Fucose Utilization in *Streptococcus pneumoniae*

by

Melanie Higgins
BSc, University of Victoria, 2006

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

Fucose Utilization in *Streptococcus pneumoniae*

by

Melanie Higgins
BSc, University of Victoria, 2006

Supervisory Committee

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

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

Dr. Martin J. Boulanger (Department of Biochemistry and Microbiology)
Departmental Member

Dr. Fraser Hof (Department of Chemistry)
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. Martin J. Boulanger (Department of Biochemistry and Microbiology)
Departmental Member
Dr. Fraser Hof (Department of Chemistry)
Outside Member

Streptococcus pneumoniae can be found in the normal flora of the throat and upper respiratory tract of humans. However, it can commonly become pathogenic causing diseases such as pneumonia and meningitis. *S. pneumoniae* is unique in that a large percentage of its genome encodes for proteins involved in carbohydrate metabolism. A number of these pathways are essential for full virulence of the bacterium, including a putative fucose utilization pathway. There are two strain-dependent varieties of fucose operons in *S. pneumoniae*. The type 1 operon consists of a putative extracellular galactosidase (Sp4GH98), intra-cellular fucosidase (GH95A), PTS relay system (EIIA, EIIB, EIIC, EIID), fucose mutarotase (FcsU), fucose isomerase (FcsI), fuculose kinase (FcsK), and fuculose 1-phosphate aldolase (FcsA). Alternatively, the type 2 operon consists of a putative extracellular galactosidase (Sp3GH98), intra-cellular fucosidase (GH29), two intra-cellular galactosidases (GH36A and B), ABC transporter system, and fucose processing enzymes (FcsI, FcsK, and FcsA).

The objective of this research is to characterize individual components from both fucose operons ultimately to generate both pneumococcal fucose utilization pathways. Specific focus on the extracellular GH98 enzymes provided evidence that these fucose pathways are initiated by the depolymerization of specific histo-blood group antigens presented on host cells. It is then proposed that the products liberated from the complex carbohydrate degradation are transported into the bacterium for further cleavage by intracellular GH enzymes releasing fucose for processing. This process is critical for *S.*

pneumoniae virulence and may be involved in bacterial internalization by host cells suggesting a novel role for this pathway in pneumococcal pathogenesis.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would first like to thank my fantastic supervisor Dr. Al Boraston who has provided an enormous amount of support and guidance through this entire process.

Thank you to my committee members Dr. Fran Nano, Dr. Marty Boulanger, and Dr. Fraser Hof as well as Dr. Paul Romaniuk and the past and present members of the Boraston and Nano Labs for all your encouragement and invaluable discussions about science.

A big thanks my family and friends, especially my Mum and Dad. I could not have done it without you!

Finally, I would like to acknowledge my funding sources: Natural Science and Engineering Council of Canada and Michael Smith Foundation for Health Research.

Chapter 1: Introduction

1.1 *Streptococcus pneumoniae*

Streptococcus pneumoniae, a known human pathogen, commonly colonizes the upper respiratory tract of health people without causing disease. As a commensal bacterium, *S. pneumoniae* is kept in check by the surrounding microbial ecosystem and by nonspecific host defences. Failure in either of these mechanisms permits the bacterium to become pathogenic, frequently leading to pneumonia but also otitis media, bacteremia, or meningitis. *S. pneumoniae* infections are usually categorized as invasive pneumococcal disease (IPD), including bacterial pneumonia and bacteraemia, or acute otitis media (AOM).

1.1.1 *S. pneumoniae* epidemiology

In 2009, there were 11.2-18 cases of *S. pneumoniae* per 100,000 people in the U.S. with an increased number of cases for young children (<5 years of age) and elderly individuals (>65 years of age) (Rosen, Thomas et al. 2011). *S. pneumoniae* infections pose a considerable problem for children where it is estimated that at least 1.2 million infant deaths annually are from pneumococcal infections (Bogaert, De Groot et al. 2004; Berkley, Lowe et al. 2005). Although the majority of these infant deaths occur in developing countries (O'Brien, Wolfson et al. 2009), pneumococcal infections still present substantial concerns in developed countries as a result of the health care burden and fatalities. In the U.S., it is estimated that about 10% of all IPD leads to death with the majority of deaths occurring in people aged 65 and above (Robinson, Baughman et al. 2001).

Presently, there are over 90 different *S. pneumoniae* serotypes that have been isolated and characterized on the basis of their capsular polysaccharide structure and immunogenicity. Of these over 90 serotypes, only 30-40 have been found to cause disease with 15 strains accountable for about 80% of pneumococcal infections to date (Bridy-Pappas, Margolis et al. 2005; Shouval, Greenberg et al. 2006). The majority of colonization occurs with one *S. pneumoniae* strain at a time; however, exposure to other serotypes can cause a switch in the colonizing *S. pneumoniae* strain.

1.1.2 *S. pneumoniae* treatment and prevention

S. pneumoniae infections are generally treated with antibiotics such as penicillin, cephalosporin, macrolides, or trimethoprim-sulfamethoxazole. The type of antibiotic treatment is important to consider due to the significant increase in antibiotic resistant strains over the last decade. In the U.S., 20-35% of *S. pneumoniae* isolates causing invasive disease are resistant to at least one antibiotic while 15-20% are multi-drug resistant with resistance to penicillin, macrolides, and trimethoprim-sulfamethoxazole (Whitney, Farley et al. 2000; Hicks, Chien et al. 2011). Also, an association between increased antibiotic resistance and high antibiotic prescribing areas is seen in the U.S. suggesting that increased antibiotic use gives rise to more resistant *S. pneumoniae* strains (Hicks, Chien et al. 2011). Worldwide, 20-50% of isolated *S. pneumoniae* strains are found to be resistant to one or more antibiotics (Oteo, Lazaro et al. 2004; Xue, Yao et al. 2010; Daoud, Kourani et al. 2011). The incidence of multi-drug resistant *S. pneumoniae* strains is not only predicted to rise but these strains are also expected to develop resistance to the always changing arsenal of antibiotics used to treat infections. Thus, over the past decade preventative measures, such as vaccines, have been an emerging research area to combat *S. pneumoniae* infections.

The first pneumococcal vaccine was a tetravalent polysaccharide vaccine; however, it was developed in 1945 which coincided with the discovery of penicillin and thus was not widely distributed. Over the years this pneumococcal polysaccharide vaccine (PPV) has advanced to become a 23-valent vaccine, which is currently available as Pneumovax® 23 (PPV23) from Merck & Co., Inc. This vaccine covers the 23 most common *S. pneumoniae* serotypes associated with disease and is approved for individuals >50 years of age and people with a high risk of pneumococcal disease that are ≥2 years of age. PPV23 is not deemed effective for children under 2 years of age and is only 62% effective in preventing invasive disease from the vaccine serotypes for children between 2-5 years of age (Fiore, Levine et al. 1999).

The more recent development of a pneumococcal conjugate vaccine (PCV), where capsular polysaccharide is conjugated to a carrier protein, have proved to be more effective for young children, especially for those under the age of 2. Improvement on a hepta-valent vaccine (PCV 7), has given rise to Prevnar 13® (PCV13) which was recently

released by Pfizer Inc. and protects against 13 major *S. pneumoniae* disease-causing strains. This vaccine is recommended for use in children from 6 weeks through 5 years of age with PCV7 demonstrating 97% effectiveness against IPD caused by vaccine serotypes (Black, Shinefield et al. 2000).

Not surprisingly, implementing PCV7 into childhood vaccine programs around the world has substantially reduced the overall number of pneumococcal infections by approximately 34-67% depending on the specific demographics (Singleton, Hennessy et al. 2007; Pilishvili, Lexau et al. 2010; Gentile and Bazan 2011; Miller, Andrews et al. 2011). This reduction in disease prevalence is generally well regarded; however the selective pressures of vaccine administration has caused a 'serotype replacement' effect, whereby non-vaccine serotypes are becoming more commonly found to cause disease (Shouval, Greenberg et al. 2006; Singleton, Hennessy et al. 2007; Gentile and Bazan 2011). This is best described by a study of IPD in Alaska Native children, where PCV7 was introduced into the childhood vaccination routine in 2000. Since 2004, there has been a 94% decrease in IPD caused by vaccine serotypes compared to a 140% increase in IPD caused by non-vaccine strains (Singleton, Hennessy et al. 2007).

Also, the administration of PCV7 has slightly decreased overall antibiotic resistance among *S. pneumoniae* isolates which is not surprising since the majority of antibiotic resistant strains are covered in the vaccine (Kyaw, Lynfield et al. 2006; Gentile and Bazan 2011). However, with the occurrence of the 'serotype replacement' effect, the non-vaccine strains, which are becoming more common in causing IPD, are also showing increased antibiotic resistance. One particular non-vaccine *S. pneumoniae* strain, 19A, appears to not only be increasing in IPD incidences but also is more rapidly gaining antibiotic resistance when compared to other *S. pneumoniae* strains (Fenoll, Aguilar et al. 2011; Hicks, Chien et al. 2011).

With the emergence of non-vaccine *S. pneumoniae* strains in IPD incidence and antibiotic resistance, the current vaccine strategies may only provide a short term solution for pneumococcal disease. Therefore, new strategies are required for a more favourable long term resolution.

1.2 *S. pneumoniae* pathogenesis

Over the past decade, interest in generating long term strategies for the treatment and prevention of *S. pneumoniae* infections has lead to exciting research providing a greater knowledge of pneumococcal pathogenesis. A number of virulence factors have been identified and their roles in pneumococcal disease characterized. Also, it has been determined that *S. pneumoniae* requires initial colonization before becoming invasive and a general model of pneumococcal internalization has been generated (Figure 1). The discussion below is a brief overview of the capsular polysaccharide, a main pneumococcal virulence factor, along with some insights into *S. pneumoniae* colonization and invasion.

1.2.1 Capsular polysaccharide

The majority of *S. pneumoniae* strains are surrounded by a capsular polysaccharide that consists of repeating carbohydrate units. The capsular polysaccharide structure along with its immunogenicity determines a specific *S. pneumoniae* serotype and over 90 different serotypes have been identified to date. There are a number of different monosaccharides (glucose, glucuronic acid, N-acetylglucosamine (glcNAc), galactose, N-acetylgalactosamine (galNAc), N-acetylfucosamine (fucNAc), and N-acetylmannosamine (manNAc)) along with other molecules (glycerol, choline, acetate, pyranose, and furanose) found in the repeating capsular polysaccharide units, which are usually covalently attached to the cell-wall peptidoglycan. The thickness of the capsule, typically between 200-400 nm, has a direct relationship with pneumococcal virulence only when the variation is within a particular strain (Kim and Weiser 1998; Weinberger, Trzcinski et al. 2009).

All *S. pneumoniae* strains contain a capsular polysaccharide biosynthesis (*cps*) locus that encodes for proteins such as glycosyl transferases (GTs), tranlocation proteins, and sugar processing enzymes. With the exception of serotypes 3 and 37, the capsule is synthesized by a Wzx/Wzy dependent biosynthesis pathway. Briefly, an initial sugar-phosphate is transferred to a membrane lipid carrier. Specific GTs attach subsequent monosaccharides to generate the capsule repeat unit at which point the Wzx flippase moves the carbohydrate chain across the cytoplasmic membrane. The Wzy transferase then links individual repeat units together until the completion of a mature polysaccharide

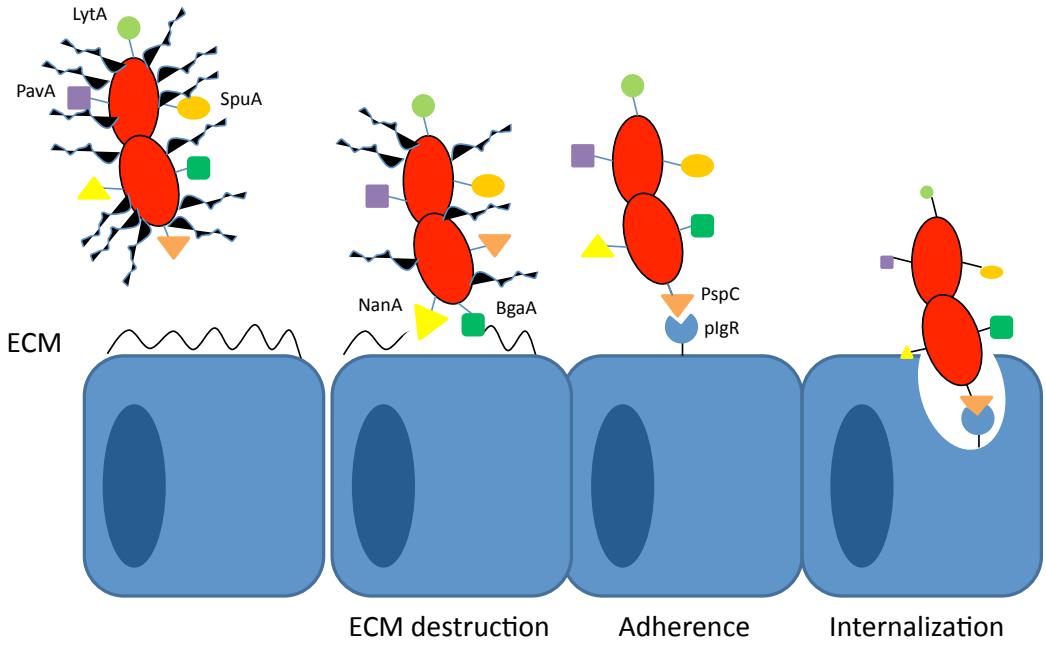


Figure 1: Model of host cell adhesion and internalization by *S. pneumoniae.*

Adapted from (Bergmann and Hammerschmidt 2006). Abbreviations are as follows: Extracellular matrix (ECM), Autolysin (LytA), Pneumococcal adherence and virulence factor A (PavA), Pullulanase (SpuA), Sialidase (NanA), Galactosidase (BgaA), Pneumococcal surface protein (PspC), and Polymeric immunoglobulin receptor (pIgR). The red circles represent *S. pneumoniae* diplococci with the coloured shapes showing cell-wall associated proteins and black protrusions showing the capsular polysaccharide.

chain which is finally translocated to the cell surface by the Wzd/Wze complex. The capsule biosynthesis mechanisms for serotypes 3 and 37 is still unknown; however, there are two genes present in the capsule locus that encode for a UDP-Glc dehydrogenase and a type 3 synthase. It is these two enzymes that appear to be the only essential components for type 3 polysaccharide synthesis (Dillard, Vandersea et al. 1995).

Almost all *S. pneumoniae* clinical isolates are encapsulated, with unencapsulated derivatives generally found to be avirulent, illustrating the importance of the capsule in pneumococcal pathogenesis. The anti-phagocytic nature of the capsule plays a significant role in the bacterium's ability to evade the host immune system. Also, the highly charged residues of the capsule increases electrostatic repulsion effects with highly sialylated mucopolysaccharides. Thus, the capsule protects the bacterium from the immune system and mucus entrapment allowing access to host epithelial cells for initial attachment for colonization and subsequent disease.

1.2.3 *S. pneumoniae* colonization

S. pneumoniae frequently asymptomatically colonizes the nasopharyngeal niche of healthy individuals, especially children. The nasopharyngeal niche consists of a number of commensal bacteria such as *Haemophillus influenzae*, *Moraxella cattarrhalis*, *Neisseria meningitidis*, *Staphylooccus aureus*, and various haemolytic streptococci species and is a dynamic environment in which there is regular turnover of resident colonizing species and serotypes (Bogaert, De Groot et al. 2004).

S. pneumoniae colonization plays an important role in the horizontal spread of the bacterium in the community and is a vital step in disease progression. Colonization not only requires adhesion of *S. pneumoniae* to non-inflamed respiratory tract epithelium but also evasion from the host immune system. As mentioned above, the capsular polysaccharide plays a crucial role in escaping host immune systems; however, due to its immense size the capsule inhibits adhesion to host cells by covering bacterial surface proteins required for attachment. Therefore, as the bacterium moves closer to host cells, the capsule is shed to allow adherence to the epithelium (Figure 1) (Hammerschmidt, Wolff et al. 2005).

There have been many factors identified that are required for colonization including a number of pneumococcal surface proteins (Rosenow, Ryan et al. 1997; Pracht, Elm et al.

2005; Cron, Bootsma et al. 2009; Kadioglu, Brewin et al. 2010; Ramos-Sevillano, Moscoso et al. 2011). *S. pneumoniae* surface proteins can be characterized as choline binding proteins (PspA, PspC, LytA), divalent metal-ion binding lipoproteins (PsaA, PiaA, PiuA), and LPXTG anchored proteins (NanA, NanB, StrH). Alternatively, some *S. pneumoniae* proteins, such as the fibronectin-binding protein PavA, plasminogen binding enolase, and the β-galactosidase BgaC are displayed on the bacterial cell-surface despite lacking a secretion signal peptide and a classical cell-wall anchoring motif (Bergmann, Rohde et al. 2001; Holmes, McNab et al. 2001; Jeong, Kwon et al. 2009). Thus, a *S. pneumoniae* also employs a non-classical method of secretion and cell-wall association and unfortunately, proteins using this route are unable to be predicted bioinformatically from a primary amino acid sequence. Furthermore, many of the surface exposed proteins are multi-functional with roles in adhesion, evasion from the immune system, and/or regulation of other colonization factors to name a few.

One such examples is the lipoprotein PsaA, a component of an ABC-type manganese permease complex (Dintilhac, Alloing et al. 1997). Strains that are lacking PsaA show significant decrease in binding to host cells *in vitro* and are completely avirulent in murine models of pneumonia and colonization (Berry and Paton 1996; Johnson, Dykes et al. 2002; Marra, Lawson et al. 2002). It has been suggested that PsaA mediates cellular adhesion indirectly through the impact of manganese uptake on gene expression of adhesin proteins since strains that are deficient in other members from the same manganese uptake operon also show decreased adhesion to host epithelium (Johnston, Myers et al. 2004; Johnston, Briles et al. 2006). However, the most distinct role in virulence may be due to the fact that manganese uptake is essential for *S. pneumoniae* resistance to oxidative stress and therefore strains deficient in this operon are unable to respond to oxidative stress caused by pneumococcal metabolism or host innate immune responses (Tseng, McEwan et al. 2002).

Pneumococcal surface proteins can also have strain specific roles, such as PpmA, a putative proteinase maturation protein. In NCTC10319 and TIGR4 *S. pneumoniae* strains, PpmA appears to have a role in adhesion to host cells; however, in the D39 strain it does not have an adhesion function but instead seems to mediate uptake by certain immune cells (Cron, Bootsma et al. 2009). This research highlights the idea that the

mechanisms for colonization and subsequent pathogenesis can vary between pneumococcal strains.

1.2.4 Invasive *S. pneumoniae*

Although there have been many pneumococcal virulence factors identified, it has been difficult to differentiate between factors that are involved in colonization with factors involved in causing invasive disease. A synergistic relationship between invasive *S. pneumoniae* and viral infections has been established providing some insight into the methods of pneumococcal pathogenesis (McCullers 2006). The Influenza virus expresses a neuraminidase that degrades terminal sialic acid residues from host cell carbohydrates and is thought to contribute to secondary pneumococcal infections (Peltola, Murti et al. 2005; McCullers 2006). It is hypothesized that the removal of terminal sialic acid residues uncovers a receptor for *S. pneumoniae* adhesion to host epithelial cells and thus could be a factor in initial colonization but perhaps not invasion.

