[c]	0.1703	0.1664
R_{free}[d]	0.1903	0.2084
Average *B*-values (Å^2)		
Protein main chain	17.89	23.74
Ligand	24.93	30.38
Water	36.67	34.94
Ramachandron plot values[e] (%)		
Favoured	98.43	98.21
Allowed	1.57	1.79
Disallowed	0.00	0.00

[a] $R_{merge} = \Sigma_h \Sigma_j |I_{h,j} - I_h| \; / \; \Sigma_b \Sigma_j |I_{h,j}|$, where $I_{h,j}$ is the *j*th observation of reflection *h* and I_h is the mean of *j* for reflection *h*

[b] Total number of protein atoms used in refinement

[c] $R_{work} = \Sigma_h ||F_{oh}| - |F_{ch}|| \; / \; \Sigma_h |F_{oh}|$, where F_{oh} and F_{ch} are the observed and calculated structure factor amplitudes, respectively, for reflection *h*.

[d] $R_{free} = R_{work}$ but for a random 5% subset of reflections

[e] defined by MolProbity analysis

Appendix II

Structural definition of hSP-D recognition of *Salmonella enterica* LPS inner core oligosaccharides reveals alternative binding modes for the same LPS

Structural definition of hSP-D recognition of *Salmonella enterica* LPS inner core oligosaccharides reveals alternative binding modes for the same LPS

Jamie R. Littlejohn[1¶], Ruben da Silva[1¶], William A. Neale[1¶], Carrie C. Smallcombe[1#¶], Howard W. Clark[2,3,4], Rose-Marie Mackay[2], Alastair S. Watson[2], Jens Madsen[2,3,4], Derek W. Hood[5], Ian Burns[1], Trevor J. Greenhough[1], and Annette K. Shrive[1*]

[1] School of Life Sciences, Keele University, Staffordshire ST5 5BG, U.K.

[2] University of Southampton, Department of Child Health, Division of Clinical and Experimental Sciences, Sir Henry Wellcome Laboratories, Southampton General Hospital, Southampton, UK

[3] Southampton NIHR Respiratory Biomedical Research Unit, Southampton General Hospital, Southampton, UK

[4] Institute for Life Sciences, University of Southampton, Southampton SO17 1BJ, UK

[5] Mammalian Genetics Unit, MRC Harwell Institute, Harwell Science and Innovation Campus, Oxfordshire, UK

[#]Current Address: Department of Pathobiology, Cleveland Clinic, 9500 Euclid Avenue, Cleveland, OH 44195, U.S.A.

*Corresponding author

E-mail: a.k.shrive@keele.ac.uk (AKS)

[¶]*These Authors contributed equally to this work*

Abstract

The crystal structures of a biologically and therapeutically active recombinant homotrimeric fragment of native human SP-D (hSP-D) complexed with the inner core oligosaccharide of the *Salmonella enterica* sv Minnesota rough strains R5 and R7 (rough mutant chemotypes Rc and Rd1) have been determined. The structures reveal that hSP-D specifically and preferentially targets the LPS inner core via the innermost conserved Hep-Kdo pair with the flexibility for alternative recognition when this preferred epitope is not available for binding. Hep-Kdo binding is achieved through calcium dependent recognition of the heptose dihydroxyethyl side chain coupled with specific interactions between the Kdo and the binding site flanking residues Arg343 and Asp325 with evidence for an extended binding site for LPS inner cores containing multiple Kdo residues. In one subunit of the R5-bound structure this preferred mode of binding is precluded by the crystal lattice and oligosaccharide is bound through the terminal inner core glucose. The structures presented here thus provide unique multiple insights into the recognition and binding of bacterial LPS by hSP-D. Not only is it demonstrated that hSP-D targets the highly conserved LPS proximal inner core Hep-Kdo motif, but also that hSP-D can recognise either terminal or non-terminal sugars and has the flexibility and versatility to adopt alternative strategies for bacterial recognition, utilising alternative LPS epitopes when the preferred inner core Hep-Kdo disaccharide is not available for binding.

Author Summary

Host recognition of bacterial surface lipopolysaccharide (LPS) by the innate immune protein human surfactant protein D (hSP-D) has previously been shown to target the LPS inner core sugar heptose with the binding pocket mechanism including both the heptose and the preceding sugar Kdo. This recognition is believed to be compromised by extended LPS structures that effectively shield vulnerable sites in the LPS core, allowing pathogens with complex LPS extensions to efficiently evade hSP-D mediated innate immune defence.

The crystal structures of the biologically and therapeutically active recombinant fragment of native hSP-D presented here act as an LPS molecular filter, providing binding sites that both accept and preclude inner core heptose-Kdo binding. Our study shows for the first time that where both an LPS inner core Heptose-Kdo motif and the hSP-D LPS binding site are accessible, recognition of bacterial LPS occurs through this motif; where this binding mode is not available and a suitable alternative LPS epitope is present, then an alternative mode of recognition is adopted. Specifically, we show that hSP-D recognition of *Salmonella enterica* strains R5 and R7 LPS occurs preferentially through the inner core heptose-Kdo, with hSP-D having the flexibility and versatility to target the terminal glucose of the R5 mutant LPS when the preferred epitope is not accessible for binding.

Introduction

The innate immune protein human surfactant protein D (hSP-D) is a collagenous C-type lectin prototypically associated with the lung surfactant but now identified throughout the human body [1]. SP-D recognises a wide range of pathogens, through their surface carbohydrate arrays, leading to rapid deactivation and clearance through activation of phagocytic cells [2-4].

Numerous studies of the binding of simple sugars to both hSP-D and rfhSP-D, a recombinant homotrimeric native fragment comprised of the neck region plus three CRDs which exhibits significant biological and therapeutic activity *in vivo* and *in vitro* [5-10] have been carried out both in solution and through 3-dimensional structural studies. Binding studies which demonstrate affinity for simple sugars such as glucose, mannose, maltose, heptose and inositol [11-13] have been confirmed crystallographically (see for example [13-18]. In all cases recognition is through CRD Ca-dependent binding of the terminal monosaccharide through a mannose-type equatorial hydroxyl pair O3' and O4' or a stereochemically equivalent pair.

A number of studies of gram-negative bacteria, including *Escherichia coli*, *Salmonella enterica* serovar Minnesota and *Haemophilus influenzae* [19,20] have reported that lipopolysaccharide (LPS) on the bacterial surface mediates hSP-D binding with LPS recognition initially thought to be via terminal carbohydrate residues through carbohydrate recognition domain Ca-dependent binding of a mannose-type equatorial hydroxyl pair on the terminal monosaccharide [14-16]. Our recent structural definition of the recognition of *H. influenzae* Eagan strains inner core LPS by a recombinant fragment of native hSP-D (rfhSP-D) [6,7,21] revealed that recognition occurs by calcium-dependent binding of the non-terminal heptose in the proximal inner core, with the Kdo positioned to interact with the binding site flanking residues Asp325 and Arg343 [15,16]. Furthermore, a recent study by Reinhardt and colleagues has identified that a number of synthetic oligosaccharide cores can be bound by wildtype hSP-D with increased affinities for heptose-rich cores and a clear preference for a terminal Hep-Kdo motif compared to Hep alone [22].

