Host glycan degradation by *Streptococcus pneumoniae*

by

Mélissa Cid

MSc, Université Louis Pasteur, 2008

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Biochemistry and Microbiology

Supervisory Committee

Host glycan degradation by *Streptococcus pneumoniae*

by

Mélissa Cid
MSc, Université Louis Pasteur, 2008

Supervisory Committee

Dr. Alisdair B. Boraston (Department of Biochemistry and Microbiology)
Supervisor

Dr. Francis E. Nano (Department of Biochemistry and Microbiology)
Departmental Member

Dr. Christopher J. Nelson (Department of Biochemistry and Microbiology)
Departmental Member

Dr. Réal Roy (Department of Biology)
Outside Member

Abstract

Supervisory Committee
Dr. Alisdair B. Boraston (Department of Biochemistry and Microbiology)
Supervisor
Dr. Francis E. Nano (Department of Biochemistry and Microbiology)
Departmental Member
Dr. Christopher J. Nelson (Department of Biochemistry and Microbiology)
Departmental Member
Dr. Réal Roy (Department of Biology)
Outside Member

Streptococcus pneumoniae is a commensal inhabitant of the human nasopharynx that can sometimes become pathogenic and cause diseases such as pneumonia, otitis media and meningitis. Carbohydrate metabolism is a critical component of *S. pneumoniae* virulence. Among the myriad of carbohydrate-specific pathways involved in the host-pneumococcus interaction, the N-glycan foraging pathway stands out because of its direct implication in numerous aspects of virulence such as fitness, adhesion/invasion and impairment of the host immune response. Much of the literature has been focussed on the importance of step-wise depolymerisation of N-glycans by the enzymes NanA, BgaA and StrH. However, the importance of the liberation of N-glycans from host glycoconjuguates and their intake by the bacterium has yet to be examined. We have identified a Carbohydrate Processing Locus (CPL) that is highly conserved throughout a large number of Firmicutes and whose individual components appear widespread in bacteria that we hypothesize is active on host N-glycans. This locus encodes for two putative α-mannosidases GH92 and GH38, a characterised α-mannosidase GH125, a putative β-hexosaminidase GH20C, a putative α-fucosidase GH29 and a ROK (Repressor, Open reading frame, Kinase) protein. The genomic context of CPL orthologues suggests that an endo-β-N-acetylglucosaminidase (EndoD) and an ABC transporter ($ABC_{N\text{-}glycan}$) are functionally associated with this locus. Based on our bioinformatic analyses and known functions of these proteins we hypothesize that the CPL encodes a concerted pathway responsible for the liberation, transport, and processing of N-glycans. The objective of this research is to characterize the putative components of this pathway and assess their

implication in virulence. Specific focus on $ABC_{N\text{-}glycan}$ demonstrated its specificity for a range of N-glycans liberated by EndoD, shedding light on a novel import system for branched N-glycans. Furthermore, we provided evidence that GH92 is an α-1,2-mannosidase that likely removes the terminal mannose residues found on high-mannose N-glycans. EndoD and GH92 are shown to participate in virulence in mice; however, their role in virulence has yet to be determined. This work will significantly advance the construction and validation of a model of N-glycan processing by *S. pneumoniae*. As the components of this model pathway are conserved amongst a wide variety of bacteria, this work is of fundamental relevance to understanding how microbes from various environments degrade and metabolize N-glycans.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to thank my outstanding supervisor Dr. Alisdair Boraston for his support and guidance throughout my degree. Thanks to my committee members Dr. Fran Nano, Dr. Chris Nelson and Dr. Réal Roy for their insight and discussion over the years. Furthermore, I would like to thank my dear friends from the Boraston lab and the Biochemistry and Microbiology department, past and present. Particularly Pola Wojnarowicz, Amanda Carew and Andrew Hettle. They have all contributed to the warm and social environment of this lab and made this grad school experience incredible.

Last but not least, thank you to my life partner, friend and co-worker Craig Robb whom I never would have done this without.

Chapter 1: Introduction

1.1 *Streptococcus pneumoniae*

Streptococcus pneumoniae (*S. pneumoniae* or pneumococcus) is a leading cause of morbidity and mortality worldwide. This bacterium is capable of asymptomatic carriage in the human nasopharynx; however, in some circumstances *S. pneumoniae* can cause severe disease ranging in severity from otitis media in children to community-acquired pneumonia, bacteremia, and even meningitis (McCullers and Tuomanen, 2001). Colonization by *S. pneumoniae* is a key early step in the development of infection, and failure of the innate immune response can allow spread of this agent to other sites and permit disease progression. Pneumococcal infections are classified into two categories: invasive disease, which is defined by the infection of a normally sterile site and includes bacteremia and meningitis, and non-invasive disease such as acute otitis media (AOM).

1.1.1 *S. pneumoniae* epidemiology

The rate of invasive pneumococcal disease (IPD) in 2009 was reported at 10-18 cases per 100 000 people in the U.S. with a high incidence in children less than 5 years old and in the elderly over 65 years old (Rosen et al., 2011). Over 1 million children under the age of 5 die every year from severe pneumococcal disease in developing and developed countries (Bogaert et al., 2004). Most IPD cases result from pneumococcal pneumonia. In the case of pneumococcal meningitis, the fatality rate remains high at 20% and 50% in developed and developing countries, respectively, and up to 60% of survivors remain permanently disabled (Shin and Kim, 2012). Although mortality and morbidity associated with *S. pneumoniae* is highest in developing countries, this bacterium is still a burden in developed countries with a death rate of about 10%, affecting mostly elderly people over 65 years old (Robinson et al., 2001). The majority of these deaths are the result of community-acquired pneumonia.

Based on its capsular polysaccharide composition, over 90 different serotypes of *S. pneumoniae* have been found. A small proportion of these serotypes are responsible for most cases of IPD, indeed approximately 10 serotypes account for about 62% of

pneumococcal disease worldwide (Bridy-Pappas et al., 2005). The prevalence of serotypes associated with severe disease differs greatly with geography and time.

1.1.2 *S. pneumoniae* treatment and prevention

In the pre-antibiotic era, the mortality rate associated with pneumococcal pneumonia was 20%, bacteremia 50% and meningitis 80-100% (Heffron, 1940). After the introduction of penicillin these rates decreased to 5, 20 and 30%, respectively (Breiman, Robert F., John S. Spika and Navarro, Paul M.Darden, 1990). Antibiotics such as penicillins, cephalosporins, macrolides, and trimethoprim-sulfamethoxazole have been commonly used to treat pneumococcal infections and are the primary clinical intervention for pneumococcal disease. However, since the rise of antibiotic resistance, judicious use of antibiotics has to be considered. Indeed, it has been demonstrated that a high rate of antibiotic use leads to a higher rate of non-susceptible IPD (Hicks et al., 2011).

S. pneumoniae has a remarkable ability to acquire drug resistance due to its recombination-mediated genome plasticity. This explains, in part, why despite the availability of antimicrobials, the burden of IPD still remains high worldwide. The rise of multidrug resistant clones constitutes a major health concern, which has led to the emergence of alternative measures, such as vaccines, to better control this pathogen. A polysaccharide-based vaccine was introduced in 1983 and covers the 23 strains associated with the most severe disease. While this 23-valent vaccine is used today to vaccinate adults at higher risk of pneumococcal infections and the elderly, the low immunogenicity of polysaccharide based vaccines in children under the age of 2, and the relatively poor protection provided to children between the age of 2-5 (62%), promoted the development of a 7-valent glycoconjuguate vaccine (PCV7) (Fiore et al., 1999). Indeed, the FDA licensed the PCV7 vaccine in 2000 for use in children under the age of 5. PCV7 includes polysaccharides from seven disease-causing serotypes conjugated to a carrier protein. Following its introduction into infant immunization routines, a significant decrease in invasive disease caused by the vaccine serotypes has been observed. In the U.S., the incidence of IPD caused by PCV7-serotypes decreased by over 90% in children under the age of 5. However, an increase in the prevalence of non-vaccine serotypes

occurred simultaneously, reducing the overall effectiveness of the vaccine. This phenomenon is called serotype replacement (Gladstone et al., 2011; Leibovitz, 2008; Pilishvili et al., 2010) and pushed the development of the 13-valent vaccine (PCV13), which includes an additional 6 strains not included in the PCV7, therefore providing broader protection. Following the introduction of the PCV13 vaccine in infant vaccination schedules in Canada, the prevalence of those vaccine-serotypes decreased from 66% to 41% in children under the age of 5 and from 54% to 43% in children over 5 years of age (Vella and Pace, 2014). In the UK, a significant decline in the incidence of IPD was observed in children under the age of 2 with a decrease as great as 75% after the introduction of the PCV13 into the infant vaccination program (Vella and Pace, 2014). However, the constant increase in the incidence of non-vaccine serotypes after the use of PCV13 requires ongoing monitoring and a 15-valent conjugate vaccine is currently being evaluated in animal models. This vaccine includes two more emerging serotypes that have been greatly responsible for IPD in children in the U.S. While pneumococcal vaccines have been shown to be effective, it is quite unlikely that *S. pneumoniae* will be permanently controlled by the use of capsule-based vaccines alone, since vaccine-induced pressure may continue to result in the emergence of new non-vaccine-serotype strains. Furthermore, while the use of vaccines has decreased antibiotic resistance among vaccine strains, the emerging non-vaccine serotypes are showing a significant increase in antibiotic resistance. In fact, serotype 19A, a non-vaccine strain that emerged after the introduction of PCV7, was found to be resistant to all FDA-approved antibiotics for children with otitis media (Fiore et al., 1999). Antibiotic resistance is often associated with serotype replacement, where *S. pneumoniae* switches its capsule to survive vaccine pressure and simultaneously acquires multidrug resistance in order to spread more easily (Fenoll et al., 2011; Hicks et al., 2011). The constant emergence of non-vaccine strains causing severe disease and the increase in antibiotic resistance suggest that new methods of treatment and prevention have to be considered in order to permanently control this worldwide pathogen.

1.2 *S. pneumoniae* pathogenesis

Despite the availability of vaccines and antibiotics, *S. pneumoniae* still remains a burden on global health care systems. This has led the scientific community to investigate the

molecular details behind *S. pneumoniae* pathogenesis in order to gain a greater understanding and develop long-term treatment and prevention strategies. Indeed, *S. pneumoniae* is a remarkable pathogen possessing a battery of virulence factors that are deployed to colonize, invade and escape the host defenses. It has been demonstrated that asymptomatic colonization is a requirement for the establishment of infection. The discussion below is an overview of *S. pneumoniae* colonization and pathogenesis, as well as a brief summary of the major characterized virulence factors, including the capsule.

1.2.1 Pathogenesis of pneumococcal pneumonia

In order to discuss the molecular details behind *S. pneumoniae* virulence, an overview of the general aspects of the pathology of pneumococcal pneumonia is necessary. Four pathological events occur during pneumococcal pneumonia. The first one is called engorgement and happens during the first 4 hours of infection. During this stage there is a rapid growth and spread of bacteria to the lungs, which results in *S. pneumoniae*-containing serous fluid filling the alveolar space. This step corresponds to the clinical onset of pneumonia (McCullers and Tuomanen, 2001). At that stage, very few leucocytes have been recruited, allowing exponential growth of the bacterium in this fluid that acts as a growth medium (Harford and Hara, 1947). Following engorgement, macrophages and erythrocytes start to reach the alveoli leading to red hepatization where the lung resembles the liver. In the next step called gray hepatisation, leucocytes fill the alveoli and phagocytosis through complement-mediated opsonisation occurs (Kline and Winternitz, 1915). In the last stage, capsule specific antibodies are produced allowing more efficient opsonisation. Resolution of the lesion begins where monocytes clear cell debris, allowing the lung to return to its normal state. Why the lung recovers so perfectly despite such an intense inflammatory process remains unclear.

1.2.2 Capsular polysaccharide

Most *S. pneumoniae* strains are encapsulated. The capsule consists of repeating carbohydrate units, and depending on the composition, over 90 different serotypes can be distinguished. The polysaccharide capsule does not participate in adherence or invasion of the host; rather its role in pathogenesis lies in its ability to prevent phagocytosis. The capsular polysaccharide unit is composed of a series of monosaccharides (glucose,

glucuronic acid, N-acetylglucosamine (GlcNAc), galactose, N-acetylgalactosamine (GalNAc), N-acetylfucosamine and N-acetylmannosamine) as well as some molecules of glycerol, choline and acetate. The capsule is attached to the peptidoglycan of the cell wall and is about 200-400 nm thick. The capsule can undergo spontaneous phase variation between an opaque and a transparent colony form, which corresponds to capsule thickness (Weiser et al., 1994). While the capsule composition, and therefore the serotype, define the degree of virulence of a particular strain, the thickness of the capsule directly correlates with virulence within a particular serotype (Kim and Weiser, 1998; Weinberger et al., 2009). A previous study has demonstrated that the capsule phase can vary from a virulent form with a thicker capsular polysaccharide to an avirulent form with low level of capsule (Kim and Weiser, 1998). Generally, *S. pneumoniae* isolates that are unencapsulated are completely avirulent, demonstrating the importance of the capsular polysaccharide in pathogenesis (Avery and Dubos, 1931). While the capsule provides protection against phagocytosis, its negatively charged nature causes electrostatic repulsion with host sialylated mucopolysaccharides. This protects *S. pneumoniae* from mucus entrapment, which allows adherence to the epithelial cell and subsequent invasion (Nelson et al., 2007). The pneumococcal vaccines available target the polysaccharide capsule, thus provide serotype specific protection. The capacity of *S. pneumoniae* to switch capsule under vaccine pressure has led to the need for a universally expressed protein target for future vaccine development.

1.2.3 *S. pneumoniae* colonization

S. pneumoniae colonizes the upper respiratory tract of 10-40% of healthy individuals and can persist for a few weeks to a few months depending on the serotype (Hodges et al., 1946). The rate of carriage can be as high as 60% in children, with up to four different serotypes present at once. Acquisition of a new serotype can lead to invasive disease (Gray et al., 2015). Adherence and colonization of the nasopharynx by *S. pneumoniae* is a prerequisite and an early key event in the development of pathogenesis. This colonization step is crucial for the horizontal spread of the bacterium in the community and requires adherence to host cells as well as evasion of the innate immune response. As the thick capsule covers the surface proteins essential for adherence, *S. pneumoniae* has to shed its

capsule as it approaches the cell surface allowing for attachment (Hammerschmidt et al., 2005). Transparent variants are better at colonizing nasopharyngeal cells, while opaque variants survive better in the blood due to their thick capsule. Colonization is also highly influenced by environmental factors such as smoking and viral infection; previous viral infection can predispose to secondary pneumococcal pneumonia (Peltola and McCullers, 2004).

A large array of pneumococcal surface proteins are implicated in colonization of the nasopharynx (Cron et al., 2009; Jeong et al., 2009; Kadioglu et al., 2010; Pracht et al., 2005; Rosenow et al., 1997; Uchiyama et al., 2009). Four different types of cell-wall attached proteins are found on the surface of the bacterium including choline binding proteins, lipoproteins, LPXTG-anchored proteins and a class of non-classically attached proteins that are found associated to the cell wall despite lacking a signal peptide (SP) or a known anchoring motif (Jeong et al., 2009; Pérez-Dorado et al., 2012). Among choline binding proteins (CBP), three autolysins are found: LytA, LytB and LytC. LytA was the first described and is the main pneumococcal autolysin (Garcõ et al., 1999) (Figure 1). Its role is to lyse *S. pneumoniae* cells during stationary phase to increase inflammation. LytB and LytC increase colonization and virulence in rats possibly through the release of other pneumococcal proteins involved in these processes after lysis (Garcõ et al., 1999; Gosink et al., 2000; Ramos-Sevillano et al., 2011). Other CPBs include PspA which is involved in complement inhibition by interfering with factor B of the alternative complement cascade, thus inhibiting complement deposition on *S. pneumoniae* and subsequent phagocytosis (Talkington et al., 1991; Tu et al., 1999). CbpA (also called PspC) is also a CBP, which has been shown to be involved in adherence to epithelial cells *in vitro* and is important for colonization of the nasopharynx in a rat model (Rosenow et al., 1997).

NanA is a well-studied pneumococcal neuraminidase which is an LPXTG-surface-bound protein and is involved in a range of virulence processes (Figure 1) (Banerjee et al., 2010; Brittan et al., 2012; Manco et al., 2006; Uchiyama et al., 2009). This enzyme removes terminal sialic acid moieties found on host glycoconjuguates and is very similar to the influenza neuraminidase. Indeed, one of the synergistic mechanisms between the

influenza virus and *S. pneumoniae* involves the degradation of sialic acid by the influenza NanA which facilitates *S. pneumoniae* adherence, thus enhancing the development of secondary pneumococcal infection (McCullers and Bartmess, 2003). The pneumococcal enzyme is involved in colonization and persistence in the nasopharynx and the middle ear, as well as in the development of otitis media in a chinchilla model (Tong et al., 2000). However it does not contribute to virulence in a mouse intraperitoneal challenge model (Berry and Paton, 2000). Another well-studied pneumococcal LPXTG-surface protein is BgaA, a β-galactosidase implicated in multiple aspects of virulence such as growth, resistance to opsonophagocytic killing and adherence (Figure 1) (Burnaugh et al., 2008; Dalia et al., 2010; Limoli et al., 2011; Singh et al., 2014). The implication of BgaA in growth and opsonophagocytic killing has been shown to be dependent on its enzymatic activity, but without BgaA increased adherence to host epithelial cells is observed. The same adherence phenotype has been shown when BgaA inhibitors were used, suggesting that BgaA acts as an adhesin to mediate cell attachment (Singh et al., 2014). Structural analysis revealed the multi-modularity of this enzyme including the discovery of two novel carbohydrate binding modules that have been shown to be involved in adherence independently of the enzymatic activity of BgaA. This reinforces the hypothesis that initial attachment of the pneumococcus to host cells can be glycan-mediated (Figure 1). The different pneumococcal surface protein examples described above highlight the fact that *S. pneumoniae* expresses a constellation of proteins at its surface that are involved in many mechanisms important for colonization and cell adherence (Figure 1). Since *S. pneumoniae* is a versatile pathogen, it is thought that it can modulate its surface proteins depending on the target cell type, which explains the various mechanisms that this bacterium possesses for colonization and host cell attachment (Bergmann and Hammerschmidt, 2006).

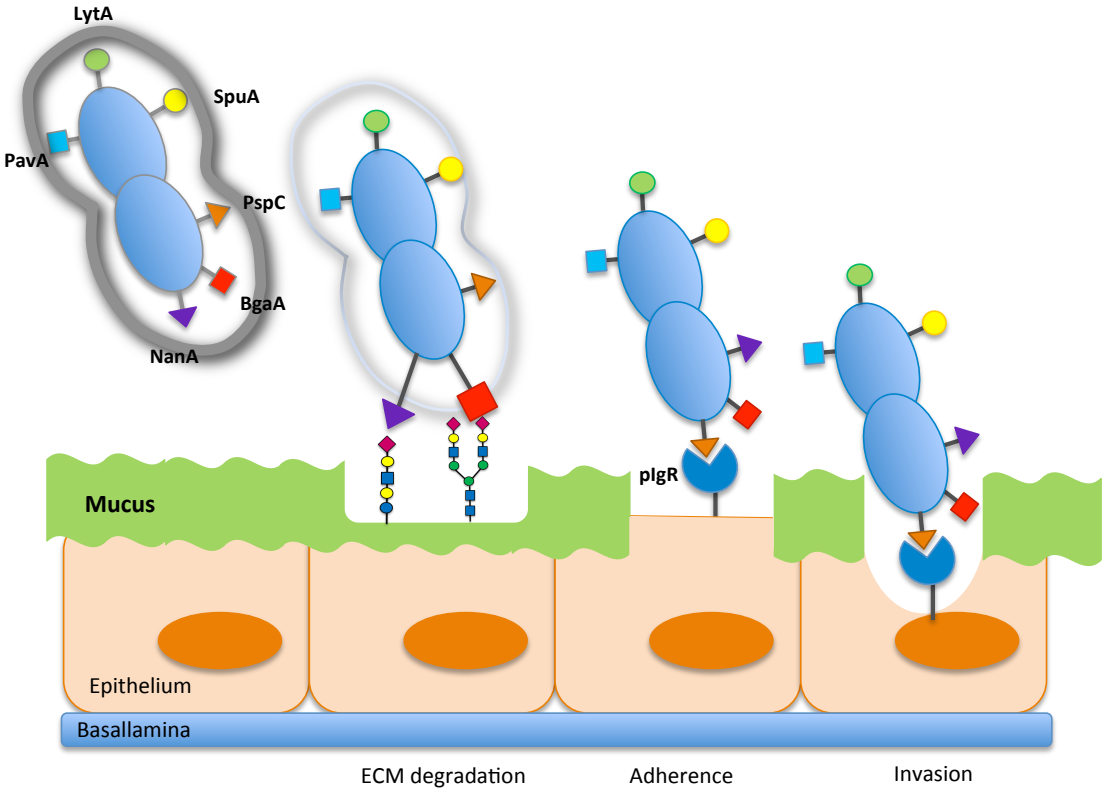


Figure 1. Schematic of mechanism of adhesion and invasion by *S. pneumoniae*.
Adapted from Bergmann and Hammerschmidt (Bergmann and Hammerschmidt, 2006). Abbreviations are as follows: Autolysin A (LytA), Pneumococcal adherence and virulence factor A (PavA), Pullulanase (SpuA), Neuraminidase A (NanA), β-Galactosidase A (BgaA), Pneumococcal surface protein (PspC) and Polymeric immunoglobulin receptor (PIgR). *S. pneumoniae* is represented as blue diplococci. Surface exposed proteins involved in ECM degradation, adherence and invasion of the host cells are represented by coloured shapes. The capsule polysaccharide is represented by grey circles and halos.

1.2.4 Pneumococcal invasive disease

Following asymptomatic colonization and adhesion, *S. pneumoniae* can infrequently cause severe disease by migrating to other sites. Indeed, the rate of carriage is much higher than the incidence of disease probably because *S. pneumoniae* can adhere to the epithelial cells lining the nasopharynx but is unable to attach to the ciliated cells of the bronchial tree. How *S. pneumoniae* goes from a passive colonizing agent of the upper respiratory tract to an invasive pathogen is presently not well understood. However, external factors such as smoking, antecedent viral infection, and inflammation promote invasive pneumococcal disease. Induction of inflammation is a key element of *S. pneumoniae* pathogenesis and further promotes disease progression. Cytokines and tumor necrosis factor (TNF) are released during inflammation and increase host cell permissiveness. Indeed, higher production of these immune molecules promotes expression of the PAF receptor, which *S. pneumoniae* can bind to through phosphorylcholine on its surface (Rijneveld et al., 2004). This process allows internalization of the bacterium through transcytosis. During the inflammatory response, an up-regulation of the polymeric immunoglobulin receptor (pIgR) is also observed (Johansen and Kaetzel, 2011). The pIgR translocates immunoglobulins across the mucosal barrier to protect mucosal surfaces against pathogens and maintain the commensal microbiota. In the case of pneumococcal infection, *S. pneumoniae* takes advantage of this internalization process by interacting with pIgR through a surface protein called PspC (also called CbpA), which induces endocytosis of the bacterium through the epithelial cells and enhances pneumococcal adhesion and invasion (Figure 1) (Agarwal et al., 2010; Zhang et al., 2000) (Figure 1). *S. pneumoniae* can also enhance inflammation through a cholesterol-dependent cytolysin called pneumolysin. This cytoplasmic toxin is released upon quorum sensing-dependent autolysis of *S. pneumoniae* and interacts with host cells through cholesterol on their surface (Mitchell and Dalziel, 2014). This process triggers pore formation and subsequent lysis of host cells, which enhances inflammation. Interaction of pneumolysin with host cells increases expression of immune regulatory molecules, further enhancing the inflammatory response (Mitchell and Dalziel, 2014). Indeed, creating a pro-inflammatory environment favours dissemination of the bacterium.

1.2.5 Control of *S. pneumoniae* colonization by the innate immune response

The innate immune system is critical to control colonization of *S. pneumoniae* in order to prevent the progression to invasive disease. However, under poorly understood circumstances *S. pneumoniae* can set in motion a series of events leading to the development of inflammation and disease. Inflammation caused by *S. pneumoniae* invasion triggers an influx of leucocytes into the alveolar space and slow clearance of bacterial cells by ineffective complement opsonisation. Complement-mediated phagocytosis is the first line of defense against pneumococcal infections and consists of the deposition of the complement component C3b on the bacterial surface, allowing its recognition and engulfment by phagocytic cells via C3b receptors on their surface (Figure 2). This innate host defense mechanism remains quite inefficient since *S. pneumoniae* has developed strategies against C3b deposition, namely capsulation (Hyams et al., 2010) and deglycosylation of complement components. Recent studies have demonstrated that disruption of three virulence associated glycoside hydrolases, NanA, BgaA and StrH results in an increase in C3b deposition and neutrophil-mediated killing, suggesting a role for these enzymes in the disruption of the alternative complement cascade (Dalia et al., 2010; Pluvinage et al., 2011; Singh et al., 2014). The alternative complement pathway is initiated by the covalent deposition of small amounts of C3b, resulting from the spontaneous hydrolysis of C3 in plasma, onto the bacterium surface (Figure 2). Surface-bound C3b then binds factor B (FB), a positive regulator, to form a C3bB complex. Factor D (FD) cleaves FB bound to C3b to form the alternative pathway C3 convertase complex, C3bFBb, which is stabilized by properdin (Walport, 2001). The C3 convertase cleaves many C3 molecules to C3b, allowing amplification of C3b deposition. Complement components are glycoproteins, and their glycan structures are required for their proper functions (Ritchie et al., 2002). Hence, degradation of their glycan decorations by pneumococcal surface enzymes decreases complement effectiveness.

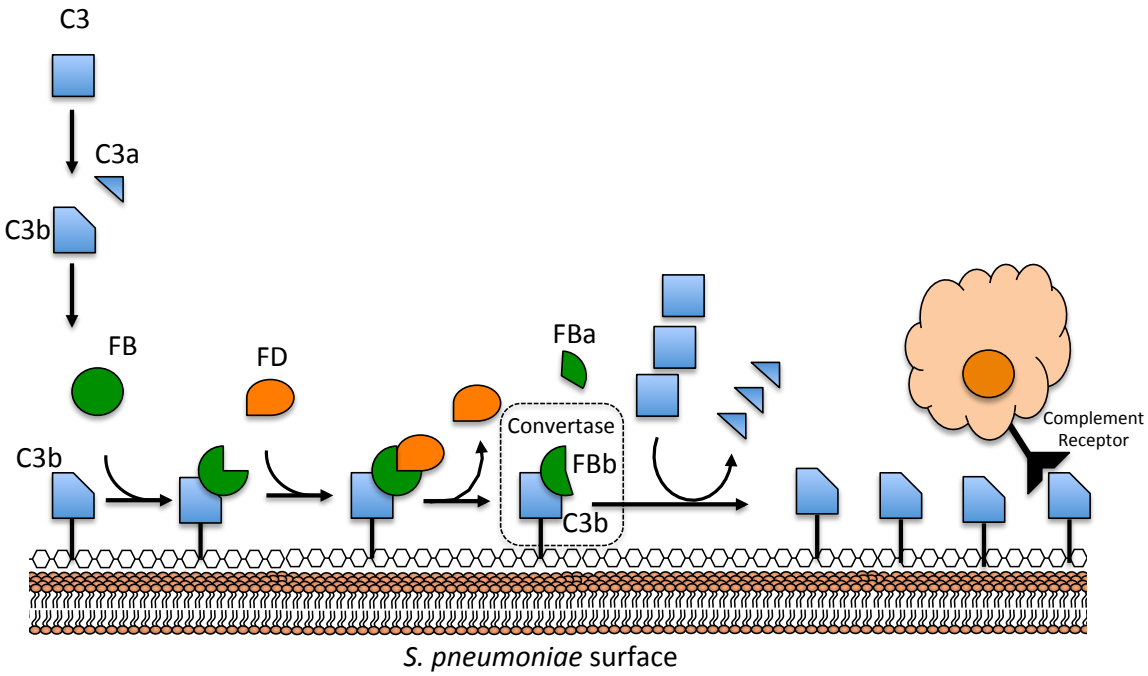


Figure 2. Alternative complement pathway.

Low amounts of plasma C3 spontaneously hydrolyse to C3b, which binds covalently to the bacterium surface through a thioester bond. Factor B (FB) binds to the surface-bound C3b and is cleaved and activated by the enzyme factor D (FD). This results in the formation of the C3 convertase enzyme C3bBb, which is stabilized by properdin (not represented here for clarity). The C3 convertase cleaves more C3, leading to the amplification of C3b deposition on the surface of the bacterium. Surface-bound C3b acts as an opsonin, which is recognized by the complement receptor on the surface of macrophages resulting in phagocytosis of the bacterium.

Over the last few decades, a great advancement in *S. pneumoniae* research has provided a better understanding of how *S. pneumoniae* shifts from a commensal colonizer to a pathogen. *S. pneumoniae* deploys many mechanisms to adhere and invade host cells, as well as for the enhancement of inflammation and evasion of the immune system. Experimental evidence suggests that host glycans are at the heart of these processes, playing a role in all aspects of the host-pathogen interaction such as deglycosylating immune molecules, acting as receptors for adherence, or simply providing a carbon source for metabolism and proliferation during infection.

1.3 Human glycans

1.3.1 Glycoconjuguates

A glycoconjuguate is a molecule that comprises a non-carbohydrate moiety covalently bound to one or more glycans. Two major types of glycoconjuguates exist, glycoproteins and glycolipids, depending on whether it is a protein or a lipid, respectively, attached to the glycan portion. Unlike protein sequences, glycans structures are not encoded by the genome but are secondary gene products. Therefore, even with the knowledge of glycan-acting enzyme expression levels, the prediction of glycan structures is very difficult. In addition, minor environmental changes can cause drastic changes in glycan structures in cells. All these reasons illustrate why human glycans are very complex, diverse, and difficult to study. As their structures vary enormously, the physiological roles of glycans are quite diverse as well. Overall, glycan roles range from subtle to absolutely crucial for the development of the organism that synthesized it. Overall, glycans in glycoproteins can have an important structural role as well as being involved in function modulation of the protein that they are attached to. Moreover, the specific recognition of glycan structures by other molecules is a key step in diverse biological processes. This section is a brief discussion of some of the different types of glycoconjuguates found in humans.

1.3.3 Glycoproteins

Depending on the mode of attachment to the protein moiety, two different types of glycoproteins are found: N-glycan, where the glycan portion is linked to an amide group of an asparagine of the protein; and O-glycan, where the glycan is attached to the protein

via the hydroxyl group of either a serine or a threonine. Many monosaccharides have been identified as starting units. O-GalNAc is of high relevance in mucin glycoproteins. Indeed, glycoproteins found in mucin are rich in serine and threonine O-glycan acceptor sites for the formation of more complex oligosaccharides (Van den Steen et al., 1998). Mucins are found in all mucous secretions, including the gastrointestinal tract, genitourinary tract and the airway, and are highly O-glycosylated with about 80% of their molecular weight being O-glycans. Based on the sugar sequence, different cores can be identified which are all represented in Figure 3B. No consensus sequence has been identified for O-glycosylation, however O-glycans in mucin are found in the presence of regions called "variable number of tandem repeats" (VNTR) which seem to be rich in O-glycan accepting serine and threonine residues.

1.3.4 N-Glycan

N-linked glycans comprise a class of oligosaccharide found on glycoconjugates and are functionally important across all of the taxonomic kingdoms. These glycans occur on many secreted and membrane-bound glycoproteins at Asn-X-Ser/Thr sequons, where X can be any amino acid other than proline. Compared to O-glycans, N-glycans are generally much more complex and are covalently attached to the protein at an asparagine by a N-glycosidic bond. Most commonly, GlcNAc is found bound to the accepting asparagine from the protein. N-glycans are essential to the proper function of glycoconjuguates by providing a range of biochemical properties important to the folding, stability and activity of their protein scaffold. Subsequently, these glycans play an essential role in a number of physiological processes ranging from immunity, cell-cell recognition, signal transduction and susceptibility to proteases as well as antigenicity (Helenius and Aebi, 2001; Jarrell et al., 2014). Defects in proper N-glycan formation on glycoproteins can lead to a variety of human diseases such as congenital disorders of glycosylation. Three different types of N-glycans exist: high mannose type with mannose residues exposed; complex type with terminal sialic acid residues; and hybrid type, in which mannose residues are found on the Man-α1,6 arm and one or sometimes two complex antennae are attached to the Man-α1,3 arm of the core (Figure 3C). Nevertheless, all three types share a common core sugar sequence: mannose-α-1,6 (mannose-α-1,3) mannose-β-1,4 N-acetylglucosamine-β-1,4 N-acetylglucosamine-β-1-

Asn ($Man_3GlcNAc_2$-Asn). N-glycan synthesis happens in the endoplasmic reticulum (ER) and starts at the ER membrane on a dolichylphosphate precursor. A series of seven mannose residues are then added by specific glycoside transferases (GT) in the cytoplasm. The glycan is then flipped to the ER lumen where the glycan is further expanded by the addition of more monosaccharides and is finally transferred to an Asn-X-Ser/Thr sequons on the nascent protein. The glycan is then re-modeled in the golgi apparatus by the action of GTs, which leads to the formation of different N-glycan types with more complex structures.

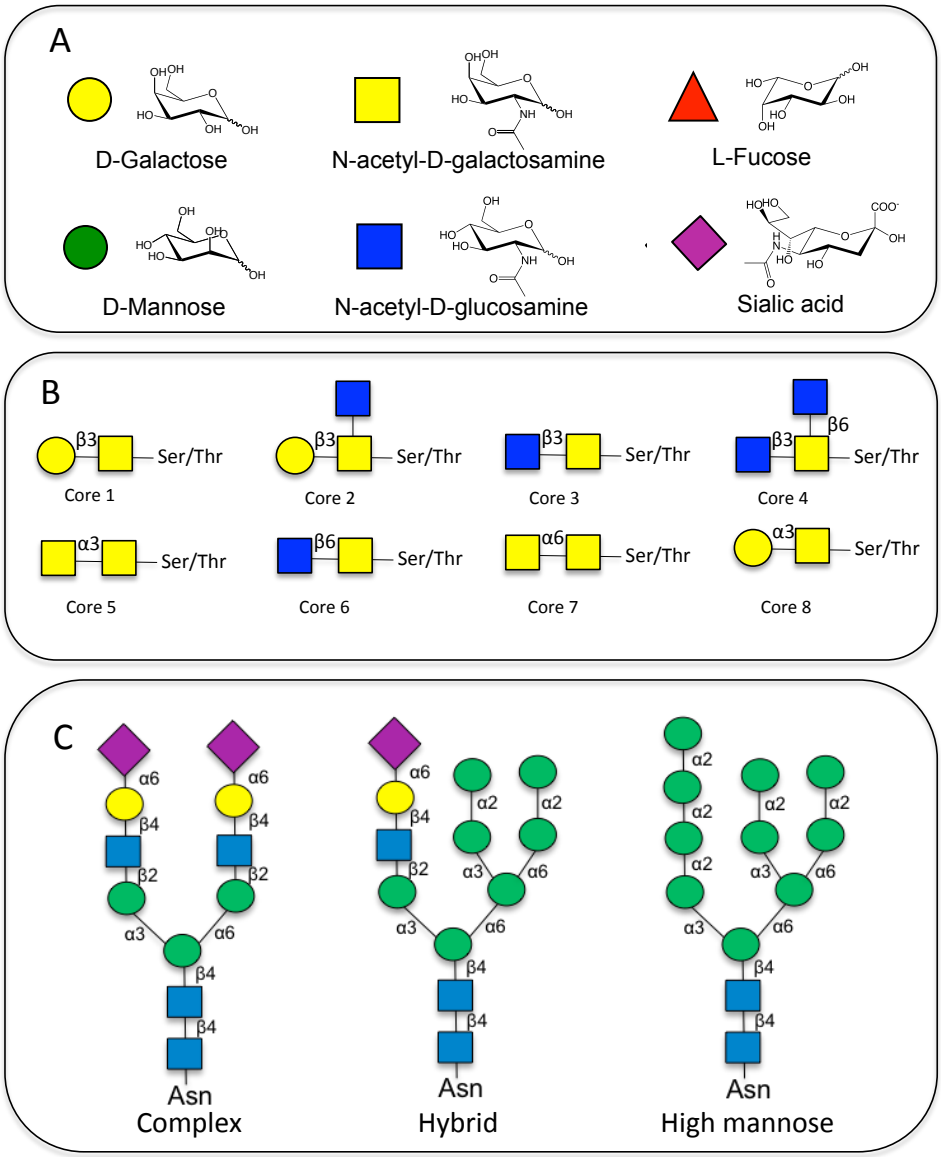


Figure 3. Symbolic representation of common human glycans.

(A) Symbolic representation of sugars. (B) Schematic representation of the O-glycan cores. (C) Schematic representation of the different type of N-Glycans.