A prominent inflammatory response, however, occurs during viral infections and is thought to have a major impact on the switch from commensal to pathogenic *S. pneumoniae*. Immune elements such as interleukin 1 and Tumor Necrosis Factor (TNF) can change the type and extent of host cell receptors that influence the ability of *S. pneumoniae* to switch from a colonizer to a pathogenic bacterium. Expression of the polymeric immunoglobulin receptor (pIgR) is up-regulated upon the presence of inflammatory cytokines and functions as an Fc receptor that translocates immunoglobulins across the mucosal barrier (Johansen and Kaetzel 2011). pIgR initially binds polymeric immunoglobulins, such as dimeric IgA molecules (dIgA), on the basolateral surface of epithelial cells (Kaetzel, Robinson et al. 1991; Mostov 1994). The pIgR-dIgA complex is internalized by endocytosis, packed into a transcytotic vesicle and transported to the apical surface. Once presented on the apical surface, the pIgR-dIgA complex is cleaved to release dIgA attached to a portion of pIgR called the secretory component. pIgR that has not been cleaved can be recycled through a retrograde pathway that works in the opposite manner to transport pIgR-dIgA from the apical surface back to the basolateral surface. *S. pneumoniae* contains the cell-surface protein PspC (also called CpbA) that binds to pIgR and appears to induce pneumococcal

internalization in a pIgR mediated mechanism to promote translocation across the mucosal barrier (Zhang, Mostov et al. 2000; Agarwal, Asmat et al. 2010).

S. pneumoniae can also induce an inflammatory response through its toxin, pneumolysin (Marriott, Mitchell et al. 2008). Pneumolysin is produced in the bacterial cytoplasm and is released upon autolysis which occurs when *S. pneumoniae* cells are at high density. Once released, pneumolysin binds to cholesterol in the host membrane and its pore-forming nature induces cell lysis causing a subsequent inflammatory response.

Over the past decade, a great deal of research has provided a better understanding of how *S. pneumoniae* switches from a commensal to pathogenic state. This process involves many bacterium-host interactions involved in adherence to host cells, evasion from the immune system, initiation of an inflammatory response, and host cell internalization. Notably, glycans play a major role in these interactions, including acting as receptors for pneumococcal adhesion and deglycosylating immune proteins promoting resistance to neutrophil-mediated opsonophagocytosis (King, Hippe et al. 2006; Dalia, Standish et al. 2010).

1.3 Human glycans

1.3.1 Glycoconjugates

Glycoconjugates can occur as glycolipids or glycoproteins, where one or more carbohydrate residues are covalently attached to either a lipid or polypeptide, respectively. Glycolipids are commonly found inserted into cell membranes with the carbohydrate moiety facing the extracellular space. With implications in many cellular functions, from membrane structural support to cell signalling and regulation functions, one main group of glycolipids are glycosphingolipids. With a glycosphingolipid, a single carbohydrate or carbohydrate chain is attached to the hydroxyl group of a ceramide backbone. When the carbohydrate chain contains one or more sialic acid residues, the glycosphingolipid is called a ganglioside. Gangliosides are abundant in the brain and are thought to play an important role in the physiological operations of the nervous system (Skaper, Leon et al. 1989). There are also glycosphingolipids with two or more monosaccharides that have a neutral charge at pH 7 and are called globosides.

Glycoproteins consist of carbohydrates covalently attached to a specific amino acid of a polypeptide chain. With the ability to incorporate different monosaccharides in a

variety of patterns, glycoproteins can be extremely diverse throughout cell types, tissues and species. This vast diversity of glycoproteins gives rise to numerous biological functions such as protein stability, where the addition of glycans can provide protection from proteases, aid in proper protein folding, and modulate enzyme activity. Glycoproteins also play a major role in signal transduction which can affect cell development and differentiation, and can provide recognition epitopes influencing immune responses and microbe or toxin binding.

There are two main modes of attachment for glycoproteins: N-linked, where the carbohydrate is linked via the amide group of an asparagine, or O-linked, where the monosaccharide is attached via the hydroxyl group from specific amino acids such as serine or threonine. N-linked glycoproteins are generally more complex with a great deal of diversity in the number of monosaccharides present on each chain and the size of the glycan branches. Representative N- and O-linked glycans are shown in Figure 2.

The synthesis of N-linked glycans begins in the endoplasmic reticulum (ER) with a dolichylphosphate precursor in the ER membrane (Helenius and Aebi 2001). Seven specific monosaccharides are sequentially transferred to the precursor by specific GT enzymes in the cytosol at which point the polysaccharide chain is flipped to the lumen. Seven more monosaccharides are then added to the expanding polysaccharide chain and transferred to the asparagine residue, part of the Asn-X-Ser/Thr consensus sequence, of a nascent growing polypeptide. The carbohydrate structure undergoes remodelling before reaching the Golgi apparatus and undergoes further remodelling until reaching its final glycan structure.

O-linked glycans are generally less complex with less branching. Most commonly, the synthesis of O-linked glycoproteins begins with the addition of a starting GalNAc residue directly to the hydroxyl group of a serine or threonine by a specific glycosyl transferase (Van den Steen, Rudd et al. 1998). However, other monosaccharides, such as GlcNAc and fucose, have been identified as starting units of O-linked glycans. Subsequent monosaccharides are then added to the O-linked glycan chain in a step-wise fashion until

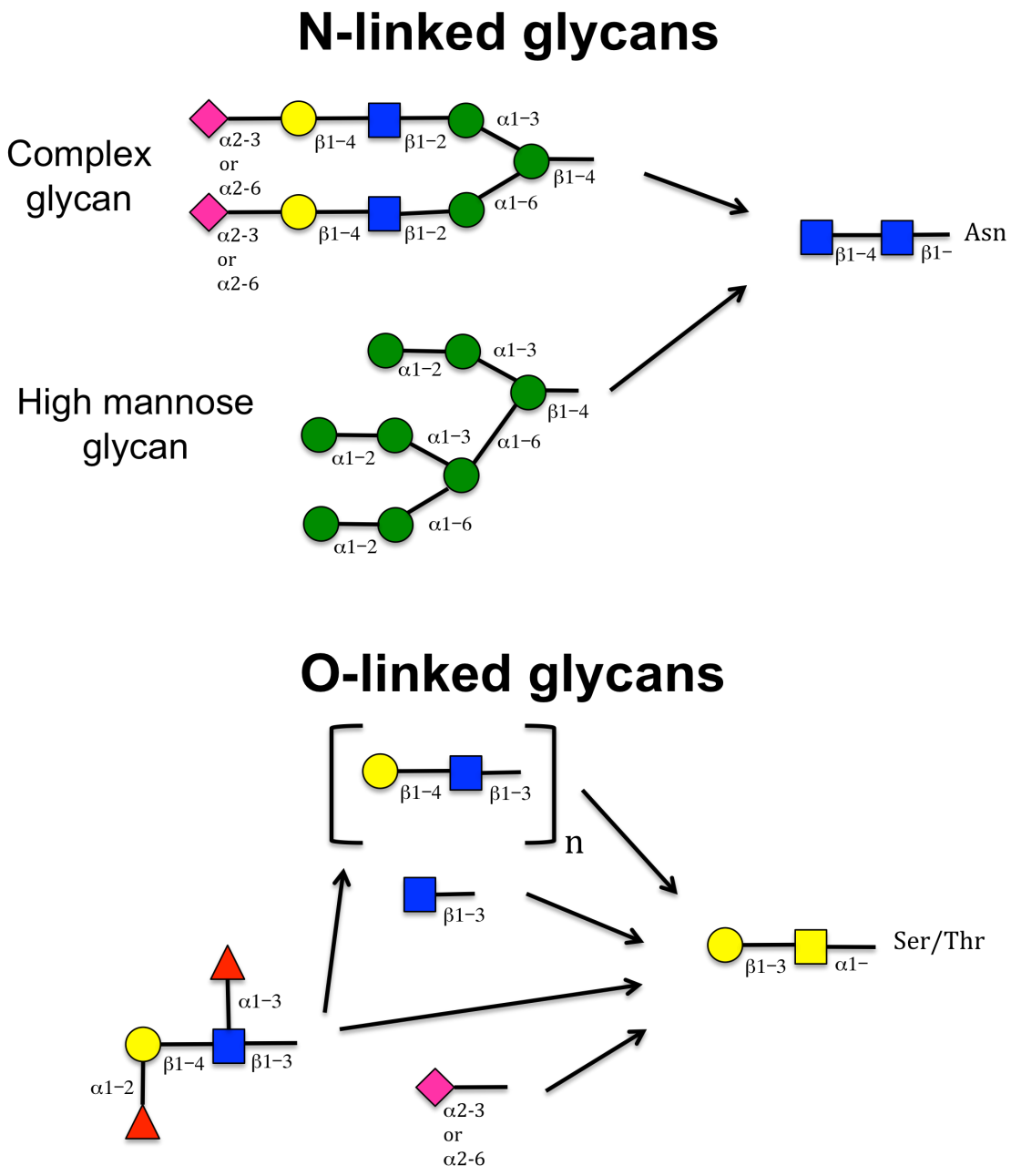


Figure 2: (Legend on next page)

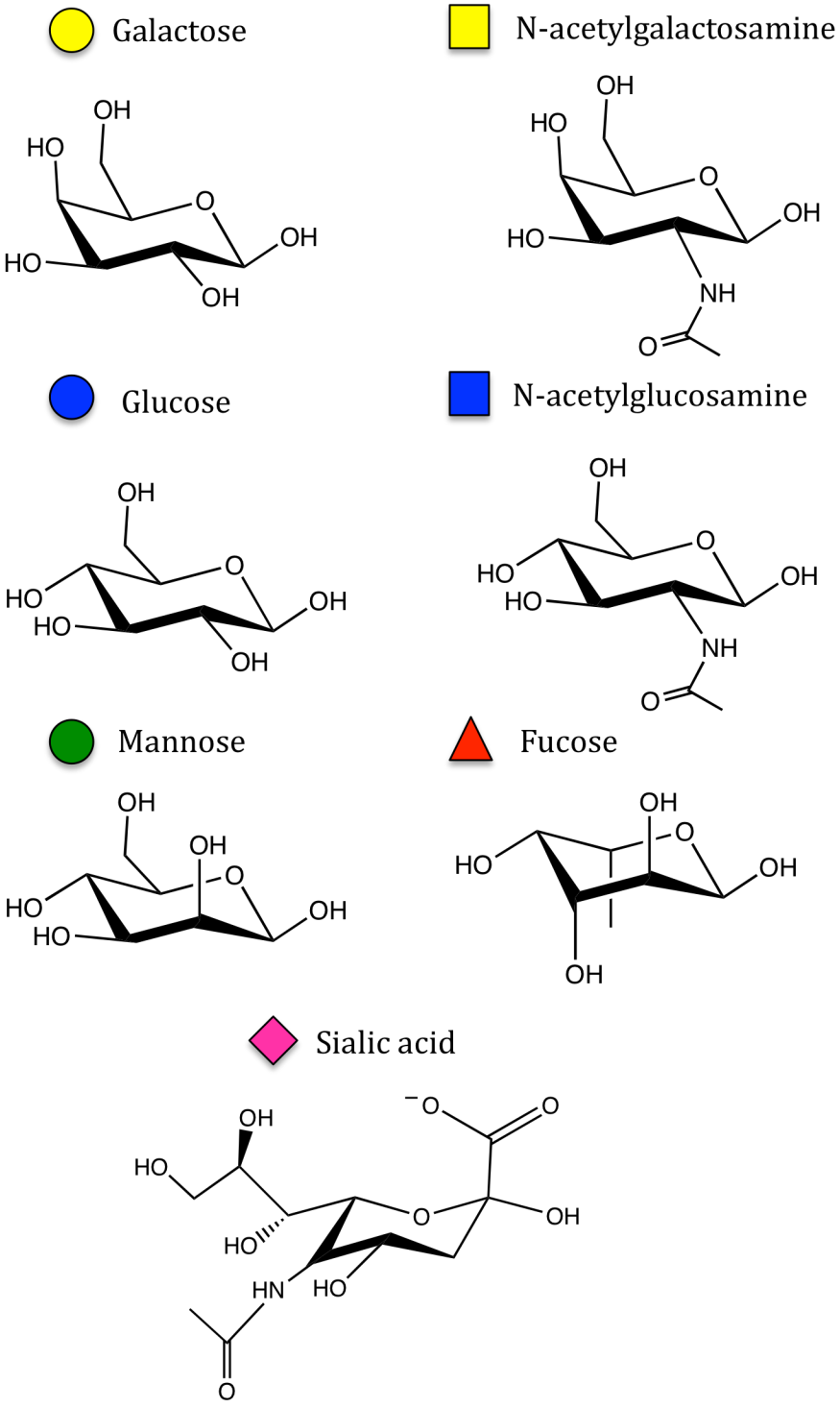


Figure 2: N- and O-linked glycans

Representative sample of N- and O-linked glycan structures found in proteoglycans.

the glycan structure is complete. Because of the diversity in accepting amino acid residues and initial monosaccharides, a consensus sequence for O-linked glycosylation sites has not been identified. However, O-linked glycans are commonly found in β-turns and regions with extended configuration, low hydrophobicity, and lacking large amino acids in the immediate vicinity.

1.3.2 Histo-blood group antigens

ABO blood groups were discovered on human red blood cells over a century ago by Karl Landsteiner and thus defined as 'blood group antigens' (Landsteiner 1900). Later on, these blood groups along with other similar structures, such as the Lewis antigens, were found on tissues and in bodily secretions and thus the new term histo-blood group antigens was created to encompass the entire family of related antigen structures and to describe their distribution more accurately.

Blood group antigens are carbohydrate moieties that are defined by the terminal portion of an oligosaccharide chain that can be linked to either a protein or a lipid anchor. There are five different core oligosaccharide structures, Types 1-5 (Figure 3). From these core structures, the histo-blood group antigens are synthesized in a stepwise manner by specific GT enzymes. Monosaccharides are added to the core structure in a certain manner, in terms of the accepting carbohydrate and the linkage for the addition, depending on the GT specificity.

There are a variety of histo-blood group antigens identified. The ABO antigens differ by their terminal galactose, galNAc, or lack thereof for the A, B, and O antigens, respectively (Figure 3). The variety of blood groups in a population are contributed by different alleles from the ABO locus that make up a persons blood type. A person who is A blood type has the *A* allele which encodes for a GT specific to transfer GalNAc to the core H-antigen while someone who is B blood type carries the *B* allele which encodes for a GT specific for galactose. If someone carries both the *A* and *B* alleles, they can present both the A and B antigens and thus have the AB blood type. Conversely, someone who lacks both of these alleles is of O blood type and thus presents the H-antigen, since they are deficient in the enzymes that transfer the terminal Gal or GalNAc residues.

In contrast, there are several Lewis antigens which include Lewis^{a}, Lewis^{b}, Lewis^{X} and Lewis^{Y} antigens (Figure 3). The Lewis antigens consist of a fucose residue appended to

the GlcNAc, GalNAc, or Glc residues of the Type 1-5 core structures that is lacking in the ABO antigens. About 10% of the European population lack the *FUT3* allele, encoding for a fucosyltransferase involved in Lewis antigen synthesis, and are designated 'Lewis negative' (Mollicone, Reguigne et al. 1994).

Classically, ABO antigens can be found in the blood, primarily on erythrocytes. ABO and Lewis antigens are frequently presented on most human epithelial tissues, and particularly on epithelial cells from all organs in direct contact with the external environment, such as in the higher respiratory or lower genito-urinary tracts (Ravn and Dabelsteen 2000; Marionneau, Cailleau-Thomas et al. 2001). Histo-blood group antigens are not present in connective tissue, muscle, spleen, or lymph nodes. Differential antigen distribution is affected by many factors including genetics, whether the individual carries functional alleles for the GTs responsible for specific antigens (Henry, Oriol et al. 1995; Marriott, Mitchell et al. 2008). The type of epithelial cells as well as the cellular differentiation with the same cell type can also impact antigen presentation (Ravn and Dabelsteen 2000).

The biological role of histo-blood group antigens is unclear, however, they are thought to play a role in host-microbe interactions where pathogens can use these carbohydrates as receptors. There are several documented polymorphisms within many of the GTs involved in histo-blood group synthesis, giving rise to differing antigen presentation throughout the human population. Therefore, the population can have diverse susceptibilities to pathogens that require histo-blood group antigens for virulence, ultimately allowing for a sustainable population during unfortunate bacterial outbreaks. The uropathogenic strains of *E. coli* provide a good example of histo-blood group specificity. Pathogenesis of these strains require binding to galactosylgloboside or their sialyated derivatives found in the urogenital tract solely in non-secretor, or *FUT2* deficient, individuals. The FUT2 enzyme, involved in blood group secretor status, is a fucosyltransferase that adds a fucose residue to galactosylgloboside masking the *E. coli* receptor. Thus, the bacterium is unable to bind its host receptor decreasing the likelihood of causing disease in people with a functional FUT2 enzyme, or secretor individuals (Stapleton, Nudelman et al. 1992). Therefore, non-secretor individuals are more sensitive to disease caused by uropathogenic *E. coli* strains than secretor individuals.

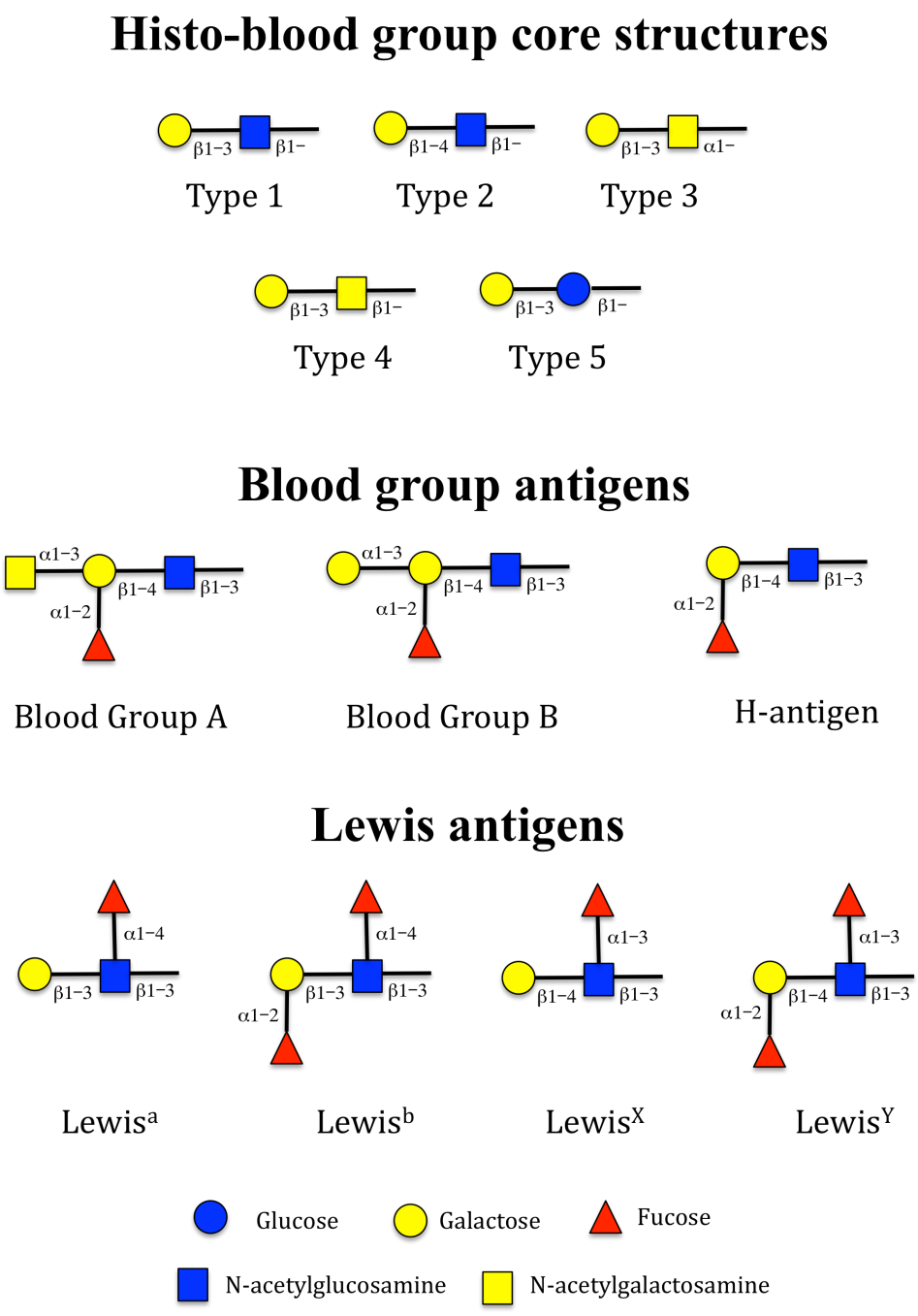


Figure 3: Histo-blood group antigens

Histo-blood group antigen carbohydrate structures. The top section shows the different core structures that can be found in blood group antigens and Lewis antigens.