Salmonella enterica is a gram-negative bacterium that causes widespread outbreaks of gastroenteritis and bacteraemia in humans, with the most severe *S. enterica* Typhi infections causing enteric fever (reviewed by Coburn *et al*. [23]). Serovar Minnesota has been identified as

the causative agent in a number of gastroenteritis outbreaks and, more recently, urosepsis in a Crohn's patient [24-26]. The LPS oligosaccharide can be divided into the core oligosaccharides, containing 3-deoxy-D-*manno*-oct-2-ulosonic acid (Kdo), L-*glycero*-D-*manno*-heptose (Hep), α-D-glucose (Glc), α-D-galactose (Gal) and N-*acetyl*-D-glucosamine (GlcNAc), and the antigen-specific O-oligosaccharides that extend in repeats away from the bacterial cell surface [27]. In *Salmonella* species the inner and, to a large extent, the outer core oligosaccharides are a conserved epitope [28].

Rough mutants of *S. enterica* Minnesota lack the O-antigen, leaving the core sugar residues exposed to the immune system leading to the possibility of recognition by SP-D in a similar manner to that reported for *H. influenzae* Eagan 4A [21]. While the LPS core is largely conserved, mutations in the synthetic enzymes responsible for assembling the LPS can result in truncated oligosaccharide cores such as those of strains R5 and R7, equivalent to the Rc and Rd1 phenotypes (Fig 1), demonstrated through lectin blotting and agglutination-inhibition studies to be recognised by hSP-D [19]. The R5 and R7 rough strain LPS exhibit similar structures with the R5 extended by an extra glucose (GlcI) bound to HepII. The R5 strain arises from a single mutation in the phosphorylation step of HepI, inhibiting and impacting on the phosphorylation of HepII and the addition of GalI and GalII to GlcI. Phosphorylation of HepI and HepII initiates the extension of GalI and GalII from GlcI, producing the Rb2 strain [30-33].

While the *H. influenzae* Eagan 4A LPS [21] includes a truncated inner core with a single Kdo and a single Hep with a single Glc extension, the *S. enterica* Minnesota strains R5 and R7provide the opportunity to investigate, through crystal structure analysis, host recognition of more complex natural LPS structures with inner cores containing multiple Kdo and Hep alongside a terminal Glc in the R5 strain. Here we show that hSP-D recognition of *Salmonella enterica* sv Minnesota rough strains R5 and R7 (rough mutant chemotypes Rc and Rd1) LPS occurs preferentially through the inner core Hep-Kdo, with hSP-D having the flexibility and

versatility to target the terminal glucose of R5 when the preferred epitope is not accessible for binding.

Results

The trimeric rfhSP-D structures comprising three carbohydrate recognition domains (CRDs) and the α-helical coiled-coil neck region have been refined to 1.65 Å and 1.75 Å resolution for the R5-bound and R7-bound structures respectively. In all but one of the CRDs the structures reveal oligosaccharide ligand in the Ca1 binding pocket. No ligand is present in subunit A of the R7-bound structure while the mode of oligosaccharide recognition in subunit A of the R5-bound structure differs from the Hep-Kdo binding seen in subunits B and C of both structures.

The bound oligosaccharide is clearly defined in the electron density of both structures; Kdo-HepI-HepII for the R7 and Kdo-HepI-HepII-Glc1 for R5 (Fig 2). The Kdo is in the anhydro form (Fig 3) reported by Shrive and co-workers [21] in the rfhSP-D –*H. influenzae* Eagan 4A structure (pdb entry 4E52) but here the 5-membered ring and all substituents are fully defined in the electron density (Fig 4) except in subunit A of the R5-bound structure where LPS binding is not via HepI and the terminal anhydro Kdo is poorly defined. The Kdo electron density clearly shows the 2-oxobutanoic acid sidechain off C4 (Fig 3b) in a beta conformation with respect to the glycosidic bond in both structures. KdoII and KdoIII (Fig 1) are not present in the electron density.

In subunits B and C, in both structures, recognition of the oligosaccharide by rfhSP-D (Fig 5 and Table 1) is primarily via protein and Ca1 coordination of the 6'OH and 7'OH of the dihydroxyethyl side chain of HepI. Further HepI interactions with the binding pocket in these four subunits include those routinely associated with carbohydrate recognition by hSP-D [21].

Table 1. Calcium and ligand binding distances (Å)

Atom 1	Atom 2	R5			R7		
		A	B	C	A	B	C
Ca1	Glu 321 OE1	2.63	2.61	2.57	2.65	2.60	2.64
	Asn 323 OD1	2.46	2.42	2.47	2.49	2.44	2.38
	Glu 329 OE1	2.42	2.49	2.44	2.51	2.41	2.34
	Asn 341 OD1	2.50	2.44	2.44	2.37	2.43	2.44
	Asp 342 OD1	2.47	2.38	2.40	2.35	2.34	2.28
	Asp 342 O	2.61	2.55	2.62	2.54	2.51	2.51
	HepI O6'	-	2.34	2.34	-	2.36	2.37
	O7'	-	2.38	2.34	-	2.35	2.34
	GlcI O3'	2.38	-	-	-	-	-
	O4'	2.36	-	-	-	-	-
HepI O6'	Glu 321 OE2	-	2.67	2.61	-	2.63	2.52
	Asn 323 ND2	-	3.01	2.92	-	3.03	2.93
O7'	Glu 329 OE2	-	2.66	2.65	-	2.64	2.57
	Asn 341 ND2	-	2.98	3.04	-	2.94	3.07
GlcI O3'	Glu 321 OE2	2.67	-	-	-	-	-
	Asn 323 ND2	2.76	-	-	-	-	-
O4'	Glu 329 OE2	2.69	-	-	-	-	-
	Asn 341 ND2	3.00	-	-	-	-	-
KdoI O1	Arg 343 NH2	-	3.09	2.87	-	-	-
KdoI O2	Arg 343 NH1	-	3.09	3.26	-	3.10	3.06
KdoI O6	Asp 325 OD2	-	2.80	2.56	-	2.76	2.69
GlcI O6'	Arg 343 NH1	3.30	-	-	-	-	-
GlcI O6'	Arg 343 NH2	3.22	-	-	-	-	-
HepI O4'	Arg 343 NH2	3.06	-	-	-	-	-

In contrast to the Ca1-HepI binding in the B and C subunits of both structures, subunit A of the R5-bound structure reveals an alternative mode of LPS recognition with calcium-dependent binding of the terminal glucose GlcI of R5, with the remaining R5 oligosaccharide chain HepII-HepI-Kdo extending over Arg343 and the SP-A (surfactant protein A), SP-D conserved Arg349 and Glu333 (Fig 6). The relative positions of the LPS for the two alternative binding modes are shown in Fig 7a where least-squares fitting of the main chain CRD in the two cases reveals the completely different LPS orientations despite the conserved Ca1-OH interactions. The position of the Ca1-bound glucose GlcI and its interactions with Ca1 and the binding pocket closely mirror those reported in a variety of structures including the rfhSP-D

maltose bound structure [14] as shown in Fig7b. The Ca1 coordinates to the terminal glucose GlcI through 3'OH (2.38 Å) and 4'OH (2.36 Å). Protein coordination to both calcium (via Glu321, Asn323, Glu329, Asn341) and the calcium-bound Glc1 hydroxyls (via Glu321, Asn323, Glu329, Asn341, Asp342) follows the established pattern for ligand-bound SP-D structures (Table 1). Calcium Ca3, the more weakly coordinated of the three calcium ions Ca1, Ca2, Ca3 [15], is absent throughout the R5-bound structure. Ca2 is only weakly defined in subunit B of the R5-bound structure and is absent in subunit C where the calcium (Ca2 and Ca3) coordinating loop between Asp297 and Tyr304 is rearranged and a water molecule replaces Ca2.