1.4 Carbohydrate metabolism

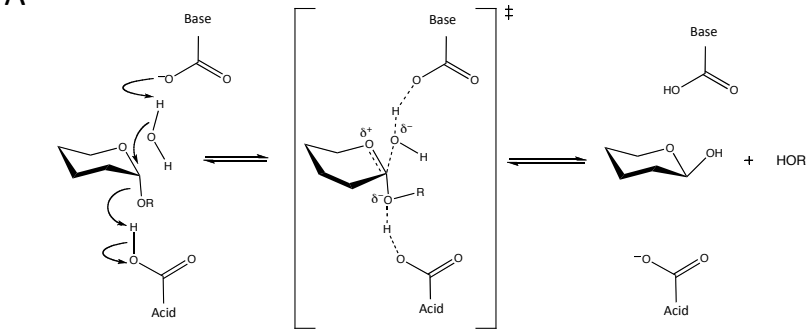
1.4.1 Glycoside hydrolases

Glycoside hydrolases are carbohydrate active enzymes (CAZymes) that can hydrolyze the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety. These enzymes are conserved in all domains of life and are crucial to a large range of biological processes. Since there is a direct relationship between sequence and folding similarities, a classification based on amino acid sequence similarities is used (www.cazy.org) (Cantarel et al., 2009). This classification allows a certain number of predictions in terms of substrate specificity, mechanisms, etc. To date, 133 families of GHs have been characterised. Although this classification is extremely useful, it is not exhaustive since there are many examples of GHs that share no sequence similarity but have similar folds. Therefore, another classification exists based on fold and mechanism. In this classification, GHs are divided into 14 clans from clan-A to –N (Davies and Henrissat, 1995). The TIM barrel fold, also known as a $(\beta/\alpha)_8$ fold, is the most conserved fold among GHs. However, their catalytic domains can also consists of a β-propeller, β-jelly roll, and $(\alpha/\alpha)_6$ arrangements. Additionally, several different active site topologies are seen: the pocket, the cleft and the tunnel (Davies and Henrissat, 1995). GHs can also possess ancillary modules, such as carbohydrate-binding modules (CBM). CBMs are non-catalytic domains whose role is to specifically bind to the substrate to position the enzyme in close proximity to its substrate, thereby increasing local enzyme concentration and enzyme efficiency (Boraston et al., 2004).

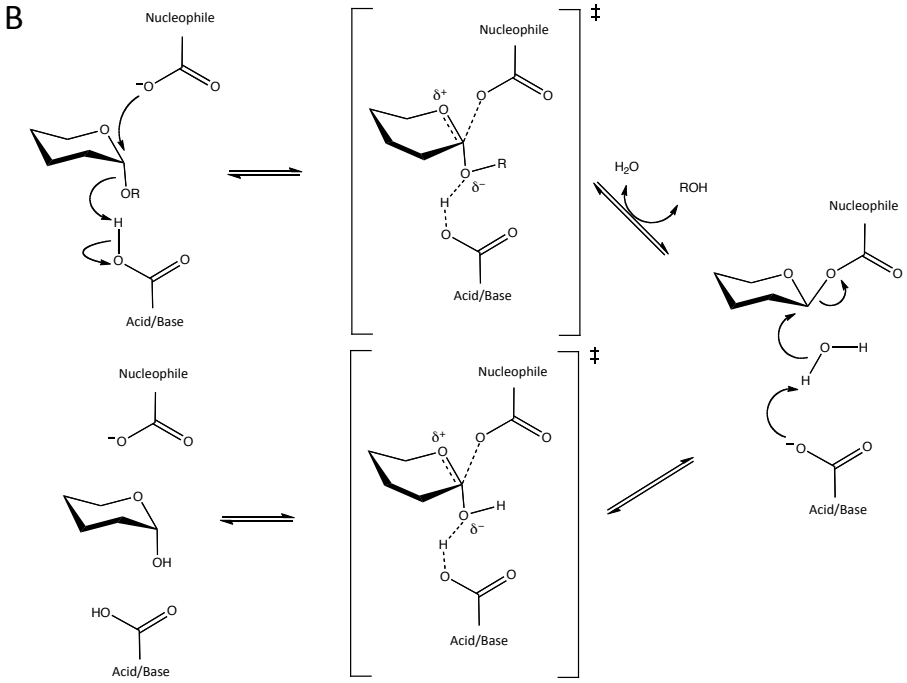
GHs can be exo- or endo-acting referring to the ability of the enzyme to cleave glycosidic bonds at the end or within the middle of a carbohydrate chain. Within the active site of GHs, subsites can be identified. Each subsite usually binds to a monosaccharide residue. They are numbered with increasingly negative numbers (-1, -2, -3, etc) away from the point of cleavage towards the non-reducing-end of the sugar, and with increasingly positive numbers (+1, +2, +3, etc) towards the reducing-end of the sugar (Davies et al., 1997). The part of the sugar at the non-reducing end of the point of cleavage is called the glycon and the part on the reducing end is called the aglycon. Two main mechanisms are used by GHs: retaining or inverting depending on whether the stereochemistry of the

anomeric carbon is retained or inverted. In the inverting mechanism, hydrolysis happens with a net inversion of the anomeric configuration through a one-step single displacement mechanism, which involves an oxocarbenium ion-like transition state (Figure 4A). Two catalytic residues, glutamate and aspartate/glutamate, are essential for an inverting catalysis and they act as a catalytic acid and base, respectively. These two catalytic residues are generally about 10 Å apart to allow proper positioning of a water molecule necessary for hydrolysis (Davies and Henrissat, 1995; McCarter and Withers, 1994). In contrast, the retaining mechanism uses a two-step double displacement mechanism involving a covalent glycosyl-enzyme intermediate and leading to net retention of the anomeric configuration (Figure 4B). Usually a glutamate and/or aspartate act as an acid/base and nucleophile, and are roughly 5.5 Å apart (Davies and Henrissat, 1995; McCarter and Withers, 1994). Each step of this mechanism goes through an oxocarbenium ion-like transition state. In the first step, one residue acts as a nucleophile and attacks the anomeric carbon to form a glycosyl-enzyme intermediate (Abbott et al., 2009). Simultaneously, the other catalytic residue acts as an acid and gives a proton to the glycosidic oxygen. In the next step of this reaction, this same residue now acts as a base and deprotonates the water molecule. Although these mechanisms are the most common, variations can be seen. The most common variation is the substrate-assisted retaining mechanism where an acetamido group from the sugar substrate itself acts as the nucleophile for the anomeric carbon attack, leading to the formation of an oxazolinium ion intermediate (Figure 4C). In this mechanism, one of the catalytic residues acts as a stabilizing residue for the oxazolinium ion intermediate.

A


Base
Base
Base
Acid
Acid
Acid
OR
R
OH
HOR

B


Nucleophile
Nucleophile
Nucleophile
Acid/Base
Acid/Base
OR
R
H2O
ROH
Nucleophile
Nucleophile
Nucleophile
Acid/Base
Acid/Base
OH

C

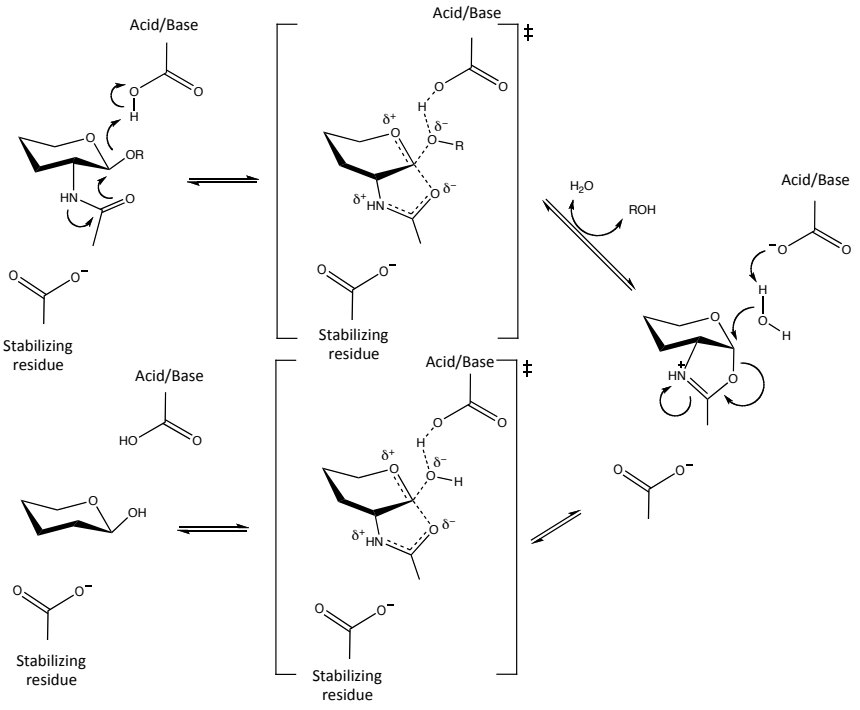


Figure 4. Glycoside hydrolase mechanisms.

Catalytic mechanisms used by GHs. (A) Inverting mechanism consists of a single-step displacement with a nucleophilic water to generate a product with an inverted stereochemistry. (B) Retaining mechanism is a double-step displacement with a covalent α-glycosyl enzyme intermediate. (C) Substrate-assisted mechanism consists of a double-step displacement with a bicyclic oxazoline as intermediate. The acetamido group of the substrate itself is the nucleophile responsible for nucleophilic attack of the anomeric carbon.

1.4.2 Glycans in host-pneumococcal interactions

1.4.2.1 *S. pneumoniae* carbohydrate metabolism

Carbohydrate metabolism is essential for the lifestyle of *S. pneumoniae*. Free sugars are rare in the human airway suggesting that habitants need to acquire carbon for growth through the catabolism of complex host glycans (Burnaugh et al., 2008; King et al., 2006; Marion et al., 2009, 2012). A large portion of the *S. pneumoniae* genome is dedicated to encoding for proteins responsible for carbohydrate metabolism, including transporters as well as CAZymes (Tettelin et al., 2001). In fact, up to 30% of all pneumococcal transport mechanisms are devoted to carbohydrate import and, according to the CAZy database, around 43 pneumococcal genes encode for GHs depending on the strain. *S. pneumoniae* possesses metabolic capabilities adapted for its environment and it has been shown that most carbohydrate-processing pathways are directed towards human derived carbohydrates (Bidossi et al., 2012; Buckwalter and King, 2012). Many of these metabolic pathways have been described and require initial depolymerisation of the glycan by extracellular pneumococcal GHs, import of the product via a transport mechanism and further processing by an intracellular metabolic pathway. A well-studied example is the glycogen-processing pathway, where a surface exposed pullulanase degrades glycogen (Abbott et al., 2010). A specific transporter then imports the product to allow further processing of the oligosaccharides by other intracellular enzymes leading to the release of glucose. Similarly, many extracellular GHs have been shown to specifically modify complex glycan from the host upper respiratory tract. Pneumococcal GHs have been shown to be active on a large variety of host sugars including N-glycans, O-glycans and glycosaminoglycan, further highlighting the remarkable capacity of *S. pneumoniae* to process human carbohydrates (King, 2010; King et al., 2006; Marion et al., 2012; Yamamoto et al., 2005). In addition, previous studies suggest that *S. pneumoniae* has evolved to utilize dietary carbohydrate that the bacterium encounters during nasopharyngeal colonization (such as inulin), frequently enough to maintain selective pressure (Hiss, 1902). This bacterium can also modify surface carbohydrate from other microbes that share the same niche, providing both a source of nutrients for *S. pneumoniae* and exposing competitive bacteria to the immune system (Shakhnovich et

al., 2002). These findings demonstrate that *S. pneumoniae* has evolved and adapted its carbohydrate metabolism properties to optimize survival in its niche.

1.4.2.2 Role of carbohydrate metabolism in *S. pneumoniae* virulence

Over the last few decades, research on *S. pneumoniae* has demonstrated the importance of carbohydrate degradation in colonization and virulence. Indeed, large-scale virulence screens such as a signature tagged mutagenesis (STM) study have identified genes encoding for CAZymes as putative virulence factors (Hava and Camilli, 2002; Orihuela et al., 2004; Polissi et al., 1998). Furthermore, microarray analyses have shown that these genes are up regulated under conditions that mimick different aspects of virulence. GHs involved in the degradation of host glycans play important roles in pathogenesis, the most evident being fitness (Figure 5). Given that the concentration of free sugar in the upper respiratory tract is low, host glycan utilization is a prerequisite for inhabitants of this niche. Indeed, *in vitro* studies have shown that *S. pneumoniae* can grow on host glycans as a sole source of carbon in a glycosidase-dependent manner. Extensive work on the pneumococcal GHs, NanA, BgaA and StrH, demonstrated their sequential action on complex N-glycans and their implication in fitness (King et al., 2006). Additionally, carbohydrate modification has been shown to be essential for adherence of *S. pneumoniae* to epithelial cells. Many studies have suggested that the neuraminidase NanA could reveal an asialylated glycan receptor on the surface of host cells, increasing *S. pneumoniae* binding to chinchilla tracheas and a human cell line (Tong et al., 2001; Uchiyama et al., 2009). These examples illustrate the key role that GHs can play in uncapping receptors to promote adherence. BgaA also plays a role in adherence to host cells and indeed, mutants deficient in either enzyme have a reduced ability to adhere to human respiratory cell lines (Limoli et al., 2011). Interestingly, the contribution of BgaA to adherence is independent of its enzymatic activity suggesting that in addition to revealing receptors, GHs could also act as adhesins to promote attachment to host cells (Singh et al., 2014)

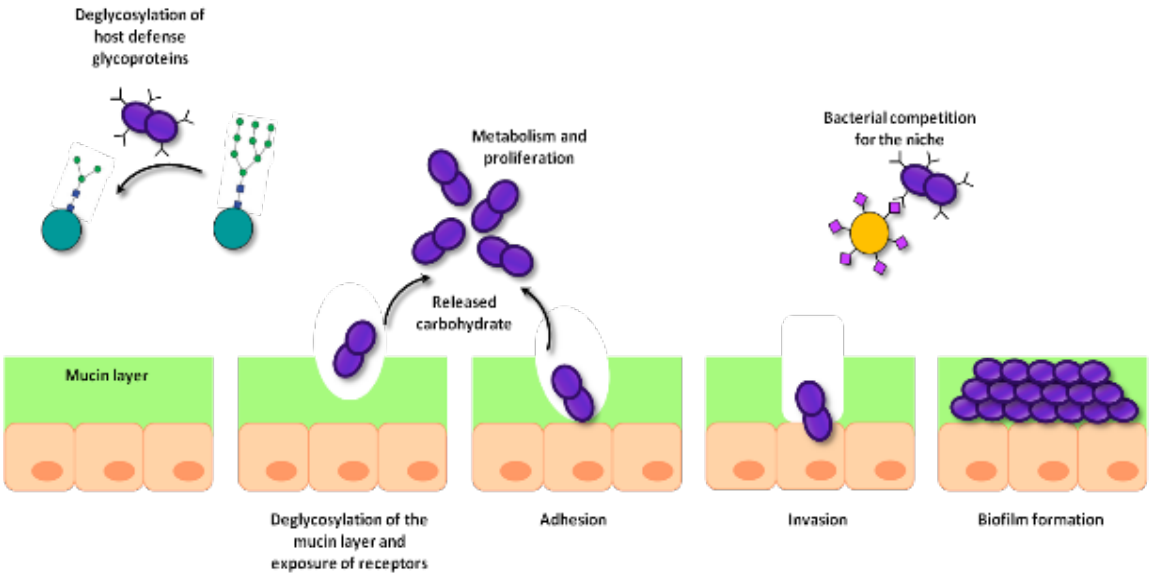


Figure 5. The proposed role of GHs in virulence.

Virulence-associated GHs from *S. pneumoniae* can have a role in various aspects of virulence such as nutrient acquisition, degradation of the mucin layer to access the surface of host cells, adhesion, invasion, biofilm formation, bacterial warfare for the same niche, and evasion of the immune system.

Modification of host carbohydrates by this bacterium can also increase resistance to the immune response. NanA, BgaA and StrH have been shown to be involved in complement modulation (Dalia et al., 2010). These enzymes depolymerize the terminal sugars on complex N-glycan of complement components triggering their inactivation, which leads to subsequent phagocytic-killing. In addition to adherence and immune evasion, further virulence-associated roles have been attributed to GHs such as participating in invasion of host cells, bacterial warfare but also biofilm formation (Figure 5).

Despite significant progress on our understanding of the role of carbohydrate modification in virulence, a full appreciation of these processes is still lacking. The *S. pneumoniae* CAZome reveals 43 GHs in the genome but only 9 have attributed specificities, suggesting that more uncharacterized pneumococcal GHs might further contribute to the full depolymerization of host glycans. Many of these uncharacterized GHs have also been identified in large-scale virulence screens suggesting that their carbohydrate modification capabilities could be linked to virulence. Currently, the known pneumococcal carbohydrate-processing systems do not permit *S. pneumoniae* to fully process the diverse host glycans encountered by the bacterium during colonization and infection, suggesting that a greater understanding of those systems is necessary to gain more insight into the molecular details of the host-pathogen interaction. Many linkages found in certain glycan structures are not presently known to be cleaved by any characterized pneumococcal enzymes. This fact, combined with the knowledge that *S. pneumoniae* encodes for additional uncharacterized GHs, suggests that this bacterium possesses the ability to process a broader diversity of host carbohydrates than is presently appreciated.

1.4.3 Glycan transport by *S. pneumoniae*

Thirty percent of pneumococcal transporters are devoted to carbohydrate transport. The most common groups of pneumococcal carbohydrate transporters are: phosphotransferase system (PTS) and ATP-Binding cassette (ABC) transporter. *S. pneumoniae* encodes for 21 PTS transporters and 10 ABC transporters. Many of them are implicated in pathogenesis further supporting the role of host carbohydrate uptake in virulence (Hava

and Camilli, 2002; Lau et al., 2001; Marion et al., 2012; McAllister et al., 2012; Obert et al., 2006; Ogunniyi et al., 2012; Orihuela et al., 2004; Polissi et al., 1998).

PTS transporters catalyze the transport and phosphorylation of their substrate. They are composed of a membrane-spanning protein and four soluble proteins. The two cytoplasmic proteins are Enzyme I (EI) and HPr. In most organisms, these phosphotransferase proteins are involved in the uptake of all carbohydrates imported by a PTS transporter. The sugar-specific enzyme II complex is composed of the components EIIA, EIIB, and EIIC. These components are responsible for the substrate specificity of the transporter (Deutscher et al., 2006; Erni, 2013). EI and HPr transfer phosphoryl groups from phosphoenolpyruvate to EIIA. EIIA and EIIB then sequentially transfer phosphates to the sugar substrate. Finally the sugar is imported across the membrane by EIIC, which creates the membrane channel. As mentioned before, 21 PTS system have been identified in *S. pneumoniae*. However, only 15 have an attributed specificity and experimental evidence supports the implication of 15 of them in virulence (Buckwalter and King, 2012).

ABC transporters are widely conserved and transport a wide variety of substrates (Berntsson et al., 2010; Procko et al., 2009). They utilize energy provided by ATP hydrolysis to translocate metabolites across a membrane. Among this class of transporters, both exporters and importers are found in bacteria. Their structure consists of two trans-membrane domains (TMDs or permeases), which dimerize to form the substrate translocation pathway, and two cytoplasmic nucleotide-binding domains (NDBs) that are responsible for ATP hydrolysis and providing energy for transport (Figure 6). Bacterial ABC importers have an additional solute-binding protein (SBP), which determines the specificity of the transporter by sequestering the ligand and delivering it to the permeases. Therefore, SBPs are found in two states: free or liganded, and usually have a high affinity for their ligand (~ low μM range) (Abbott and Boraston, 2007; Berntsson et al., 2010; Higgins et al., 2009; Quiocho et al., 1997; Suzuki et al., 2008). In Gram-positive bacteria, SBPs are lipoproteins covalently attached to the membrane via a lipo-box, while in Gram-negative bacteria they are found free-floating in

the periplasm. Because they are only found in prokaryotes, pneumococcal ABC importers and PTS transporters are appealing targets for drug development or novel drug delivery systems (Garmory and Titball, 2004; Parr and Saier, 1992).

The general mechanism of an ABC importer starts with the unliganded SBP in a closed conformation, as well as with the NBDs open in a resting state and the TMDs open to the cytoplasm (Figure 6A). Interaction of the ligand-bound SBP (Figure 6B) with the TMDs triggers ATP-dependent NBD dimerization and closure, which in turns triggers a conformational change in the TMDs (Figure 6C). The TMDs are now facing outward and are exposed to the ligand-binding cavity of the SBP. The ligand is then released by the SBP into the permease channel triggering ATP hydrolysis, which leads to the re-opening of the NBDs (Figure 6D). The open conformation of the NBDs induces the TMDs to return to facing the cytosol (Procko et al., 2009). The ligand is then release inside the cell.

About 10 ABC transporters have been identified in *S. pneumoniae,* many of which have unknown specificities. Interestingly, many pneumococcal ABC importers share an ATPase, msmK (Marion et al., 2011). It is not known if all pneumococcal ABC transporters lacking an ATPase-encoding gene use msmK. However, no other candidate gene has been identified in the *S. pneumoniae* genome. Although this is the first example of a shared ATPase among ABC transporter in pathogens, this phenomenon has been seen in other bacteria and seems to be a mechanism of genome conservation and/or regulation (Ferreira and De Sá-Nogueira, 2010; Schlosser et al., 1997; Silva et al., 2005).

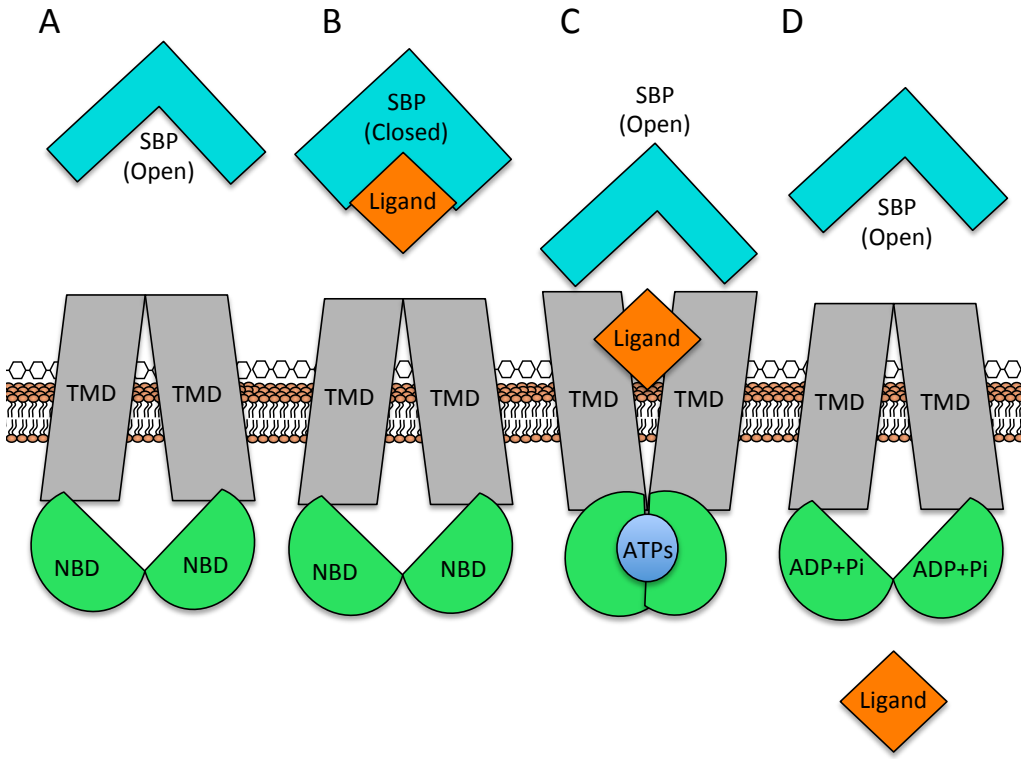


Figure 6. ABC importer mechanism.

(A) NBDs are in the resting state and the TMDs are open to the cytoplasm. The SBP is in an unliganded, open conformation. (B) The SBP binds to its ligand and adopts a closed form. (C) The SBP bound to its ligand interact with the TMDs, which triggers ATP-dependent NBD dimerization and closure. This promotes a conformational change in the TMDs, which are now facing outward. The SBP delivers the ligand to the TMDs, which triggers ATP hydrolysis. (D) ATP hydrolysis in the NBDs causes a loss of affinity for the two ATPases, causing the NDBs heterodimers to open. Dissociation of the NDBs triggers the TMDs to open facing the cytoplasm and release of the ligand inside the cell (Buckwalter and King, 2012; Procko et al., 2009).

1.4.4 N-glycan degradation by bacteria

The ability of bacteria to degrade N-glycans has been associated with important biological functions, such as nutrient acquisition in the soil, the normal mutualistic role of the gut microbiome, and bacterial pathogenesis in humans and plants, making the importance of this process well appreciated (Byers et al., 1999; Cao et al., 2014; Dupoiron et al., 2015; Renzi et al., 2011; Roberts et al., 2000; Suits et al., 2010). Consequently, the depolymerisation of N-glycan by microorganisms is relevant to understand these biological processes. However, the molecular details underlying how microbes process these sugars still remain unclear. To date, one comprehensive study on N-glycan processing has been conducted in *Xanthomonas camprestris*, where a cluster has been found that is responsible for the sequential depolymerisation of N-glycan from plant sources during host infection (Dupoiron et al., 2015). The sequential activity of numerous GHs from this genomic cluster has been demonstrated on N-glycans providing the first evidence of N-glycan degradation by a plant pathogen. *Bacteroides thetaiotaomicron* has also been described as being capable of degrading N-glycans (Martens et al., 2009). This gut microbe encodes for a Sus-like system responsible for the degradation of human high-mannose N-glycan. This system comprises an endo-β-N-acetyl-hexosaminidase from the GH18 family that cleaves high-mannose from glycoconjuguates found in the mucous layer. A series of periplasmic mannosidases from family 92 also target α-1,2-, α-1,3- and α-1,6 mannose linked residues found in these carbohydrates. Furthermore, a previous study demonstrated the ability of *Bacteroides fragilis* to utilise N-glycan from transferrin, providing an advantage to this bacterium during extraintestinal infection (Cao et al., 2014). In *Enterococcus faecalis*, an endo-β-N-acetylglucosaminidase and α-mannosidase have been shown to act on N-glycan from RNaseB and have been proposed to be involved in survival and persistence of this nosocomial pathogen *in vivo* (Roberts et al., 2000). Additionally, numerous exo-acting GHs from *Streptococcus oralis* have been shown to sequentially deglycosylate complex N-glycans to support growth. However, the enzymes responsible for this machinery have not been fully characterised (Byers et al., 1999). In *Capnocytophaga canimorsus,* a human pathogen, an enzymatic complex has been identified and proposed to be involved in the full processing of complex N-glycans (Renzi et al., 2011). This surface protein

complex comprises an endo-acting β-N-acetylglucosaminidase responsible for cleaving the chitobiose core of complex N-glycan. The released glycan is then thought to be imported to the periplasm by another component of this complex that resembles a TonB-dependent transporter. Once in the periplasm, a sialidase acts to cleave the terminal sialic acid on the N-glycan, which allows a series of putative periplasmic exoglycosidases to sequentially deglycosylate the rest of the glycan. Finally, a GH38 from *Streptococcus pyogenes* has been shown to be active on α-1,3-mannosidic linkages and therefore is presumed to be involved in N-glycan processing as well (Suits et al., 2010).

Despite significant advances towards understanding bacterial N-glycan processing, a significant gap in knowledge still exists in how microbes fully depolymerise those sugars. In this respect, the bacterium *S. pneumoniae* is an excellent model system to study bacterial degradation and metabolism of N-glycans. Extensive work has been done on the pneumococcal virulence-associated GHs, NanA, BgaA and StrH, demonstrating their sequential activity on the distal arms of complex N-glycans by trimming the terminal sialic acid, galactose and GlcNAc, respectively (Burnaugh et al., 2008). A pneumococcal endo-β-N-acetylglucosaminidase, EndoD, involved in this pathway has also been extensively characterised and shown to cleave the chitobiose core of N-glycans that has been trimmed down to at least $Man_3GlcNAc_2$ by NanA, StrH and BgaA, but also $Man_5GlcNAc_2$. However, the pneumococcal factors responsible for the degradation of the bulk of mannose in high-mannose N-glycans are completely unknown. Indeed, EndoD cannot act on high-mannose N-glycans that bears α-1,2-terminal mannose residues, suggesting that another enzyme must act prior to EndoD in order to remove those residues. Additionally, a previously studied pneumococcal α-1,6-mannosidase, GH125, has been shown to hydrolyse the α-1,6-linkage of N-glycans lacking the α-1,3-arm (Gregg et al., 2011). However, no characterised pneumococcal enzymes have been shown to cleave these α-1,3-mannoses that must be removed for GH125's action. It is thus, implied that *S. pneumoniae* encodes for additional putative mannosidases that would be responsible for the degradation of the α-1,3- and α-1,2-mannose residues on N-glycans, a prerequisite for the action of GH125 and EndoD. The fact that *S. pneumoniae* encodes for additional GHs in its genome with no attributed function led us to postulate

that some of these enzymes may target linkages found in N-glycans, and that this bacterium may encode for a pathway responsible for the complete degradation of these sugars.

1.4.5 Identification of an N-glycan-processing locus in the *S. pneumoniae* genome

In the S. *pneumoniae* genome, surrounding the SP2144 locus encoding for GH125, there are several putative CAZymes, including a gene encoding for a putative mannosidase from family 38 (Figure 7A, Table 1). Members of this family are often α-mannosidases. This GH38 encoded by the locus SP2143 shares 96% sequence identity with a characterised GH38 from *S. pyogenes* that displays α-1,3-mannosidase activity (Suits et al., 2010). This suggests that the pneumococcal GH38 could share such specificity and may hydrolyse the α-1,3-mannose linked residues found in the N-glycan core, as well as on high-mannose N-glycan mannose decorations. SP2143 and SP2144 co-occur adjacently in other firmicutes with a sequence identity ranging from 46 to 97% for SP2143 and 53 to 98% for SP2144 (Figure 7B). These two genes consistently reside in a putative carbohydrate-processing locus (CPL) encoding for a putative β-hexosaminidase (GH20C), a putative α-fucosidase (GH29), another putative α-mannosidase (GH92) and a protein belonging to the ROK family (Repressor, Open-reading-frame, Kinase) (Figure 7A, Table 1). The genomic proximity of these CPL-encoded enzymes implies a common N-glycan processing function. The putative α-mannosidase GH92 shares high sequence identity with the functionally characterised GH92s from *B. thetaiotaomicron* (Zhu et al., 2010). These enzymes target α-1,2-, α-1,3- and α-1,6 mannose linked residues from N-glycans suggesting that the pneumococcal GH92 could be active on similar linkages. However, since we suspect GH38 to be an α-1,3-mannosidase and it is well known that GH125 is a strict α-1,6-mannosidase, GH92 is likely to be active on the residual α-1,2-linkages found on high-mannose N-glycan.

This CPL locus is conserved in other Gram positives as well, with the highest sequence similarity with *Streptococcus pseudopneumoniae*, *Streptococcus mitis B6*, *Streptococcus parasanguinis* and *Streptococcus gordonii,* all of which are pathogenic bacteria (Figure 7B). Genes encoding EndoD and a hypothetical ABC transporter that are found spread

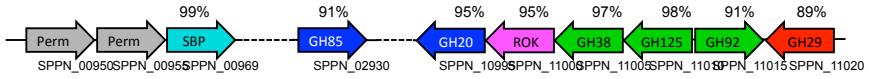
out in the *S. pneumoniae* genome, co-occur in CPL-containing firmicutes, within or in proximity of the CPL, suggesting a functional association with the CPL-components (Figure 7A-B, Table 1). The pneumococcal *endoD* and the *sbp* gene of the putative ABC transporter show quite high sequence similarity to homologues found in CPL-containing firmicutes, with percentage values from 42% to 99% for the *sbp* and 32% to 91% for *endoD* in *Lactococcus plantarum* and *S. pseudopneumoniae,* respectively. Based on this bioinformatic analysis, we hypothesise that the conserved CPL components as well as endoD and the ABC transporter encoding genes, represent a concerted N-glycan processing pathway in *S. pneumoniae.*

A

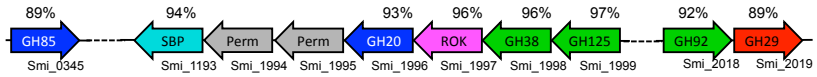
Streptococcus pneumoniae (all strains)

B

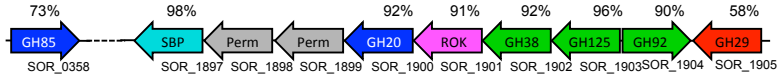
Streptococcus pseudopneumoniae IS7493



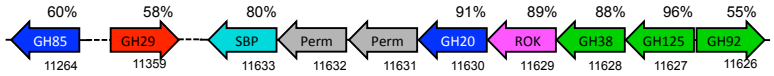
Streptococcus mitis B6



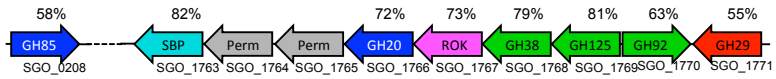
Streptococcus oralis Uo5



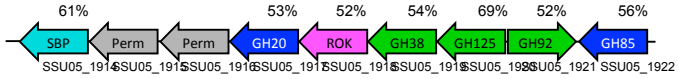
Streptococcus parasanguinis ATCC 15912



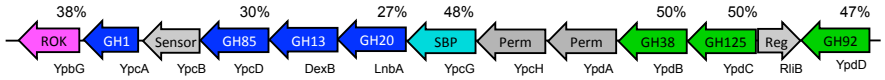
Streptococcus gordonii str. Challis substr. CH1



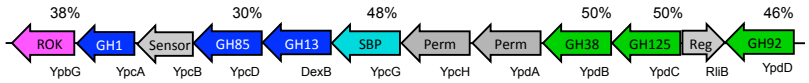
Streptococcus suis 05ZYH33



Lactococcus lactis sub. Lactis CV56



Streptococcus pasteurianus ATCC 43144



Lactobacillus plantarum subsp. plantarum ST-111

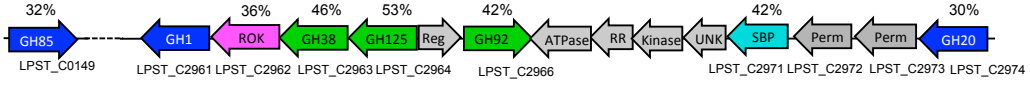


Figure 7. Organization and conservation of a Carbohydrate Processing Locus (CPL) proposed to be associated with N-glycan processing.

(A) The genetic organization of a carbohydrate-processing locus present in *S. pneumoniae* strains. GH38 and GH92 (both green) are putative α-mannosidases, GH29A (red) is predicted to be an α-fucosidase, GH20C (or GH20) (blue) is a predicted β-hexosaminidases, ROK (pink) is a predicted sugar kinase, SBP ($SBP_{N\text{-}glycan}$) (cyan) is the predicted solute-binding protein for the ABC transporter, the permeases (Perm) of the ABC transporter are shown in grey, GH125 (green) is a characterised α-1,6- mannosidase (Gregg et al., 2011). GH85 (blue) represents EndoD, an endo-β-N-acetylglucosaminidase that is active on the core of N-linked glycans (Yamamoto et al., 2005). (B) Conservation of the CPL in firmicutes. The colours are coded as in (A). Percentage values represent the sequence identity to the *S. pneumoniae* genes.

Table 1. Pneumococcal proteins encoded by the CPL or co-occuring with the CPL in firmicutes.

Locus Tag	Name	Activity	Localisation	Virulence	Closest homologue with known function
SP2141	GH20C	Putative Exo-β-hexosaminidase	UNK	-	GcnA(Harty et al., 2004; Langley et al., 2008)
SP2142	ROK	Putative Sugar kinase	UNK	V (STM)	ROK hexokinase from *Thermus thermophiles* (Nakamura et al., 2012)
SP2143	GH38	Putative Exo-α1-3-mannosidase	UNK	V+M (STM)	SpyGH38 (Suits et al., 2010)
SP2144	GH125	Exo-α-1,6-mannosidase	UNK	V (STM)	CpGH125 (Gregg et al., 2011)
SP2145	GH92	Putative Exo-α1-2-mannosidase	UNK	V (STM)	BT3390 (Zhu et al., 2010b)
SP2146	GH29	Putative Exo-α-Fucosidase	UNK	V (STM)	Smi2019 (SmiGH29)
SP0498	EndoD (GH85)	Endo-β-N-acetylglucosaminidase	Cell-wall (LPXTG)	V (STM)	EndoA (Ling et al., 2009)
SP0090-0092	$ABC_{N\text{-}Glycan}$	Putative ABC transporter	Cell wall	V+M (STM)	AlgQ1-AlgQ2/AlgM1-AlgM2/AlgS-AlgS (Nishitani et al., 2012)

UNK: Unknown; STM: identified by signature-tagged mutagenesis; M: identified by microarray analysis; LPXTG: sortase-mediated cell wall attachment (these proteins contain a classical signal peptide as well).