1.4 Carbohydrate metabolism

Carbohydrates, found as mono-, di-, oligo-, and poly-saccharides, are the most abundant natural compound found on earth. Various monosaccharides can form an enormous number of higher ordered saccharide structures, conferring diverse roles for carbohydrates in nature. Many of these functions are critical for an organism's survival, from energy acquisition and storage to structural functions such as in plant cell walls, demonstrating the value of carbohydrates. Due to the great importance and vast diversity of carbohydrates in all aspects of life, genomes for all organisms encode for many proteins and enzymes dedicated to carbohydrate biosynthesis, transport, and catabolism. The following section will focus on the enzymes that degrade polysaccharides and pathways involved in carbohydrate metabolism.

1.4.1 Glycoside hydrolases

Glycoside hydrolases (GHs) are carbohydrate-active enzymes that catalyze the hydrolysis of glycosidic linkages between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety. They are found in all domains of life and are involved in a wide variety of biological processes. There are currently 129 GH families, classified based on amino acid sequence similarity of the catalytic domain, with equivalent activities and mechanisms for catalysis frequently observed for GHs within the same family. A full list of known and predicted GHs can be find on the Carbohydrate Active Enzyme (CAZy) database (www.cazy.org) (Cantarel, Coutinho et al. 2009).

Although the catalytic modules of GHs from different families do not share amino acid sequence similarity, many share identical structural elements. Thus a new classification system was created based on conserved secondary structure folds where GHs are organized into 14 clans of related families, Clans A-N (Davies and Henrissat 1995). The majority of GHs possess a $(\beta/\alpha)_8$, also known as a TIM barrel, arrangement; however, they can also consist of β-propeller, β-jelly roll, and $(\alpha/\alpha)_6$ organizations. Furthermore, GHs display one of three active site topologies: the pocket, the cleft, or the tunnel (Davies and Henrissat 1995).

Frequently displaying multi-modularity, GHs can be relatively large in size and have up to 9 domains such as GH89 from *Clostridium perfringens* (Ficko-Blean and Boraston 2006). Along with a required catalytic module, ancillary domain types include one or

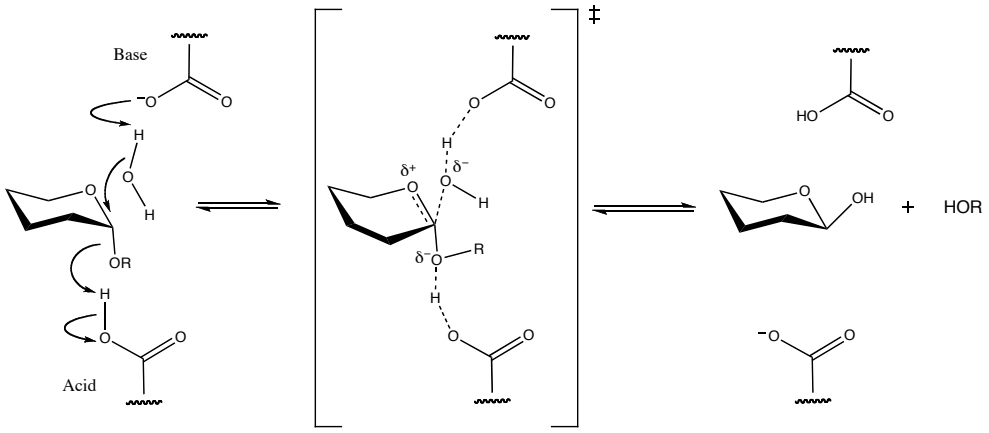
more carbohydrate binding modules (CBM), fibronectin-like modules (FN3), cohesin and dockerin modules, and modules with unknown functions. CBMs specifically recognize and bind carbohydrates, commonly identical or similar to the enzyme's substrate, and will be discussed later. Classically, cohesin and dockerin modules allow certain bacteria, such as *Clostridium thermocellum*, to form complex matrices of extracellular enzymes, called a cellulosome, that act in concert to degrade plant cell walls (Gilbert 2007). More recently noncellulosomal enzyme organizations through cohesin-dockerin interactions have been identified to degrade other complex carbohydrates (Peer, Smith et al. 2009). Lastly, FN3 domains along with unknown modules currently have unidentified functions.

GHs can be *exo*- or *endo*-cleaving enzymes, meaning they have the ability to cleave at the end or in the middle of a carbohydrate chain, respectively. GHs employ a retaining or inverting mechanism for catalysis which either retains or inverts the stereochemistry of the anomeric carbon, respectively. Generally, these mechanisms commonly utilize general acid catalysis requiring two catalytic residues, a nucleophile/base and a proton donor, and the addition of water for hydrolysis (Koshland 1953; Davies and Henrissat 1995).

An inverting mechanism uses an oxocarbenium ion-like transition state for a single displacement method (Figure 4A). Catalysis employs the side chains of glutamic or aspartic acid residues as the catalytic acid and base. Due to this one-step mechanism, the distance between catalytic residues is on average 10 Å to accommodate the critical position of the catalytic water molecule between the substrate and the base (Mccarter and Withers 1994; Davies and Henrissat 1995).

Hydrolysis by a retaining mechanism most commonly occurs via a two-step double displacement mechanism with a glycosyl-enzyme intermediate and oxocarbenium ion-like transition states (Figure 4B). The catalytic residue side chains, commonly a glutamate and/or aspartate, act as an acid/base and nucleophile and due to the different position for the catalytic water are on average 5.5 Å apart (Mccarter and Withers 1994; Davies and Henrissat 1995). Cleavage of a glycosyl linkage can also occur via a substrate-assisted retaining mechanism (Figure 4C). This mechanism is less common and can take place if the substrate contains an N-acetyl (2-acetamido group) or N-glycolyl

A) Inverting mechanism



B) Retaining mechanism

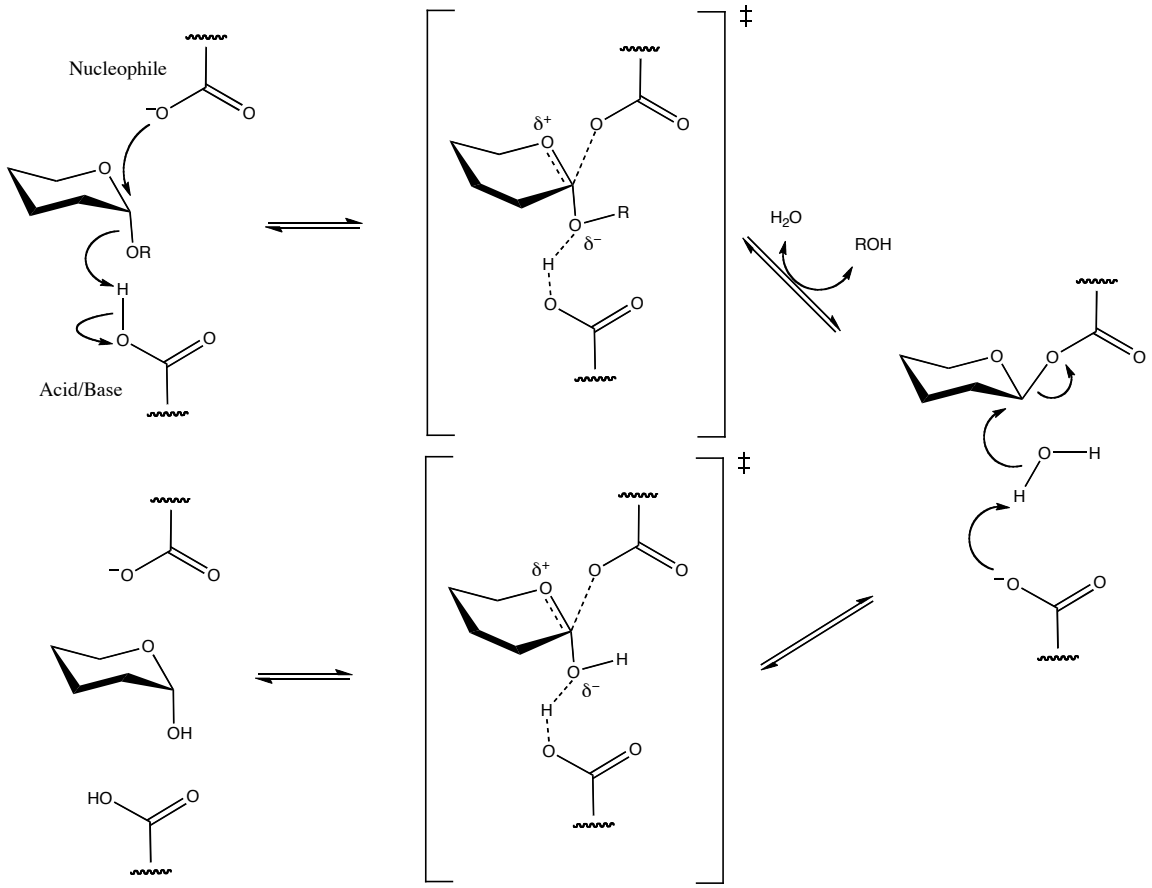
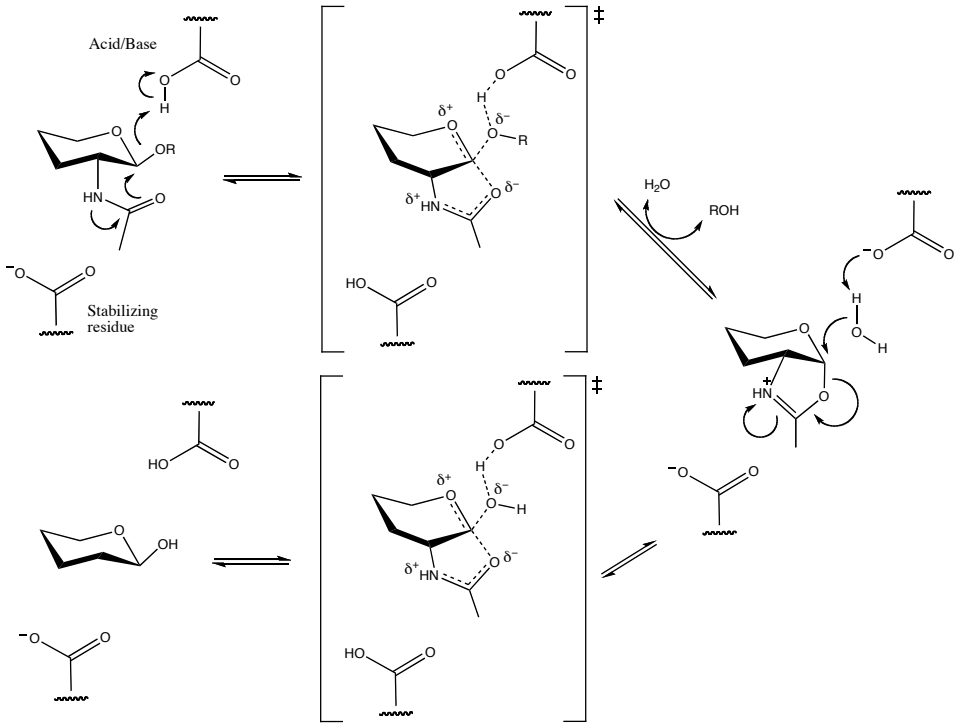


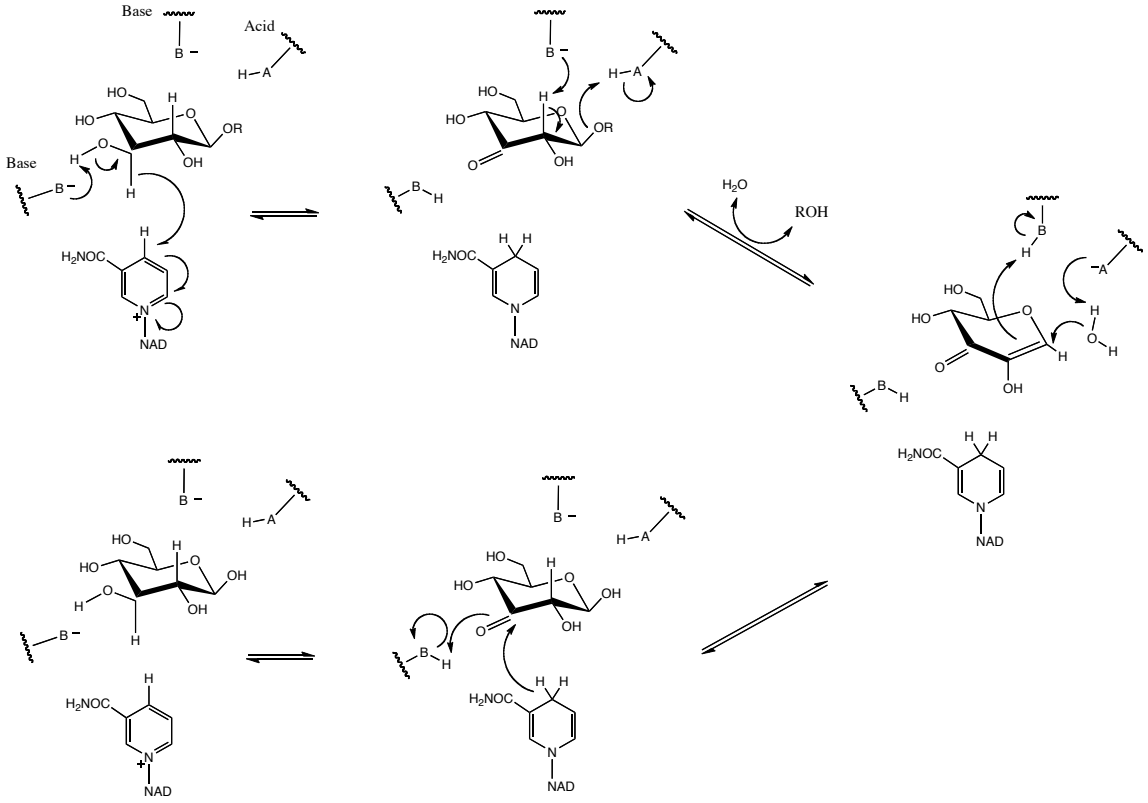
Figure 4: Glycoside hydrolase mechanisms.

Mechanisms for (A) inverting, (B) classical retaining, (C) substrate assisted, and (D) NAD-dependent GH enzymes. (C) and (D) are found on the next page.

C) Substrate-assisted mechanism



D) NAD$^+$-dependent mechanism



group at the 2-position. The 2-acetamido group acts as the nucleophile in a similar manner as the nucleophile in the retaining mechanism, and thus the enzyme only requires one catalytic residue, the acid/base (Zhang, Mostov et al. 2000; Abbott, Macauley et al. 2009). GH enzymes from families 18, 20, 25, 56, 84, and 85 all act on substrates with 2-acetamido groups and can utilize this mechanism for hydrolysis. Lastly, some GH enzymes from families 4 and 109 employ a NAD^+-dependent retaining mechanism for hydrolysis (Figure 4D). This mechanism involves an NAD^+ cofactor and proceeds via oxidation, elimination, addition and reduction steps with anionic transition states (Yip, Varrot et al. 2004; Varrot, Yip et al. 2005).

1.4.2 Carbohydrate binding modules

CBMs are non-catalytic domains that recognize and specifically bind carbohydrates. Found in all organisms, CBMs recognize a broad range of sugars (galactose, fucose, lactose, GalNac, and mannan) and polysaccharides (cellulose, glycogen, chitin, starch, xylan, pullan, and blood group antigens). There are currently 64 families of CBMs that are categorized based on amino acid sequence similarity. Extensive binding and structural studies have elucidated that CBMs from the same family tend to have comparable specificities and modes of binding to their ligands. Similar to GH families, a list of CBMs can be found on the CAZy database (Cantarel, Coutinho et al. 2009).

Along with the classification of CBMs into families based on amino acid similarity, CBMs can also be organized into 7 fold families, based on their 3-dimensional structure. (Reviewed in (Boraston, Bolam et al. 2004)). The majority of CBMs with known structures show a β-sheet topology and are assigned to the β-sandwich and β-trefoil fold families. Other fold families include OB fold, cystein knot, hevein fold, unique fold, and unique hevein-like fold. CBMs can also be classified as 3 different "types", centred around their ligand of recognition. Type A CBMs are "surface-binding" and have a planar catalytic site that recognizes crystalline polysaccharides like cellulose or chitin. "Glycan-chain-binding" CBMs, or Type B CBMs, have a groove or cleft architecture for the active site to bind soluble polysaccharide ligands. Lastly, Type C CBMs are "small-sugar-binding" and recognize free mono-, di-, or tr-saccharides.

Typically found as ancillary domains, CBMs are by definition part of a larger carbohydrate active enzyme, such as a GH or GT, which sets them apart from lectins and

sugar transport proteins. Within an enzyme, one or more CBMs can be located on the N-terminus, C-terminus, or between other domains. When multiple CBMs are involved in the same enzyme, they can have a tandem arrangement, where they are located next to one another, such as in the *Lactobacillus amylovorus* α-amylase that contains 5 tandem CBM26 modules at its C-terminus (Guillen, Sanchez et al. 2010). Conversely, CBMs can be arranged with domain separations like with a *C. perfringens* hexosaminidase that consists of three CBM32 modules, two tandem CBM32s at the N-terminus and a single CBM32 module at the C-terminus (Ficko-Blean and Boraston 2006). Remarkably, some carbohydrate active enzymes house CBMs from different families such as a *C. perfringens* sialidase toxin that contains tandem CBM32 and CBM40 modules. The presence of CBMs from two different families allows this enzyme to recognize two distinct ligands, where the CBM32 modules and the CBM40 specifically bind galactose/GalNac and sialic acid residues, respectively (Ficko-Blean and Boraston 2006).

Despite their non-catalytic activity, CBMs can play important roles in the overall activity of the enzyme. Through binding to specific ligands, CBMs can position the enzyme's catalytic module in close proximity to their substrate effectively increasing enzyme concentration. This effect has been documented particularly for Type A CBMs, or CBMs that bind crystalline ligands (Bolam, Ciruela et al. 1998). Also, CBM binding can result in a conformational change in the carbohydrate ligand necessary for the catalytic module to access the substrate. This substrate disruption by CBMs is best described by starch binding CBMs where binding induces a loop formation of the amylose chains (Southall, Simpson et al. 1999; Giardina, Gunning et al. 2001). Furthermore, the diversity of CBMs within enzymes of similar substrate activity can account for the differential recognition of these enzymes for species- or tissue-specific carbohydrates (Blake, McCartney et al. 2006). Finally, like CBM37 from *Ruminococcus albus*, CBMs within cell wall-associated enzymes can function in anchoring the bacterium to host cell walls (Ezer, Matalon et al. 2008).