In both structures and for both modes of binding (Ca1 coordination of HepI or GlcI) additional direct interactions between the protein and bound oligosaccharide are dominated by the binding-pocket flanking residues Arg343 and Asp325 (Table 1). For both ligands where HepI is coordinated to Ca1 (subunits B and C, both structures) the Kdo (anhydro) interacts with both binding site flanking residues, forming hydrogen bonds with Asp325 through Kdo O6 (2.56 – 2.80 Å) and with Arg343 through the extended 2-oxobutanoic acid sidechain of the anhydro Kdo (Fig 5 and Table 1). In subunit A of the R5-bound structure where GlcI is coordinated to Ca1, additional direct interaction between protein and oligosaccharide again involves Arg343 (Fig 6) which forms weak hydrogen bonds to both GlcI 6'OH and O4' of HepI (Table 1). The direct interactions between protein and oligosaccharide ligand in both structures and all subunits where oligosaccharide ligand is present are supplemented by extensive water molecule networks linking the ligand to the protein.

Discussion

The structures of rfhSP-D, a biologically active recombinant native fragment of hSP-D, bound to inner core structures of *S. enterica* Minnesota LPS from deep rough mutant strains reveal SP-D targeting of the proximal inner core Hep-Kdo motif through calcium-dependent

binding of the (non-terminal) Hep supplemented by interaction of both Hep and Kdo with the binding site flanking residues Arg343 and Asp325. The structures also provide a definitive demonstration for the first time that hSP-D has the versatility to recognise alternative LPS epitopes when the Hep-Kdo inner core unit is inaccessible. This flexibility and versatility of LPS recognition is supplemented by evidence of an extended recognition surface that interacts with the LPS inner core oligosaccharide. The Kdo structure itself provides the first direct confirmation that mild hydrolysis removal of the lipid A from intact LPS exhibiting 4-KdoI substituted KdoII is accompanied by the elimination of KdoII from KdoI C4 [35,36] in a manner entirely equivalent to the β-elimination of phosphate from KdoI C4 through mild acid hydrolysis [21]. This is of some significance in the interpretation of mass spectrometry studies of bacterial LPS structure.

In all subunits where the carbohydrate binding site is openly accessible in the crystal (subunits B and C, both structures), HepI is coordinated to calcium Ca1 as reported for the *H. influenzae* Eagan 4A – rfhSP-D structure [21] with KdoI extending between the two binding-site flanking residues, Asp325 and Arg343, both of which interact directly with the bound HepI-Kdo (anhydro) unit (Fig 5). Superposition of the anhydro Kdo furanose rings together with HepI from all these subunits reveals an essentially conserved position of both rings resulting in very similar Asp325 – anhydro Kdo O6 interactions across the structures (2.56 – 2.80 Å), but significant variations in the Arg343 – Kdo interactions (2.87 – 3.26 Å) (Table 1). This suggests an important flexibility inherent in the Arg343 – Kdo interaction. Combined with the well documented ability of Arg343 to respond to different ligands through conformational change in the Arg343 side chain [15], this reinforces the importance of the Arg343 -- ligand interaction as a flexible determinant of specificity.

The R5 bound structure (Fig 6) reveals for the first time that hSP-D has the flexibility and versatility to recognise and bind alternative LPS epitopes if the HepI-Kdo core motif is not

accessible or, in the physiological context, is shielded by extended oligosaccharides in the outer core and beyond [21]. In subunit A access to the carbohydrate binding site is restricted with the crystal packing precluding binding of either the R5 or R7 oligosaccharide via the HepI-Kdo pair. The crystal packing does however allow Ca1 binding of the terminal glucose of the longer R5 oligosaccharide, which binds to the SP-D subunit A binding pocket with the oligosaccharide ligand (Fig 6) oriented in a direction roughly orthogonal to the ligand when HepI is bound to Ca1 (Fig 7) as in subunits B and C of both structures (Fig 5). Interestingly, despite the completely different position of HepI with respect to the protein structure, Arg343 remains fundamental to the recognition of the LPS, stabilising the interaction through an Arg343-HepI hydrogen bond (Fig 6 and Table 1) and perhaps contributing to directing the glucose into the binding site [15,16]. While this novel structural demonstration that hSP-D can adopt alternative binding mechanisms for the same LPS is a fortuitous artefact of crystallisation, it clearly shows not only alternative recognition strategies for the same LPS but also the ability to recognise more than just the inner core heptose and Kdo residues in natural LPS ligands. This potentially provides an explanation for the extraordinary ability of hSP-D to recognise such a broad range of gram-negative bacteria by utilising Ca1-dependent binding of not just inner core HepI but also of multiple types of saccharide residue in the LPS [13,21]. The extent of the LPS, and the variety of different carbohydrate residues displayed on bacterial surfaces, provides the opportunity for versatility of recognition by the carbohydrate recognition domain.

The full *S. enterica* sv. Minnesota R5 and R7 LPS oligosaccharides (Fig 1) are known to be recognised by hSP-D [19]. Positioning the full R5 oligosaccharide, including KdoII, KdoIII and a full 6-membered KdoI ring, in the R5-bound structure (subunit C) by overlaying HepI reveals that KdoI may be oriented such that the dihydroxyethyl side chain (7'OH, 8'OH) off C6 interacts either with Asp325 or Arg343. In the orientation shown in Fig 8, KdoII sits between

KdoI and the conformationally flexible side chain of Arg343 with both KdoII and KdoIII proximal to the SP-D conserved Glu347 and the SP-A, SP-D conserved Arg349 and Glu333. KdoI O7' interacts with Asp325, KdoII O1' with Arg343, and KdoIII O8' with Glu347. The lipid A linked to KdoI O2' extends out from the CRD and the trimer overall. *Salmonella enterica* mutants with more extended cores and which are recognised by SP-D [19] can also be accommodated in this orientation. For the Rb3 mutant for example (Fig 1), the sugars linked to HepII and Glc1 all extend away from the CRD. This strongly suggests an extended inner core binding site with recognition of the Hep-Kdo motif accompanied by additional binding of inner cores with multiple Kdo structures such as those of *Pseudomonas aeruginosa*, *Escherichia coli and Klebsiella pneumoniae* which are known to be recognised by rfhSP-D [37].

Previous studies of the hSP-D/LPS interaction have shown the importance of Asp325 and Arg343 as oligosaccharide ligand selection mediators, flanking the binding site and directing calcium recognition [13-17], alongside the targeting of the LPS proximal inner core. The structures presented here combine and extend these key elements, revealing the preferred inner core Hep-Kdo binding mechanism alongside a flexibility to recognise alternative epitopes, to provide unique insights into the recognition and binding of a wide variety of gram negative bacteria by hSP-D.

Materials and methods

S. enterica **Minnesota R5 and R7 Oligosaccharide Preparation**. The oligosaccharide (delipidated) portions of *Salmonella enterica* sv. Minnesota R5 (Enzo Lifesciences-ALX-581-017-L002) and R7 (Enzo Lifesciences-ALX-581-018-L002) mutants were prepared under mild hydrolysis conditions with 2% acetic acid, in accordance with the method previously published by Masoud *et al.* [38] and our previously published procedures [21]. For the R5 LPS the molecular weight was calculated as ≈ 3036 Da by the addition of all glycoconjugates and Lipid

A. The lipid A portion accounted for ≈ 1800 Da[39], making up 60% of the LPS. 2 mg of R5 LPS were initially hydrolysed, yielding ≈0.7 mg of the oligosaccharide representing 88% hydrolysis efficiency. The R5 oligosaccharide was dissolved in deionised water to give a KdoI-HepI-HepII-GlcI concentration of ≈16 mM (accounting for the beta elimination of KdoII-KdoII (see Fig 3)) for storage and for use in crystal soaking. The same procedure was followed for the R7 LPS providing a stock solution of ≈18mM for storage and for use in crystal soaking and cryoprotection preparation.