1.5 Hypothesis and research objectives

On the basis of the predicted and known functions of the CPL-encoded proteins we propose the following hypothesis: the *S. pneumoniae* genome encodes for a concerted N-glycan processing pathway responsible for the degradation and transport of complex, hybrid and high-mannose N-glycan from host tissues. Furthermore, based on the literature, we hypothesise that the components of this pathway encoded by the CPL are required for processes implicated in pneumococcal virulence. Molecular level details of the uncharacterized functions of the enzymes/transporter encoded by the CPL will be sought using structural biology. Additionally, biochemical studies will be performed to study the specificity and activity of the CPL enzymes and transporter. Finally, we will assess the implication of these pneumococcal factors in virulence using *in vivo* assays and microbiological studies. Specific focus will be to uncover the specificity/function of the uncharacterised SBP of the ABC transporter ($ABC_{N\text{-}glycan}$) and GH92, and provide evidence for their participation in N-glycan processing. The implication in virulence of $ABC_{N\text{-}glycan}$, GH92 and EndoD will also be assessed. Lastly, a functional and structural analysis of the CPL-encoded GH20C will be undertaken in an attempt to shed light into the role of this enzyme in host carbohydrate-processing.

This work will significantly advance the construction and validation of our model of N-glycan processing by *S. pneumoniae* and give a greater appreciation for carbohydrate metabolism in pneumococcal virulence. As the components of this model pathway are conserved among a wide variety of bacteria, and even fungi, we anticipate that this work will be of fundamental relevance to understanding how microbes degrade and metabolize N-glycans in a potentially large variety of biological settings.

Furthermore, due to the rise of antibiotic resistance and the lack of effective vaccines, the threat presented by *S. pneumoniae* is greater than ever. Undeniably, this worldwide pathogen is capable of evading our presently used clinical interventions, highlighting the need to gain more insight into how *S. pneumoniae* interacts with its host. Since carbohydrate-processing systems are critical for the lifestyle and virulence of *S. pneumoniae*, the anticipated outcome of this research would at least give a greater

understanding of the host-pathogen interaction and may potentially uncover possible drug targets for the development of alternative therapeutics.

Chapter 2: Characterisation of proteins involved in N-glycan release and import and their implication in pneumococcal virulence

Melissa Cid[1], Shireen Woodiga[2], Samantha King[2], Hasan Yelsilkaya[3] and Alisdair B. Boraston[1]

[1]Biochemistry & Microbiology, University of Victoria, PO Box 3055 STN CSC, Victoria, BC, V8W 3P6, Canada

[2]Center for Microbial Pathogenesis, The Research Institute at Nationwide Children's Hospital, Columbus, OH 43205, U.S.A.

[3]Department of Infection, Immunity & Inflammation, University of Leicester, 10 Leicester, LE1 9HN, UK

Contributions: I performed the cloning, protein production, purification, crystallization, structure refinement and ITC experiment for $SBP_{N\text{-}glycan}$. The UV difference assay was performed by Alisdair B. Boraston. The *S. pneumoniae* mutant strains, growth assay, RNA preparation and RT-PCR experiment were completed by Shireen Woodiga and Samantha King. The mouse virulence assays were performed by Hasan Yelsilkaya.

2.1 Introduction

Among the pneumococcal carbohydrate-processing proteins potentially implicated in virulence are EndoD and a novel ABC transporter of unknown specificity (Hava and Camilli, 2002). Genes encoding these proteins are found genomically associated with the CPL. Indeed, homologues of these genes co-occur within or in proximity of this putative N-glycan-processing locus, in a large number of firmicutes, suggesting a functional association (Figure 7). Based on this bioinformatic analysis, we hypothesise that these two additional extracellular components are responsible for the liberation and transport of N-glycans from host glycoconjuguates. More specifically, we hypothesize that the N-glycan breakdown products obtained by EndoD are recognised by the lipid-anchored Solute Binding Protein ($SBP_{N\text{-}glycan}$) of the ABC transporter ($ABC_{N\text{-}glycan}$), permitting their uptake. EndoD was identified in a STM virulence screen study in TIGR4 (Hava and Camilli, 2002) as well as in a large-scale microarray study (Orihuela et al., 2004), while $ABC_{N\text{-}glycan}$ was only identified in one (Orihuela et al., 2004). This suggests that both components could be involved in virulence but more specifically that the release of N-

glycans by EndoD from host glycoconjuguates, and the transport of these sugars by this putative $ABC_{N\text{-}glycan}$, could be processes important for pathogenesis. Yet, why the release and import of N-glycan is important for virulence remains unknown.

The LPXTG-cell wall anchored EndoD has previously been shown to cleave the di-N-acetylchitobiose structure of N-glycans, and its activity has been demonstrated on $Man_3GlcNAc_2$-Asn and $Man_5GlcNAc_2$-Asn irrespective of whether the first GlcNAc bears a fucose (Yamamoto et al., 2005). Any modification of these structures led to a loss of EndoD activity, suggesting that other pneumococcal enzymes must trim down complex and high mannose N-glycan prior to the action of EndoD. This enzyme belongs to the family GH85 whose members are all endo-β-N-acetylglucosaminidases that are active on complex and high mannose N-glycans. In addition to cleaving the chitobiose core of N-glycans, experimental evidence suggests that certain members of this family catalyse transglycosylation reactions, rendering this group of enzymes as interesting targets for glycoprotein bioengineering (Li et al., 2006). GH85s catalyses hydrolysis through a substrate-assisted retaining mechanism. A complex structure with the inhibitor NAG-thiazoline, combined with NMR spectroscopy data, suggest that EndoD uses such a mechanism for catalysis (Abbott et al., 2009).

$ABC_{N\text{-}glycan}$ is assembled from three gene products that encode for an SBP and two permeases. No gene encoding for an ATPase has been identified in proximity of the SBP/permeases genes, which led us to think that the promiscuous ATPase msmK might be energising this putative transporter. In *S. pneumoniae,* the $SBP_{N\text{-}glycan}$ is a lipoprotein attached to the membrane through a lipo-box and will be the focus of this structural and biochemical study (Pérez-Dorado et al., 2012). SBPs are the only recognition site of the transporter and therefore dictate their specificity (Berntsson et al., 2010). Despite low sequence identities, the overall structures of SBPs are quite conserved and consist of two domains separated by a hinge region, which provides the flexibility necessary for the protein to open and close. Ligand binds between the two domains and induces a conformational change to a closed conformation through a process described as a Venus-fly trap mechanism (Mao et al., 1982). Binding of the liganded-SBP to the permeases

subsequently triggers the permease translocation channel to open and accept the ligand. Based on available crystal structures of SBPs, four different states are found: open-liganded, open-unliganded, closed-liganded and closed-unliganded (Berntsson et al., 2010; Trakhanov et al., 2005). However, in the absence of ligand SBPs are mainly found in the open conformation, and in the presence of ligand they are found primarily in the closed conformation with the ligand trapped. SBPs recognize a wide range of ligands, from metal ions to large oligosaccharides or even oligopeptides, and usually have low μM range affinity for their ligand.

To bridge the gap in knowledge regarding N-glycan processing by bacteria and uncover a potential novel transport mechanism specific to these sugars, we investigated both EndoD and $ABC_{N\text{-}glycan.}$ We speculate that EndoD is required for the release of N-glycans from glycoconjuguates through cleavage of the chitobiose core, and hypothesise that $SBP_{N\text{-}Glycan}$ specifically binds to N-glycan products permitting their uptake.

In the present chapter, we explore the function and mechanistic details behind the role of $SBP_{N\text{-}Glycan}$ and EndoD from *S. pneumoniae*. We provide compelling evidence to support the novel specificity of $SBP_{N\text{-}Glycan}$ for the reducing end of N-glycans, and show insight into the molecular mode of recognition of $SBP_{N\text{-}Glycan}$ for its ligand(s). Furthermore, we demonstrate the contribution of both $ABC_{N\text{-}Glycan}$ and EndoD to growth of *S. pneumoniae* on fetuin and explore their implication in virulence.

2.2 Material and methods

Materials

$Man_1GlcNAc$ was obtained from Toronto Research Chemicals (Toronto, ON, CA) and $Man_5GlcNAc$ was obtained from IsoSep (Tullinge, Sweden).

Bacterial strains, culture media, and chemicals

Wild-type (WT) and genetically modified strains of *S. pneumoniae* utilized in this study are described in Table 2. Broth cultures were routinely grown at 37 °C in Todd-Hewitt broth (Becton, Dickinson, and Co., Sparks, MD) supplemented with 0.2 % w/v yeast extract (Becton, Dickinson, and Co.) (THY). C media with 5 % yeast extract (C+Y) pH 8 was used for transformations (Lacks and Hotchkiss, 1960). *S. pneumoniae* was also

grown at 37 °C and 5 % CO_2 overnight on tryptic soy (TS) (Becton, Dickinson, and Co.) plates with 1.5 % agar that were spread with 5000 units of catalase (Worthington Biochemical Corporation, Lakewood, NJ) and TS plates supplemented with 5 % sheep blood (Becton, Dickinson, and Co.). Bacteria were selected on TS plates containing streptomycin (200 μg mL^{-1}) or kanamycin (500 μg mL^{-1}). Unless otherwise specified, all chemicals, substrates, and enzymes were purchased from Sigma Chemicals (St. Louis, MO).

Table 2. Strains used in this study

Strain Name	Serotype	Characteristics/genotype[a]	Source or Reference
TIGR4 Sm^r	4	Lys→Thr in RpsL [*rpsL*(*K56T*)] conferring Sm^r	Bender *et. al.* (2006)
TIGR4 Δ*endoD*	4	Δ*endoD rpsL*(*K56T*) (Sm^r)	This study
TIGR4 Δ*endoD endoD*$^+$	4	Lys→Thr in RpsL [*rpsL*(*K56T*)] conferring Sm^r	This study
TIGR4 ΔSP_0090-92	4	ΔSP_0090-92 *rpsL*(*K56T*) (Sm^r)	This study
TIGR4 ΔSP0090-92 SP0090-92$^+$	4	Lys→Thr in RpsL [*rpsL*(*K56T*)] conferring Sm^r	This study
TIGR4 ΔSP0090-92 Δ*endoD*	4	Δ*scrT rpsL*(*K56T*) (Sm^r)	This study
TIGR4 ΔSP0090-92 Δ*endoD* SP0090-92 *endoD*$^+$	4	Lys→Thr in RpsL [*rpsL*(*K56T*)] conferring Sm^r	This study
1121 Sm^r	4	Lys→Thr in RpsL [*rpsL*(*K56T*)] conferring Sm^r	Marion *et. al.* (2009)
1121 Δ*endoD*	4	Δ*endoD rpsL*(*K56T*) (Sm^r)	This study
1121 Δ*endoD endoD*$^+$	4	Lys→Thr in RpsL [*rpsL*(*K56T*)] conferring Sm^r	This study
1121 ΔSP_0090-92	4	ΔSP_0090-92 *rpsL*(*K56T*) (Sm^r)	This study
1121 ΔSP0090-92 SP0090-92$^+$	4	Lys→Thr in RpsL [*rpsL*(*K56T*)] conferring Sm^r	This study
1121 ΔSP0090-92 Δ*endoD*	4	Δ*scrT rpsL*(*K56T*) (Sm^r)	This study
1121 ΔSP0090-92 Δ*endoD* SP0090-92 *endoD*$^+$	4	Lys→Thr in RpsL [*rpsL*(*K56T*)] conferring Sm^r	This study

a. Sm^r indicates resistance to streptomycin

Construction of deletion mutants

All mutants were constructed using the Janus cassette selection system (Sung et al., 2001). This method requires two rounds of transformation. The first introduced a Janus cassette containing genes encoding for kanamycin resistance and streptomycin sensitivity (*rpsL*$^{+}$) into the genome of streptomycin-resistant (Sm^{r}) *S. pneumonia*e in place of the region to be deleted. DNA fragments flanking the region to be deleted were amplified (primers 1 and 2 and primers 4 and 6) (Table 3) and sequentially joined to the Janus cassette PCR product (primers J.F. and J.R.) using a variation of splicing by overlap extension (SOE) PCR (Burnaugh et al., 2008), first described by Horton *et al.* (Horton et al., 1989). All genomic DNA was prepared as previously described (Whatmore et al., 1999). A high-fidelity proofreading polymerase (Phusion; New England Biolabs) was used to minimize PCR-generated errors. The Janus construct was transformed into *S. pneumoniae* and the transformants were selected for on kanamycin and confirmed by PCR using primers 7 and 8, which flank the mutant construct.

The second round of transformation replaced the Janus cassette with an engineered segment of DNA consisting of the fragments flanking the deleted region. The fragments were generated using primers 1 and 3 (upstream fragment) and 5 and 6 (downstream fragment) and spliced together via SOE. The unmarked mutants were confirmed with primers flanking the construct (primers 7 and 8) and sequencing. The genetically reconstituted strains were generated by transforming the final mutants with the corresponding Janus construct and then subsequently with parental PCR products. Genetic reconstitution was confirmed by primers flanking the construct (7 and 8).

Table 3. Primers used for mutant construction

Group	No.	Primer Sequence (5' → 3')[a]	Location (accession no.)[b]
Janus	J.F	CCGTTTGATTTTTAATGGATAATG	7-30 (AY334019)
	J.R	GGGCCCCTTTCCTTATGCTT	247511-247527 (AE005672)
endoD	E.1	GTAAACCGTCTCCCATGGTCTTGATG	476146-476171 (AE005672)
	E.2	CATTATCCATTAAAAATCAAACGGAGCAGTT TCAGATGCTGTCGCTAC (1)	476778-476801 (AE005672)
	E.3	AGCAGTTTCAGATGCTGTCGCTACTC	476778-476801 (AE005672)
	E.4	AAGCATAAGGAAAGGGGCCCCCTAAATTGG AAGTTCAAGAG (2)	481098-481118 (AE005672)
	E.5	GAGTAGCGACAGCATCTGAAACTGCTCCTA AATTGGAAGTTCAAGAG (3)	481098-481118 (AE005672)
	E.6	TTATCAGATTCTTCTTTCACACGTCC	481778-481803 (AE005672)
	E.7	ACTGTTACACCAGAACTAGACTG	476025-476047 (AE005672)
	E.8	TCCAATTCAGCACCACGAGTGAC	481877-481896 (AE005672)
	E.9	CACTCGTTACGAAGATGTTGA	481948-481968 (AE005672)
	E.10	GATTTCGATACCTTGTGCTTTG	482296-482317 (AE005672)
$abc_{N\text{-}glycan}$	T.1	AGCCAGTCAGCAACTAAG	92401-92418 (AE005672)
	T.2	CATTATCCATTAAAAATCAAACGGAAGAAT AGAATCAACCAGAG (1)	93065-93084 (AE005672)
	T.3	TCACGAACGAACCAAGTATCAAGAATAGAA TCAACCAGAG (4)	93065-93084 (AE005672)
	T.4	AAGCATAAGGAAAGGGGCCCGATACTTGGT TCGTTCGTGA (2)	95948-95967 (AE005672)
	T.5	GATACTTGGTTCGTTCGTGA	95948-95967 (AE005672)
	T.6	TTCGTAGGCACCTTCAGAT	95948-95967 (AE005672)
	T.7	TGCTTCCCTAGAGTCCATT	92376-92394 (AE005672)
	T.8	CTACACTAGTATACCTGCTT	96801-96820 (AE005672)
	T.9	CCGTATCCTAGTCGCTGAC	96980-96998 (AE005672)
	T.10	CCTTGTTATCACGGACCCATT	97366-97386 (AE005672)

[a] Underlining indicates reverse complement sequence of primer J.F (1) J.R (2) E.3 (3) and T.5 (4). [b]Locations are given as nucleotide positions in the indicated accession numbers.

Growth assays

Chemically defined media (CDM) was prepared essentially as previously described (Kloosterman et al., 2006), without the addition of carbohydrate. The CDM buffer was made at 2.5 times concentration to allow addition of sufficient carbohydrate to support bacterial growth. The medium was supplemented with no sugar, glucose (12 mM), fetuin (20 mg ml^{-1}), GlcNAc (12 mM), galactose (12 mM), mannose (12 mM) and *N*-acetylneuraminic acid (12 mM). *S. pneumoniae* strains were grown in THY to an optical density at 600 nm (OD_{600}) = 0.300 ± 0.005, 2 mL aliquots were washed and resuspended in 130 µL of PBS and 130 µL of catalase (50,000 U mL^{-1}). Twenty microliter aliquots of bacterial suspensions or PBS/catalase (no bacteria control) were transferred to 180 µL of CDM supplemented with the appropriate carbon source. Medium supplemented with glucose served as a positive control, demonstrating in each experiment that bacteria were viable and that mutant strains showed no general growth defect relative to their parent strain. Medium with no added carbohydrate served as a negative control. Growth plates were incubated at 37 °C for 60 h in a BIO-TEK Synergy HT plate reader and the OD_{600} were read every 20 min. All data were adjusted to a path-length of 1 cm. As no increase in optical density was observed for any bacterial strain on no-carbohydrate medium, at each time point the average of results from triplicate wells was subtracted from results from all other wells containing the same bacterial strain. Data from at least three independent experiments were averaged, and the 95% confidence interval was calculated for each time point.

RNA preparation and reverse transcriptase RT-PCR

RNA was isolated by an acid-phenol extraction, with modifications for *S. pneumoniae*. Bacteria were grown in THY until samples reached OD_{600} of 0.3 ± 0.005. A 10 mL culture was combined with 10 mL acid phenol and 100 µL 10 % SDS and incubated at 90°C until the phases merged. Samples were then cooled on ice and centrifuged at 4 °C for 20 min at 3200 x g. Two additional extractions were performed, first with equal volumes of 1:1 acid phenol:choloroform and subsequently with chloroform. RNA was precipitated with 10 mL isopropanol and 1 mL of 3 M sodium acetate at -20 °C

overnight. RNA was concentrated and washed twice with 70 % ethanol and then vacuum dried. The resulting RNA was resuspended in 100 µL H_2O, quantified, and adjusted to 100 µg μL^{-1} to allow for a final reaction volume of 50 µL for DNase treatment. Nucleic acid was incubated with 5 µL DNase I buffer, 3 µL DNase I (New England Biolabs) and 1 µL Super RNaseIN (Promega) at 37 °C for 1 hr. Cleanup of RNA was performed with a Qiagen RNeasy mini kit as per manufacturer's instructions. DNase- and RNase-free reagents were used throughout. cDNA was generated with SmartScribe reverse transcriptase according to the manufacturer's recommendations (New England Biolabs). Parallel samples were processed without the addition of reverse transcriptase as a negative control.

The Janus system of mutant generation is designed to generate unmarked nonpolar mutants; however, as the genes encoding $ABC_{N\text{-}glycan}$ (SP0090-92) and *endoD* (SP0498) have genes downstream which may be in the same operon, we demonstrated that the mutations had no effect on the transcription of the distal gene by reverse transcriptase RT-PCR. cDNA was amplified with primers designed within the gene distal to the mutation: primers E.9 and E.10 for the *endoD* mutant (SP0499) and primers T.9 and T.10 for the $ABC_{N\text{-}glycan}$ mutant (SP0095).

In vivo analysis of pneumococcal strains

In vivo analysis of the pneumococcal strains utilized ten-week-old female MF1 outbred mice (Charles River, Margate, United Kingdom). For infections a standard inoculum was prepared for each strain. Briefly, pneumococci were grown overnight in Brain heart infusion (BHI), pelleted by centrifugation, and re-suspended in PBS (pH 7.0). A 100 mL aliquot of this suspension was administered into the peritoneal cavity. Once the animals started displaying signs of disease, blood was collected by cardiac puncture under deep anesthesia to inoculate 10 mL BHI. After overnight growth, the bacterial pellet was used to inoculate BHI containing 20% calf serum, and the growth was allowed to continue until late exponential phase (OD_{500} 1.4-1.6). At this stage, growth was ceased, and the aliquots of bacteria were stored at -80 ºC until required.

To assess the virulence of pneumococcal strains, mice were lightly anesthetized with 3% (vol/vol) isoflurane over oxygen, and an inoculum of 50 µL containing approximately 1 X 10^6 CFU in PBS was given drop-wise into the nostrils as described previously (Yesilkaya *et al.*, 2009, Hajaj *et al.*, 2012). After infection, the inoculum dose was confirmed by viable counting on blood agar plates. Mice were monitored for disease signs (progressively starry coat, hunched, and lethargic) for 7 days. When mice were lethargic, they were culled. Therefore, time to reach lethargic state has been defined as the "survival time." Mice that were alive 7 days after infection were deemed to have survived the infection.

To monitor the development of bacteremia in each mouse, approximately 20 mL of venous blood was obtained from mice at predetermined time points after infection, and viable counts were determined, as described above.

Survival times were calculated by using GraphPad Prism software and analyzed by the Mann-Whitney U test. Bacterial counts in blood were analyzed by an analysis of variance followed by the Bonferroni post-test. Statistical significance was considered to be a *P* value of <0.05.

Ethics statement

Mouse experiments at the University of Leicester were performed under appropriate project (permit no. 60/4327) and personal (permit no. 80/10279) licenses according to the United Kingdom Home Office guidelines and local ethical approval. Where appropriate, the procedures were carried out under anesthetization with isoflurane. Mice were kept in individually ventilated cages in a controlled environment and were frequently monitored after infection, to minimize suffering.

Cloning

$SBP_{N\text{-}glycan}$ is predicted to be a cell-membrane-attached lipoprotein therefore the cloned *sbp* gene was designed to lack the N-terminal secretion SP and lipobox in order to produce a soluble recombinant protein. The gene fragment encoding amino acids 24-491 of $SBP_{N\text{-}glycan}$ (SP0092) was PCR-amplified from *S. pneumoniae* TIGR4 genomic DNA using the oligonucleotide primers SBP-Fw (5'-

CATATCCCATGGGAAATTTGACAGGTAACAG-3') and SBP-Rv (5'-GGTGGTCTCGAGCTTTTTGTTTTTCAAGAATTCATC-3'). The amplified DNA fragment was cloned using standard procedures into a pET 28a plasmid (Novagen) using 5′ NcoI and 3′ XhoI restriction sites. The resulting plasmid, pSBP, encoded the desired polypeptide fused to a C-terminal six-histidine tag. The DNA sequence fidelity of all constructs was verified by bidirectional sequencing.

Protein Production and Purification.

pSBP was transformed into the expression strain, *Escherichia coli* BL21 Star (DE3). Six liters of YT broth, containing 50 μg/mL kanamycin, was inoculated with the transformed cells and incubated at 37 °C. Once an optical density of 0.6 at 595 nm was reached, protein production was induced by the addition of isopropyl 1-thio-β-d-galactopyranoside to a final concentration of 0.5 mM. Incubation of the cultures was continued overnight with shaking at 16 °C. Cells were harvested by centrifugation at 5000 x g for 10 minutes and ruptured by chemical lysis. The cleared supernatant of the cell lysate was loaded onto a Ni^{2+}-NTA immobilized metal affinity chromatography column. Polypeptide was eluted with binding buffer (20 mM Tris, pH 8.0) containing increasing concentrations of imidazole (0–500 mM). $SBP_{N\text{-}glycan}$ was concentrated and buffer-exchanged into 20 mM Tris, pH 8.0, in a stirred ultrafiltration unit (Amicon). Purified protein was further purified by size exclusion chromatography using a Sephacryl S-200 column (GE Healthcare) and concentrated using a stirred-cell ultrafiltration device (Amicon) with a 10,000 Da molecular weight cut-off membrane (Millipore).

Determination of Protein Concentration.

Protein concentration was determined by UV absorbance at 280 nm using the calculated extinction coefficient of 60280 M^{-1} cm^{-1}.

Binding analysis

Quantitative UV difference binding studies of $SBP_{N\text{-}glycan}$ were performed as described previously (Boraston et al. 2001) and the data analysed using a bimolecular binding model that accounted for ligand depletion. Difference spectra were analysed for peak and

trough wavelengths, and values at the appropriate wavelengths extracted for further analysis. The peak-to-trough heights at the wavelength pairs 289.3/301.0 nm, 289.3/294.5 nm, and 282.8/287.4 nm were calculated by subtracting the trough values from the peak values, and the dilution-corrected data were plotted against total carbohydrate concentration. Data for the three wavelength pairs were analyzed simultaneously with MicroCal Origin (v.7.0). The data reported are the averages and standard deviations of three independent titrations.

ITC was performed as previously described, with a VP-ITC (MicroCal, Northampton, MA) (Boraston et al., 2001). Briefly, protein samples were extensively dialyzed against buffer (20 mM Tris, pH 8.0), and then concentrated a stirred cell ultrafiltration device with a 5000 MWCO membrane (Millipore). Sugar solutions were prepared by mass in buffer saved from the ultrafiltration step. Both protein and sugar solutions were filtered and degassed immediately prior to use. The binding analysis with $Man_5GlcNAc$ was performed with a protein concentration of 20 μM and sugar concentration of 0.5 mM giving C value of ~20. Titrations were performed in triplicate at 25 °C. The data were fit with a single binding site model to determine K_a, n, and ΔH. ΔS was calculated using $\Delta G=\Delta H-T\Delta S$. Errors represent the standard deviations of the triplicate determinations.

Crystallization, data collection, and structure determination of SBP derivative

Crystals were obtained at 18^o C, using sitting-drop vapour diffusion for screening and hanging drop vapour diffusion for optimization. $SBP_{N\text{-}glycan}$ crystals were grown in 29% (w/v) PEG 1500, 11% (w/v) MPD and 0.1 M Tris, pH 8.5 at a concentration of 15 mg/mL. The $SBP_{N\text{-}glycan}$ structure was determined by first soaking a single crystal in the crystallisation solution containing 500 mM NaI for 30 seconds followed by 1 minute in the crystallisation solution containing NaI at 1 M. Since the soaking solution contained a sufficient amount of MPD, crystals were directly flash cooled with liquid nitrogen for data collection. Diffraction data were collected on a ‘‘home-beam’’ comprising a Rigaku R-AXIS 4++ area detector coupled to a MM-002 X-ray generator with Osmic ‘‘blue’’ optics and an Oxford Cryostream 700. Diffraction data were processed using MOSFLM and SCALA (Collaborative Computational Project, 1994; Powell, 1999). Phases were generated with auto-SHARP using the single-wavelength anomalous dispersion (SAD)

method (Sheldrick, 2007). Auto-SHARP determined the substructure of 20 I atoms, which had occupancies ranging from 1.0 to 0.0788 translating to a phasing power of 0.888 and a figure of merit of 0.34773 for acentric and 0.16097 for centric reflections. Autobuilding with ARP/wARP using the phases generated from auto-SHARP resulted in a nearly complete model (Emsley and Cowtan, 2004). Manual completion and refinement of the model was performed in Coot and Refmac (Emsley et al., 2010; Murshudov et al., 1997). In all data sets, refinement procedures were monitored by flagging 5% of all observations as ‘‘free’’ (Brunger, 1992). Model validation was performed with MOLPROBITY (Davis et al., 2007).

Crystallization, data collection, and structure determination of $SBP_{N\text{-}glycan}$ *complex structures.*

Both complex structures were obtained by co-crystallisation. The co-structure with $Man_1GlcNAc$ was obtained by co-crystallising the protein at 10 mg/mL in the presence of 5 mM of $Man_1GlcNAc$. For the structure in complex with $Man_5GlcNAc$, the protein (15 mg/mL) was co-crystallised in the presence of 1mM of $Man_5GlcNAc$. Co-crystals were grown in 23% (w/v) PEG 3350, 0.18 M sodium bromide and 0.1 M sodium citrate pH 5.5 for the $Man_1GlcNAc$ complex and 30% (w/v) PEG 400, 0.1 M cadmium chloride hydrate, 0.1 M sodium acetate trihydrate, pH 4.6, for the co-structure with $Man_5GlcNAc$. A data set of SBP/$Man_1GlcNAc$ was collected on a single crystal that was cryoprotected in the crystallisation solution supplemented with 25% (w/v) ethylene glycol while the crystallisation solution of SBP/$Man_5GlcNAc$ co-crystals was already cryoprotectant. Diffraction data were collected either on the ‘‘home-beam’’ comprising a Rigaku R-AXIS 4++ area detector coupled to a MM-002 X-ray generator with Osmic ‘‘blue’’ optics and an Oxford Cryostream 700 and at beamline 9-2 of the Stanford Linear Accelerator Center (SLAC, Stanford Synchrotron Radiation Lightsource [SSRL], CA) as indicated in Table 4. All diffraction data were processed using MOSFLM and SCALA (Collaborative Computational Project, 1994; Powell, 1999). The $SBP_{N\text{-}glycan}$ derivative structure was used as a search model to solve both complex structures using PHASER. To account for the conformational change induced by ligand binding of the SBP, only one domain of the SBP derivative structure was used as a search model. All data

collection and processing statistics are shown in Table 4. For all structures, manual model building was performed with COOT (Emsley and Cowtan, 2004) and refinement of atomic coordinates was performed with REFMAC (Murshudov et al., 1997). The addition of water molecules was performed in COOT with FINDWATERS and manually checked after refinement. In all data sets, refinement procedures were monitored by flagging 5% of all observations as ‘‘free’’ (Brunger, 1992). Model validation was performed with MOLPROBITY (Davis et al., 2007).

Table 4. Data collection and refinement statistics for $SBP_{N\text{-}Glycan}$

	SBP-NaI	SBP-Man1	SBP-Man5
Data Collection statistics			
Beamline	MM-002	SSRL BL9-2	SSRL BL9-2
Wavelength	1.54187	1.04007	0.98005
Space group	$P12_11$	P1	$P12_11$
Cell dimensions: *a, b, c* (Å)	60.91, 40.20, 93.89	39.99, 50.47, 107.89	104.18, 11.60, 105.84
Resolution (Å)	21.80-2.10 (2.21-2.10)*	45.95-3.00 (3.16-3.00)*	48.37-1.73 (1.82-1.73)*
R_{merge}	0.126 (0.503)*	0.137 (0.459)*	0.036 (0.397)*
I/σI	19.5 (5.1)*	6.6 (2.3)*	25.4 (4.0)*
Completeness (%)	99.8 (99.6)*	97.0 (97.0)*	99.9 (99.9)*
Redundancy	10.9 (10.5)*	2.2 (2.1)*	4.7 (4.6)*
Refinement			
Total reflections	291180 (40192)*	35594 (5086)	1095333 (158575)*
Unique reflections	26624 (3833)*	16399 (2413)*	234916 (34209)*
R_{work}/R_{free}	21.8/26.5	26.6/32.5	19.5/23.0
No. of atoms			
Protein	3491	6474	14567
Ligands	N/A	52	279
Solvent	226 (H2O)	10 (H2O)	1107 (H2O)
			33 (CD)
Average B-factors ($Å^2$)			
Protein	21.13	33.88	29.32
Ligands		27.90	19.06
Solvent	19.3587 (H2O)	11.822 (H2O)	31.54 (H2O)
	39.1114 (IOD)		34.25 (CD)
Root mean square deviations			
Bond lengths (Å)	0.0088	0.0053	0.0183
Bond angles (degrees)	1.1468	1.0526	1.6718
Ramachandran			
Preferred (%)	96.2 (472)	90.2 (395)	97.6 (481)
Allowed (%)	3.8 (19)	9.7 (42)	2.2 (11)
Disallowed (%)	0 (0)	0.1 (1)	0.2 (1)

*Values in parentheses are for highest resolution shell**

2.3 Results

$SBP_{N\text{-}glycan}$ interacts tightly with N-glycans

We performed a functional and structural analysis of $SBP_{N\text{-}glycan}$ using a recombinant overproduced soluble form of the protein that lacks the secretion signal peptide and lipid-anchoring motif. Binding analysis by isothermal titration calorimetry (ITC) was performed with $Man_5GlcNAc$ and the stoichiometry of binding was determined to be 1.05±0.01. The affinity constant (K_a) of (1.04± 0.01)$\times 10^6$ M^{-1} obtained is consistent with the range of affinities previously determined for other SBPs (Figure 8A) (Abbott and Boraston, 2007; Higgins et al., 2009; Quiocho et al., 1997; Suzuki et al., 2008). The change of enthalpy (ΔH) and entropy (ΔS) were determined to be -19.78 ± 0.01 kcal/mol and -38.8 ± 0.40 cal/mole/K respectively. The binding pattern determined by these thermodynamic values is in agreement with most protein-carbohydrate interactions (Higgins et al., 2009). This tight interaction, typical of SBPs for their ligands, reveals the specificity of $SBP_{N\text{-}Glycan}$ for $Man_5GlcNAc$ and implies that recognition of this oligosaccharide by $SBP_{N\text{-}Glycan}$ is biologically relevant. Additionally, a UV difference binding analysis was used with $Man_1GlcNAc$ and gave an affinity constant of 1.2 (± 0.1) $\times 10^4$ M^{-1}(Figure 8B), which is approximately two orders of magnitude lower than what was obtained for $Man_5GlcNAc$. Compared to other binding data available for SBPs, the K_a obtained for this sugar is 2 to 100 fold lower suggesting that this disaccharide may only be a fragment of the biological ligand (Berntsson et al., 2010).

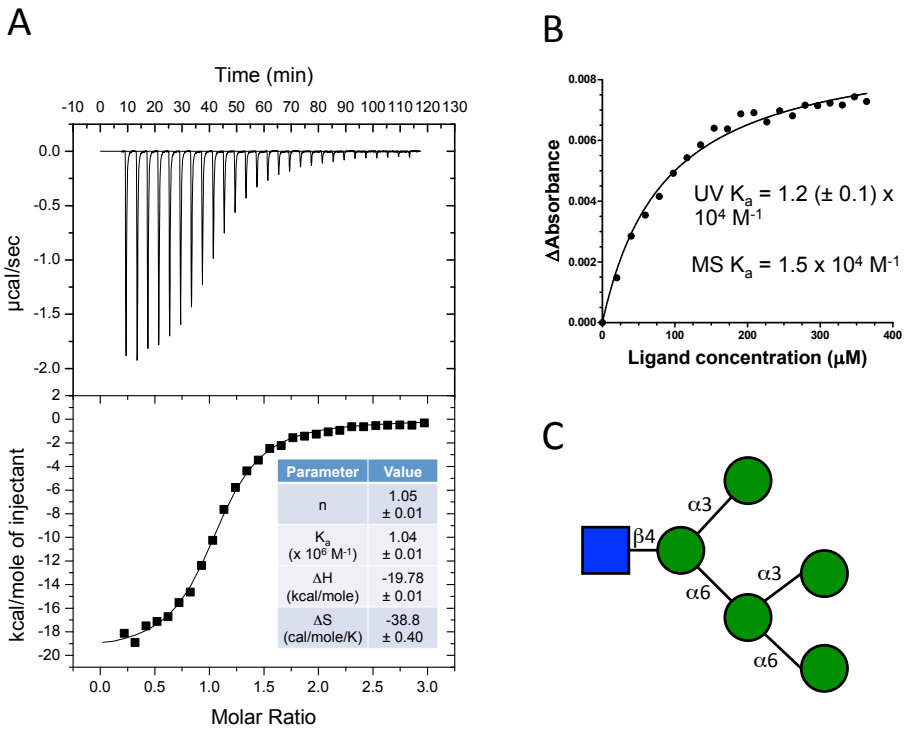


Figure 8. Binding analyses of $SBP_{Nglycan}$ with N-Glycans.

(A) ITC titration of $SBP_{N\text{-}glycan}$ with $Man_5GlcNAc$. (B) UV difference titration of $SBP_{N\text{-}glycan}$ with $Man_1GlcNAc$. (C) Structure of the $Man_5GlcNAc$ ligand that was assessed for binding by ITC.

Structural basis of N-glycan recognition

$SBP_{N\text{-}glycan}$ in its unliganded form was crystallised in the space group P1 with four protein monomers in the asymmetric unit. The structure was solved by single wavelength anomalous dispersion to a resolution of 2.1 Å using an iodine derivative (Table 4). The overall structure of $SBP_{N\text{-}glycan}$ is highly similar to other SBPs of known structure. It is composed of two α/β domains separated by a hinge region that connects the two domains with three loops from different segments of the protein sequence. The domains each consist of a central β-sheet of three β-strands flanked by α-helices (Figure 9A). The N-terminal domain possesses an additional β-sheet consisting of two β-strands that is exposed to the solvent. The strands of the N-terminal domain are numbered 1-2-3-7-8-9 and the C-terminal domain 4-5-6. The helices of the N-terminal domain are numbered 1-2-3-4-5-11-12-13-14-15-16 and 6-7-8-9-10-17-18-19 for the C-terminal domain. The polypeptide chain goes back and forth from one domain to the other three times to constitute the two globular domains that are separated by the flexible hinge region, which lacks secondary structure. In the unliganded crystal structure, $SBP_{N\text{-}Glycan}$ is trapped in an open state. The substantial degree of opening creates a deep and wide binding pocket but does not necessarily represent the average degree of opening in solution since it could be influenced by the crystal packing.