1.4.3 Bacterial carbohydrate processing enzymes

The main role of carbohydrate metabolism is to provide energy for bacterial growth and survival. Glycolysis, the classical 10-step metabolic pathway, produces two pyruvate molecules from a single glucose monosaccharide releasing ATP and NADH, or free

energy used for vital cellular functions (Figure 5). The first 5 steps, called the 'preparatory phase' involves consumption of 2 ATP molecules to convert glucose into 2 triose sugars. Each triose sugar continues through the last 5 steps, called the 'payoff phase', to produce a total of 4 ATP and 2 NADH molecules yielding a net of 2 ATP and 2 NADH molecules per one glucose molecule.

However, glucose is not always readily available in the environment so consequently bacteria have adapted alternate metabolic pathways to utilize various sugars that are present in their environmental niches. Alternate processes, such as the fructose metabolic pathway, commonly produce intermediates that can be shuttled into glycolysis for ATP and NADH production (Figure 5). Fructose is initially phosphorylated as it enters the bacterium via a fructose specific phosphotransferase system (PTS) producing fructose 1-phosphate. A fructokinase subsequently adds an additional phosphate yielding fructose 1,6-bisphosphate, a glycolysis intermediate (Fraenkel 1968; Patni and Alexander 1971). Some bacterial species can also convert fructose 1-phosphate directly to the glycolytic intermediate fructose 6-phosphate by a phosphofructomutase enzyme (Binet, Rager et al. 1998). Mannose undergoes similar manipulations; the mannose specific PTS system phosphorylates mannose producing mannose 6-phosphate which then undergoes isomerization by a phosphomannose isomerase yielding the glycolytic intermediate fructose 6-phosphate (Figure 5) (Jensen and Reeves 1998; Garcia-Alles, Zahn et al. 2002).

Furthermore, galactose metabolism occurs via the Leloir pathway that involves three enzymes from a unique galactose utilization operon (Holden, Rayment et al. 2003). Once taken into a cell by a galactose permease, galactose is phosphorylated by a galactokinase producing galactose 1-phosphate (Figure 5) (Frey 1996). The next step requires a UDP-glucose co-factor where a galactose 1-phosphate uridyltransferase transfers the phosphate from galactose 1-phosphate to UDP-glucose producing glucose 1-phosphate and UDP-galactose. UDP-galactose is recycled back to UDP-glucose by a UDP galactose epimerase, while glucose 1-phosphate is converted to the glycolytic intermediate glucose 6-phosphate by a common phosphoglucomutase enzyme.

Finally, fucose metabolism occurs via enzymes found in a fucose utilization operon. Similar to galactose, fucose is commonly transported into the bacterium by a specific

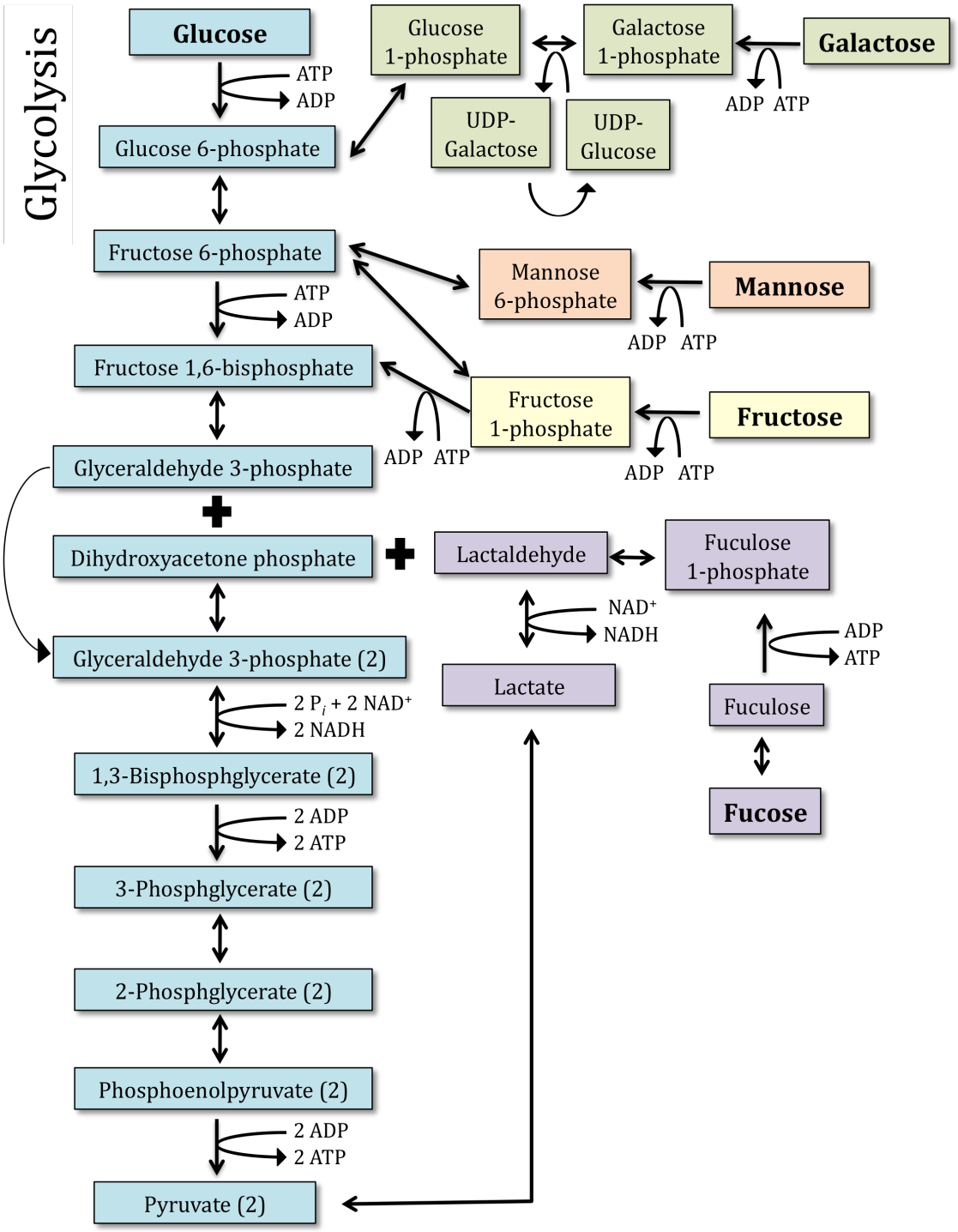


Figure 5: Network of carbohydrate metabolic pathways.

A sample of metabolic pathways that connect with central glycolysis for energy generation. Each metabolic pathway is shown in a different colour.

fucose permease. Initially fucose is converted to fuculose by a fucose isomerase, then phosphorylated by a fucose kinase to produce fuculose 1-phosphate (Figure 5). The final step involves a fuculose 1-phosphate aldolase that yields dihydroxyacetone phosphate (DHAP) and lactaldehyde. DHAP is a glycolytic intermediate and can continue through the second half of glycolysis to produce free energy while lactaldehyde is converted to either lactate and then pyruvate, the end product of glycolysis, or 1,3-propandiol for excretion from the cell (Zhu and Lin 1989).

The metabolic processes described above are only a small representation of alternate pathways that bacteria can employ for nutritional purposes; however, they highlight the bacteria's ability to adapt to their surrounding environments.

1.5 Glycans in host-pneumococcal interactions

1.5.1 *S. pneumoniae* carbohydrate metabolism

S. pneumoniae is unique in that a large percentage of its genome is dedicated to proteins involved in carbohydrate metabolism, including transporters and carbohydrate-active enzymes. The abundance of carbohydrate metabolic proteins allows *S. pneumoniae* to use many monosaccharides (glucose, mannose, and galactose), disaccharides (sucrose, lactose, cellobiose), trisaccharides (raffinose (Tyx, Roche-Hakansson et al. 2011)), and polysaccharides (glycogen and asialofetuin (Abbott, Higgins et al. 2010; Pluvinage, Higgins et al. 2011)) as carbon sources. There is a pathway of enzymes dedicated to processing each of these sugars, for example the *sus* and *scr* loci encode for proteins involved in sucrose metabolism (Iyer and Camilli 2007). Also, *S. pneumoniae* encodes for operons devoted to galactose and fructose metabolism (described in Section 1.3.3).

Although there are many metabolic pathways in place for mono- and di-saccharides, *S. pneumoniae* also has the ability to use polysaccharides for energy generation. This kind of metabolism requires initial breakdown of the polysaccharides into the individual components, which usually occurs through activity of specific GH enzymes. After GH breakdown, the remaining monosaccharides can be further processed by their individual metabolic pathways. One such carbohydrate active enzyme is the pneumococcal pullulanase SpuA, which degrades glycogen, a highly branched polymer of glucose (Lammerts van Bueren, Ficko-Blean et al. 2011). SpuA works in concert with other enzymes and proteins to cleave glycogen and transport degradation products into the cell

ultimately releasing glucose residues for further metabolism and energy generation (Abbott, Higgins et al. 2010).

According to the CAZy database, there are 41-44 glycoside hydrolase genes present in the *S. pneumoniae* genome depending on the strain. Through bioinformatics analysis, many of the pneumococcal GH enzymes appear to be single catalytic modules while others contain additional CBMs, G5 modules, or other unknown modules (Figure 6). Most of these enzymes are predicted to be either cytoplasmic or cell wall-attached by an LPXTG motif with few predicted as secreted extracellular enzymes. However, until verified some of the predicted cytoplasmic GH enzymes may be non-classically secreted and cell wall-associated thus adding to the arsenal of already characterized surface exposed GH enzymes that *S. pneumoniae* employs to degrade host carbohydrates.

1.5.2 Roles of carbohydrate metabolism in *S. pneumoniae* virulence

Many genes that encode for proteins involved in carbohydrate metabolism have not only been identified as virulence factors in the various large-scale screens (Polissi, Pontiggia et al. 1998; Hava and Camilli 2002) but also as genes that are up-regulated in response to different environmental stimuli that mimic the events in pathogenesis (Orihuela, Radin et al. 2004; Song, Connor et al. 2008). Complex carbohydrate metabolism generally contributes to three different mechanisms for pneumococcal pathogenesis: degradation of host glycoproteins, adhesion to the host cell surface, and nutrient acquisition (Shelburne, Davenport et al. 2008).

Degradation of host glycoproteins can serve as an important event promoting resistance to host immune functions. This is best described with the trio of pneumococcal enzymes NanA, BgaA, and StrH that sequentially degrade terminal N-linked glycans (King, Hippe et al. 2006). Neutrophil opsonization via the complement system is a major host defence mechanism against pneumococcal infections (van der Pol, Vidarsson et al. 2000; Brown, Hussell et al. 2002). The complement system is a complex pathway involving a number of components that act to deposit the C3 component onto bacterial cell surfaces promoting neutrophil opsonization and subsequent killing (Trouw and Daha 2011). *S. pneumoniae* strains lacking *nanA*, *bgaA*, or *strH* genes show significantly more C3

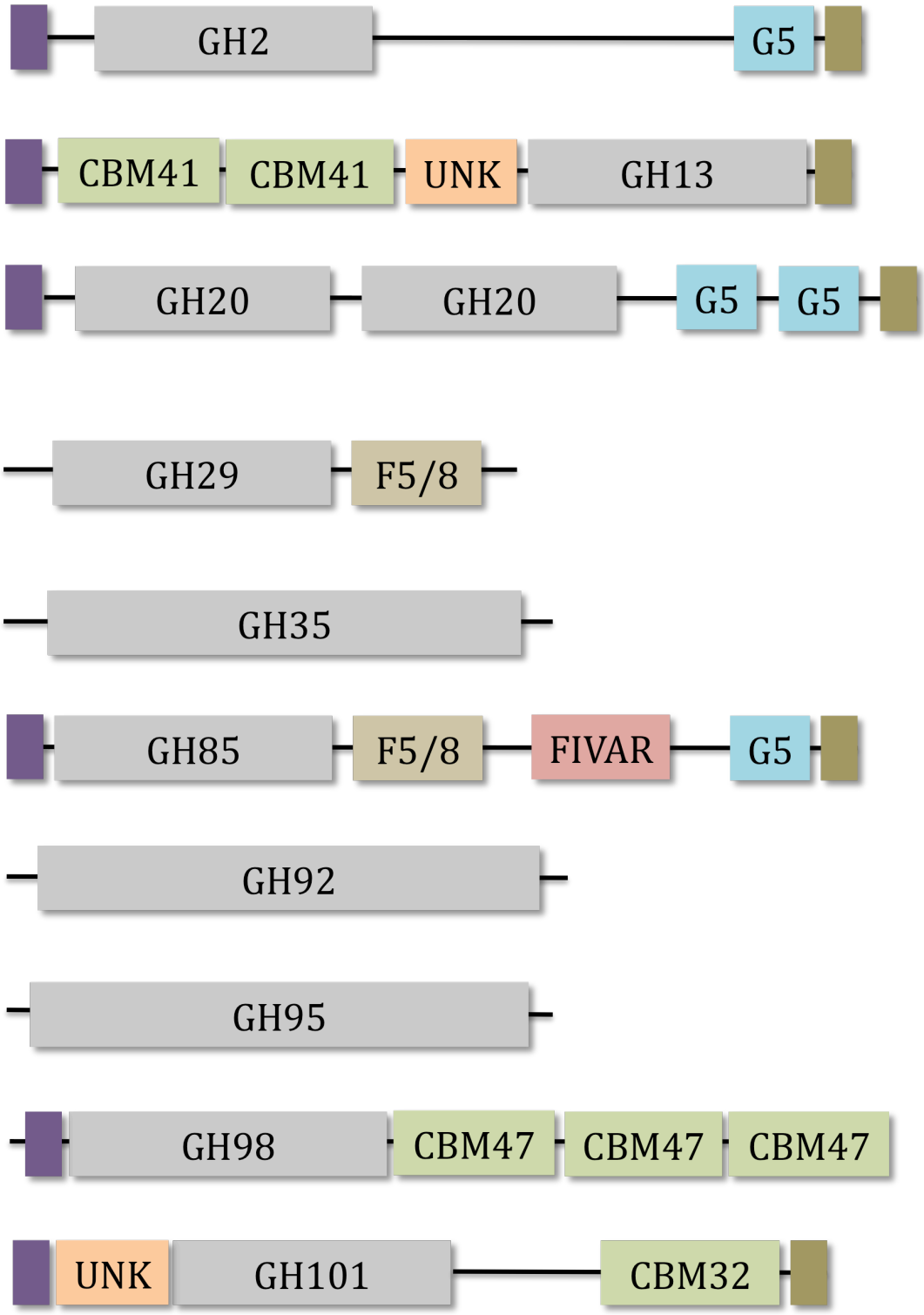


Figure 6: Modular architecture of GH enzymes from *S. pneumoniae.*

A sample of the modular architectures of GH enzymes found in *S. pneumoniae*. Purple boxes are secretion signals and olive boxes at the C terminus are LPXTG motifs.

deposition than the WT strain (Dalia, Standish et al. 2010). It is thought that these enzymes degrade the N-linked glycans found on complement glycoproteins, although the specific glycoprotein is unknown, abolishing their functions which in turn prevents C3 deposition on the bacterial surface and promotes resistance to opsonophagocytic killing by neutrophils.

Carbohydrate active enzymes, especially GHs, can act on host complex carbohydrates to uncap receptors for bacterial adhesion. This phenomenon has been well documented for neuraminidases. As mentioned earlier, expression of the Influenza neuraminidase increases pneumococcal adhesion to host epithelial cells (Peltola, Murti et al. 2005; McCullers 2006). *S. pneumoniae* also produces three distinct neuraminidases (NanA, NanB and NanC) that remove terminal sialic acid residues from glycoconjugates. NanA has been shown to contribute to pneumococcal adherence to host epithelial cells (King, Hippe et al. 2006). Furthermore, both NanA and NanB have roles in nasopharyngeal colonization (Tong, Blue et al. 2000; Manco, Hernon et al. 2006) while NanC has a more tissue specific role (Pettigrew, Fennie et al. 2006). Alternatively, carbohydrate-active enzymes can act as adhesins to promote bacterial adhesion to host cell surfaces. While the BgaA β-galactosidase activity contributes to evasion from the immune system, BgaA also appears to be involved in adhesion to host epithelial cells *in vitro*, although independent of its enzyme activity suggesting an adhesin function (Limoli, Sladek et al. 2011).

The role of carbohydrate metabolism in pneumococcal nutrition has been discussed above with emphasis on glycogen degradation. However, the sequential activity of NanA, BgaA, and StrH has also been implicated in nutrition. These enzymes are required for *S. pneumoniae* growth on glycoconjugates with an abundance of N-linked glycans (Burnaugh, Frantz et al. 2008) highlighting the multi-functionality of pneumococcal virulence factors. Although N-linked glycan degradation has been extensively studied in *S. pneumoniae* contributing to evasion from the immune system, adhesion to epithelial cells and nutrition, further investigation into the biological functions of other pneumococcal carbohydrate metabolic pathways may lead to the discovery of novel mechanisms for pathogenesis.

1.6 Fucose processing in *S. pneumoniae*

Despite *S. pneumoniae*'s inability to use fucose as a sole carbon source (Chan, O'Dwyer et al. 2003), the genome appears to contain a *fcs* locus that encodes a putative fucose utilization operon. The importance of this operon first came into light when four of the proteins from the TIGR4 strain (SP_2159, SP_2162, SP_2164, and SP_2167) were identified as virulence factors in a large scale signature-tagged mutagenesis screen in a murine model of pneumonia (Hava and Camilli 2002). Moreover, a complete operon deletion in the same TIGR4 strain showed severe attenuation in a mouse model of disease confirming the importance of fucose metabolism in pathogenesis (Embry, Hinojosa et al. 2007).

In *S. pneumoniae* TIGR4, the fucose operon consists of a putative extracellular galactosidase (GH98), intra-cellular fucosidase (GH95A), PTS system (EIIA, EIIB, EIIC, and EIID), fucose mutarotase (FcsU), fucose isomerase (FcsI), fuculose kinase (FcsK), and fuculose 1-phosphate aldolase (FcsA) (Figure 7). Interestingly, the *fcs* locus is considered a region of diversity, meaning that there is variation in the DNA sequence between *S. pneumoniae* strains (Embry, Hinojosa et al. 2007). Although most strains contain operons identical to the TIGR4 *fcs* locus, called the type 1 operon, some strains encode for a slightly different fucose operon mainly possessing diverse GHs and sugar transport systems. This operon, typified by the one found in the SP3-BS71 strain and referred to as the type 2 operon, consists of a putative extracellular galactosidase (GH98), intra-cellular fucosidase (GH29), two intra-cellular galactosidases (GH36A and B), ABC transporter system (PBP and two permeases), and fucose processing enzymes (FcsI, FcsK, and FcsA) (Figure 7).

Of the pneumococcal sequenced genomes, 101 strains harbour the type 1 fucose operon while 19 strains consist of the type 2 operon. Surprisingly, the variety of fucose operon does not appear to be serotype specific, as both type 1 and 2 are present in different strains from serotypes 6 and 14. A representative list of the fucose operon type from different *S. pneumoniae* strains is shown in Table 1.