Generation and purification of a recombinant fragment of hSP-D. A recombinant homotrimeric fragment of human SP-D (rfhSP-D) was expressed and purified as described previously [5,6,14]. Each chain contains 177 amino acids (Gly179 to Phe355) comprising a short collagen region with eight Gly-X-Y repeats, the α-helical coiled coil neck region and the globular CRD. The contaminating level of endotoxin present in the rfhSP-D preparation was minimized [6] using a Polymyxin B column (Detoxi-Gel, Pierce, UK) to remove endotoxin followed by assaying with the QCL-1000 Limulus amoebocyte lysate system (BioWhittaker, Walkersville, MD) according to manufacturer's instructions. The assay was linear over a range of 0.1–1.0 EU/ml (10 EU=1 ng of endotoxin).

Crystallisation, ligand soaking and cryoprotection. The endotoxin treated rfhSP-D at 8.8 mg/ml in PBS was used for crystallisation both directly (for the R7 studies), following addition of $CaCl_2$ to give 10mM, and after dialysis (for the R5 studies) into 50 mM Tris, 10 mM $CaCl_2$, 150 mM NaCl, 0.02% NaN_3 at pH7.4. Protein concentrations used were 7.7 – 8.9 mg/ml. Native crystals of rfhSP-D were grown in sitting drops using screens around conditions previously reported [14]. Those for soaking with the delipidated R5 LPS were grown in precipitant buffer 0.1 M tris pH 8.0, 16% PEG 6000, and for the delipidated R7 LPS in 0.1 M Tris pH 7.0, 16% PEG

10000. Cryoprotectant buffers were prepared as previously described [21] to provide increasing concentrations (5-20% for R5; 5-15% for R7) of MPD (2,4-methylpentane diol) in precipitant buffer with ligand added to provide final ligand concentrations of ≈8 mM and ≈13 mM in the crystallisation drop for R5 and R7 respectively. The crystals were soaked with ligand and prepared for cryocooling by successive addition of cryoprotectant containing increasing concentrations of MPD (5-15% for R7; 5-20% for R5) in 2 µl aliquots over three minutes followed by a final exchange of 6-8 µl of the drop with 6-8 µl of the final MPD cryoprotectant solution (15% MPD for R7; 20% MPD for R5). Crystals were flash-frozen at 100 K after a total of 24 minutes (R5) or 8 minutes (R7) soaking with ligand and cryoprotectant buffer.

Data collection, structure solution and refinement. Data for the rfhSP-D R5 and R7 complexes were collected at 100 K at Diamond Light Source, Oxfordshire, on beamlines I04-1 and I03, respectively, using Dectris Pilatus 6M hybrid photon counting detectors. Integrated intensities were calculated using the program *MOSFLM* [40]. The 1.6 Å native structure of rfhSP-D (PDB ID 1PW9 [14]) was used as a starting model for modelling both the rfhSP-D/R5 and rfhSP-D/R7 structures. Electron density maps were calculated using the CCP4 suite of programs including *AIMLESS, TRUNCATE, SORTMTZ* and, where necessary, *REINDEX* [41]. Model building was completed using *COOT* [42] and refined using maximum likelihood refinement in *REFMAC5* [43], available as part of the CCP4 package, in alternating rounds. Ligand topology and parameter files were generated using *ProDRG* [44] following fitting into the density. Final refinement statistics are presented in Table 2 and the quality of the final model was confirmed using the MolProbity web server [45]. The coordinates and structure factors for the R5-bound (PDB ID 5OXS) and R7-bound (PDB ID 5OXR) structures have been deposited with the Protein Data Bank. All main and side-chain stereochemical parameters are within accepted limits or better, with more than 98% of residues in favoured regions of the

Ramachandran plot with no outliers. Molecular figures were generated using CCP4mg molecular graphics software [46] and the PyMOL Molecular Graphics System Version 1.4 (Schrödinger, LLC, 2011).

Table 2. Data collection and processing

Data collection	*Sm R7 oligo*	*Sm R5 oligo*
Synchrotron station	DLS I03	DLS I04-1
Wavelength (Å)	0.97625	0.92819
Space group	$P2_1$	$P2_1$
Cell dimensions	a=55.33 Å, b=108.14 Å, c=55.67 Å, β=91.82□	a=55.68 Å, b=108.51 Å, c=56.17 Å, β=92.89□
Resolution range (Å)	55.3 – 1.75 (1.78-1.75)	56.10 – 1.65 (1.68-1.65)
Observations	143,142 (7,653)	230,306 (10,955)
Unique reflections	61,250 (3,299)	74,199 (3,629)
Completeness (%)	93.2 (91.2)	93.0 (93.2)
Rmerge [a]	0.066 (0.359)	0.054 (0.397)
Mean (I/σ(I))	7.3 (2.2)	11.4 (2.5)
Refinement		
Protein atoms	3483	3537
Residues chain A	205-355	204-355
Residues chain B	204-355	203-355
Residues chain C	205-355	203-355
Water molecules	479	477
Other molecules		
Subunit	A B C	A B C
Calcium ions	3 3 3	2 2 1
Oligosaccharide	- 1 1	1 1 1
R_{work} [b] (%)	17.3	16.5
R_{free} [c] (%)	19.6	18.7
r.m.s.d. bond length (Å)	0.010	0.010
r.m.s.d. bond angle (°)	1.40	1.32
Average B-values (Å^2)		
Protein	23.1	22.6
Water	34.2	34.9
Other hetero-atoms	31.1	40.2
PDB ID	5OXR	5OXS
Ramachandran plot values [d] (%)		
Favoured	97.8	97.8
Outliers	0.0	0.0

Figures in parentheses refer to the highest resolution bin

[a] $R_{merge} = \Sigma_h\Sigma_j |I_{h,j} - I_h| / \Sigma_h\Sigma_j |I_{h,j}|$, where $I_{h,j}$ is the j^{th} observation of reflection h and Ih is the mean of the j measurements of reflection h.

[b] $R_{work} = \Sigma_h ||F_{oh}| - |F_{ch}|| / \Sigma_h |F_{oh}|$ where F_{oh} and F_{ch} are the observed and calculated structure factor amplitudes, respectively, for the reflection h.

[c] R_{free} is equivalent to R_{work} for a randomly selected subset (5%) of reflections not used in the refinement.

[d] Defined according to Molprobity

Acknowledgements:

We thank Diamond Light Source for access to beamlines I03 and I04-1 (proposal numbers MX8359, MX10369, MX14692) that contributed to the results presented here. Help from the beamline scientists at the Diamond Light Source is gratefully acknowledged.