In order to gain insight into the mode of glycan recognition by $SBP_{N\text{-}Glycan}$, the crystal structure of the protein in complex with a disaccharide, $Man_1GlcNAc$, was obtained by molecular replacement at a resolution of 3.0 Å (Table 4). The ligand was modelled in an unambiguous electron density in the binding site constituted by the interface of the two α/β domains (Figure 9C). The complex structure with $Man_1GlcNAc$ reveals $SBP_{N\text{-}glycan}$ in its closed conformation with the terminal GlcNAc reducing-end tightly bound in the bottom of the binding pocket (Figure 9E). Asp 289 interacts with the equatorial O1 of GlcNAc and Trp 386 interacts with the first mannose residue through typical pyrannose ring-aromatic side chain interactions. However, another Trp (Trp81) is present in the binding pocket that does not interact with $Man_1GlcNAc$. The presence of this Trp 81, in addition to the large binding pocket observed in the closed-form $SBP_{N\text{-}glycan}$, suggest that

this protein could accommodate a larger sugar and is consistent with $Man_1GlcNAc$ only being a fragment of the biological ligand of $SBP_{N\text{-}Glycan}$.

To gain further insight into the specificity of $SBP_{N\text{-}glycan}$, we solved the crystal structure in complex with a hexasaccharide, $Man_5GlcNAc$, by molecular replacement at a resolution of 1.6 Å (Table 4). The branched oligosaccharide was modelled in a clear electron density found buried in the binding pocket with the two domains enveloping just about the entire branched sugar (Figure 9D, 9F). As seen in the complex structure with $Man_1GlcNAc$, the reducing end of $Man_5GlcNAc$ is also tightly accommodated in the depth of the binding pocket with Asp 289 interacting with the equatorial O1 of GlcNAc (Figure 10A), in the exact same position as with $Man_1GlcNAc$ (Figure 10B). Again, $SBP_{N\text{-}Glycan}$ does not make any direct hydrogen bond with the acetamido group but water-mediated hydrogen bonds are coordinating this GlcNAc chemical group. The recognition of the reducing end of the ligand suggests that extension of the sugar on the reducing end would prevent binding (Figure 9F, 10A). This in turn suggests that the chitobiose core of N-glycan must be cleaved before transport. The two aromatic residues, Trp 81 and Trp 386, create a hydrophobic cage with the ligand through classical pyranose ring-aromatic side chain interactions creating, a hydrophobic base on the N-terminus domain interface (Figure 10A). The co-structures obtained show that $SBP_{N\text{-}Glycan}$ only establishes a subset of the interactions with $Man_1GlcNAc$ compared to $Man_5GlcNAc$, which explains the lower affinity obtained for the disaccharide. All residues of the ligand form numerous putative hydrogen bonds with the binding pocket (Figure 10A). More specifically, the mannose-configured axial O2 of most mannose residues interact via hydrogen bond with the protein. Ligand binding is also mediated by indirect hydrogen bonds between the protein and an extensive water network (Figure 10D). This provides a molecular architecture for the recognition of the ligand three-dimensional structure as well as the stereochemistry of most sugar residues. $SBP_{N\text{-}Glycan}$ establishes a wide number of interactions with each carbohydrate residue and recognises specifically the non-reducing end of the branched N-glycan, which sits at the bottom of a deep binding pocket while almost completely enveloping the whole ligand at the heart of the protein (Figure 9E-F).

Ligand binding induced roughly a 14 Å shift conformational change from an open to a closed conformation (Figure 9B). Both complex structures are found in a closed conformation with an rsmd value of 1.3 Å, suggesting that the disaccharide was sufficient to induce the protein to close (Figure 10C). However, the degree of closure of the both ligand-trapped $SBP_{N\text{-}glycan}$ may be influenced by crystal packing.

$SPB_{N\text{-}glycan}$ displays a similar overall fold to most SBPs, which consist of two α/β domains with a central β-sheet of five β-strands flanked by α-helices. The hinge region can differ between SBPs, from three interconnecting segments between the two domains to having an extra domain that provides flexibility to the protein. A structural classification of SBPs based on structural features has been proposed by Bertnesson and colleagues where SBPs were grouped into six different clusters (A-F) (Berntsson et al., 2010). The main difference between each cluster is the structural organization of the hinge region. $SBP_{N\text{-}glycan}$ belongs to cluster B. Members of that cluster possess a hinge region created by three strands that allows cross-overs of the peptide chains between the two domains. This cluster comprises mainly carbohydrate specific-SBPs; however, peptides and autoinducer-specific-SBPs are also found. *Sphingomonas* sp. A1 has an alginate import system comprising an SBP that is structurally closely related to $SBP_{N\text{-}glycan}$. With only 16% sequence identity, an rsmd of 4.0 Å was calculated between the two proteins. Indeed, the overall fold of this alginate-binding protein is almost identical to $SBP_{N\text{-}glycan}$with the exception that it possesses an extra β-sheet exposed to the solvent and two antiparallel β-strands in the C-terminal domain. The size and depth of the cavity of both these SBPs is quite similar; however, the architecture of their binding pocket reveals critical differences. Overlay of the alginate binding protein with $SBP_{N\text{-}glycan}$ complexed with $Man_5GlcNAc$ reveals that a structural clash would occur between an extended loop and the α-1,3-branch of $Man_5GlcNAc$, suggesting that this SBP is not structurally designed for branched sugars. This demonstrates that besides having a similar overall fold and common binding-mechanism, the binding pocket architecture of SBPs is tailored to their specific ligand, which ultimately defines the specificity of their ABC transporter.

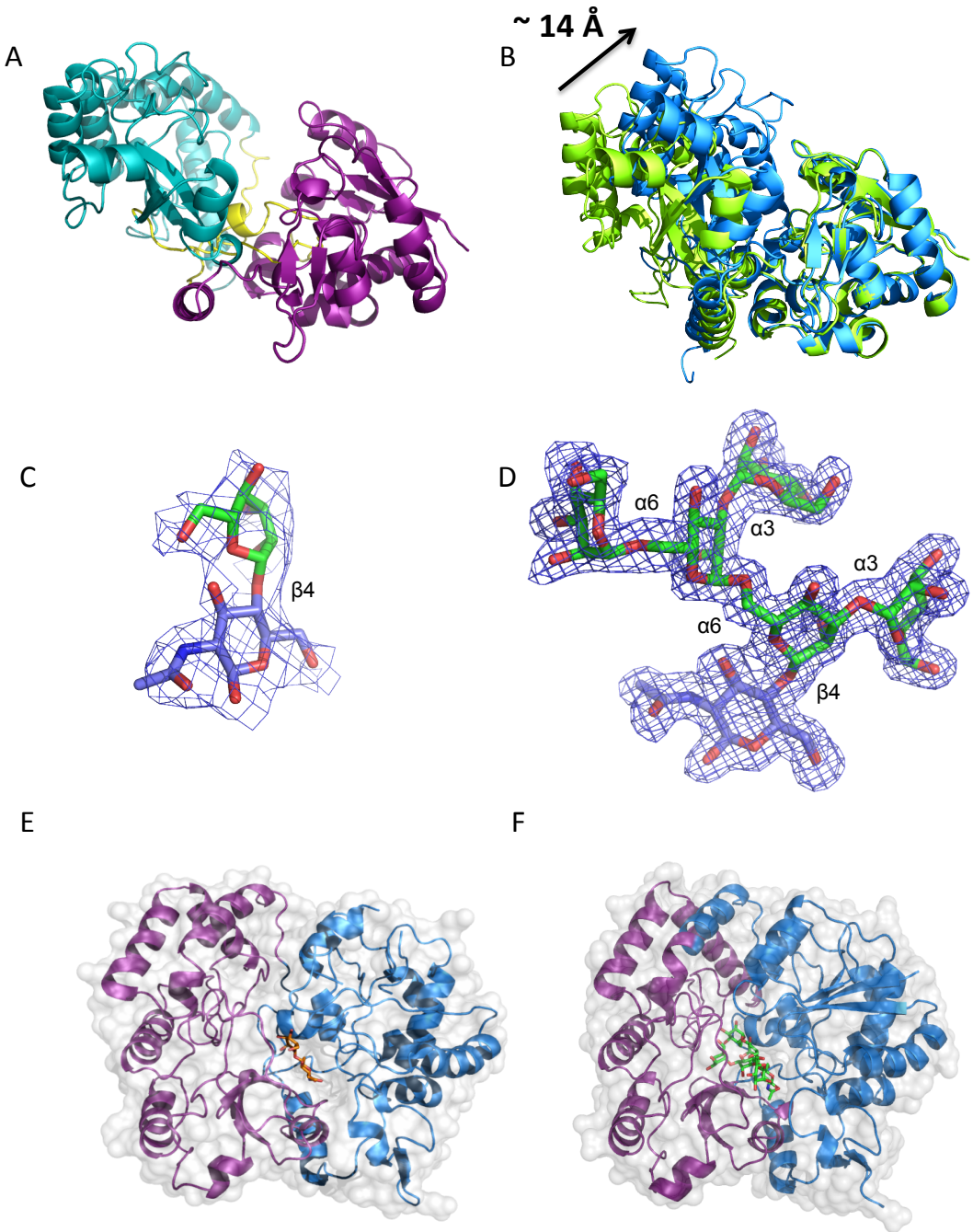


Figure 9. X-ray crystal structure of the SBP $_{N\text{-}Glycan}$ from *S. pneumoniae* TIGR4.
(A) Open form $SBP_{N\text{-}Glycan}$ showing the two domains (N-terminal domain in purple and the C-terminal domain in blue teal) separated by a hinge region represented in yellow, which creates the deep ligand-binding site. (B) Comparison of the open (in green) and

closed form (in marine blue) of $SBP_{N\text{-}Glycan}$. The sugars are found bound in the binding site formed by the interface of the N- and C-terminal domains. (C-D) The electron density of the bound ligands is shown in a σa-weighted $F_{obs}-F_{calc}$ maps generated at 1.5σ for $Man_1GlcNAc$ (C) and 2.0σ for $Man_5GlcNAc$ (D) by refinement of the $SBP_{N\text{-}Glycan}$ structures with the ligand coordinates omitted. The electron density is shown as blue mesh, while the ligands are shown in stick representation (blue, GlcNAc; green, Man). Co-structures of the $SBP_{N\text{-}Glycan}$ in complex with $Man_1GlcNAc$ (E) and $Man_5GlcNAc$ (F), the N-terminus domain is represented in purple and the C-terminus domain in blue, the solvent-accessible surface is represented in grey, the sugars are shown in stick representation, $Man_1GlcNAc$ in orange (E) and $Man_5GlcNAc$ is shown in green (F).

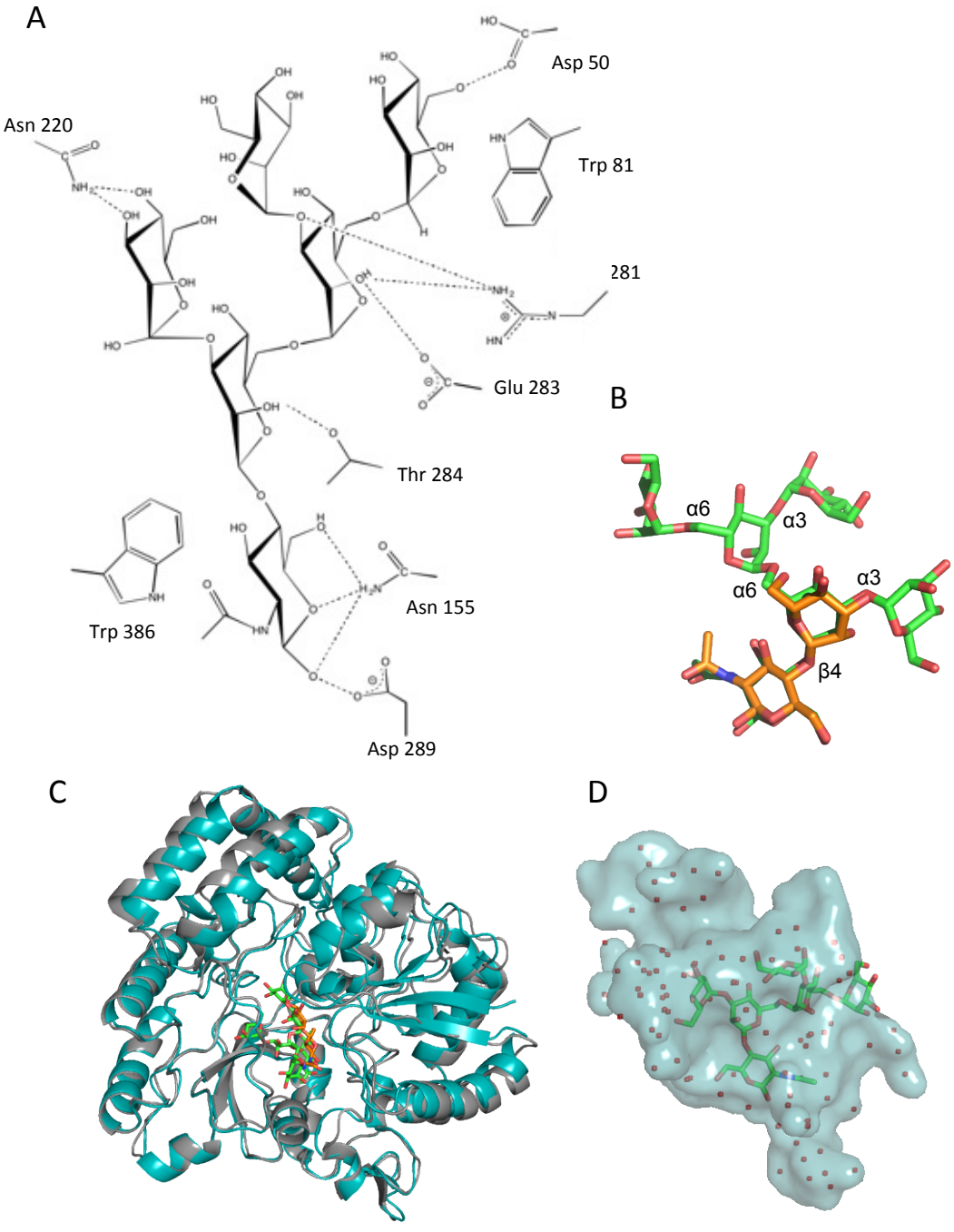


Figure 10. The structural basis of N-Glycan recognition by $SBP_{N\text{-}glycan}$.

(A) Hydrogen bonding schematic of the binding site. The direct interactions between the $SBP_{N\text{-}glycan}$ and $Man_5GlcNAc$ are represented, for clarity the water molecules interacting with the ligand via hydrogen bonding are not shown. (B) Overlay of $Man_5GlcNAc$

(represented as green sticks) and $Man_1GlcNAc$ (represented as orange sticks), both ligands are found in the same position in both structures. (C) Overlay of the closed form $SBP_{N\text{-}Glycan}$ co-structures with $Man_5GlcNAc$ ($SBP_{N\text{-}Glycan}$ cartoon represented in blue, $Man_5GlcNAc$ represented as green sticks) and $Man_1GlcNAc$ ($SBP_{N\text{-}Glycan}$ cartoon representation in grey, $Man_1GlcNAc$ represented as orange sticks). (D) The inverse solvent cavity of the binding site shows the deep and large size of the cavity in relation to $Man_5GlcNAc$. The extensive water network (waters shown as red spheres) shows the large number of waters involved with ligand binding (hydrogen bonds not represented for clarity).

EndoD and ABC $_{N\text{-}Glycan}$ contribute to growth on fetuin

Pneumococci are able to utilize monosaccharides released during sequential deglycosylation of complex N-linked glycans through the activity of pneumococcal glycosidases such as NanA, BgaA and StrH for growth (Burnaugh et al., 2008; King et al., 2006). It is thus logical that $Man_3GlcNAc$ released by EndoD would also contribute to pneumococcal growth and that $ABC_{N\text{-}glycan}$ would also be required. To address this hypothesis, deletions were constructed in *endoD*, *abc*$_{N\text{-}Glycan}$ and a double mutant was constructed lacking both loci. The growth of these strains was tested on the model glycoprotein fetuin, which bears complex N-glycan. A partial reduction in growth was observed for each of the mutants as compared to the parental and genetically reconstituted strains (Figure 11). There was no significant difference in growth of the parental strain as compared to any of the genetically reconstituted strains. The residual growth of the mutants on fetuin was expected as all strains retain the ability to release and utilize sialic acid, galactose and GlcNAc present in N-linked glycans through the action of NanA, BgaA and StrH. Furthermore, fetuin is also decorated with O-linked glycans that can be sequentially deglycosylated by *S. pneumoniae* (Marion et al., 2009).

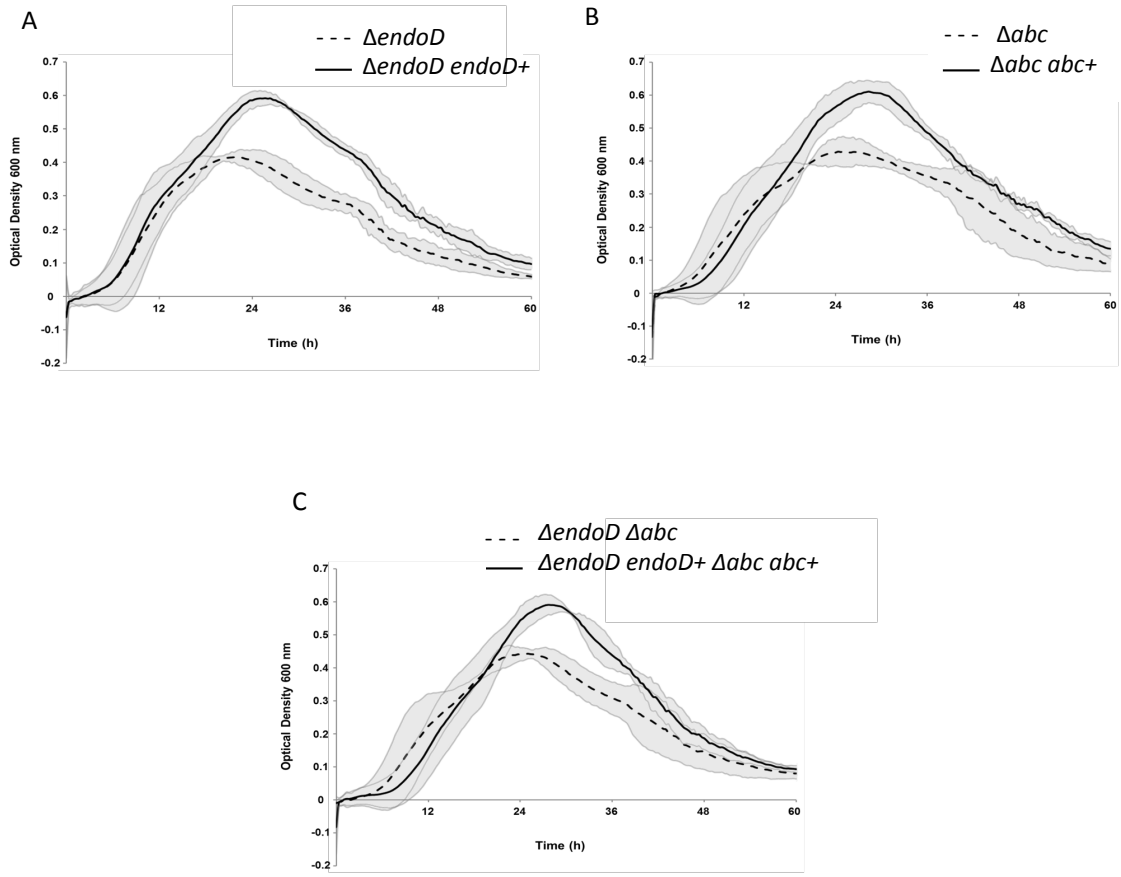


Figure 11. *endoD* and *ABC* $_{N\text{-}Glycan}$ contributes to growth on fetuin.

Growth of (A) *endoD* mutant and *endoD* genetically reconstituted strain, (B) $abc_{N\text{-}Glycan}$ mutant and $abc_{N\text{-}Glycan}$ genetically reconstituted strain and (C) *endoD* and $abc_{N\text{-}Glycan}$ double mutant and *endoD* and $abc_{N\text{-}Glycan}$ genetically reconstituted strain were grown for 60 h in CDM supplemented with 20 mg/ml of fetuin as a sole carbon source. Growth was measured as optical density at 600 nm. Data points are the means from three independent experiments performed in triplicate. Gray shading indicates a 95% confidence interval. There was no significant difference in growth profile of the genetically reconstituted strain when compared to the parental strain.

The growth of the *endoD*, $abc_{N\text{-}Glycan}$ and double mutants on fetuin were very similar with the maximal optical densities being 0.415 (Δ*endoD*), 0.442 (Δ$abc_{N\text{-}Glycan}$) and 0.428 (Δ*endoD*Δ$abc_{N\text{-}Glycan}$) (Figure 11A-C). To rule out that the substrate of $ABC_{N\text{-}Glycan}$ is a monosaccharide released by NanA, BgaA or StrH, or a monosaccharide constituent of $Man_3GlcNAc$, we tested the growth of the $abc_{N\text{-}Glycan}$ mutant strain on CDM supplemented with either 12 mM of *N*-acetylneuraminic acid, galactose, *N*-acetylglucosamine, or mannose. We saw no significant difference in growth of the $abc_{N\text{-}Glycan}$ mutant when compared to the parental and the genetically reconstituted strain (data not shown for clarity). The fact that all mutants show a similar reduction in growth supports the hypothesis that EndoD acts to release $Man_3GlcNAc$, which is transported by the $ABC_{N\text{-}Glycan}$ (Figure 12).

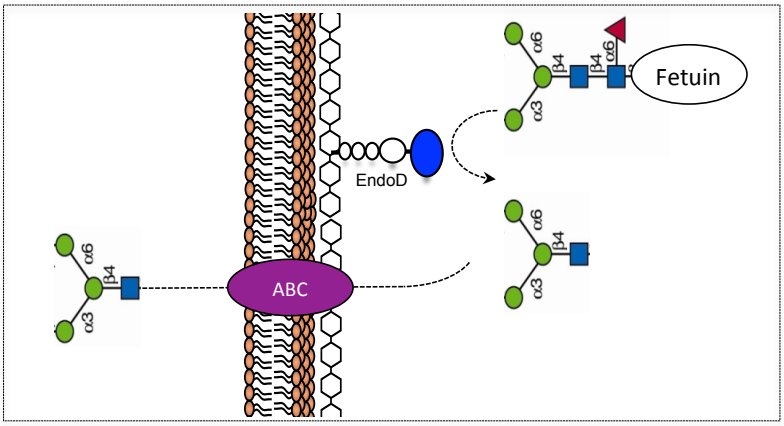


Figure 12. Schematic representation of $Man_3GlcNAc_2$ degradation and transport. EndoD is represented by white and blue pies; the $ABC_{N\text{-}Glycan}$ is represented as a purple oval.

EndoD and $ABC_{N\text{-}Glycan}$ contribution to virulence

The contributions of EndoD and $ABC_{N\text{-}Glycan}$ were tested individually, and in combination, in a mouse model of pneumococcal infection. The results showed that the median survival time of mice infected intranasally with *ΔendoD* (64 h ±33.2, n=20) was significantly longer than those of the WT (49 h ±37, n=20) and *ΔendoDΔendoD*$^{+}$ (48 h ±38.3, n=10) infected groups ($p<0.001$) (Figure 13A). Similarly, the deletion of both *endoD* and $abc_{N\text{-}Glycan}$ loci significantly reduced virulence as the cohort infected with the double *ΔendoDΔ*$abc_{N\text{-}Glycan}$ *c* mutant (85 h ±35.8, n=10) survived longer than those of the WT ($p<0.0001$) infected cohorts. However, while Δ*endoD*Δ$abc_{N\text{-}Glycan}$ infected group survived slightly longer that those of the Δ*endoD group* it was not statistically significant. Similarly, the Δ$abc_{N\text{-}Glycan}$ (57 h ±42.8, n=10) infected cohort did not survive significantly longer than the TIGR4 infected group. ($p>0.05$).

The development of bacteremia in animals infected with the pneumococcal strains was also determined (Figure 13B). It was found that at 24 and 48 h post-infection, mice infected either with Δ*endoD* ($\log_{10}$ 2.7 ±0.25, and $\log_{10}$ 4.55 ±0.36, n=20, for 24 and 48 h, respectively), or Δ*endoD*Δ$abc_{N\text{-}Glycan}$ ($\log_{10}$ 2.1 ±0.38, and $\log_{10}$ 3.7 ±0.65, n=10, for 24 and 48 h, respectively) had significantly lower mean bacterial counts in blood than WT infected cohort ($\log_{10}$ 4.2 ±0.28, and $\log_{10}$ 6.1 ±0.48, n=20, for 24 and 48 h, respectively) ($p<0.05$ for Δ*endoD*, and $p<0.01$ for Δ*endoD*Δ$abc_{N\text{-}Glycan}$ for both time points). There was no significant difference in bacterial counts of mice infected with Δ$abc_{N\text{-}Glycan}$ and the parental strain at either 24 or 48 h post infection ($p>0.05$).

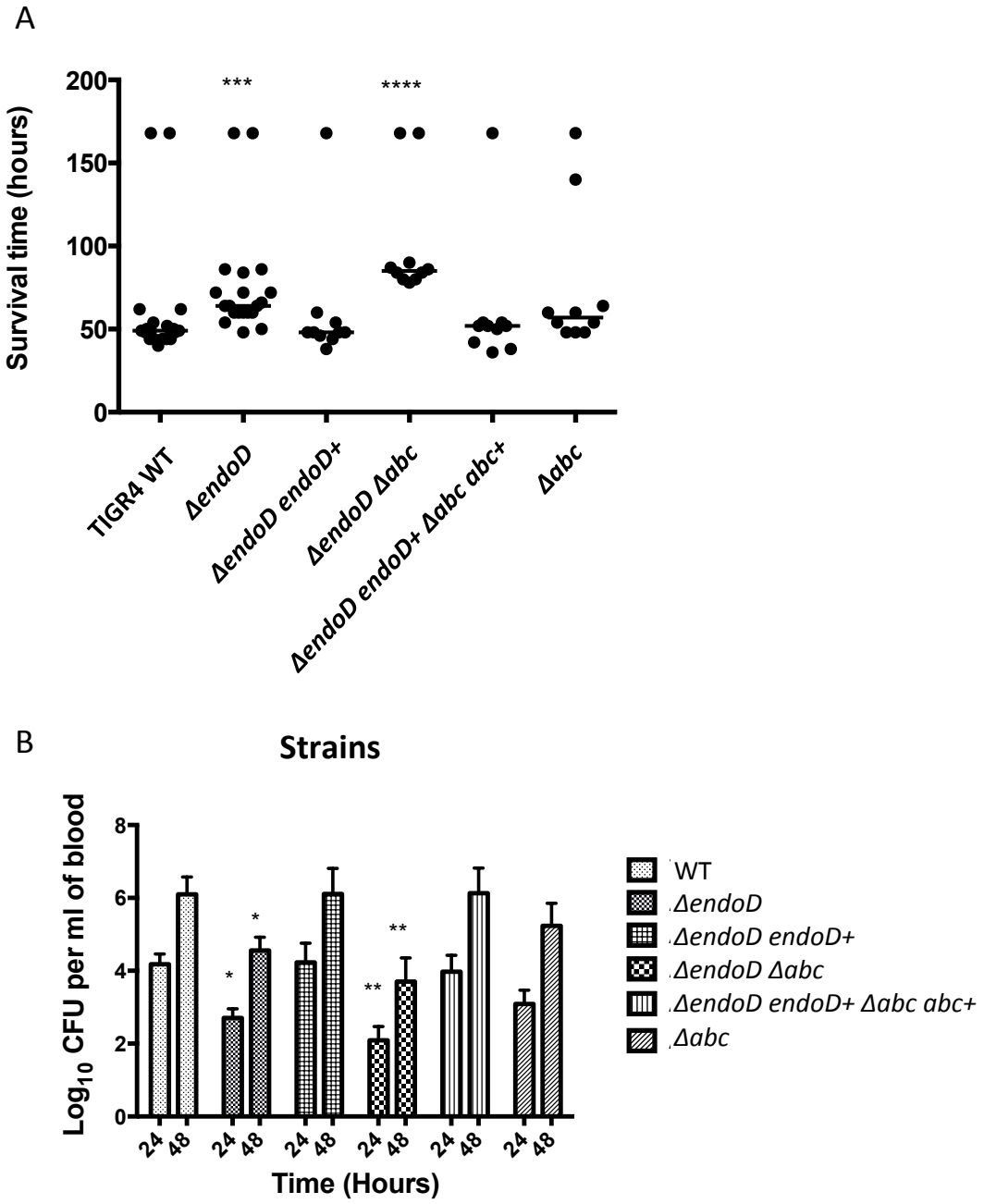


Figure 13. Contribution of *endoD* and $abc_{N\text{-}Glycan}$ to virulence.

(A). Survival of mice infected intranasally with pneumococcal strains. Graph shows the survival time of mice after infection with approximately 1 X 10^6 CFU pneumococci. Symbols show the time that individual mice became severely lethargic. The horizontal bars mark the median time to the severely lethargic state. *** indicates $p<0.01$, **** shows $p<0.0001$ relative to survival of cohort infected with TIGR4. (B) Growth of

bacteria in the blood after intranasal infection. At predetermined times an aliquot of bacteria was obtained from dorsal tail vein and CFU/ml of bacteria was determined by plating out the serial dilutions on blood agar. Each point is the mean of data from ten mice except for WT and *ΔendoD*, for which 20 mice were used. Error bars show the standard error of the mean. * indicates $p<0.05$, ** shows $p<0.01$ relative to CFU/ml of TIGR4 for same time points.

2.4 Discussion

Our structural and functional data suggest that $SBP_{N\text{-}Glycan}$ specifically binds to $Man_5GlcNAc$ with a high affinity, shedding light on the biological relevance of this ligand. The high micromolar range affinity obtained for $Man_1GlcNAc$, as well as the complex structure of the $SBP_{N\text{-}Glycan}$ in its closed conformation with this ligand, implies that this sugar is only a fragment of the biological ligand. This suggests that $SBP_{N\text{-}Glycan}$ specifically recognises the conserved reducing end found in all type of N-glycans; however, allowing some heterogeneity at the non-reducing end of the ligands would permit the transport of a range of N-glycan products. In fact, it is very common for SBPs to allow transport of a range of metabolites with a conserved feature (Berntsson et al., 2010). Furthermore our growth assay demonstrates that growth on complex N-glycans from fetuin is dependent on EndoD and $ABC_{N\text{-}Glycan}$. Since EndoD would cleave complex N-glycans trimmed down to $Man_3GlcNAc$ only, and no other identified extracellular mannosidases could further degrade the products of EndoD (Homer et al., 2001), it seems reasonable to assume that $SBP_{N\text{-}Glycan}$ would also be capable of binding to $Man_3GlcNAc$. As a result, the $ABC_{N\text{-}Glycan}$ would thus modulate the transport of all type of N-glycans including high-mannose, complex and hybrid, as they all share the same conserved core. In conclusion, these findings give insight into the mode of recognition $SBP_{N\text{-}Glycan}$ for N-glycan derived products and provide strong evidence of a novel import mechanism for branched N-glycans.

EndoD and $ABC_{N\text{-}Glycan}$ were identified in large-scale virulence screens; however, their ability to participate to disease has never been assessed. Here, we demonstrate that an *ΔendoD* mutant is significantly less virulent than the complemented or parent strains confirming for the first time that EndoD is necessary for full virulence in a mouse model. Surprisingly, no significant phenotype was observed for $ABC_{N\text{-}Glycan}$, suggesting that degrading host N-glycan-containing glycoconjugates by EndoD is more crucial for pathogenesis than the transport and utilization of these sugars for providing energy. In addition, the ability of the *endoD* mutant to cause sepsis is diminished compared to the parent strain indicating that in this model of infection, EndoD is required for the development of bacteremia. Combined, these data suggest that EndoD acts as a virulence

determinant independent of its nutrient acquisition function, as $ABC_{N\text{-}Glycan}$ does not contribute significantly to virulence.

Here we confirm for the first time that EndoD is a virulence factor, and we demonstrate that the extracellular portion of this N-glycan processing pathway contributes to full virulence. The role of this machinery in pathogenesis has yet to be fully investigated; however, from this work we propose that EndoD and $ABC_{N\text{-}Glycan}$ participate in the scavenging and utilization of N-glycans that can be used as the sole source of carbon for growth and proliferation of the bacterium. Thus, *S. pneumoniae* is likely to liberate energy from the destruction of host tissues comprising N-glycans that are found in the airway, where free sugars remain rare, and harvesting host glycans for growth is a requirement (King et al., 2006). The absence of a phenotype for $\Delta abc_{N\text{-}Glycan}$ in this virulence study, and the knowledge that EndoD and $ABC_{N\text{-}Glycan}$ play a role in nutrient acquisition from N-glycans, both strongly suggest that N-glycan degradation but not metabolism is important for pathogenesis and that EndoD is likely involved in another aspect of virulence other than providing a nutrient source for growth.

Chapter 3: Characterisation of a virulence-associated α-mannosidase from *Streptococcus pneumoniae*

Mélissa Cid[1], Hasan Yelsilkaya[2], Michael D. L. Suits[3] and Alisdair B. Boraston[1]

[1]Biochemistry & Microbiology, University of Victoria, PO Box 3055 STN CSC, Victoria, BC, V8W 3P6, Canada

[2]Department of Infection, Immunity & Inflammation, University of Leicester, 10 Leicester, LE1 9HN, UK

[3]Wilfrid Laurier University, 75 University Ave West, Waterloo, ON N2L 3C5, Canada

Contributions: I performed the cloning, protein production, purification, crystallization and the structure refinement of GH92. The data collection and structure solving was performed by Michael D. L. Suits. I performed the HPAE-PAD experiment of GH92. The mice virulence assays were performed by Hasan Yelsilkaya.

3.1 Introduction

A previous study conducted by Yamamoto *et al.* demonstrated that EndoD is active only on N-glycans that have been reduced to $Man_3GlcNAc_2$-Asn and $Man_5GlcNAc_2$-Asn (Yamamoto et al., 2005). We know from the literature that NanA, BgaA and StrH are responsible for cleaving the distal arm of complex N-glycan producing $Man_3GlcNAc_2$-Asn, which EndoD can act on. However no pneumococcal enzyme is known to cut the terminal α-1,2-mannose decoration from high-mannose N-glycan, a prerequisite for EndoD to accommodate this substrate. We hypothesise that *S. pneumoniae* must encode for an α-1,2-mannosidase that could cleave the high-mannose N-glycan decoration.

The N-glycan-processing locus (CPL) encodes for a putative α-mannosidase, GH92, of uncharacterised function. Based on its genomic context and function prediction we hypothesise that this enzyme displays α-1,2-mannosidase activity. Presently no characterised pneumococcal enzyme has been demonstrated to have such specificity strongly suggesting that GH92 could be responsible for trimming high-mannose N-glycans to $Man_5GlcNAc_2$-Asn since cleavage of these terminal residues must occur prior to the action of EndoD.

GH92 is also of particular interest as this protein was identified in previous large-scale virulence screens, possibly implicating it in the host-pathogen interaction. In STM studies conducted in TIGR4 (serotype 4) and the strain 0100993 (serotype 3) (Lau et al., 2001), *gh92* deficient mutants were attenuated in a pneumonia model of infection but not a bacteremia model (Hava and Camilli, 2002). Furthermore, other enzymes participating in the N-glycan-processing pathway such as NanA, BgaA and StrH are confirmed virulence factors and have been shown to be implicated in various aspects of virulence ranging from nutrient acquisition to complement modulation, suggesting a role for GH92 in these aspects of pathogenesis as well (Burnaugh et al., 2008; Dalia et al., 2010; Limoli et al., 2011; Manco et al., 2006; Pluvinage et al., 2011).

Based on sequence similarity and CAZyme database nomenclature, GH92 had previously been grouped into the glycoside hydrolase family 92 (Cantarel et al., 2009). These enzymes are a group of exo-acting α-mannosidases that proceed mechanistically with net inversion of stereochemistry at the anomeric carbon (Zhu et al., 2010a). Previously, a structural and functional investigation of the 23 GH92s from the prominent gut bacterium *Bacteroides thetaiotaomicron* VPI-5482 demonstrated variability with respect to enzymatic preference for α-1,2, α-1,3, α-1,4 and α-1,6 glycosidic linkages (Zhu et al., 2010). However, in the *S. pneumoniae* CPL, genes encoding for a GH38 and a GH125 would have the potential to degrade α-1,3 and α-1,6-mannosidic linkages present in N-linked glycans, respectively, leaving only the α-1,2-linkages not targeted by any known pneumococcal enzymes (Cobucci-Ponzano et al., 2010; Gregg et al., 2011; Suits et al., 2010).