The overall objective of this research is to characterize individual components from both pneumococcal fucose utilization operons, with a specific focus on GH98, FcsU, FcsI, FcsK and FcsA. A better understanding of each individual enzyme will provide the

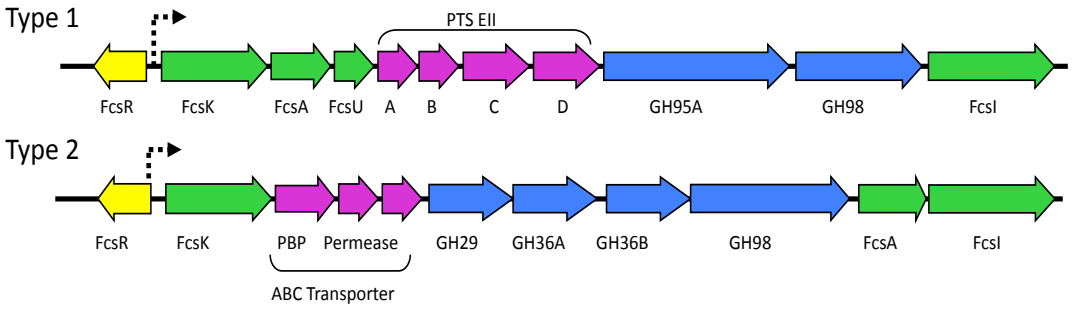


Figure 7: Genomic organization of fucose utilization operons from *S. pneumoniae*. Genes coloured in yellow encode for regulator proteins, purple encodes for transporter proteins, blue encodes for glycans degradation enzymes, and green encodes for fucose processing enzymes.

Table 1: Distribution of fucose operon types among some sequenced *S. pneumoniae* strains.

***S. pneumoniae* strain**	**Serotype**	**Operon type**
R6	Unencapsulated	1
INV104	1	1
D39	2	1
OXC141	3	2
SP3-BS71	3	2
TIGR4	4	1
70585	5	2
SP18-BS74	6	2
SP6-BS73	6	1
CDC1087-00	7F	2
CGS-P14	14	2
SP14-BS69	14	1
Hungary 19A-6	19A	1
CDC3059-06	19A	1
SP23-BS72	23	1
Spanish23F-1	23F	1

means to generate models for pneumococcal fucose metabolism. Furthermore, the biological function of fucose metabolism in *S. pneumoniae* will be investigated. This research will lead to a novel hypotheses of how fucose metabolism is involved in *S. pneumoniae* pathogenesis.

Chapter 2: Differential recognition and hydrolysis of host carbohydrate-antigens by *S. pneumoniae* family 98 glycoside hydrolases

Melanie A. Higgins[1], Garrett E. Whitworth[2], Eric Samain[3], David J. Vocadlo[2] and Alisdair B. Boraston[1]

Adapted from Differential recognition and hydrolysis of host carbohydrate antigens by *Streptococcus pneumoniae* family 98 glycoside hydrolases. J Biol Chem. (2009) 284(38):26161-73.

[1]Biochemistry & Microbiology, University of Victoria, PO Box 3055 STN CSC, Victoria, BC, V8W 3P6, Canada
[2]Department of Chemistry, Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A 1S6, Canada
[3]Centre de Recherches sur les Macromolecules Vegetales (CERMAV-CNRS), BP 53, 38041 Grenoble Cedex 9, France

Contributions: I performed the cloning, protein production and purification, crystallization, and structure refinement for both Sp3GH98 and Sp4GH98. Garrett Whitworth and David Vocadlo performed and analysed the HPAE-PAD and NMR studies.

2.1 Abstract

Streptococcus pneumoniae deploys an extensive arsenal of carbohydrate-processing enzymes that contribute to its fitness as a pathogen, yet the full array of host glycans that the bacterium can attack is presently unknown. The presence of a fucose utilization operon in the pneumococcal genome and its known importance to virulence indicates a reliance on the harvesting of host fucose-containing glycans. The biochemical and structural analysis of two family 98 glycoside hydrolases from distinctive forms of the fucose utilization operon that originate from different *S. pneumoniae* strains reveal that one enzyme, the predominant type among pneumococcal isolates, has novel *endo*-β-galactosidase activity on the Lewis^Y antigen. Altered active site topography in the other species of GH98 enzyme tune its *endo*-β-galactosidase activity to the blood group A- and B-antigens, though both enzymes, and by extension all family 98 glycoside hydrolases, used a retaining catalytic mechanism. The unique activities of the two types of GH98 enzymes, which are distributed differently among *S. pneumoniae* strains, allow for the

generation of models for the harvesting and processing of fucosylated glycans by *S. pneumoniae*.

2.2 Introduction

A suite of *S. pneumoniae* genes whose protein products are dedicated to the harvesting and processing of the sugar fucose is beginning to emerge as an important set of pneumococcal virulence factors (Embry, Hinojosa et al. 2007). These proteins are produced from a monocistronic operon whose expression is induced by the presence of fucose (Chan, O'Dwyer et al. 2003). Through a sequence based analysis of the recently available *S. pneumoniae* genomes, which span a variety of strains and serotypes, we determined that there are two different fucose utilization operons distributed amongst the pneumococcal strains. Though the organization and composition of the two operons is different, both pathways are predicted to be initiated by the action of a single family 98 glycoside hydrolase that is likely secreted. This GH98 is the same as that identified as a virulence factor in the TIGR4 strain (Hava and Camilli 2002). Remarkably, the GH98 enzymes from the two different pathways display different modular architectures and their shared catalytic modules only have approximately 30% amino acid sequence identity. Given the placement of these enzymes in a fucose utilization operon we hypothesized they have activity on fucose containing glycans; however, their divergent sequences and different modular arrangements led us to postulate that they would have different glycan substrate specificities.

Here we describe the specificity and catalytic mechanism for these two different types of *S. pneumoniae* GH98 enzymes, one from the TIGR4 strain (Sp4GH98) and one from the SP3-BS71 strain (Sp3GH98). Both enzymes act as *endo*-β-1,4-galactosidases on the galactosyl-β-1,4-*N*-acetylglucosamine linkage found in type 2 carbohydrate blood group antigens, though Sp4GH98 displays specificity for the Lewis^{Y} antigen while Sp3GH98 is specific for same linkage in the blood-group A/B-antigens. The biochemical analysis of these enzymes in combination with the determination of their structures in complex with products and substrates provides molecular-level insight to their catalytic mechanism and how they discriminate between their respective substrates. We discuss these results in the

context of the recent association of the pneumococcal fucose utilization operon with the virulence of *S. pneumoniae*.

2.3 Materials and Methods

Materials

Sodium hydroxide (50% solution) was purchased from Fisher Scientific (Pittsburgh, PA) and anhydrous sodium acetate was purchased from Fluka Chemical (Ronkonkoma, NY). D-glucose was purchased from Sigma-Aldrich, the Lewis^Y tetrasaccharide was purchased from V-LABS, inc. (Covington, La) and the blood group A and B trisaccharides were purchased from Dextra Laboratories (Reading, UK). Milli-Q (18.2 megaohms cm^{-1}) water was used to prepare all eluants, buffers and standards, while a four times phosphate buffer solution (4 x PBS) was made using salts purchased from Bioshop (Burlington, ON). A-Lewis^Y was purchased from Toronto Research Chemicals Inc. The A- and B-tetrasaccharides used to obtain the Sp3GH98 structures were obtained from Core D of the Consortium for Functional Genomics. All other sugars were produced in metabolically engineered *E. coli* strains and purified as described previously (Drouillard, Driguez et al. 2006).

Cloning

The catalytic module gene fragments from Sp4GH98 and Sp3GH98 were amplified by PCR from *S. pneumoniae* TIGR4 genomic DNA (American Type Culture Collection BAA-334D) and *S. pneumoniae* SP3-BS71 genomic DNA (kindly provided by Dr. Garth Ehrlich), respectively, using primers to introduce 5' *Nhe*I and 3' *Xho*I restriction endonuclease sites. A 'megaprimer' PCR method was employed to introduce site-directed amino acid substitutions (Barik 1996). All amplified DNA fragments were cloned into pET28a (Novagen) using primers *via* the engineered *NheI* and *Xho*I restriction sites using standard molecular biology procedures. The DNA sequence of these constructs was verified by bidirectional sequencing. A list of primers can be found in Table 2.

Table 2: Primers for Sp3GH98CM and Sp4GH98CM recombinant proteins

Primer	Sequence
Sp3GH98CM-F	5’ TTTGGAGCTAGCTCAGAACTTGATAAACGTCG
Sp3GH98CM-R	5’ CTTTCTCTCGAGTTAAAAATGAAGTTCGAATTC
Sp4GH98CM-F	5’ CACCATGGCTGATAATCGTGTTCAAATG
Sp4GH98CM-R	5’ CTCGAGAAGCTTTTATATAGACATCTCTACCATTC
Sp3GH98E558A	5’ TTCACTACGGCCAATTATTGG
Sp4GH98E158A	5’ GTTTTAAATATCGCGAATTATTGGATT

Protein production and purification

Both Sp4GH98 and Sp3GH98 were produced in *E. coli* BL21 Star (DE3) cells using LB media supplemented with 50 μg ml^{-1} of kanamycin. Cultures were grown at 37ºC until it reached an optical density of 600 nm of 0.5-0.7 then induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and further grown at 16ºC overnight. Cells were harvested by centrifugation and ruptured by a chemical lysis method (Robb, Nano et al. 2010). Target proteins were purified by Ni^{2+} immobilized metal affinity chromatography followed by size exclusion chromatography (SEC) using a Sephacryl S-200 column (GE Biosciences). SEC was performed using 20 mM Tris-HCl, pH 8.0, and 250 mM NaCl as the buffer for Sp3GH98 and 20 mM Tris-HCl, pH 8.0, and 1 mM dithiothreitol (DTT) for Sp4GH98. Selenomethionine (SeMet)-labelled Sp4GH98 was prepared using the procedures previously described (Ficko-Blean and Boraston 2005) and purified as above.

Protein concentration was determined by measuring the absorbance at 280 nm and using calculated molar extinction coefficients of 0.156440 $cm^{-1}\mu M^{-}1$ for Sp4GH98 and 0.142670 $cm^{-1}\mu M^{-1}$ for Sp3GH98 (Gasteiger, Gattiker et al. 2003).

NMR analysis

^{1}H NMR spectroscopy (600 MHz Bruker AMX spectrometer) was used to follow the progress and identify the products of the Sp4GH98 catalyzed reaction. The reaction was carried out in ~ 0.6 ml (18ºC, 1.0 mM DTT (pH 7.4) D_2O containing 2.7 mM $Lewis^{Y}$ tetrasaccharide). The reaction was initiated by the addition of 15 μl of a 35 mg ml^{-1} stock of the Sp4GH98. The hydrolysis of the $Lewis^{Y}$ tetrasaccharide was monitored until the reaction reached equilibrium. An initial spectrum (referred to as time 0) containing substrate and buffer, was acquired before the addition of enzyme, while the product spectrum containing the H-disaccharide and buffer, was acquired with no addition of enzyme.

Enzyme assays

Carbohydrates were separated by high performance anion exchange chromatography with pulsed amperometric detection (HPAE-PAD) using a Dionex ICS 3000 equipped

with ASI 100 Automated sample injector (Dionex) and an ED50 electrochemical detector (Dionex) with a gold working electrode and a Ag/AgCl reference electrode.

All assays were carried out in duplicate at 37 ºC for 30 min using a stopped assay procedure. A 50 µL assay volume was used for all standard solutions with 12.5 µL of 4 X PBS (200 mM NaH_2PO_4 and 400 mM sodium chloride) buffering the solution to pH 7.4. Assays were initiated by the addition of enzyme (2.5 µl). Upon addition of 200 µL anhydrous ethanol to halt the reaction, 10 µL of an internal standard D-glucose (from a 1 mM stock solution) was added to give 40 µM final concentration of D-glucose and the assay was stored at -20 ºC for 20 min. Precipitated protein was removed by centrifugation at 14,000 rpm and 4 ºC for 5 min in a microfuge. The supernatant containing mono-, and oligosaccharides was carefully removed from each tube, transferred to a new tube and evaporated to dryness in a vacuum centrifuge. All samples were reconstituted in 250 µL water. A 20 µL injection volume of samples were separated with a PA-100 column set (analytical + guard column), and a 100 mM NaOH isocratic gradient for 20 min.

Standard calibration curves were generated for the $Lewis^Y$ tetrasaccharide, blood group A type 2 and B type 2 pentasaccharides, H-antigen disaccharide and the blood group A- and B-trisaccharides. The standard calibration curves for each substrate was analyzed at pH 7.4 and ranged from 25 µM to 500 µM while a similar curve for the monitored products ranged from 7 µM to 500 µM.

Substrate specificity assays of Sp3GH98 (18 mg/mL) and Sp4GH98 (34 mg/mL) were carried out using the $Lewis^Y$ tetrasaccharide, and blood group A and B type 1, 2, and 4 pentasaccharides. Each substrate was individually incubated at concentrations of 100 µM and 500 µM with Sp3GH98 and Sp4GH98 at 37 ºC for 30 min using the aforementioned stopped assay procedure.

Time-dependent assays of Sp3GH98 and Sp4GH98 revealed that both enzymes were stable over a 5 min period in the PBS buffer at pH 7.4. A range of 0.05 mM to 2 mM of the substrate $Lewis^Y$ was incubated with Sp4GH98 (0.17 mg/mL) for 5 min at 37 ºC, the assay was halted, concentrated, and reconstituted as described above. For the blood group A and B type 2 pentasaccharide substrates a range from 0.025 mM to 5 mM was incubated with Sp3GH98 (0.0045 mg/mL) for 5 min at 37 ºC, the assay was halted, concentrated, and reconstituted as described above. Separation and detection followed the

previously mentioned strategy and quantification of the substrates and products were calculated relative to the calibration curves.

Crystallization

All crystals were obtained by hanging-drop vapour diffusion at 18 ^{o}C. In all cases, final diffraction quality crystals were obtained when 2-4 μl of protein was incubated with the same volume of crystallization solution at room temperature for 5-10 minutes then centrifuged for 10 minutes at 13,000 RPM prior to setting up the crystallization experiment. The highest quality crystals were grown when seeded. Crystals of SeMet-Sp4GH98 (20 mg ml^{-}1) were obtained in 19% (v/v) polyethylene glycol (PEG) 4000, 0.2 M sodium acetate, 0.1 M Tri-sodium citrate pH 5.6, and 3 mM DTT. Native Sp4GH98 crystals and mutant Sp4GH98-E158A crystals were obtained using protein at 20 mg ml^{-1} in 19% (v/v) PEG 4000, 0.2 M sodium formate, 0.1 M Tri-sodium citrate pH 5.6, and 2 mM DTT. Native Sp4GH98 crystals were soaked in mother liquor containing molar excess of Lewis^{Y} tetrasaccharide for 45 minutes to yield H-disaccharide product complex while mutant Sp4GH98E158A crystals were soaked in mother liquor supplemented with molar excess Lewis^{Y} pentasaccharide for 20 minutes to produce a substrate complex.

Native Sp3GH98 and mutant Sp3GH98-E558A, both at 20 mg ml^{-1}, were crystallized in 17% (v/v) PEG 3350, 0.2 M ammonium sulfate, and 0.1 M sodium acetate tri-hydrate, pH 4.8. Native crystals were soaked in mother liquor supplemented with a molar excess of A- or B-blood group antigen tetrasaccharides for 45 minutes to generate A- and B-blood group antigen trisaccharide product complexes, respectively, while crystals of Sp3GH98E558A were soaked with A-Lewis^{Y} antigen pentasaccharide for 20 minutes to produce a substrate complex structure.

Data collection, structure determination, and refinement

Crystals were flash cooled with liquid nitrogen in crystallization solution supplemented with 20-30% (v/v) ethylene glycol. Data was processed using Crystal Clear/d*trek (Pflugrath 1999) or MOSFLM/SCALA (1994; Powell 1999). All data collection and processing statistics are shown in Tables 3 and 4.

Table 3: Data collection and refinement statistics for Sp4GH98.

	Se-Met Sp4GH98	Sp4GH98	Sp4GH98 + Product	Sp4GH98E158A + Lewis^Y
Data collection				
Beamline	NSLS X8-C	SSRL 7.1	NSLS X8-C	CLS CMCF1
Wavelength	0.9794 Å	0.9761 Å	1.1000 Å	0.9214 Å
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions				
a, *b*, *c* (Å)	50.8, 91.7, 116.0	50.8, 90.5, 116.8	50.8, 90.8, 116.5	50.9, 90.9, 116.2
Resolution (Å)	20.00-1.60 (1.66-1.60)	20.00-1.50 (1.58-1.50)	20.00-1.70 (1.76-1.70)	40.00-2.28 (2.40-2.28)
R_{sym} or R_{merge}	0.118 (0.410)	0.060 (0.302)	0.127 (0.413)	0.143 (0.375)
$I / \sigma I$	8.5 (3.8)	20.3 (4.2)	16.9 (3.6)	11.0 (4.1)
Completeness (%)	99.9 (99.9)	84.7 (45.5)	97.4 (96.1)	97.7 (97.7)
Redundancy	6.8 (6.7)	6.2 (3.9)	3.9 (3.6)	4.8 (4.6)
Refinement				
Resolution (Å)		1.50	1.70	2.28
No. reflections		69668	55648	22735
R_{work} / R_{free}		0.126/0.160	0.146/0.197	0.160/0.227
No. atoms				
Protein		4503	4490	4498
Ligand		N/A	22	46
Water		600	642	334
B-factors (Å^2)				
Protein		12.6	10.3	12.5
Ligand		N/A	19.9	31.4
Water		24.1	22.4	16.4
R.m.s. deviations				
Bond lengths (Å)		0.016	0.015	0.010
Bond angles (°)		1.597	1.524	1.304
Ramachandran				
Preferrred (%)		99.2	99.0	99.4
Generously allowed (%)		0.4	0.6	0.2
Disallowed (%)		0.4	0.4	0.4

Values in parentheses are for highest-resolution shell.

Table 4: Data collection and refinement statistics for Sp3GH98

	Sp3GH98+A-trisaccharide	Sp3GH98+B-trisaccharide	Sp3GH98E558A +A-pentasaccharide
Data collection			
Beamline	SSRL 7.1	SSRL 7.1	CLS CMCF1
Wavelength	0.9761 Å	0.9761 Å	0.9793 Å
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$
Cell dimensions			
a, *b*, *c* (Å)	97.7, 153.8, 91.0	98.1, 154.2, 91.2	97.9, 153.7, 90.9
Resolution (Å)	40.00-1.90 (2.00-1.90)	40.00-2.00 (2.11-2.00)	40.00-1.90 (2.00-1.90)
R_{sym} or R_{merge}	0.087 (0.350)	0.063 (0.224)	0.113 (0.368)
$I / \sigma I$	13.0 (4.1)	21.9 (8.8)	14.5 (5.5)
Completeness (%)	97.0 (98.1)	84.6 (74.0)	99.1 (98.0)
Redundancy	3.6 (3.7)	6.1 (5.4)	6.7 (6.4)
Refinement			
Resolution (Å)	1.90	2.00	1.90
No. reflections	99648	74917	102091
R_{work} / R_{free}	0.167/0.206	0.163/0.211	0.165/0.210
No. atoms			
Protein	4681 (A); 4670 (B)	4681 (A); 4673 (B)	4677 (A); 4682 (B)
Ligand	36 (C); 36 (D)	33 (C); 33 (D)	60 (C); 60 (D)
Water	1024	981	974
B-factors (Å^2)			
Protein	16.2 (A); 14.6 (B)	13.4 (A); 11.5 (B)	13.2 (A); 10.9 (B)
Ligand	12.2 (C); 10.8 (D)	8.8 (C); 7.8 (D)	18.5 (C); 16.4 (D)
Water	23.9	20.3	19.9
R.m.s. deviations			
Bond lengths (Å)	0.018	0.016	0.019
Bond angles (°)	1.588	1.474	1.666
Ramachandran			
Preferrred (%)	98.5	98.4	98.4
Generously allowed (%)	0.9	0.8	0.8
Disallowed (%)	0.6	0.8	0.8

Values in parentheses are for highest-resolution shell.