References

1. Ujma, S., Horsnell, W.G.C., Katz, A., Clark, H.W., and Schäfer, G. (2017) Non-Pulmonary Immune Functions of Surfactant Proteins A and D. *J. Innate Immun.* **9**, 3–11
2. Reid, K.B.M. (1998) Interactions of surfactant protein D with pathogens, allergens and phagocytes. *Biochim. Biophys. Acta* **1408**, 290-295
3. LeVine, A.M., and Whitsett, J.A. (2001) Pulmonary collectins and innate host defense of the lung. *Microbes Infect.* **3**, 161-166
4. Crouch, E.C., and Wright, J.R. (2001) Surfactant proteins A and D and pulmonary host defence. *Annu. Rev. Physiol.* **63**, 521-554
5. Madan, T., Kishore, U., Singh, M., Strong, P., Clark, H.W., Hussain, E.M., Reid, K.B.M., and Sarma, P.U. (2001) Surfactant proteins A and D protect mice against pulmonary hypersensitivity induced by *Aspergillus fumigatus* antigens and allergens. *J. Clin. Invest.* **107**, 467-475
6. Singh, M., Madan, T., Waters, P., Parida, S.K., Sarma, P.U., and Kishore, U. (2003) 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
7. Clark, H.W., and Reid, K.B.M. (2002) Structural requirements for SP-D function in vitro and in vivo: Therapeutic potential of recombinant SP-D. *Immunobiology* **205,** 619-631

8. Madan, T., Kishore, U., Singh, M., Strong, P., Hussain, E.M., Reid K.B.M. and Sarma, P.U. (2001b) Protective role of surfactant protein D in a murine model of Invasive Pulmonary Aspergillosis. *Infect. Immun.* **69**, 2728-2731
9. Hickling, T.P., Bright, H., Wing, K., Gower, D., Martin, S.L., Sim, R.B. and Malhotra, R. (1999) A recombinant trimeric surfactant D carbohydrate recognition domain inhibits respiratory syncytial virus infection *in vitro* and *in vivo*. *Eur. J. Immunol.* **29**, 3478- 3484
10. Strong, P., Townsend, P., Mackay, R., Reid, K.B.M. and Clark, H.W. (2003) A recombinant fragment of human SP-D reduces allergic responses in mice sensitised to house dust mite allergens. *Clin. Exp. Immunol.* **134**, 181-187
11. Kishore, U., Wang, J-Y., Hoppe, H.J. and Reid, K.B.M. (1996) The α-helical neck region of human lung surfactant protein D is essential for the binding of the carbohydrate recognition domains to lipopolysaccharides and phospholipids. *Biochem. J.* **318**, 505-511
12. Crouch, E., Tu, Y., Briner, D., McDonald, B., Smith, K., Holmskov, U. and Hartshorn, K. (2005) Ligand specificity of human surfactant protein D: expression of a mutant trimeric collectin that shows enhanced interactions with influenza A virus. *J. Biol. Chem.* **280(17)**, 17046-17056
13. Wang, H., Head, J., Kosma, P., Brade, H., Müller-Loennies, S., Sheikh, S., McDonald, B., Smith, K., Cafarella, T., Seaton, B., and Crouch, E. (2008) Recognition of Heptoses and the Inner Core of Bacterial Lipopolysaccharides by Surfactant Protein D. *Biochemistry* **47**, 710–720
14. 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
15. 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., and Greenhough, T.J. (2009) Structural Characterisation of

Ligand-Binding Determinants in Human Lung Surfactant Protein D: Influence of Asp325. *J. Mol. Biol*. **394**, 776–788

16. Crouch, E., Hartshorn, K., Horlacher, T., McDonald, B., Smith, K., Cafarella, T., Seaton, B., Seeberger, P. H., and Head, J. (2009) Recognition of mannosylated ligands and influenza a virus by human surfactant protein D: Contributions of an extended site and residue 343. *Biochemistry* **48**, 3335–3345
17. Crouch, E., McDonald, B., Smith, K., Roberts, M., Mealy, T., Seaton, B., and Head, J. (2007) Critical role of Arg/Lys343 in the species-dependent recognition of phosphatidylinositol by pulmonary surfactant protein D. *Biochemistry* **46**, 5160–5169
18. Crouch, E., McDonald, B., Smith, K., Cafarella, T., Seaton, B. & Head, J. (2006a) Contributions of phenylalanaine 335 to ligand recognition by human surfactant protein D. *J. Biol. Chem*. **281**, 18008–18014
19. Kuan, S.F., Rust, K., and Crouch, E. (1992) 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
20. 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 or D Deficiency During Bacterial Infection on the Lung. *J. Immunol.* **165**, 3934-3940
21. Clark, H. W., Mackay, R.-M., Deadman, M. E., Hood, D. W., Madsen, J., Moxon, E. R., Townsend, J.P., Reid, K.B.M., Ahmed, A.D., Shaw, A.J., Greenhough, T.J., and Shrive, A.K. (2016) Crystal structure of a complex of Surfactant Protein D and *Haemophilus influenzae* lipopolysaccharide reveals shielding of core structures in SP-D resistant strains. *Infect. Immun.* **84**, 1585-1592

22. Reinhardt, A., Wehle, M., Geissner, A., Crouch, E.C., Kang, Y., Yang, Y., Anish, C., Santer, M., and Seeberger, P.H. (2016) Structure binding relationship of human surfactant protein D and various lipopolysaccharide inner core structures. *J. Struct. Biol.* **195**, 387–395
23. Coburn, B., Grassl, G., and Finlay, B.B. (2007) Salmonella, the host and disease: A brief review. *Immunol. Cell Biol.* **85**, 112-118
24. McCracken, D. A. (1954) *Salmonella* minnesota Infection in Northamptonshire. *J R Sanit Inst* **74**, 1091–1094
25. Borbolla, L., Guerra Chabau, A., Prieto, E., and Valledor, T. (1956) [Meningitis caused by *Salmonella* minnesota: case report] Translated from Spanish. *Rev. Cubana. Pediatr.* **10**, 611–618
26. Steinebrunner, N., Sandig, C., Zimmermann, S., Stremmel, W., Eisenbach, C., and Mischnik, A. (2013) *Salmonella enterica* serovar Minnesota urosepsis in a patient with Crohn's disease in the absence of recent or current gastrointestinal symptoms. *J. Med. Microbiol.* **62**, 1360–1362
27. Holst O. (2011) Structure of the lipopolysaccharide core region, in *Bacterial Lipopolysaccharides: Structure, Chemical Synthesis, Biogenesis and Interaction with Host Cells* (Knirel, Y. A., and Valvani, M. A., eds) pp. 21–39, Springer, Vienna
28. Holst, O., and Brade, H. (1991) Structural studies of the core region of the lipopolysaccharide from *Salmonella* minnesota strain R7 (rough mutant chemotype Rd1). *Carbohydr. Res.* **219**, 247–251
29. Mansfield, L.P., and Forsythe, S.J. (2001) Demonstration of the Rb-1 lipopolysaccharide core structure in *Salmonella* strains with the monoclonal antibody M105. *Journal of Medical Microbiology* **50(4)**, 339-344

30. Muhlradt, P. (1971) Biosynthesis of Salmonella lipopolysaccharide. Studies on the transfer of glucose, galactose, and phosphate to the core in a cell free system. *Eur. J. Biochem.* **18(1)**, 20-27

31. Schnaitman, C.A., and Klena, J.D. (1993) Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiological Reviews* **57**, 655–682

32. Jousimies, H., and Mäkelä, P.H. (1974) Genetic analysis of *Salmonella* minnesota R mutants with defects in the biosynthesis of the lipopolysaccharide core*. J. Bacteriol*. **119(3)**, 753-759

33. Raetz, C.H.R., and Whitfield, C. (2002) Lipopolysaccharide Endotoxins. *Annu. Rev. Biochem.* **71**, 635–700

34. Auzanneau, F-I., Charon, D., and Szabó, 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

35. Yethon, J.A., Vinogradov, E., Perry, M.B., and Whitfield, C. (2000) Mutation of the Lipopolysaccharide Core Glycosyltransferase Encoded by *waaG* Destabilizes the Outer Membrane of *Escherichia coli* by Interfering with Core Phosphorylation. *Journal of Bacteriology* **182(19)**, 5620-5623

36. Olsthoorn, M.A., Haverkamp, J., and Thomas-Oates, J.E. (1999) Mass Spectrometric Analysis of *Klebsiella pneumoniae* ssp. *pneumoniae* Rough Strain R20 ($O1^{-}$:$K20^{-}$) Lipopolysaccharide Preparations: Identification of Novel Core Oligosaccharide Components and Three 3-Deoxy-D-*manno*-oct-2-ulopyranosonic Artifacts. *J. Mass Spectrom*. **34**, 622–636