Three GH92 family members have been structurally characterized: Bt2199 and Bt3990 from *B. thetaiotaomicron* and an *exo*-mannose-6-phosphate hydrolase CcGH92_5 from *Cellulosimicrobium cellulans* (CcMan92_5) (Tiels et al., 2012; Zhu et al., 2010). These enzymes all display a two-domain structure with a small β-sandwich N-terminal domain and a C-terminal domain that adopts a $(\alpha/\alpha)_6$ barrel fold. The active site resides in a shallow binding pocket created by both C- and N-terminal domains. These enzymes hydrolyze the glycosidic bond through a single displacement mechanism which leads to

inversion of the anomeric carbon generating β-mannose products (Figure 14). They also require a Ca^{2+} metal ion for proper catalysis. This metal ion bridges the O2 and O3 of the -1 subsite mannose residue leading to substrate distortion that is favorable for catalysis (Figure 14). Generally, enzymes of this family possess two catalytic residues, a glutamic acid that acts as a general acid and an aspartate that acts as a base.

GH92 from *S. pneumoniae* does not possess a signal peptide (SP) or a known anchoring motif implying intracellular localisation. However, non-classically secreted pneumococcal proteins with no SP or cell-wall anchoring motif have been previously identified rendering the localisation prediction of pneumococcal proteins difficult (Jeong et al., 2009; Pérez-Dorado et al., 2012). Since EndoD is a cell-wall attached protein through an LPXTG motif, GH92 is most likely found extracellularly in order to act before EndoD.

In this chapter, structural and functional insight of GH92 will be discussed. Additionally the potential implication of GH92 in virulence in mice will be assessed. Our findings indicate that this enzyme is a Ca^{2+} dependent α-1,2-mannosidase with similar structural features to other enzymes of the same family. Furthermore, we demonstrate that GH92 is important for virulence in a mouse model of pneumonia and sepsis.

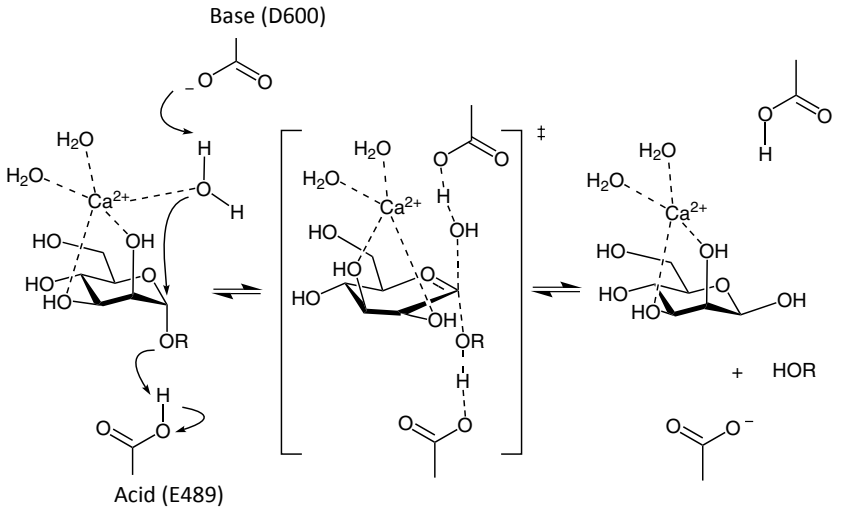


Figure 14. Inverting catalytic mechanism of GH92 (Adapted from Zhu et al.)

Putative reaction mechanism for an inverting α-mannosidase from family 92. The Asp 600 would act as a catalytic base and Glu489 act as an acid. Ca^{2+} bridges the O2 and O3 of the substrate while orienting the nucleophilic water.

3.2 Experimental procedure

Cloning, protein production and purification, crystallization and structural solution

The PCR primers used for amplification of GH92 (residues 1-694) from TIGR4 (SP2145) were GH92-Fw (5'-CATATCGCTAGCATGAAACCACTAC-3') and GH92-Rv (5'-GGTGGTCTCGAGTTAGTGACTTGGTAAC-3'). *gh92* (SP2145) was cloned via standard PCR, restriction endonuclease, and ligation procedures into pET28 expression plasmid using *Nhe*I and *Xho*I sites. Recombinant GH92 fused to a thrombin cleavable N-terminal histidine-affinity tag was produced via autoinduction method (Studier, 2005), by shaking inoculated 1L cultures supplemented with kanamycin for 96 hours at 37^{o}C. Cells were harvested by centrifugation and disrupted by chemical lysis. Clarified lysates were purified via nickel-affinity and anion exchange chromatography to homogeneity. GH92 in 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl was concentrated using a stirred cell amicon to 12 mg/ml. GH92 crystallized by mixing equal volumes of protein with a solution consisting of 0.5 M magnesium formate dihydrate and 0.1 M 4-(2-hydroxyethyl)-1-piperazinethanesulfonic acid (pH 7.5) using the sitting-drop vapour-diffusion method. Rhombohedral shaped crystals developed after ~3 months, which were cryo-protected with 30% glycerol for diffraction experiments.

Diffraction data were collected at beamline 9-2 of the Stanford Synchrotron Radiation Lightsource, processed with XDS and scaled with Scala (Evans, 2006; Kabsch, 2010). The *B. thetaiotaomicron* BT3990 structure (PDBID: 2WVY) (Zhu et al., 2010) was used as a molecular replacement model in Phaser (McCoy, 2006), which unambiguously found four molecules in the asymmetric unit. The structure of GH92 (residues 2 – 694) was iteratively improved with cycles of manual building with Coot and refinement with Refmac (Emsley and Cowtan, 2004; Murshudov et al., 1997). Data collection, phasing, and refinement statistics are presented in Table 5.

Table 5. Data collection and refinement statistics for GH92

	SpGH92
Data Collection statistics	
Beamline	SSRL BL9-2
Wavelength	1.0000
Space group	$P2_12_12_1$
Cell dimensions: *a, b, c* (Å)	128.8, 161.8, 208.5
Resolution (Å)	50.00-2.50 (2.59-2.50)*
R_{merge}	0.075 (0.409)*
I/σI	26.4 (5.0)*
Completeness (%)	99.4 (96.3)*
Redundancy	10.7 (8.5)*
Refinement	
Resolution (Å)	2.50
No. of reflections Total, Unique	1613893, 150345
R_{work}/R_{free}	17.42 / 0.21.13
No. of atoms	
Protein Chain A, B, C, D	21939
Calcium	8
Ligands	N / A
Solvent	737
Average B-factors ($Å^2$)	
Protein	43
Calcium Ions	49
Ligands	
Solvent	39
Root mean square deviations	
Bond lengths (Å)	0.0135
Bond angles (degrees)	1.4091
Ramachandran	
Preferred (%)	2676 (96.4)
Allowed (%)	100 (3.6)
Disallowed (%)	0 (0)

*Values in parentheses are for highest resolution shell**

GH92 activity assay

GH92 was tested for activity against various oligomannoses using high performance anion exchange chromatography using pulsed amperometric detection (HPAEC-PAD) to monitor glycoside cleavage. Assays were performed in 20 µL volumes at 37 °C in PBS at pH 7.5 with 2 mM calcium chloride. Reactions were initiated by the addition of 500 nM enzyme to 0.5 mM of each carbohydrate. Reactions were stopped by the addition of 140 µL of 100 mM sodium hydroxide. After centrifugation at 5,000 rpm for 5 min, 20 µL of the samples were analyzed by HPAEC-PAD using a Dionex ICS 3000HPLC equipped with an ASI 100 Automated sample injector and an ED50 electrochemical detector (Dionex) with a gold working electrode and an Ag/AgCl reference electrode. Products were analyzed using a PA-20 column set (analytical plus guard column) using an isocratic gradient of 100 mM NaOH.

Construction of deletion mutants

Δ*gh92 and* Δ*gh92*Δ*gh92+* mutant strains were constructed using the Janus cassette selection system as described previously in section 2.2. for the Δ*endoD* and Δ$abc_{N\text{-}glycan}$ mutant strains.

In vivo analysis of pneumococcal strains

In vivo analysis of the pneumococcal strains, utilized ten to twelve-week-old female MF1 outbred mice (Charles River, Margate, United Kingdom). For infections, a standard inoculum was prepared for each strain by inoculating fresh BHI serum broth (80% [vol/vol] BHI and 20% [vol/vol] bovine serum) with overnight grown pneumococcal culture. When the OD_{500} reached 1.6, bacteria were stored at −80°C in small aliquots until needed. Immediately before infection, bacteria were diluted to $2X10^7$ CFU/ml in PBS (pH 7.0).

To assess the virulence of pneumococcal strains, mice were lightly anesthetized with 3% (vol/vol) isoflurane over oxygen, and an inoculum of 50 µL containing approximately 1 x 10^6 CFU in PBS was given drop-wise into the nostrils as described previously (Gaspar et al., 2014; Yesilkaya et al., 2009). After infection, the inoculum dose was confirmed by

viable counting on blood agar plates to determine the actual administered dose. Mice were monitored for disease signs (progressively starry coat, hunched, and lethargic) for 7 days, and disease signs were evaluated independently by two experienced operators. When mice were lethargic, they were culled. Therefore, time to reach lethargic state has been defined as the "survival time." Mice that were alive 7 days after infection were deemed to have survived the infection.

To monitor the development of bacteremia in each mouse, approximately 20 μL of venous blood was obtained from mice at predetermined time points after infection. Viable counts in blood were determined by serial dilution in sterile PBS and plating onto blood agar plates supplemented with 5% (vol/vol) defibrinated horse blood.

Survival times were calculated by using GraphPad Prism software and analyzed by the Mann-Whitney U test. Bacterial counts in blood were analyzed by an analysis of variance followed by the Bonferroni post-test. Statistical significance was considered to be a P value of <0.05.

Ethics statement

Mouse experiments at the University of Leicester were performed under appropriate project (permit no. 60/4327) and personal (permit no. 80/10279) licenses according to the United Kingdom Home Office guidelines and local ethical approval. Where appropriate, the procedures were carried out under anesthetization with isoflurane. Mice were kept in individually ventilated cages in a controlled environment and were frequently monitored after infection, to minimize suffering.

3.3 Results

Overall structure of GH92

GH92 crystallized in the space group $P2_12_12_1$ with four molecules in the asymmetric unit and diffracted to a resolution of 2.5 Å (Table 5). The structure of GH92 consists of two domains (Figure 15). A shallow surface depression located at the domain interface houses the active site. A calcium ion was bound in the active site suggesting that, like other enzymes from this family, catalysis by GH92 might be Ca^{2+} dependent. The N-terminal

domain is a β-sandwich that terminates with an extended α-helix. The larger C-terminal domain consists of an $(\alpha/\alpha)_6$ barrel, which is a conserved and widespread fold adopted by other glycoside hydrolases from family 37, 63, 15, 65, 125, 8 and 48 (Cantarel et al., 2009). However, the $(\alpha/\alpha)_6$ barrel C-terminal domain of GH92 contains a particular feature that was also observed in the Bt3990 structure (Zhu et al., 2010). Four strands from the C-terminal domain and one strand derived from the N-terminal β-sandwich domain participate in forming the five-stranded β-sheet found at the C-terminus of the $(\alpha/\alpha)_6$ domain. The function of this unusual β-sheet is presently unknown but it may be important for the folding of GH92. The active site is formed by extended loops derived from the N-terminal end of helices 1, 9, 12, 13, 16 from the $(\alpha/\alpha)_6$ barrel and from the N-terminal end of β-strand 3.

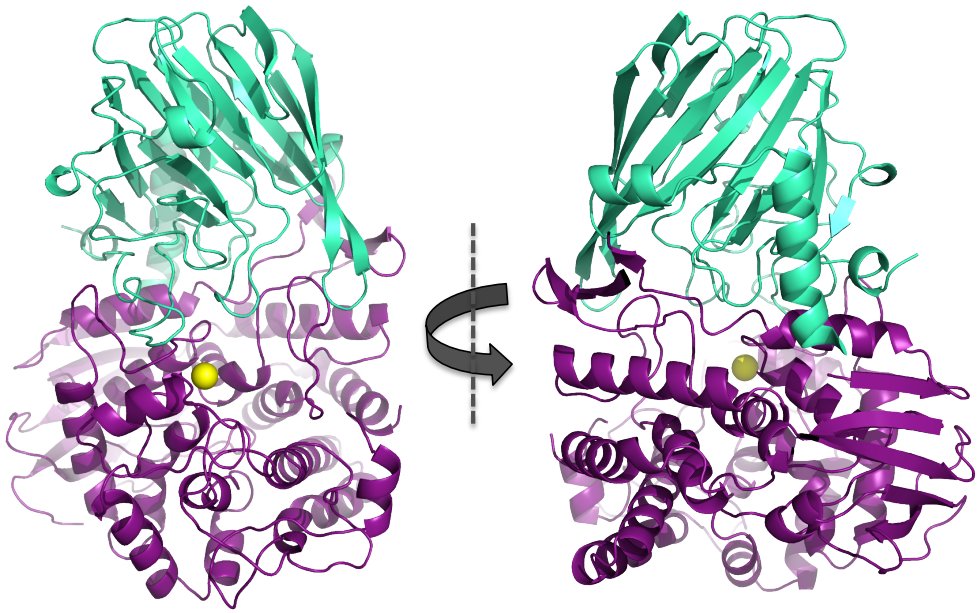


Figure 15. Overall structure of GH92

Cartoon representation of GH92, the N-terminal domain is represented in green-cyan and the C-terminal domain in purple. The catalytic calcium is shown as a yellow sphere.

Structure comparison with other GH92s

To date, only 3 other GH92 structures are available and all show a very similar bi-modular fold with the active site formed by the interface of both domains. With 32% sequence identity with Bt3990, 31% with Bt2199, and 21% with CcGH92, the structure of GH92 shows high structural similarity with these proteins (Table 6). Despite lower sequence identity with CcGH92 and higher rsmd value, the overall fold of these enzymes is almost identical with no significant structural differences (Figure 16). On the other hand, the main difference between Bt3990 and GH92 is the presence of two extended β-strands, 11 and 12 in the N-terminal domain, which appears to be much shorter in the Bt3990 structure. Despite these differences, the structural alignment statistics revealed a low rmsd value. Similarly, structural alignment with Bt2199 reveals an insertion of extra α-helices projecting away from the $(\alpha/\alpha)_6$ core, which is absent in all other members of this family with known structures including GH92 (Figure 16). The only structures available of an enzyme from this family in complex with a substrate-like compound is Bt3990 (Zhu et al., 2010). The structure of Bt3990 was solved in complex with a thio-linked substrate mimic, which shed light into the catalytic mechanism of GH92s. The active site of GH92 is conserved and occupies essentially the same space compared to the Bt3990 active pocket further suggesting that the specificities of these enzymes are similar (Figure 17A-B). The catalytic calcium is octa-coordinated by the side chain carbonyl of Asn557 and Gln558 and by one of the carboxylate oxygens of Asp600. Three water molecules and the two oxygens of an ethylene glycol molecule are also participating in the coordination of the calcium. The conservation of the active site of GH92 supports a Ca^{2+} dependent inverting mechanism of catalysis where the calcium would coordinate the O2 and O3 of the -1 subsite mannose residue in order to distort the substrate and prepare it for catalysis by the two conserved catalytic residues (Glu489 and Asp600). Asp600 would act as a base and deprotonate the nucleophilic water while the Glu489 would act as an acid, which donates a proton to the leaving group. The metal ion would bridge the O2 and O3 hydroxyls of the -1 mannoside in order to aid distorting it towards a skewed boat or boat to facilitate catalysis. This metal ion would also be responsible for orienting the nucleophilic water for optimal nucleophilic attack of the substrate C1. The structural

architecture of GH92 active site, and its similarity to Bt3990, also supports a specific activity for α-1,2-mannose-linked residues (Figure 17A-B).

Table 6. Superposition statistics of family 92 glycoside hydrolases with GH92

Protein	Number of Cα aligned over 693 residues	RMSD (Å)	Identity (%)
Bt3990	689	1.8	30
Bt2199	664	2.2	29
CcGH92	381	2.3	21

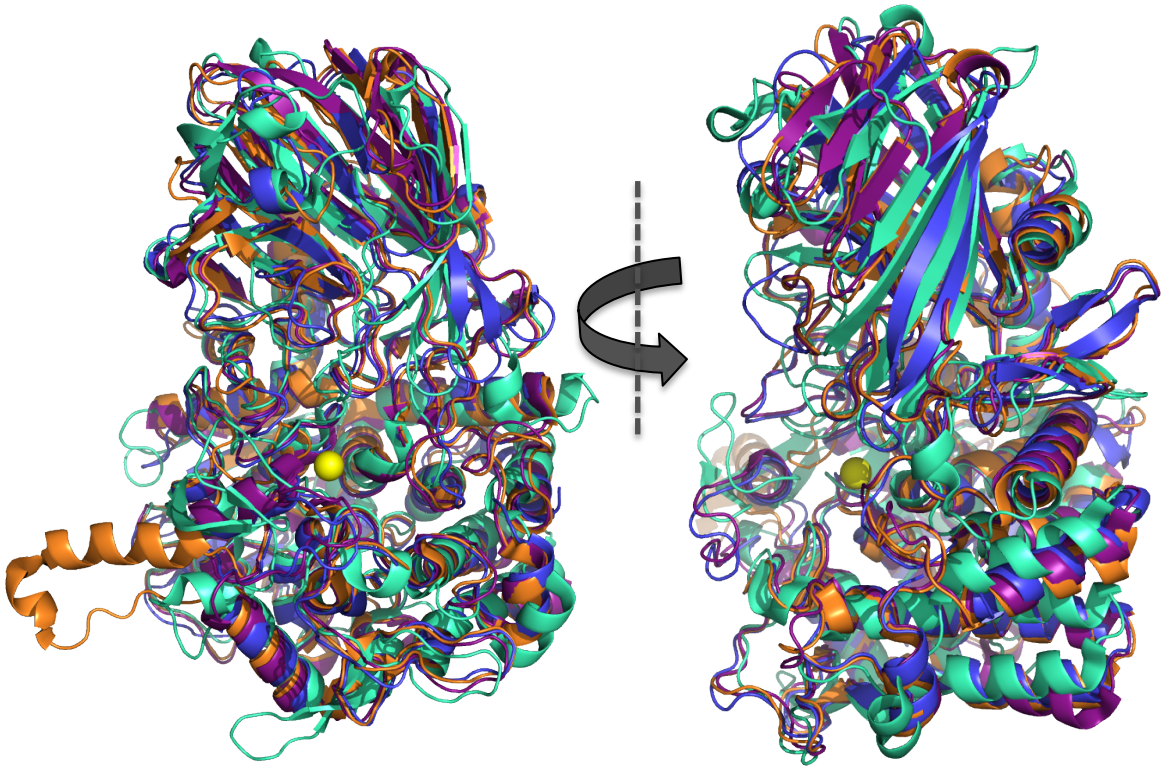


Figure 16. Overlay of all GH92 structures available.

The cartoon representation of GH92 structure is represented in blue, in green-cyan for CcGH92 (PDBID: 2XSG), in purple for Bt3990 (PDBID: 2WW1) and orange for Bt2199 (PDBID: 2WVY). The catalytic calcium from GH92 is represented by a yellow sphere.

GH92 is an a-1,2-mannosidase

HPAEC-PAD analysis of various oligomannoses showed that GH92 displays activity only on α-1,2-mannobiose and produces mannose (Figure 18). GH92 activity was only observed with the addition of Ca^{2+} indicating that this enzyme is Ca^{2+} dependent, as expected. This analysis demonstrates the ability of this enzyme to act on isolated fragments of N-glycans; however, its ability to process entire high mannose N-gycan still remains untested. An analysis of the *B. thetaiotaomicron* GH92s from a previous study suggest that a His-Glu and a distant Trp comprised a signature motif necessary for α-1,2-mannosidase activity. These residues are found in GH92 with His 540, Glu 541 and the distant Trp 70. An overlay of GH92 active site with Bt 3390 in complex with a α-1,2-thio-mannobiose demonstrate the potential contribution of these residues to specificity at the +1 subsite (Figure 17B). Trp 70 forms a hydrophobic platform accommodating the aglycone residue, while the His 540 and Glu 541 pair would interact via hydrogen bonds with O3 and O4. Modelling of α-1,2-thio-mannobiose in the active site of GH92 suggests that the same type of interaction would be possible. The aglycone residue of α-1,2-thio-mannobiose would lay on the hydrophobic Trp 70 while Glu 541 and His 540 would interact via hydrogen bonds with the O3 of the same sugar residue. Comparison of GH92 modelled with an α-1,3-and α-1,6-mannobiose suggests that the active site of GH92 is better tailored for the accommodation of α-1,2-linked mannose residues, which is in agreement with the HPAEC-PAD data (Figure 17C-D). While the aglycone mannoside of α-1,6-mannobiose would sit further away from the critical Trp 70 residue, the orientation of the sugar ring of the aglycone mannose also suggests that stacking interaction with Trp 70 would be unlikely. The aglycone mannose residue of α-1,3-mannobiose could possibly establish a stacking interaction with Trp 70; however, because of longer distances interactions with His 540 and Glu 541via hydrogen bonding would be compromised. Poor accommodation of α-1,3 and α-1,6- mannobiose seems to be due to insufficient interactions with the active site side chains, not steric hindrance. GH92 did not exhibit any activity on 4-nitrophenyl α-D-mannoside suggesting that the aglycone residue that sits in the +1 subsite of the active site is essential for the specificity of the enzyme. Therefore optimum accommodation of the aglycone sugar residue should be critical for favouring binding and catalysis. This modelling analysis supports the

activity data showing that GH92 is an exo-acting enzyme specific for α-1,2-mannobiose and would be able to cleave terminal α-1,2-mannose residues from high mannose N-glycans.

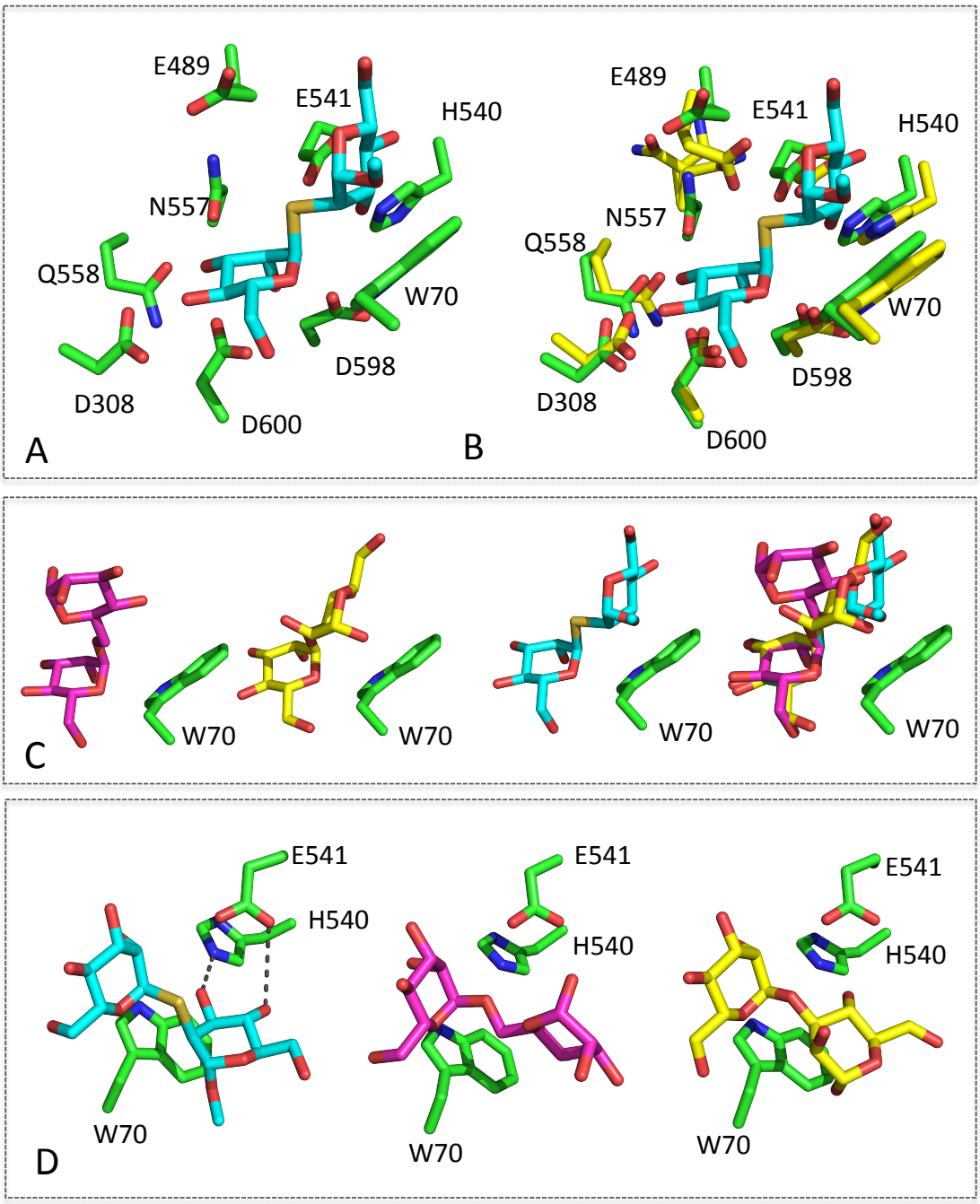


Figure 17. Structural basis for α-1,-2-mannosidase activity of GH92.

(A-D) The active site side chain residues of GH92 and Bt3990 are represented in green and yellow sticks, respectively. (A-D) α-1,2-thio-mannobiose represented in blue sticks from the Bt3990 complex structure was modelled in the GH92 active site. (C-D) α-1,6- and α-1,3-mannobiose represented in magenta and yellow sticks respectively were modelled in the active centre of GH92.

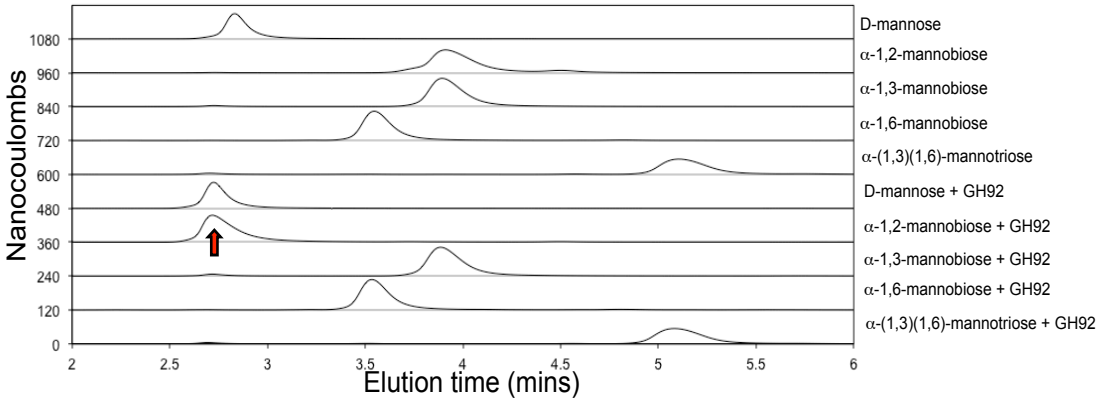


Figure 18. HPAEC-PAD analysis of GH92 activity

The top five traces show carbohydrate standards while the bottom five traces show these standards digested with GH92. Only activity on α-1,2-mannobiose was observed.

The contribution of GH92 to virulence

The contribution of GH92 to virulence was assessed by creating a *gh92* deficient mutant in *S. pneumoniae* TIGR4 and testing its ability to cause disease in a mouse model of infection. The mutants were tested in acute pneumonia and sepsis models that develop after intranasal infection. The results showed that the median survival time of mice infected intranasally with Δ*gh92* (168 h ±46.2, n=10) was significantly longer than those of the TIGR4 WT (47.5 h ±38.3, n=10) and Δ*gh92*Δ*gh92*+ (52 h ±49.5, n=10) infected groups ($p<0.001$) (Figure 19A). Both the WT and genetically complemented strain infected mice survived significantly longer than the mutant infected cohort in spite of the higher inoculum given to the mutant infected cohort (3,520,000 CFU/mouse) compared to WT (1,100,000 CFU/mouse). Consistent with the survival assay, bacteremia was more severe in wild type and genetically complemented infected groups. It was found that at 24- and 36 h post-infection, respectively, mice infected with Δ*gh92* (log10 0.90 ±0.46, and log10 1.44 ±0.72, n=10), had significantly lower mean bacterial counts in blood than WT (log10 3.46 ±0.64, and log10 6.26 ±0.74, n=10) and Δ*gh92*Δ*gh92*+ (log10 2.52 ±0.87, and log10 4.85 ±0.90, n=10) infected cohorts ($p<0.05$ and $p<0.0001$ for 24- and 36 h post-infection) (Figure 19B). There was no significant difference in bacterial counts of mice infected with Δ*gh92*Δ *gh92*+ and TIGR4 at either 24- or 36 h post infection ($p>0.05$).

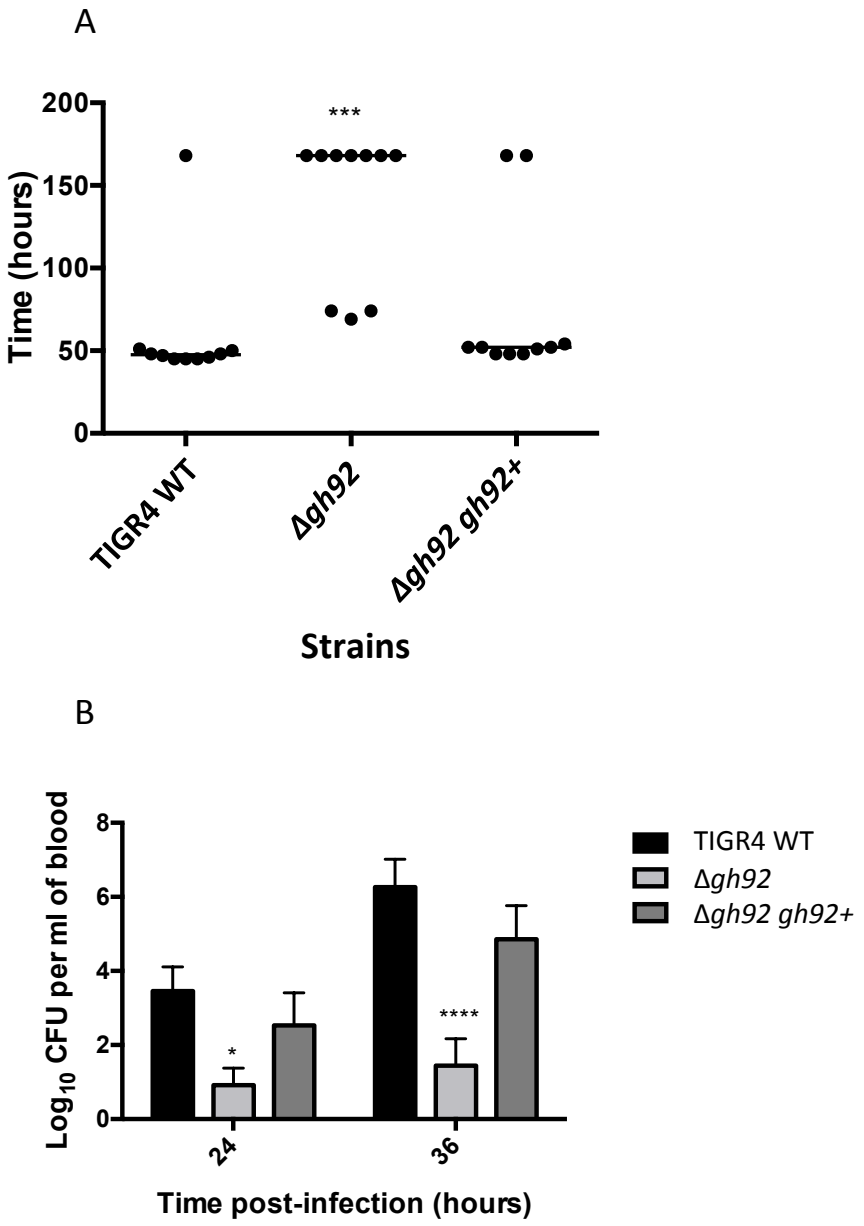


Figure 19. Impaired virulence of *gh92* defective strain following intranasal infection. (A) Survival time of mice after infection with approximately 1.1 to 3.5 X 10^6 CFU pneumococci. Symbols show the times mice became severely lethargic. The horizontal bars mark the median times to the severely lethargic state. (B) Growth of bacteria in the blood. Each point is the mean of data from ten mice. Error bars show the standard error of the mean. Symbols: * p<0.05; *** p<0.001; **** p<0.0001.

3.4 Discussion

The structural and functional analysis of GH92 reveals that it is an α-1,2-mannosidase that likely targets the terminal α-1,2-mannoside present in high mannose N-glycans. The activity of GH92, combined with the fact that no other pneumococcal α1,2-mannosidase has been identified, led us to speculate that GH92 is extracellular and acts first on high mannose N-glycan permitting the cell-wall attached EndoD to cleave the chitobiose core and release this glycan from its host glycoprotein scaffold. From this study we propose that GH92 is part of the N-glycan degradation pathway of *S. pneumoniae* and is essential to initiate high mannose N-glycan degradation by trimming the α-1,2- terminal mannoses.

Moreover, our mice study confirms for the first time that GH92 is necessary for virulence and participates in the development of pneumonia and bacteremia in mice. Indeed, as demonstrated, virulence in mice infected with a *gh92* deficient strain was strongly attenuated. As previously concluded for EndoD, the lack of virulence attenuation for the $ABC_{N\text{-}glycan}$ suggests that the role of EndoD and GH92 in virulence is independent of nutrient acquisition since uptake of the product of these enzymes does not affect pathogenesis. This suggests that the effect of these enzymes observed in our mice virulence model can be attributed to participation in a process other than providing a carbon source for growth.

Future direction would be focused on demonstrating α-1,2-mannosidase activity of GH92 on high mannose N-glycan. Furthermore, demonstrating that growth of *S. pneumoniae* on α-1,2-mannobiose as sole source of carbon is dependent on GH92 would be of high relevance. Indeed, no other α-1,2-mannosidases have been identified in the *S. pneumoniae* genome suggesting *S. pneumoniae* would grow on this substrate only in a GH92-dependent manner. Since performing a growth assay using purified high-mannose N-glycan would be too costly and *S. pneumoniae* does not grow well on glycoproteins bearing high mannose N-glycans, a growth assay on α-1,2-mannobiose will be the only feasible approach to test this hypothesis. Lastly, a long-term future direction would include determining the potential different aspects of virulence that GH92 is involved in

using *S. pneumoniae gh92* deficient mutants and assess their phenotype in assays that mimic aspects of virulence.

Chapter 4: A second β-hexosaminidase encoded in the *Streptococcus pneumoniae* genome provides an expanded biochemical ability to degrade host glycans.

Mélissa Cid[1], Craig S. Robb[1] and Alisdair B. Boraston[1]

[1]Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, V8W 3P6, Canada

Contributions: I performed the cloning, protein production, purification, crystallization, kinetics and inhibition study of GH20C. I also performed the structure refinement of the native structure of GH20C and the structures in complex with GlcNAc, GalNAc, Gal-NGT and Gal-PUGNAc. Craig S. Robb performed the structure refinement of the GH20C structures in complex with PUGNAc and NGT.

4.1 Introduction

The S. pneumoniae genome encodes for an exo-β-D-N-acetylglucosaminidase, StrH, which was first characterized by Yamashita et al. (Yamashita et al., 1981). This cell-wall attached enzyme is capable of releasing terminal β-linked GlcNAc residues from complex N-glycans and is implicated in virulence, possibly by participating in complement modulation (Dalia et al., 2010; King et al., 2006; Pluvinage et al., 2011). Interestingly, StrH possesses two tandem catalytic modules belonging to glycoside hydrolase family 20 (GH20A and GH20B). The structural features of these two catalytic modules give them an extremely refined specificity for terminal GlcNAc residues in complex N-linked glycans (Pluvinage et al., 2011).

GHs from family 20 are β-N-acetylhexosaminidases that hydrolyze non-reducing terminal β-linked GlcNAc and/ or GalNAc residues with retention of the anomeric carbon configuration in a substrate-assisted manner (Bøhle et al., 2011; Jiang et al., 2011; Prag et al., 2000; Sumida et al., 2009; Wang et al., 2012; Williams et al., 2002). Indeed, extensive studies on enzymes of this family suggest that they use a two-step substrate-assisted catalytic mechanism in which the acetamido group of the hexosamine acts as a nucleophile resulting in the formation of a transient oxazoline intermediate. The

catalytic residues in GH20 enzymes are found as a conserved Asp/Glu pair. The Glu residue acts as the general catalytic acid/base while the Asp is thought to stabilize the oxazoline intermediate. This family includes over 100 characterized enzymes, including 14 members for which X-ray crystal structures have been solved (Cantarel et al., 2009). Despite these advancements, the molecular details behind the glycon and aglycon (leaving group on reducing side of the terminal sugar) specificity is presently not entirely understood. Indeed, all GH20s are strict β-N-acetylhexosaminidases and usually have a preference for GalNAc or GlcNAc. Furthermore, enzymes of this family can be endo- or exo-acting. While β-N-acetylhexosaminidase activity is easily predictable among this family, aglycone specificity is, however, impossible to predict. The human hexosaminidases A and B are excellent examples to illustrate the importance of aglycon specificity in the biological function of the enzyme. HexA and HexB are responsible for the cleavage of terminal GalNAc from the G_{M2} ganglioside (GalNAcβ(1–4)-[NANAα(2–3)-]-Galβ(1–4)-Glc-ceramide) within the lysosome. Mutations of these enzymes results in the lysosomal storage disorders called Tay-Sachs disease and Sandhoff disease (in the case of HexA and HexB, respectively) (Fernandes et al., 1997; Lemieux et al., 2006; Mark et al., 2003). Interestingly, StrH can degrade β-linked GalNAc from disaccharides. However, its strict aglycon specificity does not allow for cleavage of terminal GalNAc from biologically relevant glycans further illustrating the emphasis of understanding the aglycon specificity of these enzymes (Pluvinage et al., 2011).