The structure of SeMet-Sp4GH98 structure was solved by an single anomalous dispersion experiment using an X-ray wavelength optimized for the f'' of selenium (determined by a fluorescence scan). The positions of 16 of the 18 selenium atoms expected for the single Sp4GH98 molecule in the asymmetric unit were determined using ShelxC/D (Schneider and Sheldrick 2002). Initial phases were produced by refinement of the selenium substructure parameters with SHARP (Evans and Bricogne 2002) followed by phase improvement with DM (Cowtan and Zhang 1999). Using the phases output from DM, ARP/wARP (Perrakis, Morris et al. 1999) was able to build a nearly complete model of Sp4GH98 with docked sidechains. This initial model was used as a starting point for the building and refinement of Sp4GH98 using the 1.5 Å resolution native data set. The model was completed using COOT (Emsley and Cowtan 2004) followed by refinement using REFMAC (Murshudov, Vagin et al. 1997).

The native Sp4GH98 coordinates were used to solve the structure of Sp3GH98 in complex with the A-blood group antigen trisaccharide by molecular replacement (MR). An MR solution comprising the two Sp3GH98 molecules in the asymmetric unit was found with PHASER (McCoy, Grosse-Kunstleve et al. 2005). This initial model was manually corrected by successive rounds of model building using COOT followed by refinement using REFMAC. In all cases, the addition of water molecules was performed by the REFMAC implementation of ARP/wARP and manually checked. In all datasets, refinement procedures were monitored by flagging 5% of all observation as 'free' (Brunger 1992). Model validation was performed with SFCHECK (Vaguine, Richelle et al. 1999) and PROCHECK (Laskowski, Moss et al. 1993). All model statistics are shown in Tables 3 and 4.

2.4 Results and Discussion

S. pneumoniae GH98 enzymes

The carbohydrate-processing enzymes of *S. pneumoniae* were originally the subject of intense interest due to their utility in characterizing the structures of eukaryotic glycans. With the recent observation that many of these proteins are involved in pneumococcal virulence, interest in these enzymes has been re-invigorated, driven by the desire to

understand pneumococcal pathogenesis and, perhaps, exploit these proteins as vaccine components and/or therapeutic targets (Bongaerts, Heinz et al. 2000; King, Hippe et al. 2006; Burnaugh, Frantz et al. 2008; Caines, Zhu et al. 2008). The pneumococcal fucose utilization pathway is a relatively newly discovered carbohydrate-processing pathway in the bacterium whose study hold the potential for new insight into *S. pneumoniae*-host relationship. Specifically, detailed knowledge of the specificity, mechanism, and structure of the family 98 glycoside hydrolases, which feature prominently in the fucose utilization pathway as the only predicted extracellular components, should reveal the nature of the host glycan(s) targeted by this pathway. To this end, we have singled out the GH98 enzymes from *S. pneumoniae* TIGR4 (Sp4GH98) and SP3-BS71 (Sp3GH98) for structural and mechanistic studies, the first for any family 98 glycoside hydrolases.

Sp4GH98 and Sp3GH98 are 1038 and 1005 amino acid, respectively, multimodular proteins with predicted classical N-terminal Gram-positive secretion signal sequences. Sp4GH98 possesses three C-terminal family 47 carbohydrate-binding modules (CBMs) (Boraston, Wang et al. 2006) while Sp3GH98 has two N-terminal modules that have sequence identity with family 51 CBMs (Gregg, Finn et al. 2008). The N-terminus of Sp4GH98 and the C-terminus of Sp3GH98 contain a domain of ~650 amino acids that is predicted on the basis of sequence alignments with a known family 98 glycoside hydrolase, the GH98 *end*o-galactosidase E-ABase from *C. perfringens* (Anderson, Ashida et al. 2005), to house the catalytic activity. An amino acid sequence-based comparison of the catalytic modules reveals that Sp4GH98 shows 34% amino acid identity with *C. perfringens* E-ABase, whereas Sp3GH98 shows 60% amino acid sequence identity with E-ABase. The catalytic modules of Sp4GH98 and Sp3GH98 show only 35% amino acid identity with each other. Homologues of Sp4GH98 with 99% or greater amino acid sequence identity are found in 101 of the 120 sequenced genomes of *S. pneumoniae*; the remaining 19 strains have identical homologues of Sp3GH98. Also, the distribution of the two types of GH98 enzymes among pneumococcal strains appears to be independent of strain serotype (Table 1).

To facilitate the biochemical and structural analysis of Sp4GH98 and Sp3GH98 we used our amino acid sequence alignments to precisely define the catalytic domain in these enzymes. Guided by this information we cloned the DNA fragments encoding only the

catalytic domains and overproduced the proteins in *E. coli*. For simplicity, we will continue to refer to these truncated constructs as Sp4GH98 and Sp3GH98.

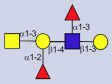
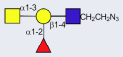
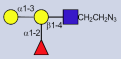
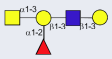
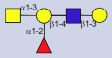
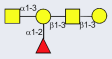
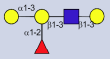
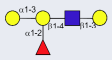
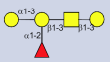
GH98 enzymes are active on blood group antigens in vitro

Given the different modular structures of Sp3GH98 and Sp4GH98 and the presence of the genes encoding these divergent GH98 enzymes within fucose utilization operons having different gene contents and organizations (Higgins, Abbott et al. 2009), we suspected that these enzymes might have different, yet related, substrate specificities for human cell surface fucose-containing glycans. Furthermore, previous studies found that E-ABase cleaved the A- and B-blood group glycotopes from protein substrates and cell surfaces (Anderson, Ashida et al. 2005). Accordingly, we set out to evaluate the substrate specificity of these two *S. pneumoniae* enzymes against select fucose-containing blood group antigens including pentasaccharides of blood group glycotypes A and B containing type 1, 2, and 4 core chains, the type 2 H-trisaccharide and the Lewis^y tetrasaccharide (Table 5).

To assess the activity of Sp3GH98 and Sp4GH98 against these oligosaccharides we used high performance anion exchange chromatography with pulsed amperometric detection (HPAE-PAD) (Table 4). Of these substrates, Sp4GH98 cleaved only the Lewis^Y tetrasaccharide to generate the H-antigen disaccharide (Fucα1-2Gal) and Fucα1-3GlcNAc, suggesting the active site of this enzyme cannot accommodate the terminal GalNAc or Gal moiety of the A and B blood group antigens. In contrast, we found that Sp3GH98 did not cleave the Lewis^Y tetrasaccharide but did process the pentasaccharide blood group glycotypes A and B depending, however, on the structure of the core chain. In keeping with its high amino acid sequence identity with *C. perfringens* E-ABase, Sp3GH98 is specific for the Galβ1-4GlcNAc (type 2 core chain) linkage over either the Galβ1-3GlcNAc (type 1 core chain) linkage or the Galβ1-3GalNAc (type 4 core chain) and liberates the terminal GalNAcα1-3(Fucα1-2)Gal trisaccharide of the A-antigen and the terminal Galα1-3(Fucα1-2)Gal trisaccharide of the B-antigen.

To quantitatively evaluate the processing of these substrates by Sp3GH98 and Sp4GH98 we carried out enzyme kinetics studies using an HPAE-PAD assay following

Table 5: GH98 specificity

Glycan	Core Type	Structure	Sp3GH98 Activity	Sp4GH98 Activity
H-trisaccharide	2		Not tested	-[1]
Lewis^{Y} tetrasaccharide	2		-	+
Lewis^{Y} pentasaccharide	2		-	+
Lewis^{B} tetrasaccharide			-	-
A-Lewis^{Y} pentasaccharide	2		Not tested	Not tested
A-antigen tetrasaccharide (synthetic tail)	2		+[2]	-
B-antigen tetrasaccharide (synthetic tail)	2		+[2]	-
A-antigen pentasaccharide	1		-	-
A-antigen pentasaccharide	2		+	-
A-antigen pentasaccharide	4		-	-
B-antigen pentasaccharide	1		-	-
B-antigen pentasaccharide	2		+	-
B-antigen pentasaccharide	4		-	-

Structures of substrates used in this study and the activity of GH98 enzymes on them. + indicates activity, - the absence of activity. [1] activity tested by thin layer chromatography; [2] activity observed in crystals. All other activities were tested by HPAE-PAD. The symbols used are as follows: yellow circle is Gal, yellow square is GalNAc, blue square is GlcNAc, and red triangle is Fuc.

the production of the A- or B-trisaccharides (for Sp3GH98) or the H-disaccharide (for Sp4GH98) in conjunction with an internal standard. Time-dependent assays revealed that both enzymes were stable at 37 ºC over the assay period and, by varying the substrate concentrations, we were able to establish kinetic parameters governing the enzyme catalyzed hydrolysis of these substrates (Table 6). For the Sp3GH98-catalyzed hydrolysis of the A- and B-antigens we found clear saturation kinetics (Figure 8A and 8B) yielding very comparable kinetic parameters (Table 6). Interestingly, for the Sp4GH98-catalyzed hydrolysis of the Lewis^Y tetrasaccharide we did not observe substrate saturation (Figure 8C), making it possible only to determine the second order rate constant, k_{cat}/K_M. The absence of saturation even at concentrations of 2 mM Lewis^Y tetrasaccharide suggests the enzyme may have additional binding subsites in its active site that are not exploited by the tetrasaccharide, that the concentration of substrate in the natural environment when bound to the cell surface is much higher, or there may be a different physiological saccharide that is processed more efficiently. Alternatively, both intact Sp4GH98 and Sp3GH98 possess CBMs that likely function *in vivo* to target the enzymes to their respective tissue-presented carbohydrate substrates. By analogy to the function of CBMs found as components of plant-cell wall degrading glycoside hydrolases, by virtue of their affinity for glycans, the CBM's in Sp4GH98 and Sp3GH98 may enhance the apparent K_m of the enzymes for immobilized glycans thus aiding in overcoming the inherently low K_ms of the catalytic modules (Boraston, Bolam et al. 2004).

Structure of GH98 and specific glycon recognition

A small number of glycoside hydrolases are now known to be active on blood group antigens (Calcutt, Hsieh et al. 2002; Anderson, Ashida et al. 2005; Liu, Sulzenbacher et al. 2007). Other than E-ABase, these are typically *exo*-glycosidases that remove the non-reducing terminal A- or B-antigen determining N-acetylgalactosamine or galactose residues, respectively, converting the antigen to the H-type (O-type). The specificity of Sp3GH98 is thus quite different, and more akin to that of *C. perfringens* E-ABase, while the activity of Sp4GH98 on the Lewis^Y antigen is entirely unique. This prompted us to explore the structural basis of substrate specificity in GH98 enzymes using X-ray crystallography.

Table 6: Kinetic analysis of GH98 activity

Enzyme	Substrate[1]	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (μM^{-1} min^{-1})
Sp4GH98	$Lewis^Y$	ND	ND	0.026 (± 0.000)
Sp3GH98	A-antigen	280 (± 10)	440 (± 40)	0.64 ± (0.10)
	B-antigen	670 (± 20)	570 (± 60)	1.2 (± 0.10)

[1] The $Lewis^Y$ antigen used was the $Lewis^Y$ pentasaccharide and the A/B antigens were the type 2 pentasaccharides.

ND, not determined

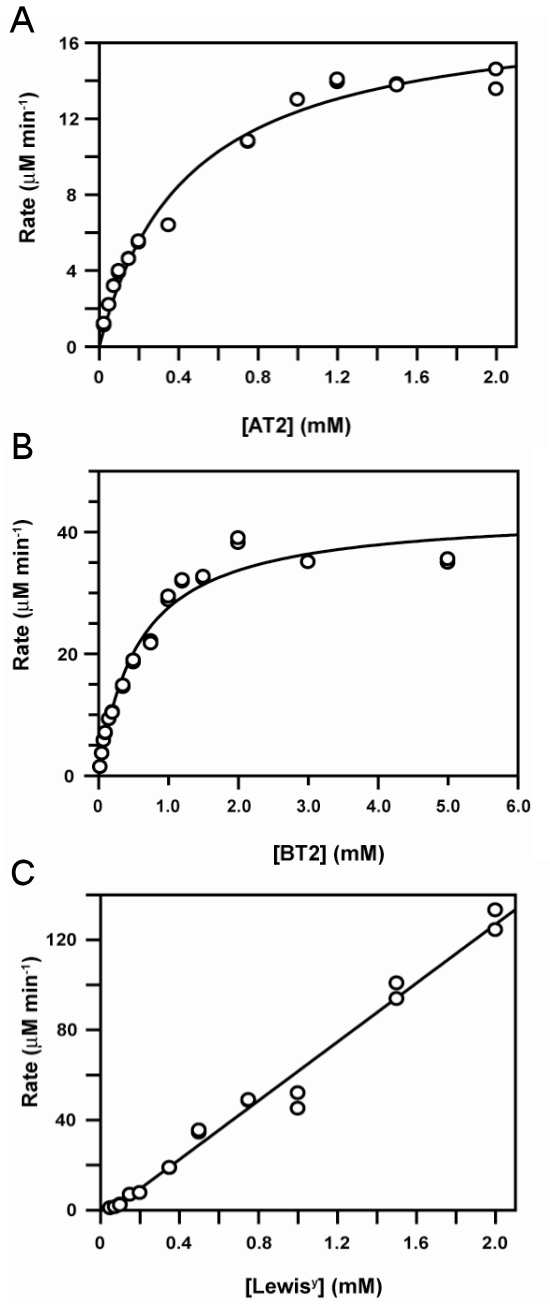


Figure 8: Kinetic properties of GH98 enzymes.

Kinetic plots of the activity of Sp3GH98 on the type 2 A-pentasaccharide (A) and the type 2 B-pentasaccharide (B). A corresponding analysis of the activity of Sp4GH98 on the Lewis^{Y} pentasaccharide (C).

The X-ray crystal structure of Sp4GH98 was determined by optimized selenium single anomalous dispersion to 1.6 Å resolution. The resulting initial model was subsequently used to complete and refine a native structure to 1.5 Å resolution. The N-terminal domain of the crystallized construct comprises a classical $(\alpha/\beta)_8$-barrel (Figure 9A). This module is followed by an 11 stranded β-sandwich domain comprising a β-sheet of 5 anti-parallel β-strands packed against a β-sheet of 6 anti-parallel β-strands. An α-helical insertion between β-strands 7 (strand 18 in the complete structure) and 8 (19 in the complete structure) extends out and packs against the $(\alpha/\beta)_8$ barrel. Likewise, a small insertion between β-strand 7 and α-helix 8 of the $(\alpha/\beta)_8$ barrel interacts with the extension from the β-sandwich. The substantial interactions between the $(\alpha/\beta)_8$ barrel and the β-sandwich create what appears to be a relatively rigid structure. The structure of Sp3GH98 is extremely similar and overlaps with Sp4GH98 with an root mean square deviation of 1.4 Å over 511 matched Cα atoms (determined by secondary structure matching (Krissinel and Henrick 2004); not shown). Indeed, the relative placements of the constituent domains are nearly identical for the two enzymes suggesting they function as a single, rigid structural unit.

To provide insight into the glycon specificities of Sp4GH98 and Sp3GH98 complexes of these enzymes with their reaction products were obtained by soaking crystals of catalytically active proteins with their respective substrates. The structure of Sp4GH98 incubated with the Lewis^Y antigen revealed unmistakable electron density for a disaccharide in the active site. This electron density could be easily modelled as the H-disaccharide (Fucα1-2Gal) but not as any other component of the Lewis^Y antigen (Figure 9B). Likewise, the structures of Sp3GH98 incubated with the blood group A- and B-tetrasaccharides yielded clear electron density for the blood group A- and B-trisaccharides, respectively, in the active site (Figure 9C; only the A-trisaccharide complex is shown as representative data). These results are consistent with the product analysis by HPAE-PAD and provides additional supporting evidence that the Galβ1-4GlcNAc linkage common to carbohydrate-antigens having the core 2 chain is hydrolyzed by these GH98 enzymes.

The galactosyl residue of the H-disaccharide product in the active site of Sp4GH98 occupies what must be the -1 subsite while the α-1,2-linked fucosyl residue occupies the

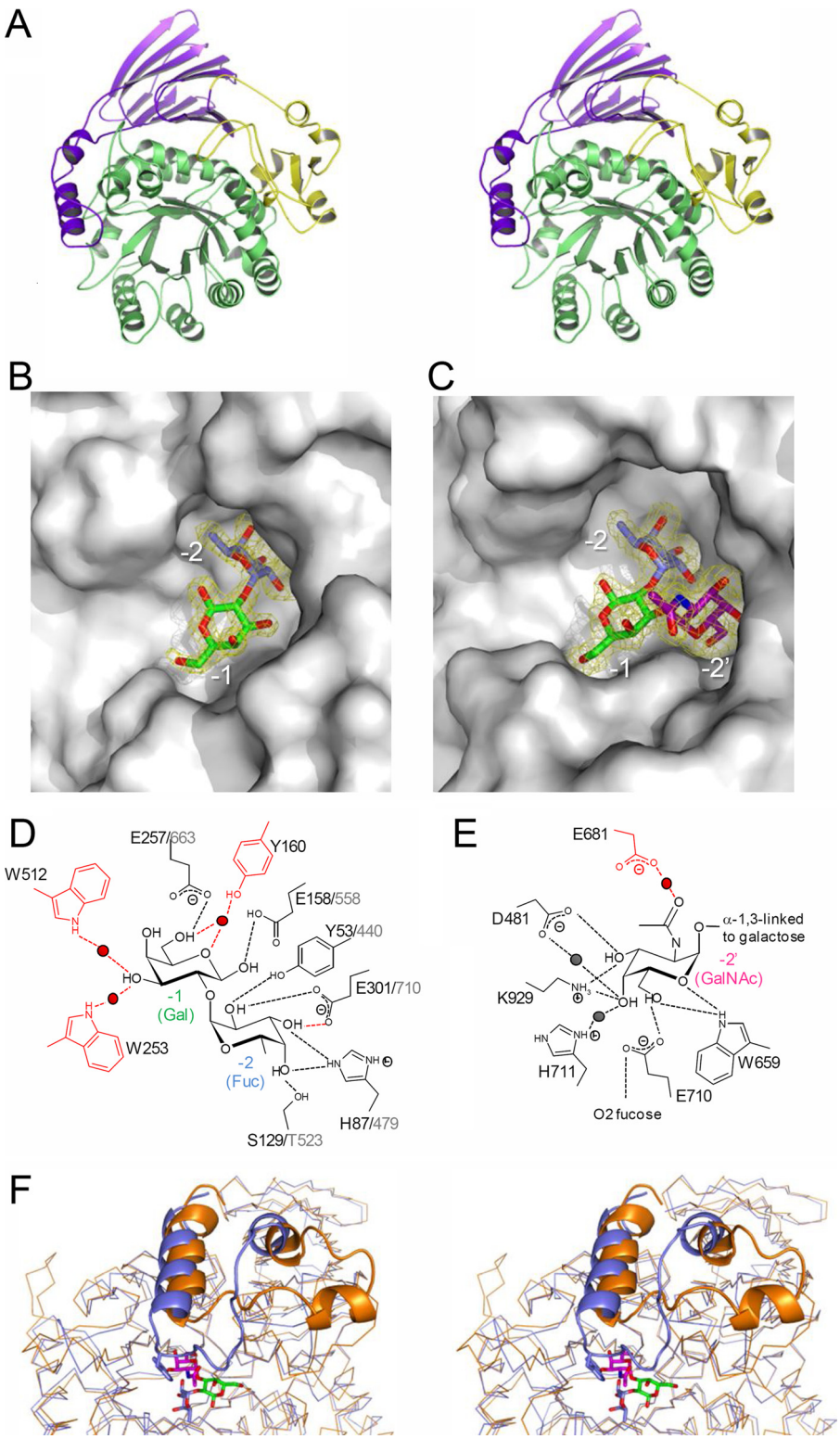

A
B
C
-2
-1
-2
-1
-2'
D
W512
E257/663
Y160
E158/558
Y53/440
E301/710
-1
(Gal)
W253
-2
(Fuc)
H87/479
S129/T523
E
E681
D481
α-1,3-linked
to galactose
-2'
(GalNAc)
K929
H711
W659
E710
O2 fucose
F

Figure 9: Structures features of Sp4GH98 and Sp3GH98

(On previous page) (A) Divergent stereo "cartoon" representation of the structure of Sp4GH98. The N-terminal $(\alpha/\beta)_8$ barrel housing the catalytic residues in shown in green, the C-terminal β-sandwich domain in purple, and the smaller intervening domain in yellow. Solvent accessible surface representations (grey) of the active site of Sp4GH98 containing the H-disaccharide (shown in stick representation) (B) and Sp3GH98 containing the A-trisaccharide (shown in stick representation) (C). Galactose is colored green, fucose colored blue, and N-acetylgalactosamine magenta. The yellow mesh shows the maximum-likelihood (Murshudov, Vagin et al. 1997)/σ_a-weighted (Read 1986) F_{obs}-F_{calc} omit-map for the carbohydrates contoured at 3σ (0.24 electrons/Å^3 for Sp4GH98 and 0.20 electrons/Å^3 for Sp3GH98). Subsites of the active site are labelled in white. (D) Schematic of the interactions in the -1 and -2 subsites of Sp4GH98 and Sp3GH98. Residues in black are conserved in both enzymes and labelled in black (Sp4GH98) or grey (Sp3GH98). Red residues are resent only in Sp4GH98. (E) Schematic of the -2' subsite of Sp3GH98. Black residues are shown for interactions that are conserved for the galactose of the B-antigen and the N-acetylgalactosamine of the A-antigen. The red residue represents an interaction that is unique to the A-antigen. (F) Divergent stereo representation of a structural overlay of the active sites of Sp4GH98 (blue) and Sp3GH98 (orange). The A-trisaccharide product in the Sp3GH98 active site is shown in sticks and coloured as in panel C. The loop protruding from the β-sandwich domain is shown in "cartoon" representation. W512 of Sp4GH98 that blocks the -2' subsite in this protein is shown in stick representation.