37. Lim, B.L., Wang, J.Y., Holmskov, U., Hoppe, H.J., and Reid, K.B.M. (1994) Expression of the carbohydrate recognition domain of lung surfactant protein D and demonstration of its binding to lipopolysaccharides of gram-negative bacteria. *Biochem Biophys Res Commun*. **202(3)**, 1674-1680

38. Masoud, H., Altman, E., Richards, J. C., and Lam, J. S. (1994) General strategy for structural analysis of the oligosaccharide region of lipooligosaccharides. Structure of the oligosaccharide component of Pseudomonas aeruginosa IATS serotype 06 mutant R5 rough-type lipopolysaccharide. *Biochemistry* **33**, 10568–10578
39. Que, N.L.S., Lin, S., Cotter, R.J., and Raetz, C.R.H. (2000) Purification and Mass Spectrometry of Six Lipid A Species from the Bacterial Endosymbiont Rhizobium etli. *J. Biol. Chem.* **275(36)**, 28006–28016
40. 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 26. Daresbury Laboratory, Warrington, UK
41. Winn, M.D. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr* **D67**, 235-242
42. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010) Features and Development of Coot. *Acta Crystallogr* **D66**, 486-501
43. Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1977) Refinement of Macromolecular Structures by the Maximum-Likelihood method. *Acta Crystallogr* **D53**, 240-255
44. Schuettelkopf, A.W., and van Aalten, D.M.F. (2004) PRODRG – a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr* **D60**, 1355–1363
45. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity□: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr* **D66**, 12–21
46. McNicholas, S., Potterton, E., Wilson, K. S., and Noble, M. E. M. (2011) Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr* **D67**, 386–394

FIGURE LEGENDS

Fig 1. Core LPS structure of *Salmonella enterica* rough mutant strains. The LPS phenotypes (Ra, Rb1, Rb2, Rb3, Rc, Rd1, Rd2, and Re) and the mutant strains R5 and R7 are shown (modified from Mansfield and Forsythe, 2001 [29]). Glc: glucose; GlcNAc: N-acetyl-glucosamine; Gal: galactose; Hep: L-D-Heptose; P: phosphate; PEtn: phosphoethanolamine; Kdo: 3-deoxy-D-manno-oct-2-ulosonic acid.

Fig 2. Structure of the product of mild acid hydrolysis of *S. enterica* Minnesota rough mutants R5 and R7. Glc: glucose; Hep: L-D-Heptose; Kdo: 3-deoxy-D-manno-oct-2-ulosonic acid.

Fig 3. Structure of the 4,7 closure furanoid derivative (anhydro Kdo) of Kdo following LPS delipidation by mild acid hydrolysis and β-elimination of the Kdo O4 substituent, reported by Shrive and co-workers (PDB ID 4E52, ligand KD5). Anhydro Kdo is proposed to be a racemic mixture (Auzanneau *et al*., 1991 [34]) with the 2-oxobutanoic acid side chain at C4 both alpha and beta to the glycosidic bond (at C5). (a) Original Kdo numbering retained. (b) Numbering according to pdb entry 4E52 (alpha) and used for the structure here.

Fig 4. Electron density for the *Salmonella enterica* oligosaccharides bound to rfhSP-D (subunit B) with HepI in the Ca1 binding site. The calcium ion is in green and the four calcium and ligand coordinating residues Glu321, Asn323, Glu329 and Asn341 are shown in cpk. (a) R7 oligosaccharide KdoI (anhydro)-HepI-HepII (b) R5 oligosaccharide KdoI (anhydro)-HepI-HepII-Glc1. Maps are 2mFo-DFc contoured at 1σ.

Fig 5. *S. enterica* R5 oligosaccharide bound to rfhSP-D (subunit C) with the inner core HepI in the Ca1 binding site. KdoI (anhydro) interacts with both Asp325 and Arg343. Similar binding is

observed in subunit B and in subunits B and C of the R7-bound structure which lacks the terminal glucose GlcI. Calcium coordinating bonds are represented by dots and protein-ligand interactions by dashes. The interactions of Glu333 are also indicated.

Fig 6. *S. enterica* R5 oligosaccharide bound to rfhSP-D (subunit A) with the terminal glucose GlcI in the Ca1 binding site. HepI interacts with Arg343. No ligand was observed in subunit A of the R7-bound structure due to crystal lattice constraints. Calcium coordinating bonds are represented by dots and protein-ligand interactions by dashes. The interactions of Glu333 are also indicated.

Fig 7. (a) Superposition of subunit A (GlcI bound) in cyan and subunit B (HepI bound) in yellow of the R5 LPS-bound structure following a least-squares fit of the CRD protein main chain atoms. (b) Superposition of subunit A (GlcI bound) in cyan of the R5 LPS-bound structure and subunit A (Glc1 bound) in coral of the maltose-bound structure [14] following a least-squares fit of the CRD protein main chain atoms.

Fig 8. Model of complete *S. enterica* R5 oligosaccharide bound to rfhSP-D with the inner core HepI in the Ca1 binding site. KdoI (intact) is positioned to interact with Asp325, with KdoII and KdoIII interacting with Arg343 and Glu347 respectively. (a) subunit C, R5-bound structure (b) the rfhSP-D trimer with R5 similarly placed in each subunit and the link to lipid A from KdoI O2' indicated by ★.

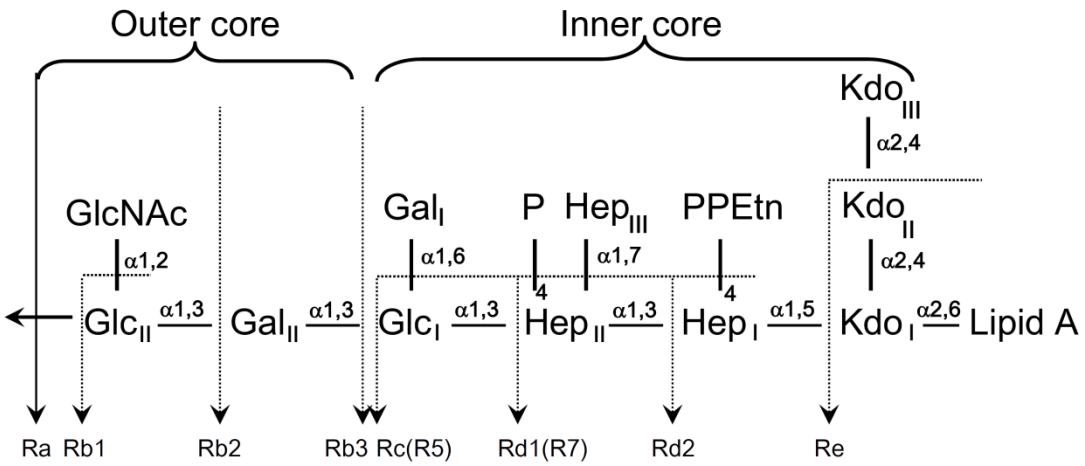


Fig 1. Core LPS structure of *Salmonella enterica* rough mutant strains. The LPS phenotypes (Ra, Rb1, Rb2, Rb3, Rc, Rd1, Rd2, and Re) and the mutant strains R5 and R7 are shown (modified from Mansfield and Forsythe, 2001 [29]). Glc: glucose; GlcNAc: N-acetyl-glucosamine; Gal: galactose; Hep: L-D-Heptose; P: phosphate; PEtn: phosphoethanolamine; Kdo: 3-deoxy-D-manno-oct-2-ulosonic acid.