The *S. pneumoniae* genome encodes for an additional family 20 GH that is present in all sequenced *S. pneumoniae* strains within the previously identified CPL. This enzyme shares ~70% sequence identity to the structurally characterized *S. gordonnii* GcnA which possesses both N-acetylgalactosaminidase and N-acetylglucosaminidase activity (Harty et al., 2004; Langley et al., 2008). The high sequence similarity between GH20C and GcnA suggests similar glycon specificity; however, the aglycon specificity of both enzymes remains completely unknown. To explore why *S. pneumoniae* encodes for two β-N-acetylhexosaminidases from the same family and their potential overlapping functions, we investigated the specificity of GH20C through biochemical and structural analysis. The data show that GH20C is a promiscuous enzyme that cleaves terminal β-

linked GlcNAc and GalNAc with very limited aglycon specificity. These results suggest that rather than being redundant with StrH, GH20C is likely responsible for the cleavage of terminal N-acetylhexosamine residues that StrH cannot hydrolyze and, in particular, provides N-acetyl-β-D-galactosaminidase activity that does not appear to be displayed by any other known or predicted *S. pneumoniae* enzymes.

4.2 Materials and Methods

Materials

Oligosaccharides were obtained from V-labs (Covington, LA). PUGNAc and NGT were synthesised by Dr. David Vocadlo and Gal-NGT and Gal-PUGNAc were synthesised by Dr. Keith Stubbs. All other reagents including GlcNAc and GalNAc were purchased from Sigma unless otherwise specified.

Cloning

The gene fragment encoding GH20C (SP2141) (amino acids 1-626) was PCR-amplified from *S. pneumoniae* TIGR4 genomic DNA using the oligonucleotide primers GH20C-Fw (5'-CATATCGCTAGCATGGTAAGATTTACAG) and GH20C-Rv (5'-GGTGGTCTCGAGTTAAGTCGTATAAATC). The amplified DNA fragment was cloned using standard procedures into pET 28a (Novagen) using 5′ NheI and 3′ XhoI restriction sites, respectively. The resulting plasmid, pGH20C, encoded the desired polypeptide fused to an N-terminal six-histidine tag by a thrombin protease cleavage site. PCR site-directed mutagenesis procedures using the "megaprimer" PCR method were used to introduce the E223Q substitution into pGH20C using the oligonucleotide primers GH20CE223Q-Fw (5'-ATCGGGATGGACCAGGCCCACTTGGTT) and GH20CE223Q-Rv (5'-AACCAAGTGGGCCTGGTCCATCCCGAT) (Barik, 1996). The DNA sequence fidelity of all constructs was verified by bidirectional sequencing.

Protein production and purification

All constructs were transformed into the expression strain *Escherichia coli* BL21 Star (DE3) strain. Six liters of YT broth, containing 50 μg/mL kanamycin, was inoculated with the transformed cells and incubated at 37 °C. Once an optical density of 0.6 at 595 nm was reached, protein production was induced by the addition of isopropyl 1-thio-β-d-

galactopyranoside to a final concentration of 0.5 mM. Incubation of the cultures was continued overnight with shaking at 16 °C. Cells were harvested by centrifugation at 5000 x g for 10 minutes and ruptured by chemical lysis. The cleared supernatant of the cell lysate was loaded onto a Ni^{2+}-NTA immobilized metal affinity chromatography column. Polypeptide was eluted with binding buffer (20 mM Tris, pH 8.0) containing increasing concentrations of imidazole (0–500 mM). GH20C was concentrated and buffer-exchanged into 20 mM Tris, pH 8.0, in a stirred ultrafiltration unit (Amicon). Purified protein was further purified by size exclusion chromatography using a Sephacryl S-200 column (GE Healthcare) and concentrated using a stirred-cell ultrafiltration device (Amicon) with a 10,000 Da molecular weight cut-off membrane (Millipore).

Determination of Protein Concentration

Protein concentration was determined by UV absorbance at 280 nm using the calculated extinction coefficient of 94200 M^{-1} cm^{-1}.

Enzyme kinetics and inhibition

All steady state kinetic studies were performed at 37 °C in a SpectraMax M5 plate reader (Molecular Devices). The pH optimum was determined using end-point assay (stopped with 100 mM of NaOH) in McIlvaine buffer for pHs ranging from 2.25 to 9.2; the optimum activity was found at 6.5. Determination of kinetic constants was done in triplicate in 50 mM citrate buffer, pH 6.5, in 1 mL volumes containing 0.1% BSA, 2 nM of GH20C and 0–0.9 mM 4-nitrophenyl-N-acetylglucosaminide (pNP-GlcNAc) or 0–0.45 mM 4-nitrophenyl-N-acetylgalactosaminide (pNP-GalNAc). Nitrophenolate production was monitored by absorbance at 400 nm; rates were calculated using an extinction coefficient, ε400nm, of 6090 M^{-1} cm^{-1}. Michaelis-Menten parameters were determined using GraphPad by nonlinear curve fitting. The K_i values for the inhibition by *O*-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino *N*-phenyl carbamate (PUGNAc), *O*-(2-acetamido-2-deoxy-D-galactopyranosylidene)amino *N*-phenyl carbamate (Gal-PUGNAc), *N*-acetyl-D-glucosamine-thiazoline (NGT) and *N*-acetyl-D- galactosamine-thiazoline (Gal-NGT) were determined by Dixon plot analysis using pNP-GlcNAc as a substrate.

Activity screen

Qualitative analysis of GH20C activity on di- and trisaccharides was done by thin layer chromatography (TLC) using a variety of oligosaccharides containing non-reducing terminal hexosamine. Reactions were done in 20 μL final volume containing 50 mM Tris, pH 8, 30 μM of GH20C and 3 mM sugar. Reactions were incubated at 37 °C for an hour and then spotted on a silica plate. TLC was run for two hours in n-propanol/H_2O/ethanol (7/2/1) solvent and oligosaccharides were revealed by using 5% sulfuric acid and a 30 minute incubation at 110 °C.

General crystallography procedures

Crystals of GH20C were obtained at 18^{o} C, using sitting-drop vapour diffusion for screening and hanging drop vapour diffusion for optimization. For data collection single crystals were flash cooled with liquid nitrogen in crystallization solution previously dehydrated to increase cryoprotectant content for each crystal form as given below. Diffraction data was collected either on a ‘‘home-beam’’ comprising a Rigaku R-AXIS 4++ area detector coupled to a MM-002 X-ray generator with Osmic ‘‘blue’’ optics and an Oxford Cryostream 700, Beamline 9-2 of the Stanford Linear Accelerator Center (SLAC, Stanford Synchrotron Radiation Lightsource [SSRL], CA), or CLS ID-1 at the Canadian Light Source (CLS, Saskatoon, Saskatchewan) as indicated in Table 7. Most diffraction data was processed using MOSFLM and SCALA (Collaborative Computational Project, 1994; Powell, 1999). However, the NGT and PUGNAc complexes were processed with HKL2000 and XDS, respectively. All data collection and processing statistics are shown in Table 7. For all structures, manual model building was performed with COOT (Emsley and Cowtan, 2004) and refinement of atomic coordinates was performed with REFMAC (Murshudov et al., 1997). The addition of water molecules was performed in COOT with FINDWATERS and manually checked after refinement. In all data sets, refinement procedures were monitored by flagging 5% of all observations as ‘‘free’’ (Brunger, 1992). Model validation was performed with MOLPROBITY (Davis et al., 2007).

GH20C structure and complexes determination

All crystals of GH20C (20 mg/ml) were grown in 12% (w/v) PEG 3350, 0.2 M $MgCl_2$, 0.05 M CAPS, pH 10.2 and 1% glycerol, although the protein crystallized in a variety of PEG 3350 concentrations, salts, and pHs. Crystals were soaked for 5 min to several hours in the crystallization solution containing an excess of reaction products and inhibitors. Data sets were collected as above on a single crystal that was cryoprotected by dehydration of the drop to increase glycerol concentration. The GcnA structure (PDBID: 2EPO) was used to solve the native structure of GH20C by molecular replacement using PHASER (Langley et al., 2004). For all the complex structures, the GH20C native structure was used as a search model for molecular replacement using PHASER.

Table 7. Data collection and refinement statistics for GH20C

	GH20C						
	Native	GlcNAc	GalNAc	NGT	Gal-NGT	PUGNAc	Gal-PUGNAc
Data collection statistics							
Wavelength mtz	0.97949	1.03322	0.97946	0.97926	0.97965	1.12709	1.12709
Beamline	MM-002	CLS ID-1	SSRL BL9-2	CLS ID-1	CLS ID-1	SSRL BL9-2	SSRL BL9-2
Processing Software	MOSFLM	MOSFLM	MOSFLM	HKL2000	MOSFLM	XDS	MOSFLM
Space group	P212121	P 21 21 21	P212121	P212121	P212121	P212121	C2221
Resolution (Å)	50.70 - 1.94	46.60 - 1.84	79.54 - 2.08	78.12 - 2.40	95.69 - 2.10	38.66 - 1.78	47.90-2.27
	(2.05 - 1.94)	(1.94 - 1.84)	(2.19 - 2.08)	(2.49 - 2.40)	(2.21 - 2.10)	(1.81 - 1.78)	(2.40 - 2.27)
Cell dimension a, b, c (Å)	92.8, 136.9, 202.3	87.7, 98.1, 132.9	95.6, 99.1, 133.5	83.5, 96.7, 132.6	95.7, 99.4, 133.2	93.7, 136.9, 201.4	85.2, 95.8, 131.2
α, β, γ (Å)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Rmerge overall,outer shell	0.106 (0.337)	0.079 (0.323)	0.101 (0.389)	0.097 (0.918)	0.115 (0.243)	0.042 (0.329)	0.067 (0.377)
Completeness (%)	97.4 (96.9)	100.0 (100.0)	98.2 (96.0)	95.2 (94.9)	92.9 (84.5)	99.6 (93.2)	87.3 (88.0)
⟨I/σI⟩	9.2 (3.9)	18.0 (6.2)	20.1 (6.2)	22.9 (3.6)	11.1 (6.0)	28.2 (3.9)	5.5 (5.7)
Redundancy	3.8 (3.7)	8.1 (8.0)	11.1 (9.8)	7.9 (8.1)	6.4 (5.9)	7.8 (4.8)	2.9 (3.0)
Total reflections	704,154 (98,506)	816,738 (115,612)	833,007 (102,853)	310,528 (30,999)	444,859 (52,848)	1,943,027 (54,993)	119,195 (18,065)
Unique reflections	184,285 (26,532)	100,651 (14,523)	74,999 (10,549)	39566 (3,827)	69,114 (9,010)	248,507 (11,400)	21,777 (3,161)
Refinement statistics							
Rwork/Rfree (%)	21.37/25.80	0.1567/20.47	15.72/20.14	23.52/29.65	16.66/22.31	16.39/2.22	22.90/27.47
Root mean square deviations							
Bond lengths (Å)	0.0060	0.045	0.018	0.005	0.017	0.019	0.013
Bond angles (degrees)	1.075	2.686	1.773	0.8995	1.743	1.819	1.464
Average B-factors (Å2)							
Overall							
Protein molecules	19.9	18.1	17.9	52.2	14.1	24.8	47.9
Solvent molecules	23.0 (H20)	20.6 (H2O)	19.2 (H2O)	44.4 (H2O)	21.2 (H2O)	32.3 (H2O)	36.5 (H2O)
	25.5 (EDO)	35.1 (GOL)		71.9 (EDO)		23.8 (MG)	
Ligand		33.5	13.3	40.9	8.7	22.5	46.8
No. of atoms							
Protein atoms	19641	9941	10071	9218	10011	19405	4380
Solvent atoms	1618 (H2O)	592 (H2O)	621 (H2O)	53 (H2O)	1095	1956 (H20)	38 (H20)
	4 (EDO)	4 (EDO)		4 EDO		3 (MG)	
		12 (GOL)					
Ligand atoms		30	29	28	27	100	24
Ramachandran statistics							
Most favored (%)	96.2 (602)	96.4 (603)	96.6 (604)	96.4	96.6 (604)	96.8	93.9 (588)
Additional allowed (%)	3.7 (23)	3.4 (21)	3.2 (20)	4.8	3.2 (20)	3	6.1 (38)
Disallowed (%)	0.1 (1)	0.2 (1)	0.2 (1)	0.2	0.2 (1)	0.2	0 (0)

Values in parentheses are for the highest resolution shell. RMSd, root-mean-square deviation

4.3 Results

Activity and specificity of GH20C

Recombinant GH20C protein showed activity on both pNP-GlcNAc and pNP-GalNAc. We performed a kinetic characterization using both these synthetic substrates. Using pNP-GlcNAc we obtained a K_M of 560 (±60) μM, a k_{cat} of 3650 (±330) min^{-1}, and a k_{cat}/K_M of 6.5 (±0.8) min^{-1} μM^{-1}. The kinetic analysis with pNP-GalNAc revealed a K_M of 262 (±14) μM, a k_{cat} of 1100 (±59) min^{-1}, and a k_{cat}/K_M of 4.5 (±0.3) min^{-1} μM^{-1}. Although the kinetic parameters obtained suggest some differences in the catalysis with either substrate, the $k_{cat/km}$ values suggest a similar overall efficiency of cleavage for both substrates.

Given the lack of preference of this enzyme for the gluco- or galacto-configuration, we explored its potential aglycone specificity by performing qualitative analysis using thin-layer chromatography to detect potential activity on a range of biologically relevant oligosaccharide bearing terminal GlcNAc and GalNAc. The results reveal the ability of the protein to release both GlcNAc and GalNAc from a variety of substrates with galactose (Gal), mannose (Man), or GalNAc as aglycon residues linked β1-2, β1-3, β1-4, or β1-6, indicating broad substrate cleavage capabilities (Table 8). Notably, GH20C was able to release GlcNAc from the branched β-D-GlcNAc-(1→6)-[β-D-Gal-(1→3)]-a-D-GalNAc, suggesting the ability to accommodate complex aglycons (Table 8). The only substrate that was not observed to be cleaved was β-D-GalNAc-(1→4)-β-D-Gal. Owing to the β-1,4-linkage and axial configuration of C4 in the Gal residue of this disaccharide, this sugar would adopt a kinked conformation with a significant angle between the planes of the pyranose rings. This feature separates it from the other tested sugars where the composite pyranose rings would be roughly coplanar resulting in an extended substrate conformation. These results suggest that GH20C is quite a promiscuous enzyme that could potentially increase host-glycan degradation ability of *S. pneumoniae.*

Table 8. Activity screen of GH20C

Oligosaccharide	Activity
β-D-GlcNAc-(1→6)-[β-D-Gal-(1→3)]-a-D-GalNAc	+
β-D-GalNAc-(1→3)-β-D-Gal	+
β-D-GalNAc-(1→4)-β-D-Gal	-
β-D-GlcNAc-(1→3)-β-D-GalNAc	+
β-D-GlcNAc-(1→6)-β-D-GalNAc	+
β-D-GlcNAc-(1→3)-β-D-Man	+
β-D-GlcNAc-(1→2)-β-D-Man	+
β-D-GlcNAc-(1→3)-β-D-Gal	+

Structure of GH20C

To provide molecular details beyond substrate recognition and catalysis of GH20C we performed structural studies by X-ray crystallography. GH20C readily crystallized in multiple crystal forms. An unliganded X-ray crystal structure of the enzyme was determined to 1.9 Å resolution by molecular replacement using the *S. gordonii* GcnA structure as a search model (PDBID: 2EPO) (Langley et al., 2008). The final refined model of GH20C comprised four molecules of the protein in the asymmetric unit. Electron density for residues 275-286 was consistently absent in chains A, B and C. In Chain D, electron density for residues 273-288 and 325-326 was absent as well.

The GH20C structure comprises three domains. Domain I consists of a series of a β-sheet of parallel β-strands that are exposed to the solvent and two almost parallel α-helixes situated against in outer helical surface of domain II (Figure 20A). This domain is present in all structurally characterized members of this family expected for *Aggregatibacter actinomycetemcomitans* DspB and the two GH20s found in StrH, where domain I is replaced by a FIVAR-like (found in various architectural regions) domain (Ramasubbu et al., 2005). The central catalytic domain (Domain II) is a typical $(\beta/\alpha)_8$ TIM barrel which is not only conserved among all enzymes of this family but is also common among GHs in general. This domain houses the active site situated at the C-terminal end of the β-strands. The active site residues Asp 222 and Glu 223 are located at the base of the substrate-binding cavity (Figure 20C). A small hydrophobic pocket is created by three tryptophan side chains (Trp 266, Trp 306 and Trp 373) that are thought to be involved in positioning the 2-acetamido groups of the substrate for catalysis (Figure 20C). The two conserved catalytic residues are found bordering the hydrophobic pocket. The binding pocket is open and solvent exposed at the C-terminal end of the TIM barrel, which is consistent with exo-glycosidase activity of GH20C. The active site of GH20C is highly conserved. An overlay of GH20C and GcnA active site shows that all active site residue side chains occupy almost the same space, further suggesting some similarity in terms of substrate specificity (Figure 20C). Domain III is an arrangement of helixes with four anti-parallel sheets organized in a near planar arrangement. This domain is absent in all GH20s except for GcnA. Indeed, the appended N- and C-terminal domains to the TIM

barrel of GcnA and GH20C are highly conserved resulting in a nearly identical overall architecture with an rmsd of 0.79 Å over 619 matched Cαs.

The four molecules present in the asymmetric unit were found as two identical homodimers (Figure 20B). The C-terminal domain (domain III) directly participates in dimerization, where the bundle of α-helices from one monomer stackes against the same domain of the other molecule within the same dimer. The C-terminal end of the last α-helix extends into the $(\beta/\alpha)_8$ barrel of the other dimer subunit participating in the dimer contact. The dimer interface is over 3000 $Å^2$ of buried surface area and involves about 45 hydrogen bonds and 19 salt bonds. The GH20C dimer is similar to the GcnA dimer with an rmsd of 0.85 Å over 1228 matched Cαs (Langley et al., 2008). These results suggest that dimerization may be a conserved feature of GH20 enzymes possessing domain III. However, presently no evidence suggests that dimerization is necessary for catalysis.

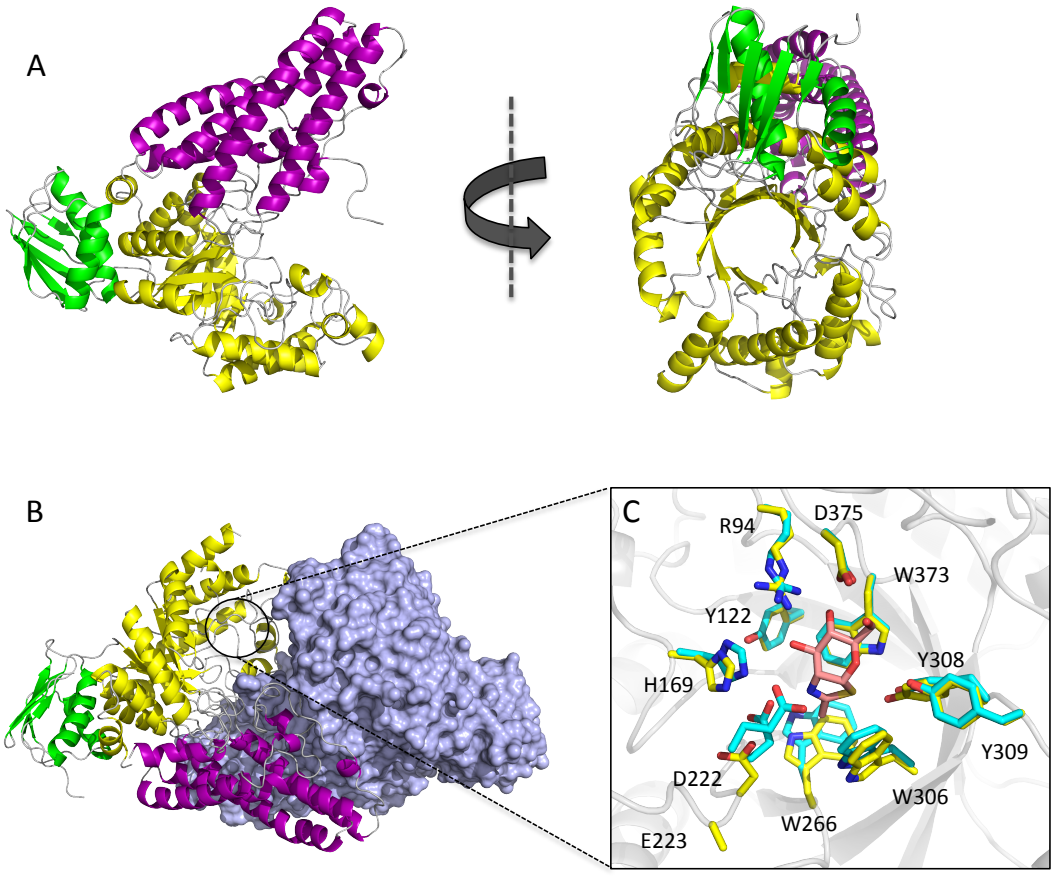


Figure 20. GH20C overall architecture.

(A) GH20C monomer with the loops represented in grey and the domains represented in green cartoon for the N-terminal domain, yellow for the central domain and purple for the C-terminal domain. (B) GH20C dimer with one monomer colored as in (A) and the other monomer represented as a slate blue surface. (C) Overlay of the GH20C active site in its retracted form (yellow stick representation for side chains) with GcnA (PDBID: 2EPN) (cyan stick representation for side chains) in complex with NAG-thiazoline (Salmon stick representation).

Structure of GH20C in complex with reaction products

In order to understand GH20C interaction with its substrates, we determined its structure in complex with relevant carbohydrates and carbohydrate analogues of both gluco- and galacto-configurations. Initially, we soaked crystals of GH20C with GlcNAc in order to obtain a product complex with this sugar. Only one dimer was present in the asymmetric unit and F_o-F_c electron density consistent with GlcNAc was found in both active sites of the dimer, however the electron density suggested only a partial occupancy or disorder (Figure 21D). GlcNAc could not be modeled in the anticipated chair conformation but instead a conformation resembling a ^4E half-envelope expected for the Michaelis-complex was observed in both active sites (Figure 21A) (Liu et al., 2011a; Pluvinage et al., 2011). In this half-envelope conformation, the acetomido group of GlcNAc is bent underneath the pyranose ring with O7 moving towards its trajectory for nucleophilic attack on C1. Both active sites of the dimer have different conformations referred to as the engaged and retracted form (Figure 21B). In the engaged form, the two catalytic residues are close to the substrate with E223 approximately 2.7 Å away from the O1 of the GlcNAc to act as the acid base and D222 interacting with the acetamido N2. In the active site of the other monomer, the loop from residue 219-225 bearing the two catalytic residues is retracted from the product, which appears to be an inadequate conformation for catalysis. Indeed, the retracted form was observed in all four monomers of the native structure. Interestingly, both retracted and engaged conformations were observed in two different crystal forms of GcnA suggesting that a subset of GHs from family 20 might undergo a conformational shift between engaged and retracted that could correspond to an active and inactive form. In the engaged active site, three tryptophans line the binding pocket creating a hydrophobic cage, which is responsible for carbohydrate-protein stacking interactions as well as creating a solvent free environment for hydrolysis to happen in a retaining manner. Furthermore, no water molecules are observed on the tryptophan-facing side, which ensures that the nucleophile attack by water occurs on the correct face of the sugar and yield a final product with a retained stereochemistry. The GlcNAc polar groups are coordinated by a series of direct and water mediated hydrogen bonds between the sugar and enzyme (Figure 21A). All of these interactions, except those

mediated by Asp222 and Glu223, are maintained in the active site even when the catalytic residues are not engaged.

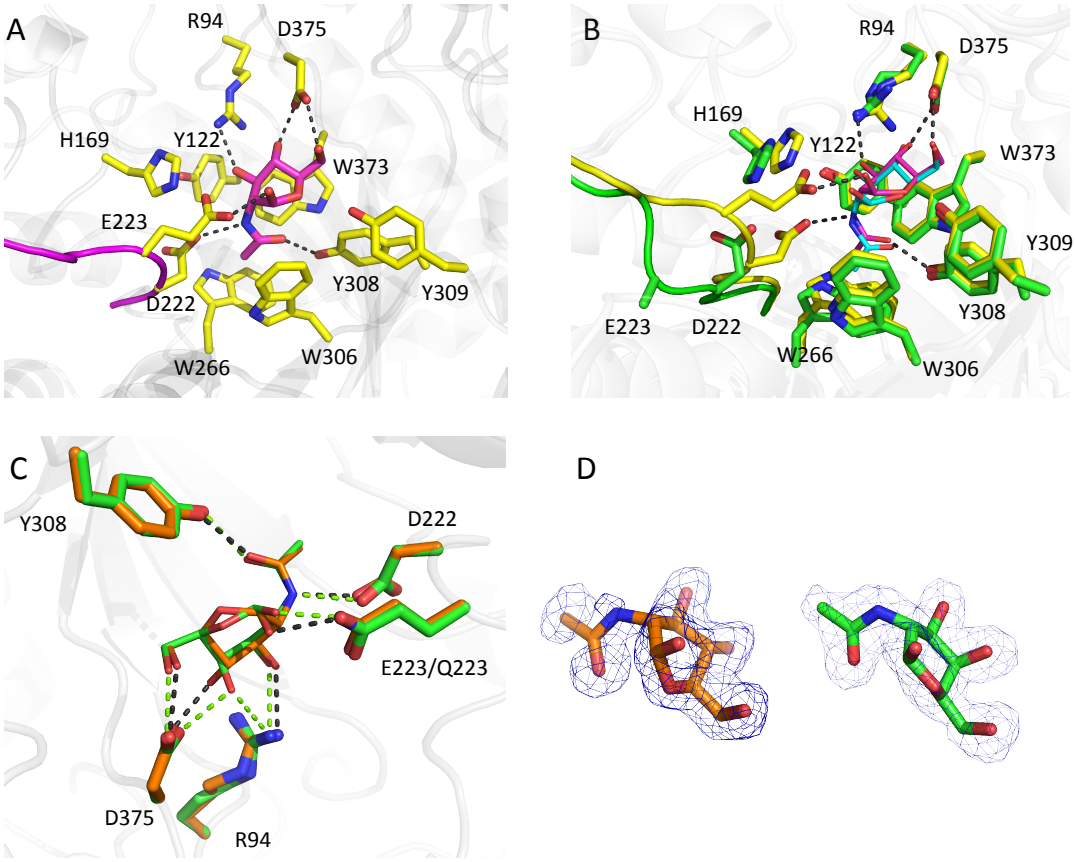


Figure 21. Structure of GH20C in complex with reaction products.

(A) Specific interactions between GH20C active site (side chains represented as yellow sticks) in its engaged form and GlcNAc (in magenta stick representation). The loop carrying the catalytic residues is represented as magenta cartoon. (B) Overlay of GH20C active site in its engaged (side chains represented as yellow sticks) and retracted (side chains represented as green sticks) form. GlcNAc from the engaged form is represented as magenta sticks and in the retracted from as cyan sticks. The moving loop carrying the catalytic residues are represented in yellow cartoon in the engaged form and green cartoon in the retracted form. (C). Overlay of GH20C complexes with GlcNAc and GalNAc. The active site of GH20C active site bound to GlcNAc is represented as orange sticks. The active site of GH20C bound to GalNAc is represented in green sticks. Hydrogen bonds between GH20C and the reaction products are represented by green dashed lines for the GalNAc complex and black dashed lines for the GlcNAc complex. (D) The F_o-F_c electron density maps for GlcNAc (represented in orange sticks) and GalNAc (represented in green sticks) is contoured at 1.5σ and 2.7σ, respectively.

To further understand GH20C aglycon specificity, we attempted to obtain non-hydrolyzed substrate complexes by catalytically inactivating the enzyme. We generated a conservative mutation of the catalytic acid/base Glu223 to a glutamine. The GH20CE223Q mutant showed low activity on aryl-glycoside substrates and soaking GH20CE223Q crystals with β-D-GalNAc-(1,3)-β-D-Gal resulted in a product complex with a GalNAc that could be modeled in both active sites of the dimer. The unambiguous F_o-F_c electron density cleanly terminated at O1 of the GalNAc, indicating that, rather than the aglycon being disordered, the substrate was likely hydrolyzed over the 30 second timescale of the crystal soaking experiment, despite the reduced activity of the mutant (Figure 21D). Both GalNAc molecules were in the expected chair conformation and fully engaged the active site residues (Figure 21C). The series of interactions observed with GalNAc are for the most part the same as the one observed with the distorted GlcNAc. Asp375 seems to be directly involved in substrate selectivity since it is interacting with the O4 of both hexosamides. Indeed this residue establishes a hydrogen bond with both axial and equatorial O4 of the hexosamides (Figure 21C).

Structure of GH20C in complex with inhibitor

Aside from differences in K_M and k_{cat} values for pNP-GlcNAc and pNP-GalNAc, GH20C seems to have an overall efficiency of hydrolysis that is similar for both substrates that only differ by their C4 stereochemistry. To further understand the role of the stereochemistry of C4, independent of catalysis, we investigated the ability of two pairs of inhibitors that only differ in their C4 configuration to inhibit GH20C using pNP-GlcNAc as a substrate.

Two of these compounds, NGT and Gal-NGT, adopt an sp^2-hybridised center at C-1 that geometrically resembles the planar oxocarbenium ion making them intermediate analogues, whereas the potency of PUGNAc and Gal-PUGNAc lies in their ability to mimic the transition state of the substrate assisted catalysis. Both sets of inhibitor tested here only differ by being C4 epimers. A Dixon plot analysis demonstrated that all four compounds were potent competitive inhibitors of GH20C (Figure 22). The PUGNAc-based compounds were the most potent with K_i values of 7.9 (±0.7) nM for PUGNAc and an approximately 8-fold better K_i for Gal-PUGNAc of 1.1 (±1.0) nM. NGT and Gal-NGT

yielded K_i values of 126 (±23) nM and 1130 (±110) nM, respectively. The inhibition constant obtained for all compounds tested were in agreement with previous inhibition data obtained for the same inhibitors against other homologues of GH20C (Sumida et al., 2012). Interestingly the K_i of Gal-PUGNAc was 8 times better than that for PUGNAc, but the K_i for GalNGT was 10 times worse than for NGT. This observation suggests that GH20C does not consistently favour an axial or equatorial O4 in these pairs of inhibitors rendering the role of the O4 stereochemistry in substrate recognition still unclear; however, it suggests that the O4 is not the only discriminating feature among these pairs of inhibitors. Therefore, to further understand GH20C selectivity for these inhibitors, we determined the X-ray crystal structure of GH20C in complex with all four compounds.

Both Gal-NGT and NGT were modeled in unambiguous electron density (Figure 23C). In the NGT complex, one monomer had a retracted active site while the other was found in the expected active form where the catalytic residues engaged the inhibitor. Both of these inhibitors were modeled in a $^4C_1$ conformation. As these two compounds are mimicking the intermediate state of the substrate assisted catalysis reaction, this conformation is to be expected. Both intermediate mimicking thiazoline rings laid in a hydrophobic crater created by Trp 266, Trp 306, Trp 308, and Trp 373. In both engaged forms of the NGT and Gal-NGT complexes, the active site residues were similarly positioned with the exception of Arg 94 side chain, which was positioned differently when bound to NGT in order to accommodate hydrogen bonding with the NGT equatorial O4 (Figure 23A).

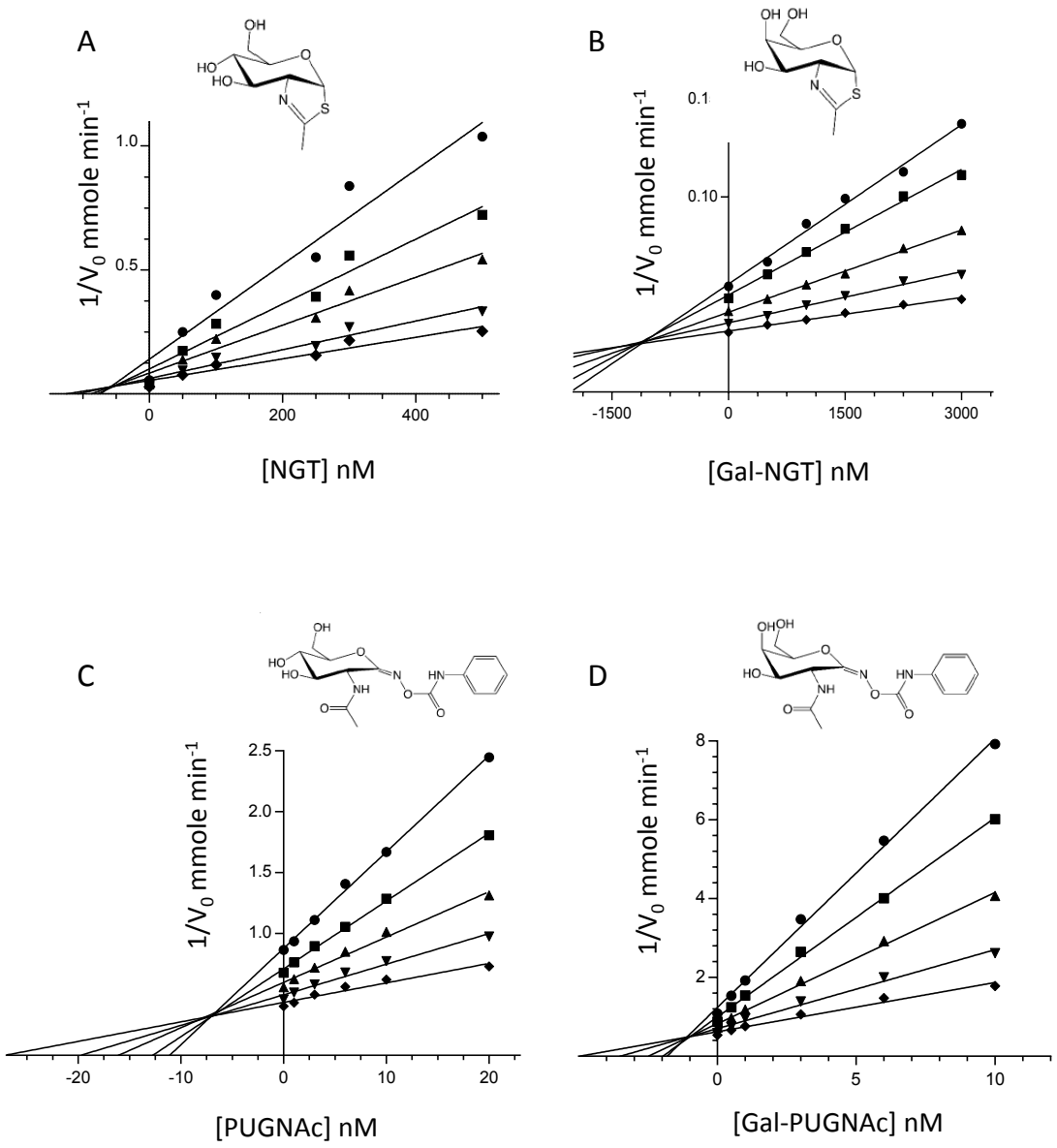


Figure 22. Inhibition of GH20C.

Dixon plot analyses used to determine the inhibition constant of (A) NGT, (B) Gal-NGT, (C) PUGNAc, (D) Gal-PUGNAc.