-2 subsite (Figure 9B, subsite nomenclature is according that established by Davies et al. (Davies, Wilson et al. 1997)). The chemical groups of this disaccharide make a number of hydrogen bonding and van der Waals interactions with the active site (Figure 9D). The active site of Sp3GH98 also accommodates the H-determining Fucα1-2Gal motif in -1 and -2 subsites that are very similar those of Sp4GH98 (Figure 9D). The most striking difference between the active sites of Sp3GH98 and Sp4GH98 is the presence of an additional subsite found in Sp3GH98, which enables binding of the branched blood group A- and B-trisaccharide antigens (Figure 9C and 9E). This subsite accommodates the α-1,3-linked terminal N-acetylgalactosamine (Group A) or galactose (Group B) residue. This subsite, which we term the -2′ subsite, is absent in Sp4GH98 and is clearly the key determinant of glycon specificity for these two enzymes and, more broadly, within family GH98. In Sp4GH98 this -2′ subsite is occluded by a tryptophan sidechain (W512) that extends from an α-helical insertion between β-strands 7 and 8 of the β-sandwich domain, preventing binding of group A- or B-antigens (Figure 9F). In contrast, the same insertion in the β-sandwich domain of Sp3GH98 assumes a conformation that extends at roughly right-angles to the active site, re-contouring the active site topography and thereby forming the -2′ subsite that makes specific interactions with the terminal residue of the A- or B-antigen determinant (Figure 9F). Thus, somewhat surprisingly, the formation of the -2′ subsite relies on the contribution of amino acid residues from the C-terminal β-sandwich domain, revealing the critical role of this domain in determining substrate specificity. This role played by the C-terminal domain may be what drives its conservation among family 98 glycoside hydrolases (Rigden 2005).

Amino acid side chains in the -2′ subsite form five hydrogen bonds with either the N-acetylgalactosamine of the A-antigen or the galactose of the B-antigen; none of these hydrogen bonds involve the substituent at C2 of the pyranose ring (Figure 9E). The 2-acetamido group of GalNAc, which is larger than the hydroxyl of galactose is readily accommodated because this substituent is solvent exposed. Relative to the galactose of the B-antigen, only one additional water mediated hydrogen bond between O7 of the GalNAc and the sidechain of E681 is made (Figure 9E). This lack of strong structural selectivity in the -2′ subsite provides a solid structural rationale for the ability of

Sp3GH98 to process the blood group A- and B-antigens with quite similar kinetic parameters.

Substrate complexes reveal differential aglycon recognition

Our analyses of Sp4GH98 revealed that it was able to hydrolyze the galactosyl-β-1,4-N-acetylglucosamine linkage in the Lewis^{Y} antigen but not the type 2 H-trisaccharide indicating a requirement for an aglycon having an α-1,3-fucosylated N-acetylglucosamine. Sp3GH98, however, displayed no such requirement thus suggesting that the two enzymes also have different aglycon specificities in addition to their divergent glycon preferences. To probe this difference we sought to inactivate the enzymes through site directed mutagenesis to allow substrate complexes to be trapped. The product complexes showed that E158 in Sp4GH98, and the analogous residue E558 in Sp3GH98, is positioned approximately 2.6 Å from O1 of the galactose in the -1 subsite strongly suggesting that this residues may function as a general acid, aiding departure of the aglycon moiety. On the basis of our predictions we sought to inactivate Sp4GH98 and Sp3GH98 through substitution of this residue with alanine in both enzymes; deletion of the carboxylate of the general acid catalyst has been shown, almost invariably, to have deleterious effects on the glycoside hydrolase mediated processing of substrates having a carbohydrate leaving group (Vocadlo and Davies 2008). Generation of these mutant proteins resulted in enzymes that crystallized with sufficiently low activity in the crystalline state so as to enable us to determine the structures of Sp4GH98E158A and Sp3GH98E558A in complex with intact Lewis^{Y} and blood group A-Lewis^{Y} antigen substrates, respectively. Clear electron density for both sugars allowed the substrates to be readily modelled (Figure 10A and 10B). The structures of the substrate complexes revealed the structural details of two additional subsites, +1 and +1'. The +1 subsite in both enzymes is conserved and a tryptophan residue (W161 in Sp4GH98 and W561 in Sp3GH98) that interacts with the GlcNAc residue of the substrate plays a key role by forming a classic protein:carbohydrate interaction whereby the pyranose ring lies parallel to the plane of the indole ring (Figure 10C). The O6 of this sugar residue hydrogen bonds with Y160 in Sp4GH98 or Y560 SP3GH98 (Figure 10D). The +1' subsite

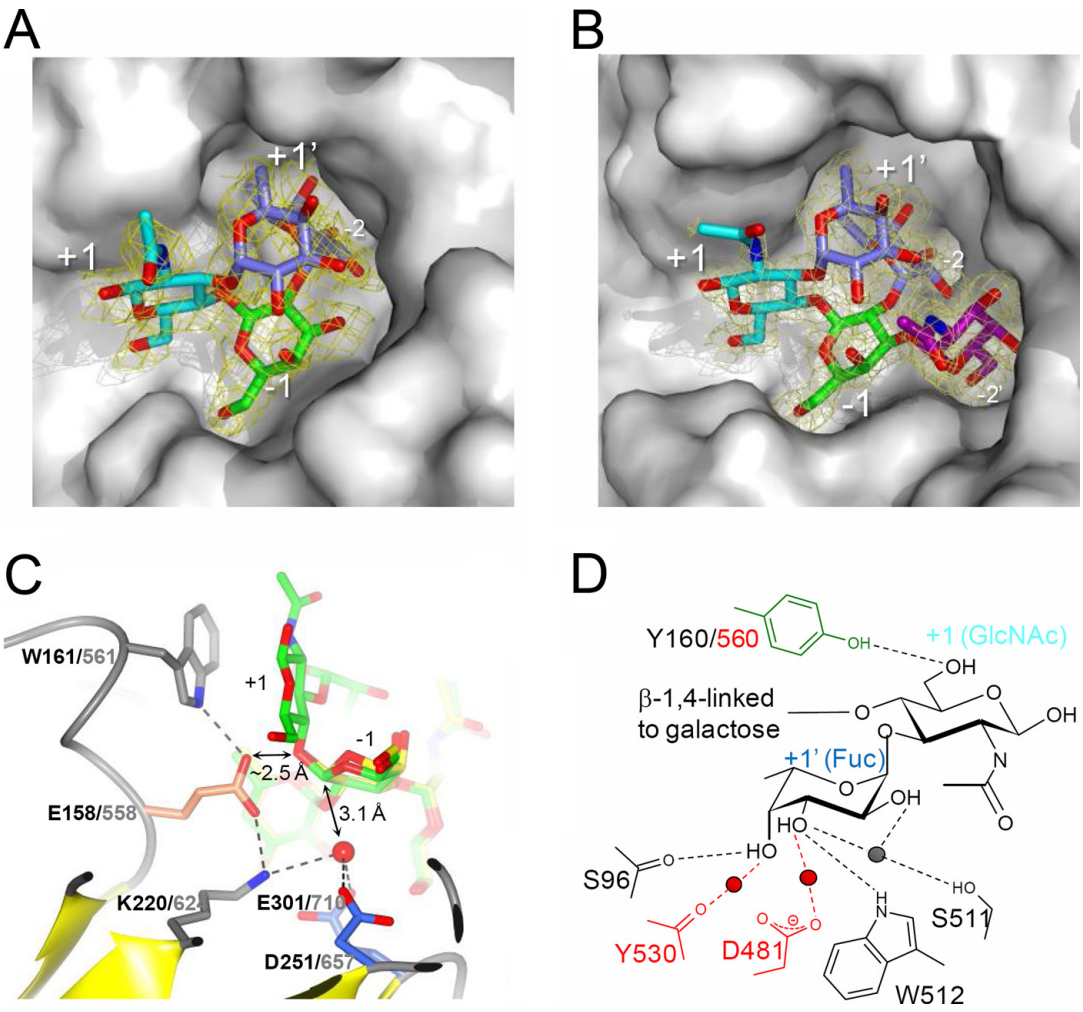


Figure 10: Aglycon recognition in GH98 enzymes.

Solvent accessible surface representation (grey) of the active site of Sp4GH98E158A containing the Lewis^Y tetrasaccharide substrate (shown in stick representation) (A) and Sp3GH98E558A containing the A-Lewis^Y pentasaccharide (shown in stick representation) (B). Galactose is coloured green, fucose coloured blue, N-acetylglucosamine in cyan, and N-acetylgalactosamine magenta. The yellow mesh shows the maximum-likelihood/σ_a-weighted F_{obs}-F_{calc} omit-maps for the carbohydrates contoured at 2.5σ (0.16 electrons/Å^3) for Sp4GH98E158A and 3σ (0.22 electrons/Å^3) for Sp3GH98E558A. Subsites of the active site are labelled in white. (C) Representative overlay of the GH98 active site. All GH98 structures were overlaid. Because of the virtually identical positioning of the active site structures the Sp3GH98 A-trisaccharide

complex was chosen as a reference point to display key features. The backbone of the Sp3GH98 A-trisaccharide product complex is shown in "cartoon" with relevant active site residues shown in stick representation. The A-trisaccharide sugar is shown as yellow sticks. The A-Lewis^{Y} pentasaccharide from the Sp3GH98E558A substrate complex is shown as green sticks. Residues in Sp3GH98 are labelled in grey and analogous residues in Sp4GH98, which were identically positioned, labelled in black. Sugar residues in the +1 and -1 subsites of Sp4GH98 were also virtually identically positioned to those of the Sp3GH98 complexes. Relevant inter-residue hydrogen bonds and protein-substrate distances are shown. The putative catalytic acid is coloured pink and putative catalytic bases coloured blue. (D) Schematic of the interactions in the +1 and +1' subsites. Interactions conserved between Sp4GH98 and Sp3GH98 are shown with green amino acids. Black amino acids are those only in Sp4GH98 and red amino acids are those only in Sp3GH98.

accommodates the α-1,3-linked fucose of the Lewis^Y antigen of both potential substrates, Lewis^Y and blood group A-Lewis^Y. This subsite in Sp4GH98 snugly houses this fucose residue and makes numerous van der Waals interactions (Figure 10A), in particular, the C6-methyl group of the fucose fits into a hydrophobic pocket formed by the apolar groups of the W161, A130, I98, and T95 sidechains (not shown). This close fit is complemented by both direct and water-mediated hydrogen bonds (Figure 10D). Likewise, the blood group A-Lewis^Y antigen complex of Sp3GH98 reveals that the α-1,3-linked fucose residue is comfortably accommodated in the +1' subsite of this enzyme; however, this subsite is particularly spacious in Sp3GH98 (Figure 10B), making limited van der Waals interactions and no direct hydrogen bonds with the fucose residue (Figure 10D). We speculate that this space provides Sp3GH98 with the flexibility to act on the both A/B antigens and the A/B-Lewis^Y antigens. For Sp4GH98, however, the secure fit of the Lewis^Y determining fucose residue suggests that the presence of an α-1,3-linked fucose in the substrate aglycon may play an important role in substrate recognition or catalysis although further studies will be required to address these finer details.

In summary, these two subtypes of GH98 enzymes recognize the type 2 H-trisaccharide core of their respective substrates through very similar interactions at the +1, -1, and -2 subsites while unique structural aspects of the -2' and +1' subsites impart precise substrate specificity to these two enzymes. In the Sp3GH98, the -2' subsite specifically accommodates the terminal GalNAc and galactose residues that define the blood group A and B antigens, respectively. This subsite is occluded in Sp4GH98 preventing binding of the A/B-glycotopes. In Sp4GH98 the +1' subsite specifically interacts with an α-1,3-linked fucose, a residue that is a defining group of the Lewis^Y antigen. In contrast, Sp3GH98 can accommodate this α-1,3-linked fucose residue but, because of an altered +1' subsite, makes limited interactions with it and does not require this residue on its substrate for activity.

Family 98 glycoside hydrolases use an inverting catalytic mechanism

One defining feature of glycoside hydrolases is the catalytic mechanism that is used (Vocadlo and Davies 2008). Many families of glycoside hydrolases use a retaining

catalytic mechanism in which the substrates is cleaved to generate a hemiacetal product with retained stereochemistry at the anomeric center. Others use an inverting mechanism, which as the name implies, results in a product having inverted stereochemistry at the anomeric center (see Section 1.4,1). Knowledge of the catalytic mechanism is useful to engineering these enzymes for biotechnology (Shaikh and Withers 2008) and to design effective inhibitors (Lillelund, Jensen et al. 2002; Vasella, Davies et al. 2002). For Sp4GH98 and Sp3GH98 the identical arrangement of conserved active site residues within the –1 subsite is most consistent with a single-step catalytic mechanism in which a molecule of water attacks the anomeric center, displacing the aglycon leaving group and generating a product with inverted stereochemistry at C1 of the glycon. K624, E710, and D657 in Sp3GH98 coordinate a water molecule that lies ~3.1 Å directly beneath the C1 of the galactose residue bound in the -1 subsite, which is perfectly poised to attack this anomeric center and displace the leaving group (Figure 10C). This region of the Sp4GH98 active site is identically organized and all the residues mentioned are conserved (Figure 10C). More detailed mechanistic studies will be required to precisely identify the functional roles of these three residues. These potential catalytic residues, however, are separated by ~6.5-7.5 Å (for example, in Sp3GH98, E558 to E710 or E558 to D657), which is a distance that is between what is generally expected to separate the acid/base and nucleophile in a retaining enzyme (~5.5 Å) and the acid and base in an inverting enzyme (~11 Å) (Mccarter and Withers 1994). Despite this ambiguity, the coordination and appropriate positioning of a putative catalytic water by E301 and D251 in Sp4GH98 and E710 and D657 in Sp3GH98 suggest an inverting catalytic mechanism in which these residues are candidate general bases.

The general acid catalytic residue in these enzymes are very likely E158 in Sp4GH98 and E558 in SP3GH98, since they are clearly position to fulfil this role and their deletion results in enzymes apparently lacking activity as discussed above. Indeed, Oε1 of E558 in Sp3GH98 is positioned ~2.5 Å from O1 of the reducing end galactose of the A-trisaccharide product, forming a strong hydrogen bond and allowing delivery of the proton to the glycosidic oxygen from the *syn*-trajectory (Figure 10C) (Vocadlo and Davies 2008). The enzyme-product complexes reveal in both cases, an identical and somewhat unusual conformation of the scissile glycosidic linkage of the galactosyl-β-1,4-

N-acetylglucosaminyl disaccharide fragment of the substrate that bridges the +1 and -1 catalytic subsites. The planes of the two pyranose rings of this portion of the substrate are related by roughly right angles with, as discussed above, the glycosidic oxygen accepting a hydrogen bond form the likely general acid (Figure 10C). The positioning of this catalytic group is stabilized by hydrogen bonding interactions with a tryptophan residue in the +1 subsite and K220 (Sp3GH98) or the equivalent K624 (Sp4GH98). These same lysine residues also interact with the nucleophilic water molecule and may help to enhance its nucleophilicty through electrostatics or controlling its orientation. This lysine residue likely also modulates the pKa values of the adjacent catalytic groups although its precise role awaits further detailed mechanistic studies.

To probe our gross mechanistic proposal that GH98 enzymes use an inverting mechanism we carried out analysis of the stereochemical outcome of the reaction catalyzed by Sp4GH98 using ^{1}H-NMR. A time course of the hydrolysis of the Lewis^{Y} tetrasaccharide (Fig. 11A) reveals production of the α-hemiacetal of the H-disaccharide (Fig. 11B) precedes that of the β-hemiacetal (Fig. 11C). These results are entirely consistent with GH98 enzymes using an inverting catalytic mechanism, in keeping with the structural predictions, and differing from previous tentative bioinformatics proposals (Rigden 2005).