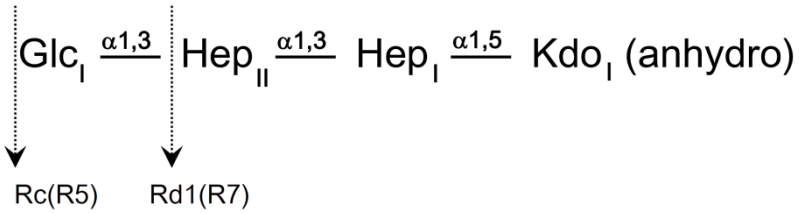


Fig 2. Structure of the product of mild acid hydrolysis of *S. enterica* Minnesota rough mutants R5 and R7. Glc: glucose; Hep: L-D-Heptose; Kdo: 3-deoxy-D-manno-oct-2-ulosonic acid.

Figure 3

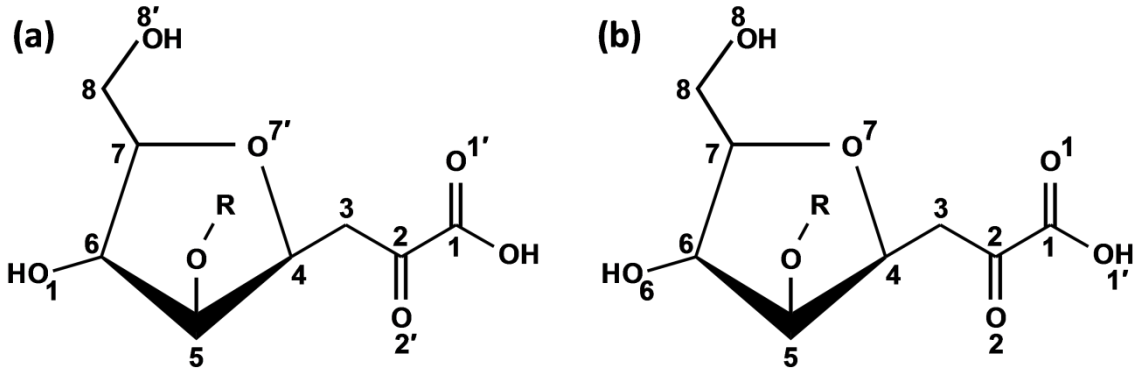
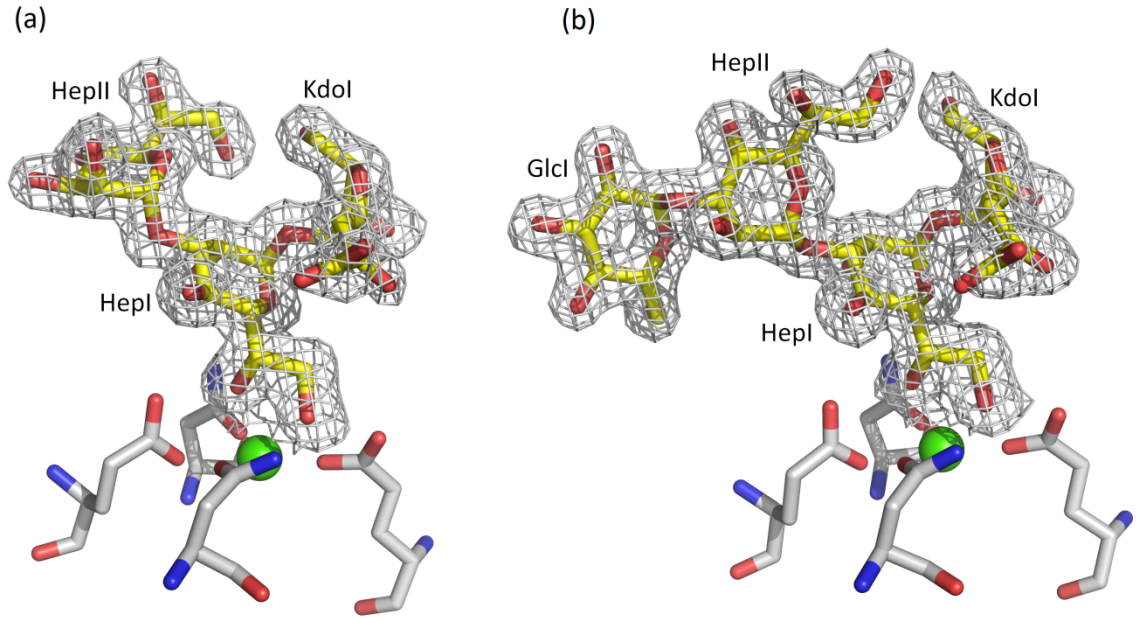


Fig 3. Structure of the 4,7 closure furanoid derivative (anhydro Kdo) of Kdo following LPS delipidation by mild acid hydrolysis and β-elimination of the Kdo O4 substituent, reported by Shrive and co-workers (PDB ID 4E52, ligand KD5). Anhydro Kdo is proposed to be a racemic mixture (Auzanneau *et al.*, 1991 [34]) with the 2-oxobutanoic acid side chain at C4 both alpha and beta to the glycosidic bond (at C5). (a) Original Kdo numbering retained. (b) Numbering



according to pdb entry 4E52 (alpha) and used for the structure here.

Fig 4. Electron density for the *Salmonella enterica* oligosaccharides (in yellow) bound to rfhSP-D (subunit B) with HepI in the Ca1 binding site. The calcium ion is in green and the four calcium and ligand coordinating residues Glu321, Asn323, Glu329 and Asn341 are shown in cpk. (a) R7 oligosaccharide KdoI (anhydro)-HepI-HepII (b) R5 oligosaccharide KdoI (anhydro)-HepI-HepII-Glc1. Maps are 2mFo-DFc contoured at 1σ.

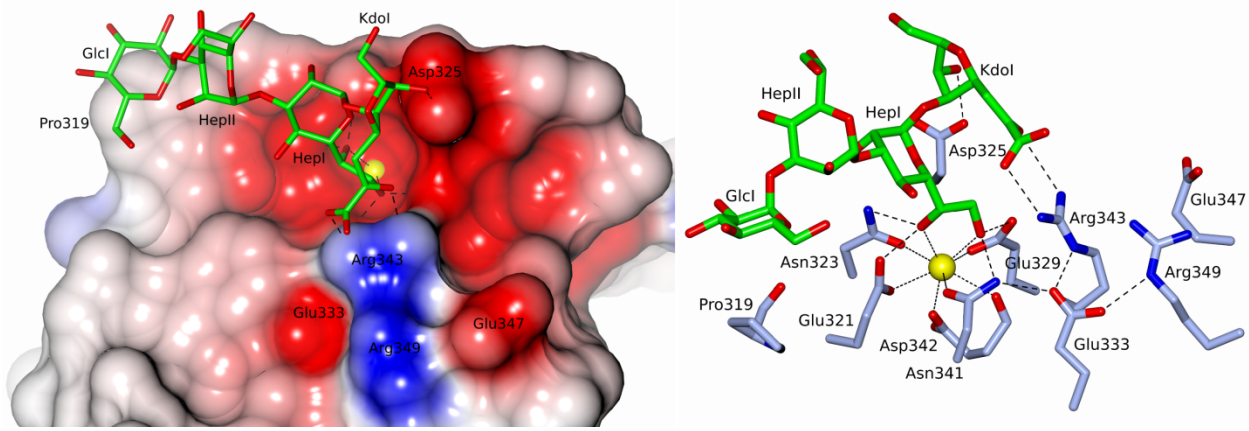


Fig 5. *S. enterica* R5 oligosaccharide bound to rfhSP-D (subunit C) with the inner core HepI in the Ca1 binding site. KdoI (anhydro) interacts with both Asp325 and Arg343. Similar binding is observed in subunit B and in subunits B and C of the R7-bound structure which lacks the terminal glucose GlcI. Calcium coordinating bonds are represented by dots and protein-ligand interactions by dashes. The interactions of Glu333 are also indicated.