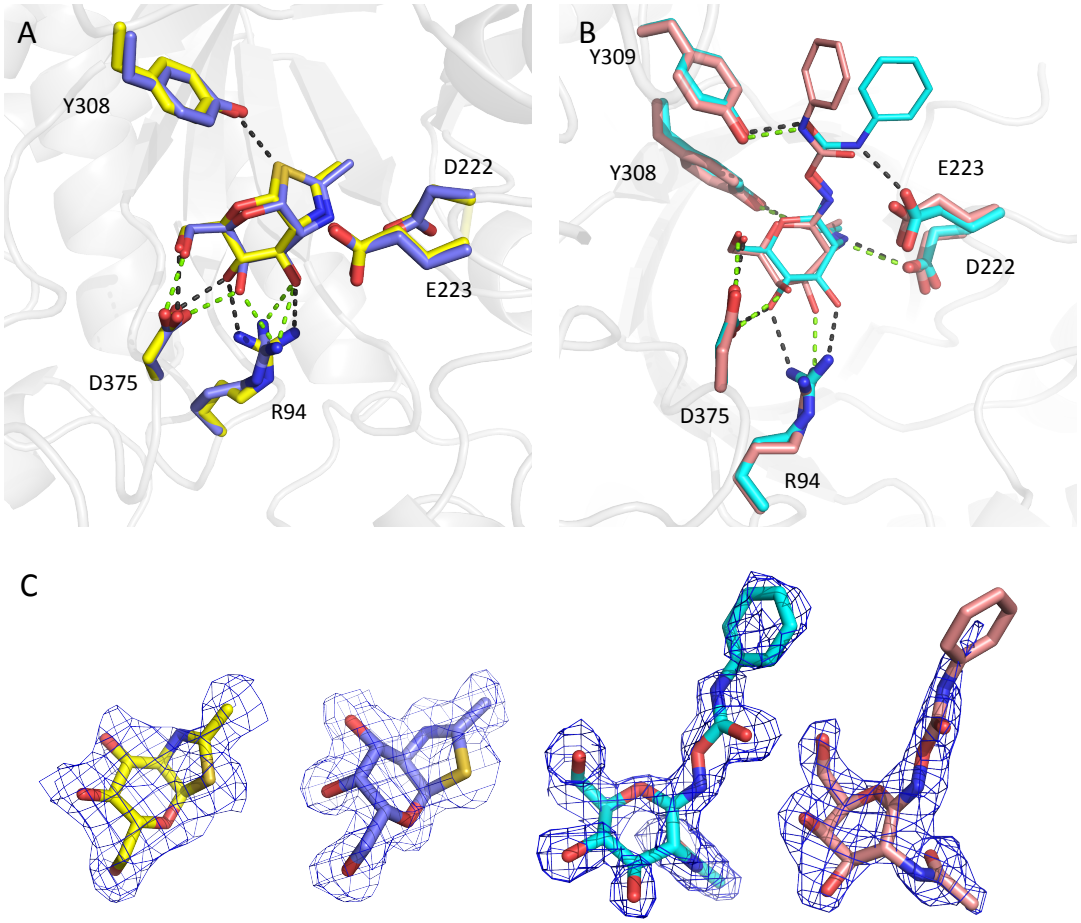


Figure 23. Structure of GH20C in complex with inhibitors.

(A) Overlay of the active site of GH20C bound to NGT (yellow sticks) and Gal-NGT (blue sticks). Hydrogen bonds are represented by green dashed lines for the NGT complex structure and black dashed lines for the Gal-NGT complex structure. (B) Overlay of the active site of GH20C bound to PUGNAc (cyan sticks) and Gal-PUGNAc (salmon sticks). Hydrogen bonds are represented by black dashed lines for the PUGNAc complex structure and green dashed lines for the Gal-PUGNAc complex structure. (C) Electron density maps for NGT, Gal-NGT, PUGNAc and Gal-PUGNAc are contoured at 1.5σ, 1.9σ, 1.5σ and 2.8σ, respectively.

The Gal-PUGNAc complex only had one monomer in the asymmetric unit and had its active site engaging the inhibitor. Gal-PUGNAc was modeled in clear electron density and appeared in the expected ^{4}E conformation (Figure 23C) (Liu et al., 2011a; Pluvinage et al., 2011). The same series of interactions seen with GalNAc and Gal-NGT was observed with the GalNAc portion of Gal-PUGNAc and GH20C (Figure 23B). The high affinity can be attributed to the mimicry of the transition state conformation as well as to the additional hydrogen bonds established between the phenyl-carbamate moiety and both active site residues Tyr 309 and Glu 223. In the PUGNAc complex structure, 3 of the 4 molecules found in the asymmetric unit had an active site in the retracted form and the phenyl-carbamate group of these three molecules were found in the same orientation as the Gal-PUGNAc (Figure 23B). However, since the catalytic residues are not engaging these three inhibitors, fewer hydrogen bonds were observed, notably the hydrogen bond interaction with Glu 223 present in the Gal-PUGNAc complex was absent here and only Tyr 309 established additional hydrogen bond interactions with this phenyl-carbamate group. The fourth molecule of the asymmetric unit was found in the active form and the fourth PUGNAc was modeled with the phenyl-carbamate group in the reverse orientation compared to the other PUGNAc molecules of this complex. However a similar hydrogen bond pattern was observed. The lack of clear electron density for the phenyl group of this compound suggested disorder of this region indicating the possibility of the inhibitor to be accommodated in more than one conformation (Figure 23C). These extra interactions between the phenyl N-carbamate group of Gal-PUGNAc/PUGNAc and the protein explain the higher affinity obtained for this pair of inhibitors compared to the thiazoline inhibitors. However, this kinetic and structural analysis does not resolve the ambiguity in the role of the O4 stereochemistry in substrate turnover by GH20C.

4.4 Discussion

GH20C from *S. pneumoniae* is an exo-acting-β-hexosaminidase that catalyses the cleavage of both β-GlcNAc and β-GalNAc. Most β-N-acetyl-hexosaminidases are active on both gluco- and galacto-configurations but these enzymes usually display preferences for one configuration or the other (Bøhle et al., 2011; Liu et al., 2011b; Pluvinage et al., 2011; Ramasubbu et al., 2005). GH20C is rather unique in that sense, due to its ability to process both types of substrates with the same overall efficiency. This was supported by

our structural study, which demonstrated that the -1 subsite conveniently accommodates both substrates with the active site residues interacting via hydrogen bond with both gluco- and galacto-configurated β-hexosamines. The arrangement of the glycon binding site (-1 subsite) residues is common to the GH20 enzyme family and includes a conserved aromatic cage. Furthermore, specificity for the acetamido group is determined by interactions between Asp 222 and Tyr 308 with N2 and O7 of the sugar. Asp 375 provides hydrogen bonding to both stereochemistries at C4. The Asp 375 side chain provides hydrogen bonding potential with both the axial O4 of GalNAc and the equatorial O4 of GlcNAc, which is consistent with the observed activity profile. Overall, the structure supports the observation that this enzyme is able to cleave both β-GlcNAc and β-GalNAc.

In addition, GH20C cleaved di- and trisaccharides with terminal GlcNAc or GalNAc suggesting that this enzyme has very limited selectivity for the aglycon moiety. This implies that unlike StrH, which is strictly specific for terminal b-(1,2)-linked GlcNAc residues from N-glycans, GH20C could target a wider range of substrates. The GH20B catalytic domain of StrH has a narrow active site responsible for the constrained specificity of this enzyme. On the other hand, the active site architecture of GH20C is open beyond the glycon binding site meaning that this protein would not likely establish interactions with sugar residues preceding the glycon. This explains the promiscuity of this enzyme observed in the activity screen. Thus, it is becoming clear that rather than having overlapping function, GH20C and StrH likely act on different biological substrates. The only substrate that GH20C did not display any activity for was β-D-GalNAc-(1,4)-β-D-Gal. This disaccharide possesses a bent conformation, which is predicted to sterically clash with the acid/base catalytic residue, Glu 223.

The open active site of GH20C explains the wide range of aglycon residues that this enzyme can accommodate. However, because of dimer formation, the entrance of the active site is actually a tunnel of about 25 Å, until the glycon binding site is reached. Indeed, a small soluble substrate such as the one used in this study would easily access the active site but processing longer sugars may be more problematic for this enzyme and

may influence its biological specificity. While the limited aglycon specificity of this enzyme allows the hydrolysis of a broad spectrum of substrates, the depth and architecture of the binding site entrance would dictate the size and three-dimensional shape of the sugars that GH20C is active on. The long 25 Å tunnel entrance suggests that an extended sugar of at least five monosaccharide residues would be preferred suggesting that GH20C may be active on sugars that extend away from their protein scaffold. In conclusion, while GH20C appears to be promiscuous *in vitro*, it is possible that *in vivo*, its substrate range may be more specific. StrH, *in vitro*, releases GlcNAc from a variety of disaccharides and displays limited N-acetylgalactosaminidase activity; however, on biologically relevant substrates this enzyme is strictly specific to releasing GlcNAc β-1,2-linked to mannose found in complex N-glycan (Pluvinage et al., 2011). Indeed, its overall active site architecture only selects for these large glycans. While we are still unsure of the biological specificity of GH20C, it is becoming clear that it is at least different than StrH.

The architecture and depth of the GH20C active site could also suggest that this enzyme would be only active on small soluble sugars. However, such sugars do not exist in the environment of *S. pneumoniae* suggesting that the bacterium could only acquire them from the degradation of host glycoproteins, which would require the action of an endo-acting enzyme. Presently, no pneumococcal enzyme is known to have such specificity. Furthermore the bacterium appears to favour an approach whereby host glycans are trimmed from the non-reducing ends by exo-acting GHs. Additionally if GH20C was intracellular, its substrate(s) would have to be imported by a carbohydrate transporters that would display specificity for a potentially wide range of substrates. Based on this, we propose that GH20C is a non-classically secreted enzyme, where its extracellular location would allow easy access to the non-reducing terminal GlcNAc or GalNAc residues of a wide variety of substrates of at least five residues without the need for any endo-acting enzymes or transporters. The lack of a SP and known cell-wall anchoring motif in the GH20C sequence suggests that this enzyme would be secreted/cell-wall-anchored through an unknown mechanism. Indeed, many examples of pneumococcal enzymes

lacking SP or anchoring motif, are found in the extracellular compartment (Jeong et al., 2009; Pérez-Dorado et al., 2012).

The structural analysis of GH20C demonstrates that this enzyme undergoes a change in the active site upon binding both products of catalysis and inhibitors, whereby the catalytic residues go from a retracted state to engaged state in a manner consistent with being positioned for catalysis. An identical conformational change was observed for the *S. gordonii* homologue of GH20C, GcnA, whereby the catalytic residues go from a retracted state to engaging the inhibitor in a manner consistent with being positioned for catalysis. Indeed, the loop in GH20C containing the catalytic residues (the catalytic loop) goes from retracted and disordered to engaged and ordered. The fact that the retracted form was also observed in product and inhibitor bound forms of the enzyme suggests that engagement of the catalytic loop is not necessarily a prerequisite for the enzyme to bind its ligands. This suggests that GH20C may proceed via a trajectory of substrate recognition that may go from the retracted form, to the substrate-occupied retracted form, to the substrate-stabilized engaged form.

Based on the observed k_{cat}/K_M, GH20C hydrolysed pNP-GlcNAc and pNP-GalNAc with similar efficiency, however different, yet compensating, kinetic parameters were obtained. To provide insight into this, we explored the inhibition of the enzyme using pairs of inhibitors that only differed by their galacto- and gluco-configuration. The pattern of K_is obtained correlated with previously observed data for other GH20s: Gal-PUGNAc < PUGNAc < NGT < Gal-NGT. As we observed with the substrates, GH20C does not have a consistent preference for the axial or equatorial C4 hydroxyl. However, the data suggest that when an aglycon is present, such as the phenylcarbamate of Gal-PUGNAc and PUGNAc, the axial O4 of the galacto-configuration is preferred for binding. On the other hand, when the aglycon is absent, such as for the thiazoline inhibitors, the equatorial O4 is favoured. Overall, the specificity of GH20C is not likely determined by the configuration of O4 but is rather determined by the size restrictions of the active site tunnel.

The biological role of GH20C and its possible implication in virulence are presently unknown. While many carbohydrate-degrading enzymes from *S. pneumoniae* play a role in various aspects of virulence, such as adherence and complement modulation, a common role of these enzymes is their participation in nutrient acquisition. Since free sugars are scarce in the airway, this process is thought to help support microbial persistence in the host. Based on this study we can at least hypothesise that GH20C complements the very specific StrH. Indeed, a more promiscuous enzyme such as GH20C would potentially expand the overall ability of *S. pneumoniae* to degrade the hexosaminide sugars from host glycans, thus greatly increasing the nutrient reservoir in the host.

Chapter 5: Discussion

Host glycan utilization is a key process in the establishment of colonization and disease by pathogens (Ficko-Blean et al., 2012; Limoli et al., 2011; Meng et al., 2010). Numerous studies have demonstrated the central role of host glycan metabolism in various aspects of *S. pneumoniae* virulence ranging from colonization to complement modulation (Banerjee et al., 2010; Dalia et al., 2010; Limoli et al., 2011; Pluvinage et al., 2011; Uchiyama et al., 2009). It is well known that GHs from pathogens play an essential role in the development and severity of disease. One of the best examples of such enzymes is the neuraminidase A from the influenza virus. This enzyme is responsible for the degradation of terminal sialic acid often found on host glycans (Crusat and De Jong, 2007). Influenza neuraminidase activity is essential for the spread of the virus in the respiratory tract and therefore has been the target of many inhibitors developed to reduce the severity of disease.

It is well established that *S. pneumoniae*'s ability to cause disease relies greatly on its host carbohydrate degradation capabilities (Banerjee et al., 2010; Brittan et al., 2012; Dalia et al., 2010; Pluvinage et al., 2011). In the last few decades, numerous studies have revealed that *S. pneumoniae* depends on complex carbohydrate processing machineries for virulence. Despite the effort undertaken to characterize these systems, a complete molecular understanding of these pathways is still lacking. The research presented here sheds light on a novel pneumococcal carbohydrate processing system involved in pneumococcal virulence. Indeed, through bioinformatic analyses we have identified a carbohydrate-processing locus (CPL) that encodes for multiple GHs and a ROK protein. This virulence-associated carbohydrate-degrading "hot spot" is highly conserved throughout all *S. pneumoniae* strains and through a large number of firmicutes, and co-occurs in association with genes encoding for EndoD and a novel ABC transporter, suggesting a functional association. Based on the genomic context of these components as well as their known or predicted functions, we hypothesised the CPL to encode for a

concerted pathway responsible for the harvest and complete degradation of host N-glycan.

In this work we have characterised the putative extracellular components of the pneumococcal N-glycan degrading pathway. Our structural and functional analysis indicated that the GH92 α-1,2-mannosidase is likely responsible for removing the terminal α-1,2-mannose residues from high mannose N-glycans. Based on the fact that GH92 would need to release these sugars to allow the cell-wall anchored EndoD to accommodate $Man_5GlcNAc_2$-Asn, we strongly suggest that GH92 is a non-classically secreted pneumococcal protein lacking a SP or a known motif for cell-wall anchoring. While GH92 is active on α-1,2-mannobiose, its activity on high-mannose N-glycan type sugars still has to be demonstrated in order to further validate its role in this metabolic pathway. In parallel, it has been demonstrated that NanA, BgaA and StrH sequentially remove the terminal sialic acid, galactose and GlcNAc of the distal arms of complex N-glycans (King et al., 2006). It has been shown that growth of *S. pneumoniae* on fetuin, which bears complex N-glycans, is dependent on these three exo-acting enzymes. A previous study demonstrated that EndoD is strictly active on complex and high-mannose N-glycans that lacked the terminal α-1,2-mannose decoration and the arms of complex N-glycans, respectively. In our model, degradation of high-mannose N-glycan or complex N-glycans by GH92 and NanA, BgaA and StrH, respectively, yields a substrate for the cell-wall-anchored EndoD to hydrolyse and release (Yamamoto et al., 2005). Once the core glycans are free from the polypeptide chain, one of two events can take place: either the sugars are further degraded by additional extracellular mannosidases or the products of EndoD are shuttled inside the bacterium for further processing. Insight into the specificity of the CPL-associated ABC transporter would shed light on the events that occur after the action of EndoD. Our structural and functional analysis of the SBP of this transporter indicates that the sugars released by EndoD are imported without any further extracellular degradation and that $ABC_{N\text{-}glycan}$ modulates the transport of a range of N-glycans. Indeed, $SBP_{N\text{-}glycan}$ recognises a conserved feature of all types of N-glycans, their reducing-end. Thus, this study revealed a novel import mechanism for branched N-glycans in *S. pneumoniae* that further suggests that processing of the glycan core occurs

within the cell. Based on homology, we expect this import mechanism to be conserved across Firmicutes. Finally, the predicted α-mannosidase GH38 and the characterised GH125 would work sequentially in order to first process the α-1,3-branches, followed by the degradation of the α-1,6-linked mannose residues. Gregg *et al.* previously demonstrated that GH125 is active on the α-1,6-linked mannose residues only when the α-1,3- arm is lacking (Gregg et al., 2011). From this we can infer that GH38 acts prior to GH125. While the activity of GH38 has never been demonstrated, we suspect this enzyme to be active on this type of linkage because of first, high homology with the GH38 from *S. pyogenes* (which display α-1,3-mannosidase activity) and second, by a process of elimination. GH38 is the only uncharacterised enzyme predicted to be an α-mannosidase and its genomic synteny with GH125 strongly suggests that they work together within the cytoplasm to degrade the core of N-glycans. Finally, we postulate that the ROK protein encoded by the CPL would phosphorylate the final disaccharide product prior to further metabolism. The bacterial ROK family contains transcriptional repressors, sugar kinases and as yet uncharacterised open reading frames that all share a similar overall fold. However, ROK repressors possess an additional helix-turn-helix DNA-binding domain that is absent in the sugar kinase. While the function of the CPL-encoded ROK protein is still unknown, its crystal structure has been solved (PDBID: 2GUP) by the structural genomics consortium. The absence of DNA-binding domain suggests that this protein is likely to be a sugar kinase responsible for phosphorylating the final product of the N-glycan processing pathway. No additional GH has been identified that would be specific for Man-β-1,4-GlcNAc suggesting that ROK may phosphorylate this disaccharide. Another possibility is that ROK is active on monosaccharides produced by the degradation of Man-β-1,4-GlcNAc by an additional enzyme whose gene is not encoded by the CPL. On the basis of the previous studies and the work presented in this dissertation, we propose an N-glycan processing pathway represented in figure 24, that would be responsible for the complete degradation, liberation and transport of complex, hybrid and high-mannose N-glycans by *S. pneumoniae*. Future work would involve uncovering the function of the uncharacterised CPL components, such as GH38 and ROK, as well as determining the cellular localisation of all CPL components that lack a SP or a known cell-wall anchoring motif, which includes GH92, GH38 and GH125.

These future directions would help provide a more exhaustive description of how *S. pneumoniae* degrades these glycans.

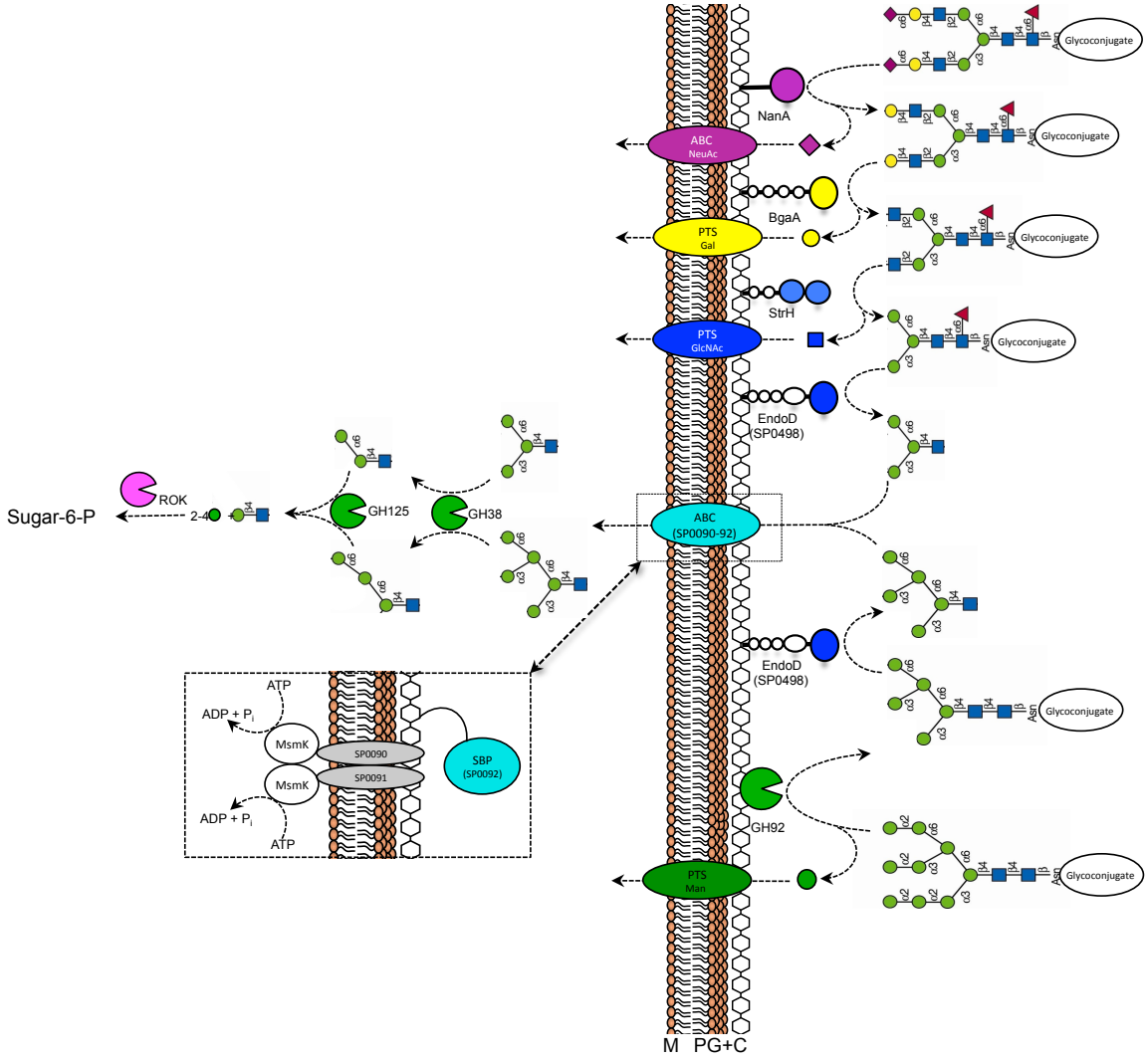


Figure 24. Proposed N-glycan processing pathway in *S. pneumoniae*

Flow of substrates through the N-glycan pathway. Enzymes are represented by pies, which are color-coded according to their general known or proposed activities: purple – sialidase, yellow – β-galactosidase, blue - β- hexosaminidase, green – α-mannosidase, and pink – sugar kinase. Substrate specificity is only known for NanA, BgaA, EndoD, and GH125 based on the literature; and for GH92 based on this study. Cellular location is only known for NanA, BgaA, StrH, and EndoD, which bear Gram-positive LPXTG cell wall anchoring motifs. The functions of the PTS transporters and the sialic acid transporter are derived from the literature but the function of the putative N-glycan specific ABC transporter is based on this study. The proposed architecture of the ABC_N-

$_{glycan}$ is shown in the inset. M refers to the lipid bilayer and PG+C a combination of the peptidoglycan and capsule.

The conservation of the CPL in other bacteria from diverse niches suggests that microbial N-glycan foraging is relevant in various biological processes (Byers et al., 1999; Homer et al., 2001; Langley et al., 2004, 2008). The organization and conservation of the CPL in firmicutes suggest that these Gram-positive bacteria also possess an N-glycan utilisation pathway probably nearly identical to the one described in *S. pneumoniae*. Indeed, all CPLs found in these firmicutes encode for homologues of the pneumococcal N-glycan-processing proteins described here. Indeed, these organisms possess a GH92, a GH85, a GH38 and a GH125 suggesting that the sequential degradation of N-glycan may occur in a similar manner. Homologues of $ABC_{N\text{-}glycan}$ are also found in other CPLs suggesting that this import mechanism is conserved throughout these firmicutes and that their SBPs likely have a similar specificity and structural features to the one found in *S. pneumoniae.* Most of these CPL-encoding firmicutes such as *S. mitis*, *S. oralis*, *S. gordonii*, etc, are human pathogens, which reinforces the idea that N-glycan depolymerisation is of high importance during infection.

Few examples of N-glycan degrading machinery have been identified in Gram-negative bacteria such as *B. thetaiotaomicron and C. canimorsus* (Cuskin et al., 2015; Renzi et al., 2011). The best described is the N-glycan-processing system from the prominent gut microbe *B. thetaiotaomicron*. This bacterium possesses a polysaccharide utilisation loci (PUL) that encodes for a Sus-like (Starch Utilisation System) system responsible for the degradation of human high-mannose N-glycan (Bolam and Koropatkin, 2012; Cuskin et al., 2015). These Sus-like systems found in *Bacteroides* species are tuned to sensing and capturing human and dietary glycans found in the very competitive environment of the gut. A previous study conducted by *Cuskin et al,* described the *B. thetaiotaomicron* Sus-like system responsible for high mannose N-glycan utilization (Cuskin et al., 2015). This system starts with an endo-acting-N-acetyl-glucosaminidase from family 18, which cleaves high-mannose N-glycan from their polypeptide. In contrast to EndoD from *S. pneumoniae*, this GH18 is active on native high-mannose N-glycan that still bears the terminal α-1,2-mannose decoration meaning that no extracellular α-mannosidase is required. The product, $Man_9GlcNAc$, is held on the surface of the bacterium through a mannose-binding protein, while another glycan-binding lipoprotein (SusD) recognizes

the GlcNAc residue at the reducing end of the glycan. This is thought to orientate the glycan into the outer membrane porin that resembles a TonB-dependant transporter (SusC) for transport into the periplasm, where α-mannosidases further depolymerise the glycan.

The *S. pneumoniae* versus *B. thetaiotaomicron* N-glycan-degrading machinery displays major differences. The pneumococcal N-glycan-processing pathway possesses only one recognition site with the $SBP_{N\text{-}glycan}$ dictating the specificity and trapping almost completely the glycan for transport. Conversely the Sus-like N-glycan system possesses many recognition elements including SusD and the surface mannose-binding protein that are thought to possess shallow binding pockets (Bolam and Koropatkin, 2012). The role of these glycan-binding lipoproteins found in Sus-like systems is to offer many surface binding sites, usually with subtle specificity differences, to compensate for large structural variety in glycans and dietary fluctuation. Furthermore, the cellular organisation and strategy of these two N-glycan degrading systems varies. In *B. thetaiotaomicron*, only one GH is found outside the cell and is responsible for releasing the sugar from host tissues meaning that intact high mannose is transported into the periplasm where the rest of the degradation occurs. In *S. pneumoniae*, many extracellular enzymes sequentially degrade N-glycan for the specialised ABC transporter to modulate the transport of a range of smaller N-glycan derived products. Both N-glycan degrading systems are adapted to the respective niches of these bacteria. As a matter of fact, the pneumococcal carbohydrate-processing pathways are tailored to the degradation of consistent source of host glycans for colonization and/or infection, while the Sus-like systems from *Bacteroides*, have evolved to help these species sequestering variable dietary glycans, as well as host glycans, at their surface and taking them away from competitive bacteria.

It appears that critical differences can be seen between N-glycan degradation systems from Gram-positive and Gram-negative bacteria. Indeed, bioinformatic analyses suggest that in Gram-positive bacteria the N-glycan utilization machinery probably resembles the pneumococcal pathway, where extracellular exo-acting GHs trim N-glycans. Trimming is

followed by an endo-acting GH that releases the oligosaccharide from the protein scaffold. The sugar is then imported via an ABC transporter. Whereas, generally Gram-negative N-glycan-processing systems seem to have only one endo-acting extracellular GH that can accommodate native N-glycans (Cuskin et al., 2015; Renzi et al., 2011). Furthermore, both N-glycan processing pathways from *B. thetaiotaomicron and C. canimorsus* comprise several surface glycan-binding lipoproteins responsible for sequestrating the sugar at the surface of these bacteria allowing the endo-acting-N-acetyl-glucosaminidase to detach them from their protein scaffold. This endo-acting GH from the *C. canimorsus* system can also accommodate native complex-N-glycan bearing the terminal sialic acid, galactose and GlcNAc. The released N-glycans from both Gram-negative systems are then imported via a TonB-dependant-like transporter and further degraded by the sequential action of multiple exo-acting periplasmic GHs. Further characterisation of these N-glycan degradation systems from diverse organisms will provide a better model for N-glycan processing by bacteria and will shed light on the role of N-glycan degradation in a potentially enormous variety of biological settings.

Most components of the pneumococcal N-glycan pathway have been identified in large-scale virulence screens, suggesting the importance of host N-glycan degradation in pneumococcal pathogenesis. In support of the hypothesis that these proteins are virulence factors, we have demonstrated the importance of EndoD and GH92 in the development of pneumococcal disease in mice. We confirmed for the first time that GH92 and EndoD are required for full virulence and participate in the progression of pneumonia and bacteremia in mice. These results, combined with the lack of virulence attenuation for the $ABC_{N\text{-}glycan}$ deficient strain, suggest that the outcome of N-glycan depolymerisation in virulence is not coupled to the transport of these sugars. This implies the participation of GH92 and EndoD in pathogenic pathways that do not involve utilization of N-glycan for growth. Furthermore, the importance of EndoD and GH92 in the development of bacteremia in mice has been examined and suggest that these enzymes might be involved in the epithelial cell-blood progression and/or survival in the blood. It has been demonstrated that NanA, BgaA and StrH are involved in providing protection against phagocytic-killing through modulation of the complement. This innate host defence

mechanism is quite inefficient since *S. pneumoniae* has developed strategies against C3b deposition on its surface (Hyams et al., 2010; Jarva, 2003). One possible mechanism would be through deglycosylation of complement components from the alternative pathway. Although it is not clear whether EndoD and GH92 play a direct role in this process, we know that disruption of NanA, BgaA and StrH results in an increase in C3b deposition and neutrophil-mediated killing, suggesting the importance of complex N-glycan degradation in complement modulation by *S. pneumoniae* (Dalia et al., 2010; Pluvinage et al., 2011). Complement components are glycoproteins whose glycan portion are essential for their proper function; thus, deglycosylation of these molecules impairs complement effectiveness and prevents C3b binding to the bacterium surface (Ritchie et al., 2002). More experimental work needs to be performed in order to further elucidate the individual contribution of the N-glycan-processing components to virulence and shed light on their role in the host-pathogen interaction; however, processing of glycans on complement proteins in order to compromise the innate immune response is one likely mechanism by which these GHs could act as virulence factors.

The CPL encodes for other enzymes, such as GH20C, that do not fit into the N-glycan processing system. We believe this enzyme is a generalist uncapping enzyme whose activity is essential to remove terminal sugar units of host glycans and expose them to other enzymes that are part of other glycan-degrading systems. Most of the linkages cleaved by GH20C are found in O-glycans, glycosphingolipids and N-glycans, which may imply a role for GH20C in the breakdown of these carbohydrates. This enzyme is predicted to be intracellular; however, based on the possibility of non-classically secreted proteins by *S. pneumoniae* and the known terminal linkages targeted by this enzyme, we believe that GH20C is secreted or found attached to the cell-wall through an uncharacterised secretion mechanism.

Interestingly, GH20C is active on an abundance of linkages found in glycopshingolipids. These glycan structures constitute up to 80% of the brain's glycans. Despite the yet unknown biological specificity of GH20C, the role of this enzyme in the development of meningitis should be investigated. Meningitis can be the result of *S. pneumoniae*

progressing from the nasopharynx directly to the nervous system through axonal transport into olfactory nerves which, in mice, constitute about 47% of the nasal surface and thus by-passing the blood brain barrier (van Ginkel et al., 2003). The potential degradation of these glycans from the olfactory nerves by GH20C could facilitate pneumococcal progression to the brain. In spite of the lack of identification of GH20C in previously performed large-scale virulence screens, the potential role of this enzyme in disease should be explored, particularly since false negatives frequently occur in large scale STM and microarray studies and since genomically associated enzymes are known to participate in pneumococcal virulence. We therefore suspect that GH20C is part of other carbohydrate-degrading machinery, possibly involved in virulence, and may play a role in glycosphingolipid degradation. Future directions would include uncovering novel pneumococcal glycan-processing systems that would involve GH20C. Seeing as bacterial genomes are size restricted, encoding for a promiscuous glycoside hydrolase with a large substrate range is advantageous for survival in an environment where free sugars are sparse and complex carbohydrates very abundant, hence perhaps increasing colonization and invasion capabilities of *S. pneumoniae.*

The threat presented by *S. pneumoniae* due to the rise of antibiotic resistance is greater than ever. This has reinvigorated efforts toward gaining a greater understanding of the molecular details behind the pathogenesis of *S. pneumoniae* in order to develop ways of better controlling this pathogen. One approach is to target specific virulence associated mechanisms that would prevent the bacterium from becoming pathogenic, allowing the immune system to control colonization and infection. Rather than being bactericidal and disturb the natural microbiome of the nasopharynx, these anti-infective therapies would reduce severe disease and help the immune system to clear infection as well as strongly reducing selective pressure. Extracellular GHs participate in virulence-associated glycan degradation, and so are of interest in the development of novel therapeutics against *S. pneumoniae*. One strong advantage for targeting such enzymes is that they are often conserved throughout all *S. pneumoniae* strains. Unlike the serotype-dependent vaccines, targeting conserved pneumococcal proteins would provide protection against all strains and overcome the phenomenon of serotype replacement. These enzymes often initiate

virulence-associated carbohydrate processing pathways; thus inhibiting them using specific inhibitors would impede the entire pathway thus reducing virulence. Therefore, understanding their mechanism of catalysis and specificity is essential for the development of potential inhibitors. This approach has been used already and shown to be effective in the case of the anti-viral drug Tamiflu, a neuraminidase inhibitor that is also effective at reducing pneumococcal infection (Crusat and De Jong, 2007; Trappetti et al., 2009). Lastly, since the role of GH92 and EndoD in virulence has been confirmed, these enzymes could potentially become targets for the development of novel inhibitors. However, careful drug design would be necessary to avoid inhibition of human homologues. Anti-infective therapies used in combination with conventional therapeutics such as antibiotics could become an effective approach to treat and control pneumococcal infection and have a huge impact on reducing selective pressure caused by overuse of antibiotics.

The incredible ability of *S. pneumoniae* to adapt to any clinical interventions previously tried obligates us to gain a greater understanding of how this pathogen interacts with its host. Since carbohydrate degradation is essential for the lifestyle of this bacterium, as well as for causing disease and evading the immune response, it is essential to gain further insight into the ability of this bacterium to utilise these structures and to understand the role of these processes in pathogenesis. In the research presented here we have bridged a gap in knowledge on how *S. pneumoniae* processes N-glycans by characterizing novel components of this pathway that are essential for diverse aspects of virulence. This original work is of fundamental relevance in the understanding of carbohydrate processing in the host-pathogen interaction but also in other bacteria from diverse environments that exploit similar glycans, such as the ones found in the gut microbiome. Although drug development is not our direct goal, an anticipated outcome of this research is to provide critical molecular detail insights for the future development of novel therapeutics targeting key processes involving carbohydrate-processing enzymes.

Bibliography

Abbott, D.W., and Boraston, A.B. (2007). Specific recognition of saturated and 4,5-unsaturated hexuronate sugars by a periplasmic binding protein involved in pectin catabolism. J. Mol. Biol. *369*, 759–770.

Abbott, D.W., Macauley, M.S., Vocadlo, D.J., and Boraston, A.B. (2009). Streptococcus pneumoniae endohexosaminidase D, structural and mechanistic insight into substrate-assisted catalysis in family 85 glycoside hydrolases. J. Biol. Chem. *284*, 11676–11689.

Abbott, D.W., Higgins, M. a, Hyrnuik, S., Pluvinage, B., Lammerts van Bueren, A., and Boraston, A.B. (2010). The molecular basis of glycogen breakdown and transport in Streptococcus pneumoniae. Mol. Microbiol. *77*, 183–199.

Agarwal, V., Asmat, T.M., Dierdorf, N.I., Hauck, C.R., and Hammerschmidt, S. (2010). Polymeric immunoglobulin receptor-mediated invasion of Streptococcus pneumoniae into host cells requires a coordinate signaling of SRC family of protein-tyrosine kinases, ERK, and c-Jun N-terminal kinase. J. Biol. Chem. *285*, 35615–35623.

Avery, O., and Dubos, R. (1931). The protective action of a specific enzyme against type 3 pneumococcus infection in mice. J. Exp. Med. *54*, 73–90.

Banerjee, A., Van Sorge, N.M., Sheen, T.R., Uchiyama, S., Mitchell, T.J., and Doran, K.S. (2010). Activation of brain endothelium by pneumococcal neuraminidase NanA promotes bacterial internalization. Cell. Microbiol. *12*, 1576–1588.

Barik, S. (1996). Site-directed mutagenesis in vitro by megaprimer PCR. Methods Mol. Biol. *57*, 203–215.

Bergmann, S., and Hammerschmidt, S. (2006). Versatility of pneumococcal surface proteins. Microbiology *152*, 295–303.

Berntsson, R.P., Smits, S.H.J., Schmitt, L., Slotboom, D.-J., and Poolman, B. (2010). A structural classification of substrate-binding proteins. FEBS Lett. *584*, 2606–2617.