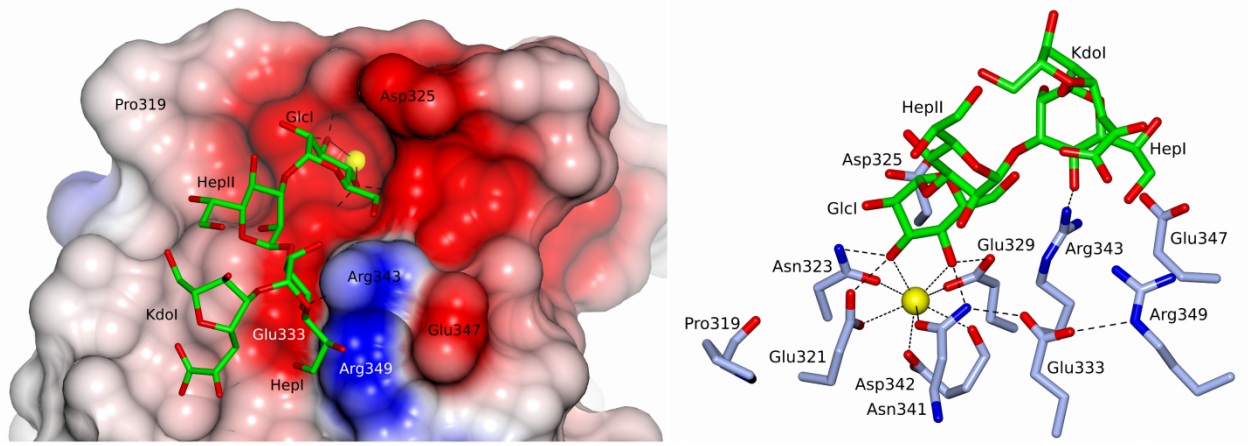


Fig 6. *S. enterica* R5 oligosaccharide bound to rfhSP-D (subunit A) with the terminal glucose GlcI in the Ca1 binding site. HepI interacts with Arg343. No ligand was observed in subunit A of the R7-bound structure due to crystal lattice constraints. Calcium coordinating bonds are represented by dots and protein-ligand interactions by dashes. The interactions of Glu333 are also indicated.

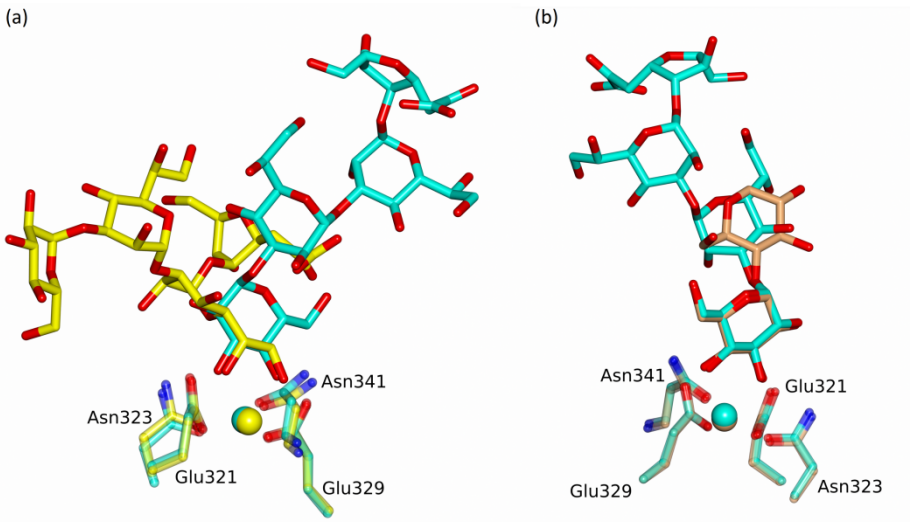


Fig 7. (a) Superposition of subunit A (GlcI bound) in cyan and subunit B (HepI bound) in yellow of the R5 LPS-bound structure following a least-squares fit of the CRD protein main chain atoms. (b) Superposition of subunit A (GlcI bound) in cyan of the R5 LPS-bound structure and subunit A (Glc1 bound) in coral of the maltose-bound structure [14] following a least-squares fit of the CRD protein main chain atoms.

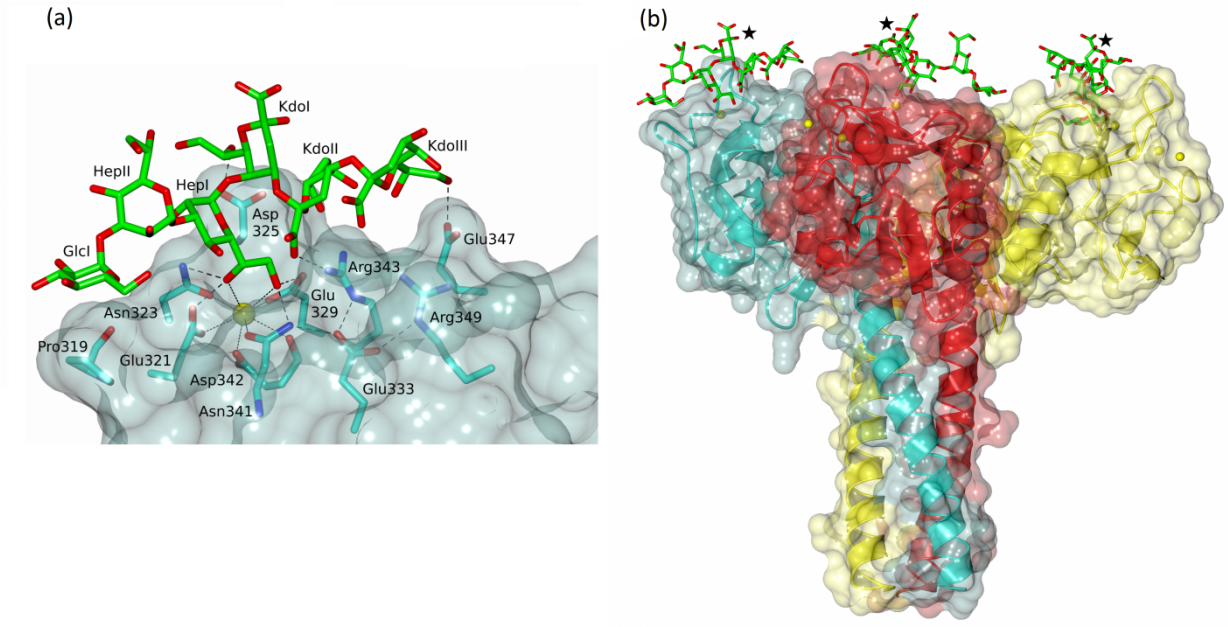


Fig 8. Model of complete *S. enterica* R5 oligosaccharide bound to rfhSP-D with the inner core HepI in the Ca1 binding site. KdoI (intact) is positioned to interact with Asp325, with KdoII and KdoIII interacting with Arg343 and Glu347 respectively. (a) subunit C, R5-bound structure (b) the rfhSP-D trimer with R5 similarly placed in each subunit and the link to lipid A from KdoI O2' indicated by ★.