Berry, a. M., and Paton, J.C. (2000). Additive Attenuation of Virulence of Streptococcus pneumoniae by Mutation of the Genes Encoding Pneumolysin and Other Putative Pneumococcal Virulence Proteins. Infect. Immun. *68*, 133–140.

Bidossi, A., Mulas, L., Decorosi, F., Colomba, L., Ricci, S., Pozzi, G., Deutscher, J., Viti, C., and Oggioni, M.R. (2012). A functional genomics approach to establish the complement of carbohydrate transporters in Streptococcus pneumoniae. PLoS One *7*, e33320.

Bogaert, D., Groot, R. De, and Hermans, P.W.M. (2004). Streptococcus pneumoniae colonisation Streptococcus pneumoniae colonisation : the key to pneumococcal disease. *4*, 144–154.

Bøhle, L.A., Mathiesen, G., Vaaje-Kolstad, G., and Eijsink, V.G.H. (2011). An endo-β-N-acetylglucosaminidase from Enterococcus faecalis V583 responsible for the hydrolysis of high-mannose and hybrid-type N-linked glycans. FEMS Microbiol. Lett. *325*, 123–129.

Bolam, D.N., and Koropatkin, N.M. (2012). Glycan recognition by the Bacteroidetes Sus-like systems. Curr. Opin. Struct. Biol. *22*, 563–569.

Boraston, A.B., Creagh, a. L., Alam, M.M., Kormos, J.M., Tomme, P., Haynes, C. a., Warren, R.A.J., and Kilburn, D.G. (2001). Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from Thermotoga maritima xylanase 10A. Biochemistry *40*, 6240–6247.

Boraston, A.B., Bolam, D.N., Gilbert, H.J., and Davies, G.J. (2004). Carbohydrate-binding modules : fine-tuning polysaccharide recognition. *781*, 769–781.

Breiman, Robert F., John S. Spika, V.J., and Navarro, Paul M.Darden, C.P.D. (1990). Pneumococcal Bacteremia in Charleston County, South Carolina. A Decade Later. Arch. Intern. Med. *150*, 1401–1405.

Bridy-Pappas, A.E., Margolis, M.B., Center, K.J., and Isaacman, D.J. (2005). Streptococcus pneumoniae: description of the pathogen, disease epidemiology, treatment, and prevention. Pharmacotherapy *25*, 1193–1212.

Brittan, J.L., Buckeridge, T.J., Finn, a, Kadioglu, a, and Jenkinson, H.F. (2012). Pneumococcal neuraminidase A: an essential upper airway colonization factor for Streptococcus pneumoniae. Mol. Oral Microbiol. *27*, 270–283.

Buckwalter, C.M., and King, S.J. (2012). Pneumococcal carbohydrate transport: Food for thought. Trends Microbiol. *20*, 517–522.

Burnaugh, A.M., Frantz, L.J., and King, S.J. (2008). Growth of Streptococcus pneumoniae on human glycoconjugates is dependent upon the sequential activity of bacterial exoglycosidases. J. Bacteriol. *190*, 221–230.

Byers, H.L., Tarelli, E., and Homer, K.A. (1999). Sequential deglycosylation and utilization of the N-linked , complex-type glycans of human α 1 -acid glycoprotein mediates growth of Streptococcus oralis. *9*, 469–479.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. *37*, D233–D238.

Cao, Y., Rocha, E.R., and Smith, C.J. (2014). Efficient utilization of complex N-linked glycans is a selective advantage for Bacteroides fragilis in extraintestinal infections. Proc. Natl. Acad. Sci. U. S. A. *111*, 12901–12906.

Cobucci-Ponzano, B., Conte, F., Strazzulli, A., Capasso, C., Fiume, I., Pocsfalvi, G., Rossi, M., and Moracci, M. (2010). The molecular characterization of a novel GH38 α-mannosidase from the crenarchaeon Sulfolobus solfataricus revealed its ability in de-mannosylating glycoproteins. Biochimie *92*, 1895–1907.

Cron, L.E., Bootsma, H.J., Noske, N., Burghout, P., Hammerschmidt, S., and Hermans, P.W.M. (2009). Surface-associated lipoprotein PpmA of Streptococcus pneumoniae is involved in colonization in a strain-specific manner. Microbiology *155*, 2401–2410.

Crusat, M., and De Jong, M.D. (2007). Neuraminidase inhibitors and their role in avian and pandemic influenza. Antivir. Ther. *12*, 593–602.

Cuskin, F., Lowe, E.C., Temple, M.J., Zhu, Y., Cameron, E. a, Pudlo, N. a, Porter, N.T., Urs, K., Thompson, A.J., Cartmell, A., et al. (2015). Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. Nature *517*, 165–169.

Dalia, A.B., Standish, A.J., and Weiser, J.N. (2010). Three surface exoglycosidases from Streptococcus pneumoniae, NanA, BgaA, and StrH, promote resistance to opsonophagocytic killing by human neutrophils. Infect. Immun. *78*, 2108–2116.

Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. Structure *3*, 853–859.

Davies, G.J., Wilson, K.S., and Henrissat, B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochem. J. *321 (Pt 2*, 557–559.

Deutscher, J., Francke, C., and Postma, P.W. (2006). How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. Microbiol. Mol. Biol. Rev. *70*, 939–1031.

Dupoiron, S., Zischek, C., Ligat, L., Carbonne, J., Boulanger, A., Dugé de Bernonville, T., Lautier, M., Rival, P., Arlat, M., Jamet, E., et al. (2015). The N-Glycan cluster from Xanthomonas campestris pv. campestris: a toolbox for sequential plant N-glycan processing. J. Biol. Chem. *290*, 6022–6036.

Emsley, P., and Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. Acta Crystallogr. Sect. D Biol. Crystallogr. *60*, 2126–2132.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. Sect. D Biol. Crystallogr. *66*, 486–501.

Erni, B. (2013). The bacterial phosphoenolpyruvate: Sugar phosphotransferase system (PTS): An interface between energy and signal transduction. J. Iran. Chem. Soc. *10*, 593–630.

Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallogr. Sect. D Biol. Crystallogr. *62*, 72–82.

Fenoll, A., Aguilar, L., Vicioso, M.-D., Gimenez, M.-J., Robledo, O., and Granizo, J.-J. (2011). Increase in serotype 19A prevalence and amoxicillin non-susceptibility among paediatric Streptococcus pneumoniae isolates from middle ear fluid in a passive laboratory-based surveillance in Spain, 1997-2009. BMC Infect. Dis. *11*, 239.

Fernandes, M.J., Yew, S., Leclerc, D., Henrissat, B., Vorgias, C.E., Gravel, R. a, Hechtman, P., and Kaplan, F. (1997). Identification of candidate active site residues in lysosomal beta-hexosaminidase A. J. Biol. Chem. *272*, 814–820.

Ferreira, M.J., and De Sá-Nogueira, I. (2010). A multitask ATPase serving different ABC-type sugar importers in Bacillus subtilis. J. Bacteriol. *192*, 5312–5318.

Ficko-Blean, E., Stuart, C.P., Suits, M.D., Cid, M., Tessier, M., Woods, R.J., and Boraston, A.B. (2012). Carbohydrate recognition by an architecturally complex α-N-acetylglucosaminidase from Clostridium perfringens. PLoS One *7*, e33524.

Fiore, A.E., Levine, O.S., Elliott, J.A., Facklam, R.R., and Butler, J.C. (1999). Effectiveness of Pneumococcal Polysaccharide Vaccine for Preschool- Age Children with Chronic Disease. *5*, 828–831.

Garcõ, E., Garcõ, P., and Garcõ, Â.L. (1999). The molecular characterization of the first autolytic lysozyme of Streptococcus pneumoniae reveals evolutionary mobile domains. *33*, 128–138.

Garmory, H.S., and Titball, R.W. (2004). ATP-Binding Cassette Transporters Are Targets for the Development of Antibacterial Vaccines and Therapies MINIREVIEW ATP-Binding Cassette Transporters Are Targets for the Development of Antibacterial Vaccines and Therapies. Society *72*, 6757–6763.

Gaspar, P., Al-Bayati, F. a Y., Andrew, P.W., Neves, A.R., and Yesilkaya, H. (2014). Lactate dehydrogenase is the key enzyme for pneumococcal pyruvate metabolism and pneumococcal survival in blood. Infect. Immun. *82*, 5099–5109.

Van Ginkel, F.W., McGhee, J.R., Watt, J.M., Campos-Torres, A., Parish, L. a, and Briles, D.E. (2003). Pneumococcal carriage results in ganglioside-mediated olfactory tissue infection. Proc. Natl. Acad. Sci. U. S. A. *100*, 14363–14367.

Gladstone, R.A., Jefferies, J.M., Faust, S.N., and Clarke, S.C. (2011). Continued control of pneumococcal disease in the UK - the impact of vaccination. J. Med. Microbiol. *60*, 1–8.

Gosink, K.K., Mann, E.R., Guglielmo, C., Tuomanen, E.I., and Masure, H.R. (2000). Role of novel choline binding proteins in virulence of streptococcus pneumoniae. Infect. Immun. *68*, 5690–5695.

Gray, B.M., Converse III, G.M., and Jr., H.C.D. (2015). Epidemiologic Studies of Streptococcus pneumoniae in Infants : Acquisition , Carriage , and Infection during the First 24 Months of Life. *142*, 923–933.

Gregg, K.J., Zandberg, W.F., Hehemann, J.-H., Whitworth, G.E., Deng, L., Vocadlo, D.J., and Boraston, A.B. (2011). Analysis of a new family of widely distributed metal-independent alpha-mannosidases provides unique insight into the processing of N-linked glycans. J. Biol. Chem. *286*, 15586–15596.

Hammerschmidt, S., Wolff, S., Hocke, A., Rosseau, S., Rohde, M., Müller, E., and Rohde, M. (2005). Illustration of pneumococcal polysaccharide capsule during adherence and invasion of epithelial cells. Infect. Immun. *73*, 4653–4667.

Harford, C., and Hara, M. (1947). Pulmonary edema in influenzal pneumonia of the mouse and the relation of fluid in the lung to the inception of pneumococcal pneumonia. J. Exp. Med. *91*, 245–260.

Harty, D.W.S., Chen, Y., Simpson, C.L., Berg, T., Cook, S.L., Mayo, J. a, Hunter, N., and Jacques, N. a (2004). Characterisation of a novel homodimeric N-acetyl-beta-D-glucosaminidase from Streptococcus gordonii. Biochem. Biophys. Res. Commun. *319*, 439–447.

Hava, D.L., and Camilli, A. (2002). Large-scale identification of serotype 4 Streptococcus pneumoniae virulence factors. Mol. Microbiol. *45*, 1389–1406.

Heffron, R. (1940). Pneumonia, with special reference to pneumococcus lobar pneumonia. Can. Med. Assoc. J. 204.

Helenius, A., and Aebi, M. (2001). Intracellular Functions of N-Linked Glycans. *291*, 2364–2370.

Hicks, L. a, Chien, Y.-W., Taylor, T.H., Haber, M., and Klugman, K.P. (2011). Outpatient antibiotic prescribing and nonsusceptible Streptococcus pneumoniae in the United States, 1996-2003. Clin. Infect. Dis. *53*, 631–639.

Higgins, M. a, Abbott, D.W., Boulanger, M.J., and Boraston, A.B. (2009). Blood group antigen recognition by a solute-binding protein from a serotype 3 strain of Streptococcus pneumoniae. J. Mol. Biol. *388*, 299–309.

Hiss, P.H. (1902). A contribution to the physiological differentiation of Pneumococcus and Streptococcus, and to methods of staining capsules. J. Exp. Med. *6*, 317–342.

Hodges, R.G., MacLeod, C.M., and Bernhard, W.G. (1946). Epidemic pneumococcal pneumonia III. Pneumococcal carrier studies. Am. J. Epidemiol. *44*, 207–230.

Homer, K. a, Roberts, G., Byers, H.L., Tarelli, E., Whiley, R. a, Philpott-Howard, J., and Beighton, D. (2001). Mannosidase production by viridans group streptococci. J. Clin. Microbiol. *39*, 995–1001.

Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K., and Pease, L.R. (1989). Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. Gene *77*, 61–68.

Hyams, C., Camberlein, E., Cohen, J.M., Bax, K., and Brown, J.S. (2010). The Streptococcus pneumoniae capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. Infect. Immun. *78*, 704–715.

Jarrell, K.F., Ding, Y., Meyer, B.H., Albers, S.-V., Kaminski, L., and Eichler, J. (2014). N-linked glycosylation in Archaea: a structural, functional, and genetic analysis. Microbiol. Mol. Biol. Rev. *78*, 304–341.

Jarva, H. (2003). Complement resistance mechanisms of streptococci. Mol. Immunol. *40*, 95–107.

Jeong, J.K., Kwon, O., Lee, Y.M., Oh, D.-B., Lee, J.M., Kim, S., Kim, E.-H., Le, T.N., Rhee, D.-K., and Kang, H.A. (2009). Characterization of the Streptococcus pneumoniae BgaC protein as a novel surface beta-galactosidase with specific hydrolysis activity for the Galbeta1-3GlcNAc moiety of oligosaccharides. J. Bacteriol. *191*, 3011–3023.

Jiang, Y.-L., Yu, W.-L., Zhang, J.-W., Frolet, C., Di Guilmi, A.-M., Zhou, C.-Z., Vernet, T., and Chen, Y. (2011). Structural basis for the substrate specificity of a novel β-N-acetylhexosaminidase StrH protein from Streptococcus pneumoniae R6. J. Biol. Chem. *286*, 43004–43012.

Johansen, F.-E., and Kaetzel, C.S. (2011). Regulation of the polymeric immunoglobulin receptor and IgA transport: new advances in environmental factors that stimulate pIgR expression and its role in mucosal immunity. Mucosal Immunol. *4*, 598–602.

Kabsch, W. (2010). XDS. Acta Crystallogr. Sect. D Biol. Crystallogr. *66*, 125–132.

Kadioglu, a, Brewin, H., Härtel, T., Brittan, J.L., Klein, M., Hammerschmidt, S., and Jenkinson, H.F. (2010). Pneumococcal protein PavA is important for nasopharyngeal carriage and development of sepsis. Mol. Oral Microbiol. *25*, 50–60.

Kim, J., and Weiser, J. (1998). Association of intrastrain phase variation in quantity of capsular polysaccharide and teichoic acid with the virulence of Streptococcus pneumoniae. J. Infect. Dis. 368–377.

King, S.J. (2010). Pneumococcal modification of host sugars: a major contributor to colonization of the human airway? Mol. Oral Microbiol. *25*, 15–24.

King, S.J., Hippe, K.R., and Weiser, J.N. (2006). Deglycosylation of human glycoconjugates by the sequential activities of exoglycosidases expressed by Streptococcus pneumoniae. Mol. Microbiol. *59*, 961–974.

Kline, B.S., and Winternitz, M.C. (1915). Studies upon experimental pneumonia in rabbits. VIII. Intra vitam staining in experimental pneumonia, and the circulation in the pneumonic lung. J. Exp. Med. 311–319.

Kloosterman, T.G., Bijlsma, J.J.E., Kok, J., and Kuipers, O.P. (2006). To have neighbour's fare: extending the molecular toolbox for Streptococcus pneumoniae. Microbiology *152*, 351–359.

Lacks, S., and Hotchkiss, R.D. (1960). A study of the genetic material determining an enzyme in pneumococcus. Biochim. Biophys. Acta *39*, 508–518.

Langley, D.B., Harty, D.W.S., Graham, S.C., Guss, J.M., Hunter, N., and Collyer, C. (2004). Crystallization of GcnA, an N-acetyl-beta-D-glucosaminidase, from Streptococcus gordonii. Acta Crystallogr. D. Biol. Crystallogr. *60*, 1910–1911.

Langley, D.B., Harty, D.W.S., Jacques, N. a, Hunter, N., Guss, J.M., and Collyer, C. a (2008). Structure of N-acetyl-beta-D-glucosaminidase (GcnA) from the endocarditis pathogen Streptococcus gordonii and its complex with the mechanism-based inhibitor NAG-thiazoline. J. Mol. Biol. *377*, 104–116.

Lau, G.W., Haataja, S., Lonetto, M., Kensit, S.E., Marra, a, Bryant, a P., McDevitt, D., Morrison, D. a, and Holden, D.W. (2001). A functional genomic analysis of type 3 Streptococcus pneumoniae virulence. Mol. Microbiol. *40*, 555–571.

Leibovitz, E. (2008). The effect of vaccination on Streptococcus pneumoniae resistance. Curr. Infect. Dis. Rep. *10*, 182–191.

Lemieux, M.J., Mark, B.L., Cherney, M.M., Withers, S.G., Mahuran, D.J., and James, M.N.G. (2006). Crystallographic structure of human beta-hexosaminidase A: interpretation of Tay-Sachs mutations and loss of GM2 ganglioside hydrolysis. J. Mol. Biol. *359*, 913–929.

Li, B., Song, H., Hauser, S., and Wang, L.X. (2006). A highly efficient chemoenzymatic approach toward glycoprotein synthesis. Org. Lett. *8*, 3081–3084.

Limoli, D.H., Sladek, J. a, Fuller, L. a, Singh, A.K., and King, S.J. (2011). BgaA acts as an adhesin to mediate attachment of some pneumococcal strains to human epithelial cells. Microbiology *157*, 2369–2381.

Ling, Z., Suits, M.D.L., Bingham, R.J., Bruce, N.C., Davies, G.J., Fairbanks, A.J., Moir, J.W.B., and Taylor, E.J. (2009). The X-ray Crystal Structure of an Arthrobacter protophormiae Endo-β-N-Acetylglucosaminidase Reveals a (β/α)8 Catalytic Domain, Two Ancillary Domains and Active Site Residues Key for Transglycosylation Activity Zhenlian. J. Mol. Biol. *389*, 1–9.

Liu, T., Zhang, H., Liu, F., Chen, L., Shen, X., and Yang, Q. (2011a). Active-pocket size differentiating insectile from bacterial chitinolytic β-N-acetyl-D-hexosaminidases.pdf. Biochem. J. *438*, 467–474.

Liu, T., Zhang, H., Liu, F., Wu, Q., Shen, X., and Yang, Q. (2011b). Structural determinants of an insect beta-N-Acetyl-D-hexosaminidase specialized as a chitinolytic enzyme. J. Biol. Chem. *286*, 4049–4058.

Manco, S., Hernon, F., Yesilkaya, H., Paton, J.C., Andrew, P.W., and Kadioglu, A. (2006). Pneumococcal neuraminidases A and B both have essential roles during infection of the respiratory tract and sepsis. Infect. Immun. *74*, 4014–4020.

Mao, B., Pear, M.R., and McCammon, A.J. (1982). Hinge-bending in L-Arabinose-binding Protein. *257*, 1131–1133.

Marion, C., Limoli, D.H., Bobulsky, G.S., Abraham, J.L., Burnaugh, A.M., and King, S.J. (2009). Identification of a pneumococcal glycosidase that modifies O-linked glycans. Infect. Immun. *77*, 1389–1396.

Marion, C., Aten, A.E., Woodiga, S. a, and King, S.J. (2011). Identification of an ATPase, MsmK, which energizes multiple carbohydrate ABC transporters in Streptococcus pneumoniae. Infect. Immun. *79*, 4193–4200.

Marion, C., Stewart, J.M., Tazi, M.F., Burnaugh, A.M., Linke, C.M., Woodiga, S. a, and King, S.J. (2012). Streptococcus pneumoniae can utilize multiple sources of hyaluronic acid for growth. Infect. Immun. *80*, 1390–1398.

Mark, B.L., Mahuran, D.J., Cherney, M.M., Zhao, D., Knapp, S., and James, M.N.G. (2003). Crystal Structure of Human β-Hexosaminidase B: Understanding the Molecular Basis of Sandhoff and Tay–Sachs Disease. J. Mol. Biol. *327*, 1093–1109.

Martens, E.C., Koropatkin, N.M., Smith, T.J., and Gordon, J.I. (2009). Complex glycan catabolism by the human gut microbiota: The bacteroidetes sus-like paradigm. J. Biol. Chem. *284*, 24673–24677.

McAllister, L.J., Ogunniyi, A.D., Stroeher, U.H., and Paton, J.C. (2012). Contribution of a genomic accessory region encoding a putative cellobiose phosphotransferase system to virulence of streptococcus pneumoniae. PLoS One *7*, 1–9.

McCarter, J.D., and Withers, S.G. (1994). Mechanisms of enzymatic glycoside hydrolysis. Curr. Opin. Struct. Biol. *4*, 885–892.

McCoy, A.J. (2006). Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr. Sect. D Biol. Crystallogr. *63*, 32–41.

McCullers, J. a, and Bartmess, K.C. (2003). Role of neuraminidase in lethal synergism between influenza virus and Streptococcus pneumoniae. J. Infect. Dis. *187*, 1000–1009.

McCullers, J.A., and Tuomanen, E.I. (2001). Molecular pathogenesis of pneumococcal pneumonia. Front. Biosci. *6*, 877–889.

Meng, B., Marriott, A.C., and Dimmock, N.J. (2010). The receptor preference of influenza viruses. Influenza Other Respi. Viruses *4*, 147–153.

Mitchell, T.J., and Dalziel, C.E. (2014). The biology of pneumolysin. *80*, 145–160.

Murshudov, G.N., Vagin, A. a., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. Sect. D Biol. Crystallogr. *53*, 240–255.

Nakamura, T., Kashima, Y., Mine, S., Oku, T., and Uegaki, K. (2012). Characterization and crystal structure of the thermophilic ROK hexokinase from Thermus thermophilus. J. Biosci. Bioeng. *114*, 150–154.

Nelson, A.L., Roche, A.M., Gould, J.M., Chim, K., Ratner, A.J., and Weiser, J.N. (2007). Capsule enhances pneumococcal colonization by limiting mucus-mediated clearance. Infect. Immun. *75*, 83–90.

Nishitani, Y., Maruyama, Y., Itoh, T., Mikami, B., Hashimoto, W., and Murata, K. (2012). Recognition of heteropolysaccharide alginate by periplasmic solute-binding proteins of a bacterial ABC transporter. Biochemistry *51*, 3622–3633.

Obert, C., Sublett, J., Kaushal, D., Hinojosa, E., Barton, T., Tuomanen, E.I., and Orihuela, C.J. (2006). Identification of a Candidate Streptococcus pneumoniae core genome and regions of diversity correlated with invasive pneumococcal disease. Infect. Immun. *74*, 4766–4777.

Ogunniyi, A.D., Mahdi, L.K., Trappetti, C., Verhoeven, N., Mermans, D., Van der Hoek, M.B., Plumptre, C.D., and Paton, J.C. (2012). Identification of genes that contribute to the pathogenesis of invasive pneumococcal disease by in vivo transcriptomic analysis. Infect. Immun. *80*, 3268–3278.

Orihuela, C.J., Radin, J.N., Sublett, J.E., Gao, G., Kaushal, D., and Tuomanen, E.I. (2004). Microarray analysis of pneumococcal gene expression during invasive disease. Infect. Immun. *72*, 5582–5596.

Parr, T.R., and Saier, M.H. (1992). The bacterial phosphotransferase system as a potential vehicle for the engry of novel antibiotics. Res. Microbiol. *143*, 443–447.

Peltola, V.T., and McCullers, J. a (2004). Respiratory viruses predisposing to bacterial infections: role of neuraminidase. Pediatr. Infect. Dis. J. *23*, S87–S97.

Pérez-Dorado, I., Galan-Bartual, S., and Hermoso, J. a (2012). Pneumococcal surface proteins: when the whole is greater than the sum of its parts. Mol. Oral Microbiol. *27*, 221–245.

Pilishvili, T., Lexau, C., Farley, M.M., Hadler, J., Harrison, L.H., Bennett, N.M., Reingold, A., Thomas, A., Schaffner, W., Craig, A.S., et al. (2010). Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. J. Infect. Dis. *201*, 32–41.

Pluvinage, B., Higgins, M. a, Abbott, D.W., Robb, C., Dalia, A.B., Deng, L., Weiser, J.N., Parsons, T.B., Fairbanks, A.J., Vocadlo, D.J., et al. (2011). Inhibition of the pneumococcal virulence factor StrH and molecular insights into N-glycan recognition and hydrolysis. Structure *19*, 1603–1614.

Polissi, A., Pontiggia, A., Feger, G., Altieri, M., Mottl, H., Ferrari, L., and Simon, D. (1998). Large-Scale Identification of Virulence Genes from Streptococcus pneumoniae. *66*, 5620–5629.

Pracht, D., Elm, C., Gerber, J., Bergmann, S., Rohde, M., Seiler, M., Kim, K.S., Jenkinson, H.F., Nau, R., and Hammerschmidt, S. (2005). PavA of Streptococcus pneumoniae modulates adherence, invasion, and meningeal inflammation. Infect. Immun. *73*, 2680–2689.

Prag, G., Papanikolau, Y., Tavlas, G., Vorgias, C.E., Petratos, K., and Oppenheim, a B. (2000). Structures of chitobiase mutants complexed with the substrate Di-N-acetyl-d-glucosamine: the catalytic role of the conserved acidic pair, aspartate 539 and glutamate 540. J. Mol. Biol. *300*, 611–617.

Procko, E., O'Mara, M.L., Bennett, W.F.D., Tieleman, D.P., and Gaudet, R. (2009). The mechanism of ABC transporters: general lessons from structural and functional studies of an antigenic peptide transporter. FASEB J. *23*, 1287–1302.

Quiocho, F. a, Spurlino, J.C., and Rodseth, L.E. (1997). Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. Structure *5*, 997–1015.

Ramasubbu, N., Thomas, L.M., Ragunath, C., and Kaplan, J.B. (2005). Structural analysis of dispersin B, a biofilm-releasing glycoside hydrolase from the periodontopathogen Actinobacillus actinomycetemcomitans. J. Mol. Biol. *349*, 475–486.

Ramos-Sevillano, E., Moscoso, M., García, P., García, E., and Yuste, J. (2011). Nasopharyngeal colonization and invasive disease are enhanced by the cell wall hydrolases LytB and LytC of Streptococcus pneumoniae. PLoS One *6*, e23626.

Renzi, F., Manfredi, P., Mally, M., Moes, S., Jenö, P., and Cornelis, G.R. (2011). The N-glycan glycoprotein deglycosylation complex (Gpd) from Capnocytophaga canimorsus deglycosylates human IgG. PLoS Pathog. *7*, e1002118.

Rijneveld, A.W., Weijer, S., Florquin, S., Speelman, P., Shimizu, T., Ishii, S., and van der Poll, T. (2004). Improved host defense against pneumococcal pneumonia in platelet-activating factor receptor-deficient mice. J. Infect. Dis. *189*, 711–716.

Ritchie, G.E., Moffatt, B.E., Sim, R.B., Morgan, B.P., Dwek, R. a, and Rudd, P.M. (2002). Glycosylation and the complement system. Chem. Rev. *102*, 305–320 – 19.

Roberts, G., Tarelli, E., Homer, K.A., Philpott-howard, J., Microbiology, J., Thomas, S., and London, K.C. (2000). Production of an Endo-β-N-Acetylglucosaminidase Activity Mediates Growth of Enterococcus faecalis on a High-Mannose-Type Glycoprotein. *182*, 882–890.

Robinson, K.A., Baughman, W., Rothrock, G., Barrett, N.L., Pass, M., Lexau, C., Stefonek, K., Barnes, B., Patterson, J., Zell, E.R., et al. (2001). Epidemiology of Invasive Streptococcus pneumoniae Infections in the United States, 1995-1998. *285*, 1729–1735.

Rosen, J.B., Thomas, A.R., Lexau, C. a, Reingold, A., Hadler, J.L., Harrison, L.H., Bennett, N.M., Schaffner, W., Farley, M.M., Beall, B.W., et al. (2011). Geographic variation in invasive pneumococcal disease following pneumococcal conjugate vaccine introduction in the United States. Clin. Infect. Dis. *53*, 137–143.

Rosenow, C., Ryan, P., Weiser, J.N., Johnson, S., Fontan, P., Ortqvist, A., and Masure, R. (1997). Contribution of novel choline-binding proteins to adherence , colonization and immunogenicity of Streptococcus pneumoniae. *25*, 819–829.

Schlosser, a., Kampers, T., and Schrempf, H. (1997). The Streptomyces ATP-binding component MsiK assists in cellobiose and maltose transport. J. Bacteriol. *179*, 2092–2095.

Shakhnovich, E. a., King, S.J., and Weiser, J.N. (2002). Neuraminidase Expressed by Streptococcus pneumoniae Desialylates the Lipopolysaccharide of Neisseria meningitidis and Haemophilus influenzae: a Paradigm for Interbacterial Competition among Pathogens of the Human Respiratory Tract. Infect. Immun. *70*, 7161–7164.

Sheldrick, G.M. (2007). A short history of SHELX. Acta Crystallogr. Sect. A Found. Crystallogr. *64*, 112–122.

Shin, S.H., and Kim, K.S. (2012). Treatment of bacterial meningitis: an update. Expert Opin. Pharmacother. *13*, 2189–2206.

Silva, Z., Sampaio, M., Henne, A., Gutzat, R., Boos, W., Costa, M.S., Santos, H., and Bo, A. (2005). The High-Affinity Maltose / Trehalose ABC Transporter in the Extremely Thermophilic Bacterium Thermus thermophilus HB27 Also Recognizes Sucrose and Palatinose The High-Affinity Maltose / Trehalose ABC Transporter in the Extremely Thermophilic Bacterium Th. *187*, 1210–1218.

Singh, A.K., Pluvinage, B., Higgins, M. a, Dalia, A.B., Woodiga, S. a, Flynn, M., Lloyd, A.R., Weiser, J.N., Stubbs, K. a, Boraston, A.B., et al. (2014). Unravelling the multiple functions of the architecturally intricate Streptococcus pneumoniae β-galactosidase, BgaA. PLoS Pathog. *10*, e1004364.

Van den Steen, P., Rudd, P.M., Dwek, R. a, and Opdenakker, G. (1998). Concepts and principles of O-linked glycosylation. Crit. Rev. Biochem. Mol. Biol. *33*, 151–208.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. *41*, 207–234.

Suits, M.D.L., Zhu, Y., Taylor, E.J., Walton, J., Zechel, D.L., Gilbert, H.J., and Davies, G.J. (2010). Structure and kinetic investigation of Streptococcus pyogenes family GH38 alpha-mannosidase. PLoS One *5*, e9006.

Sumida, T., Ishii, R., Yanagisawa, T., Yokoyama, S., and Ito, M. (2009). Molecular cloning and crystal structural analysis of a novel beta-N-acetylhexosaminidase from Paenibacillus sp. TS12 capable of degrading glycosphingolipids. J. Mol. Biol. *392*, 87–99.

Sumida, T., Stubbs, K. a, Ito, M., and Yokoyama, S. (2012). Gaining insight into the inhibition of glycoside hydrolase family 20 exo-β-N-acetylhexosaminidases using a structural approach. Org. Biomol. Chem. *10*, 2607–2612.

Sung, C.K., Li, H., Claverys, J.P., and Cnrs-universite, U.M.R. (2001). An rpsL Cassette , Janus , for Gene Replacement through Negative Selection in Streptococcus pneumoniae. *67*, 5190–5196.

Suzuki, R., Wada, J., Katayama, T., Fushinobu, S., Wakagi, T., Shoun, H., Sugimoto, H., Tanaka, A., Kumagai, H., Ashida, H., et al. (2008). Structural and Thermodynamic Analyses of Solute-binding Protein from Bifidobacterium longum Specific for Core 1 Disaccharide and Lacto-N-biose I. *283*, 13165–13173.

Talkington, D.F., Crimmins, D.A.N.L., Voellinger, D.C., Yother, J., and Briles, D.E. (1991). A 43-Kilodalton Pneumococcal Surface Protein , PspA : Isolation , Protective Abilities , and Structural Analysis of the Amino-Terminal Sequence. *3*, 1285–1289.

Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J. a, Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., et al. (2001). Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. Science *293*, 498–506.

Tiels, P., Baranova, E., Piens, K., De Visscher, C., Pynaert, G., Nerinckx, W., Stout, J., Fudalej, F., Hulpiau, P., Tännler, S., et al. (2012). A bacterial glycosidase enables mannose-6-phosphate modification and improved cellular uptake of yeast-produced recombinant human lysosomal enzymes. Nat. Biotechnol. *30*, 1225–1231.

Tong, H.H., Blue, L.E., James, M.A., and Maria, T.F.D.E. (2000). Evaluation of the Virulence of a Streptococcus pneumoniae Neuraminidase-Deficient Mutant in Nasopharyngeal Colonization and Development of Otitis Media in the Chinchilla Model. *68*, 921–924.

Tong, H.H., James, M., Grants, I., Liu, X., Shi, G., and DeMaria, T.F. (2001). Comparison of structural changes of cell surface carbohydrates in the eustachian tube epithelium of chinchillas infected with a Streptococcus pneumoniae neuraminidase-deficient mutant or its isogenic parent strain. Microb. Pathog. *31*, 309–317.

Trakhanov, S., Vyas, N.K., Luecke, H., Kristensen, D.M., Ma, J., and Quiocho, F.A. (2005). Ligand-Free and -Bound Structures of the Binding Protein (LivJ) of the Escherichia coli ABC Leucine / Isoleucine / Valine Transport System : Trajectory and Dynamics of the Interdomain Rotation and Ligand Specificity †. 6597–6608.

Trappetti, C., Kadioglu, A., Carter, M., Hayre, J., Iannelli, F., Pozzi, G., Andrew, P.W., and Oggioni, M.R. (2009). Sialic acid: a preventable signal for pneumococcal biofilm formation, colonization, and invasion of the host. J. Infect. Dis. *199*, 1497–1505.

Tu, A.T., Fulgham, R.L., Crory, M.A.M.C., Briles, D.E., and Szalai, A.J. (1999). Pneumococcal Surface Protein A Inhibits Complement Activation by Streptococcus pneumoniae. *67*, 4720–4724.

Uchiyama, S., Carlin, A.F., Khosravi, A., Weiman, S., Banerjee, A., Quach, D., Hightower, G., Mitchell, T.J., Doran, K.S., and Nizet, V. (2009). The surface-anchored NanA protein promotes pneumococcal brain endothelial cell invasion. J. Exp. Med. *206*, 1845–1852.

Vella, M., and Pace, D. (2014). Glycoconjugate vaccines: an update. Expert Opin. Biol. Ther. 1–18.

Walport, M.J. (2001). Complement First of two parts. *344*, 1058–1066.

Wang, Y., Liu, T., Yang, Q., Li, Z., and Qian, X. (2012). A modeling study for structure features of β-N-acetyl-D-hexosaminidase from Ostrinia furnacalis and its novel inhibitor allosamidin: species selectivity and multi-target characteristics. Chem. Biol. Drug Des. *79*, 572–582.

Weinberger, D.M., Trzciński, K., Lu, Y.-J., Bogaert, D., Brandes, A., Galagan, J., Anderson, P.W., Malley, R., and Lipsitch, M. (2009). Pneumococcal capsular polysaccharide structure predicts serotype prevalence. PLoS Pathog. *5*, e1000476.

Weiser, J.N., Austrian, R., Sreenivasan, P.K., and Masure, H.R. (1994). Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization. Infect. Immun. *62*, 2582–2589.

Whatmore, A.M., Barcus, V. a, Christopher, G., and Dowson, C.G. (1999). Genetic Diversity of the Streptococcal Competence (com) Gene Locus Genetic Diversity of the Streptococcal Competence (com) Gene Locus. *181*, 3144–3154.

Williams, S.J., Mark, B.L., Vocadlo, D.J., James, M.N.G., and Withers, S.G. (2002). Aspartate 313 in the Streptomyces plicatus hexosaminidase plays a critical role in substrate-assisted catalysis by orienting the 2-acetamido group and stabilizing the transition state. J. Biol. Chem. *277*, 40055–40065.

Yamamoto, S., Muramatsu, H., and Muramatsu, T. (2005). Mutational studies on endo-beta-N-acetylglucosaminidase D which hydrolyzes core portion of asparagine-linked complex type oligosaccharides. Glycoconj. J. *22*, 35–42.

Yamashita, K., Ohkura, T., Yoshima, H., and Kobata, A. (1981). Substrate specificity of Diplococcal beta-N- acetylhexosaminidase, a useful enzyme for the structural studies of complex type asparagine- linked sugar chains. *100*, 226–232.

Yesilkaya, H., Spissu, F., Carvalho, S.M., Terra, V.S., Homer, K. a, Benisty, R., Porat, N., Neves, A.R., and Andrew, P.W. (2009). Pyruvate formate lyase is required for pneumococcal fermentative metabolism and virulence. Infect. Immun. *77*, 5418–5427.

Zhang, J.R., Mostov, K.E., Lamm, M.E., Nanno, M., Shimida, S., Ohwaki, M., and Tuomanen, E. (2000). The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. Cell *102*, 827–837.

Zhu, Y., Suits, M.D.L., Thompson, A.J., Chavan, S., Dinev, Z., Dumon, C., Smith, N., Moremen, K.W., Xiang, Y., Siriwardena, A., et al. (2010). Mechanistic insights into a Ca2+-dependent family of alpha-mannosidases in a human gut symbiont. Nat. Chem. Biol. *6*, 125–132.