Type 2 carbohydrate-antigen degradation by S. pneumoniae

Two independent studies have shown that the fucose utilization operon is critical to the virulence of the *S. pneumoniae* TIGR4 strain (Hava and Camilli 2002; Embry, Hinojosa et al. 2007); however, the biological target of the pathway encoded by this operon has not been elucidated. Interestingly, Sp4GH98 is the only predicted extracellular component of this pathway and thus very likely initiates the pathway by action on a host glycan. Here we have provided compelling biochemical and high-resolution structural evidence that Sp4GH98, and therefore most likely the entire pathway, is tuned to harvesting and processing the terminal H-disaccharide fragment from the Lewis^{Y} antigen. As we have noted, however, this particular pathway is only conserved in its entirety in 101 of the 120

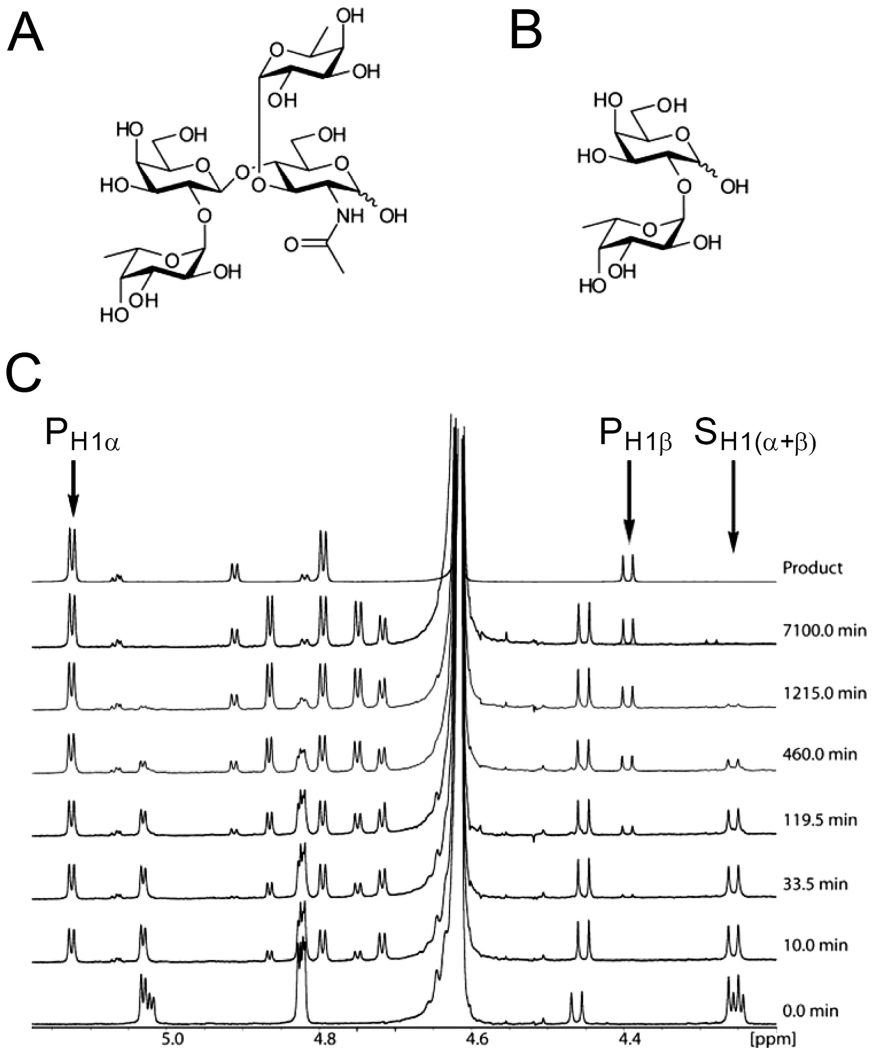


Figure 11: NMR analysis of GH98 catalytic mechanism.

(A) Structure of the Lewis^Y tetrasaccharide substrate. (B) Structure of the fucosylα1-2galactopyranosyl (H-disaccharide) product. (C) The hydrolysis of Lewis^Y tetrasaccharide measured as a function of time by ^{1}H NMR spectroscopy. Peaks corresponding to the chemical shifts of the proton on the anomeric carbon for subsrate (S) and product (P) are labelled for the α and β anomers.

available pneumococcal genomes. The other 19 genomes contain a variation of this fucose utilization pathway that is initiated by another family 98 glycoside hydrolase having a distinctive domain organization, here represented by Sp3GH98, which is active towards the type 2 A- and B- blood group glycotopes. Knowledge of the specificity of the two types of GH98 enzymes now allows us to propose a model for the harvesting and processing of fucosylated glycans antigens by *S. pneumoniae* (Figure 12). In general, it is evident that the remaining components of each pathway elegantly meet the unique requirements of transporting and processing the different fucosylated glycans liberated by the two types of GH98 enzymes.

Despite the divergent substrate specificity of the GH98 enzyme initiating these two pathways, the absolute conservation in all 120 pneumococcal strains of the components dedicated to processing the monosaccharide fucose suggests that fucose metabolism might be a generally important feature in the pneumococcus-host relationship. The interaction of *S. pneumoniae* with the host involves, of course, both colonization and invasion. The observation that the operon did not contribute to the host colonization abilities in the TIGR4 strain (Embry, Hinojosa et al. 2007) suggests that it likely features more prominently in the invasive component of *S. pneumoniae's* relationship with the host. We acknowledge, however, that the importance of the fucose utilization pathway to virulence in other pneumococcal strains remains to be demonstrated. Nevertheless, it is clear from our analysis of the GH98 enzymes from these two different pathways that the pneumococcus displays a strain dependent specificity for harvesting fucosylated glycans from human carbohydrate antigens bearing a type 2 core linkage. Though we have elucidated the likely biological targets of the GH98 enzymes, the precise impact of their activity on the biology and virulence of *S. pneumoniae* awaits further study.

A
Bacterial cell surface
Host-cell surface
M
PG+C
EIID
EIIC
EIIA
EIIB
Sugar independent components of PTS phosphorelay system
Sp4GH98
Central metabolism
DHAP + lactaldehyde
Fuculose-1-P
Fuculose
FcsK
FcsA
FcsI
FcsU
GH95A
6-P
B
Bacterial cell surface
Host-cell surface
Metabolism
Central metabolism
DHAP + lactaldehyde
Fuculose-1-P
Fuculose
FcsK
FcsA
FcsI
Cytoplasm
GH36A
GH29
GH36B
GH29
ATP
ADP + Pi
1
2
SBP
Sp3GH98
Lewis^Y
A-antigen
B-antigen
α1-2
β1-4
α1-3
Galactose
N-acetylglucosamine
Fucose
N-acetylgalactosamine

Figure 12: Schematics of the two proposed pathways for fucose utilization.

(On previous page). (A) the fucose utilization pathway involving the Sp4GH98-type enzyme; (B) the fucose utilization pathway involving the Sp3GH98-type enzyme. Enzymes are represented by ovals and colour-coded by general function: fucose processing (green), carbohydrate transport (purple), and glycans hydrolysis (blue). FcsA, -O, -K, and –U enzymes putatively acting as an aldolase, isomerase, kinase, and mutarotase, respectively. The EII components are those of a PTS transporter. SBP, 1, and 2 represent the solute binding protein (FcsSBP (Higgins, Abbott et al. 2009)) and two permease components, respectively, of an ABC transporter. The ATPase component of the ABC transporter is unidentified and is represented by gray ovals. GH95 and GH29, putative α-1,2-fucosidases; GH36A, a putative α-N-acetylgalactosaminidase; GH36B, a putative a-galactosidase. PG+C, peptidoglycan and capsule layers; M, membrane. Details of the sugar notation and glycans structures are shown below B. Further details of the homology based prediction of component functions are given in the Appendix.

Chapter 3: The overall architecture and receptor binding of pneumococcal carbohydrate-antigen hydrolyzing enzymes.

Melanie A. Higgins[1], Elizabeth Ficko-Blean[1], Caitlin Wright[1], Peter A. Meloncelli[2], Todd L. Lowary[2], and Alisdair B. Boraston[1]

Adapted from The overall architecture and receptor binding of pneumococcal carbohydrate-antigen hydrolyzing enzymes. J Mol Biol. (2011) 411(5):1017-36.

[1]Biochemistry & Microbiology, University of Victoria, PO Box 3055 STN CSC, Victoria, BC, V8W 3P6, Canada.
[2]Alberta Ingenuity Centre for Carbohydrate Science and Department of Chemistry, The University of Alberta, Gunning-Lemieux Chemistry Centre, Edmonton, Alberta, Canada.

Contributions: I performed the cloning, protein production and purification, crystallography, protein preparation for the microarray, and ITC experiments. Elizabeth Ficko-Blean collected and analysed the SAXS data.

3.1 Abstract

The TIGR4 and SP3-BS71 strains of *Streptococcus pneumoniae* each produce family 98 glycoside hydrolases, called Sp4GH98 and Sp3GH98, respectively, which have different modular architectures and substrate specificities. Sp4GH98 degrades the Lewis^Y antigen and possesses three C-terminal family 47 CBMs that bind to this substrate. Sp3GH98 degrades the blood-group A/B antigens and has two N-terminal family 51 CBMs that are of unknown function. Here we examine the complex carbohydrate-binding specificity of the family 51 CBMs from Sp3GH98 (referred to as CBM51-1 and CBM51-2), the structural basis of this interaction, and the overall solution conformations of both Sp3GH98 and Sp4GH98, which are shown to be fully secreted proteins. Through glycan microarray binding analysis and isothermal titration calorimetry CBM51-1 is found to bind specifically to the blood-group A/B antigens. However, due to a series of relatively small structural rearrangements that were revealed in structures determined by X-ray crystallography, CBM51-2 appears to be incapable of binding carbohydrates. Analysis of small-angle X-ray scattering data in combination with the available high-resolution X-ray crystals structures of the Sp3GH98 and Sp4GH98 catalytic modules and their CBMs yielded models of the biological solution structures of the full-length enzymes. These

show the complex architectures of the two enzymes and suggest that carbohydrate-recognition by the CBMs and the activity of the catalytic modules are not directly coupled.

3.2 Introduction:

S. pneumoniae encodes for a putative fucose utilization operon that has been relatively uncharacterized. Bioinformatic methods suggest a common final product; however, it has recently been shown that the enzymes predicted to initiate the pathways, both family 98 glycoside hydrolases (GH98), are active on different substrates (Higgins, Whitworth et al. 2009). The GH98 from the TIGR4 strain, called Sp4GH98, is specific for the Lewis^Y antigen and, through hydrolysis of the internal type 2 Galβ1-4GlcNAc linkage, releases the H-disaccharide (Fucα1-2Gal). The GH98 from the SP3-BS71 strain, called Sp3GH98, is specific for the blood-group A and B (BGA and BGB) antigens and by hydrolysis of the same type 2 linkage releases the A-trisaccharide (GalNAcα1-3[Fucα1-2]Gal) and B-trisaccharide (Galα1-3[Fucα1-2]Gal), respectively.

Both Sp3GH98 and Sp4GH98 are multimodular, each comprising multiple CBMs and a single catalytic module (Figure 13). Sp4GH98 has an N-terminal family 98 catalytic module, whose structure is known (Higgins, Whitworth et al. 2009), and three C-terminal family 47 CBMs, whose structures and Lewis^Y antigen binding have been investigated (Boraston, Wang et al. 2006). Sp3GH98 possesses a different modular architecture with two N-terminal family 51 CBMs (CBM51-1 and CBM51-2) that are uncharacterized and a C-terminal family 98 catalytic module, whose structure is known (Higgins, Whitworth et al. 2009). Both proteins contain predicted N-terminal Gram-positive secretion signal sequences.

The modules of the pneumococcal GH98 enzymes, with the exception of the CBM51 modules from Sp3GH98, are individually well characterized at the structural and functional levels, which gives an appreciation of their isolated functions: carbohydrate hydrolysis or carbohydrate binding. What is presently lacking, however, is an appreciation of how these functions are integrated in the context of the intact protein. Recent evidence indicates that the activity of the catalytic module might be coordinated

with carbohydrate binding by the CBMs simply by the relative orientation of the modules (Ficko-Blean, Gregg et al. 2009). In a more profound example, the overall organization

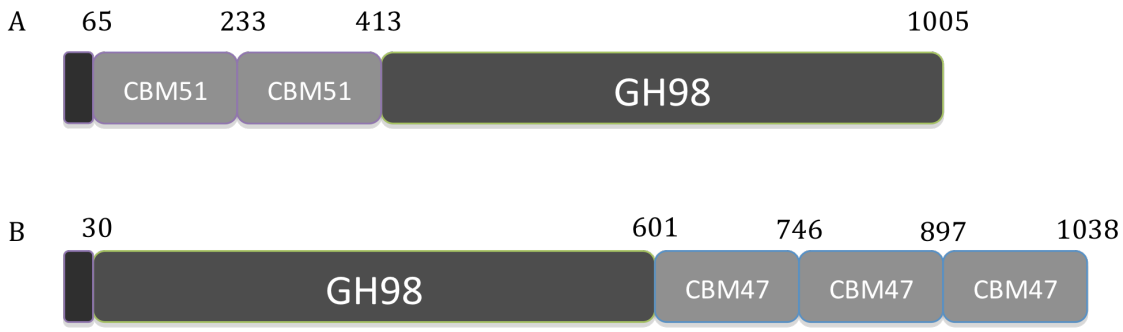


Figure 13: Modular architecture of Sp4GH98 and Sp3GH98

Schematic representations of the modular structure of Sp3GH98 (A) from *S. pneumoniae* SP3-BS71 and Sp4GH98 (B) from *S. pneumoniae* TIGR 4. Amino acid numbers denoting the modular boundaries are shown. The white boxes at the N-termini represent predicted Gram-positive secretion signal peptides.

of the enzyme can result in the CBMs comprising an integral part of the active site (Lammerts van Bueren, Ficko-Blean et al. 2011). Moreover, the role of the CBMs and catalytic modules may be influenced by where the protein is localized: in proteins that are attached to the bacterial cell surface CBMs may function as carbohydrate specific adhesins whereas CBMs in soluble extracellular proteins likely target the individual enzyme, rather than the bacterium, to appropriate glycans. To begin to address these issues we sought to study the cellular locations of the Sp3GH98 and Sp4GH98 enzymes and the overall architectures of the two proteins. The latter goal first required a structural and functional analysis of the CBM51 modules in Sp3GH98, which was pursued using glycan microarray screening for specificity, isothermal titration calorimetry for quantitative binding analysis, and X-ray crystallography to determine the molecular basis of ligand binding. The overall solution structures of Sp3GH98 and Sp4GH98 were determined by small-angle X-ray scattering in combination with the high-resolution X-ray structures of the isolated modules from these proteins. These data reveal a picture of architecturally complex molecules that are deployed as soluble extracellular proteins that are likely targeted to host-glycans by their CBMs followed by degradation of the glycan by the catalytic module.

3.3 Materials and Methods

Strains and growth conditions

S. pneumoniae TIGR4 was obtained from American Type Culture Collection (ATCC BAA-334) and *S. pneumoniae* SP3-BS71 was kindly provided by Dr. Garth Ehrlich. Both strains were grown on Tryptic soy agar (TSA) plates supplemented with 17% defibrinated sheep blood (BioMerieux) and in AGCH minimal media (Lacks 1966) containing 0.5% yeast extract (AGCHY) supplemented with specific monosaccharides as carbon sources and grown at 37 °C in a candle jar.

Cloning and protein production and purification

The gene fragments encoding CBM51-1.2 (amino acids 65-413), CBM51-1 (amino acids 65-233), and CBM51-2 (amino acids 248-413) were amplified by PCR from *S. pneumoniae* SP3-BS71 genomic DNA (kindly provided by Dr. Garth Ehrlich) and cloned into pET28a(+) (Novagen) via engineered *NheI* and *XhoI* restriction sites. A list of

Table 7: Primer list for Sp3GH98, Sp4GH98, and CBM51 recombinant proteins.

Primer name	Primer sequence	Construct
Sp3GH98-F	5’ CATATGGCTAGCATGACATATCTTTCAGAT	Sp3GH98, CBM51-1, CBM51-1.2
Sp3GH98-R	5’ GTGGTGCTCGAGTTAAAAATGAAGTTCGAATTC	Sp3GH98
CBM51-1-R	5’ GTGGTGCTCGAGTTAAGTACTAACAGTTTTAAT	CBM51-1
CBM51-2-F	5’ CATATGGCTAGCGGAGTATATCTTTCTGAT	CBM51-2
CBM51-2-R	5’ GTGGTGCTCGAGTTAACCATTTGCCATAAACAA	CBM51-2, CBM51-1.2
Sp4GH98-F	5’CATATGGCTAGCGCTGATAATCGTGTTCAAATG AGAACG	Sp4GH98
Sp4GH98-R	5’GTGGTGCTCGAGTTACTTATAAGTATAAACCTCT ACCTC	Sp4GH98

Underlined DNA sequences show the restriction endonuclease sites.

primers can be found in Table 7. The DNA sequence fidelity of the clones was verified by bidirectional sequencing. All three proteins were produced in E. coli BL21 Star (DE3) cells (Novagen) using LB media supplemented with 50 μg mL^{-1} of kanamycin. Cultures were grown at 37°C to an optical density of 0.5-0.7 at 600 nm, induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) and further incubated at 16°C overnight with shaking. Cells were harvested by centrifugation and ruptured by chemical lysis. Target proteins were purified by Ni^{2+} immobilized metal affinity chromatography (GE Healthcare). Protein containing fractions were pooled and dialyzed into 20 mM Tris-HCl, pH 8.0, and concentrated in a stirred ultracentrifugation device (Amicon, Beverly, MA) using a 5 kD molecular weight cut-off membrane (Filtron, Northborough, MA). Protein used for crystallization was thrombin digested to remove the $(His)_6$ tag and further purified by size exclusion chromatography using a Sephacryl S-200 column (GE Biosciences) equilibrated in 20 mM Tris-HCl, pH 8.0.

Gene fragments encoding mature Sp3GH98 (amino acids 65-1005) and Sp4GH98 (amino acids 31-1038) were PCR amplified from *S. pneumoniae* SP3-BS71 and TIGR4 genomic DNA, respectively. These fragments were cloned and the proteins ultimately produced in *E. coli* as described above. Primers are listed in Table 7. Sp3GH98 was purified by Ni^{2+} immobilized metal affinity chromatography and size exclusion chromatography using a Sephacryl S-200 column (GE Biosciences) equilibrated in 20 mM Tris-HCl, pH 8.0 and 250 mM NaCl. Sp4GH98 was initially purified using Ni^{2+} immobilized metal affinity chromatography followed by further purification over a resin with immobilized fucose (EY Laboratories) and then size exclusion chromatography using a Sephacryl S-200 column (GE Biosciences) equilibrated in 20 mM Tris-HCl, pH 8.0 and 1 mM DTT (Sigma).

Protein concentration was determined by UV absorbance at 280 nm using calculated molar extinction coefficients (Gasteiger, Gattiker et al. 2003) of 42860 $cm^{-1}M^{-1}$, 19940 $cm^{-1}M^{-1}$, and 22920 $cm^{-1}M^{-1}$, for CBM51-1.2, CBM51-1 and CBM51-2, respectively and 191030 $cm^{-1}M^{-1}$ and 213280 $cm^{-1}M^{-1}$ for Sp3GH98 and Sp4GH98, respectively.

Carbohydrate microarray

CBM51-1 and CBM51-2 proteins were labelled by coupling to Alexo Fluor® 488 labelled streptavidin via a biotin-NTA:Ni^{2+} linker. Briefly, streptavidin conjugated to Alexo Fluor® 488 (Invitrogen S11223) was complexed with an excess of biotin-NTA:Ni^{2+} and purified by gel filtration chromatography using Sephadex G-25 (GE Healthcare). Recombinant His-tagged protein was then complexed with an excess of the biotin-NTA:Ni^{2}/streptavidin complex before being used to probe the printed glycan arrays following the standard procedure of Core H of the Consortium for Functional Glycomics (www.functionalglycomics.org/).

Isothermal titration calorimetry. Isothermal titration calorimetry was performed as previously described (Ficko-Blean and Boraston 2006) using a VP-ITC (Microcal, Northampton, MA). Proteins were dialysed simultaneously and extensively against 20 mM Tris-HCl, pH 8.0, to ensure that all the proteins were in identical buffer systems. Buffer saved from the dialysis was used to prepare carbohydrate solutions by mass; all solutions were filtered and degassed immediately prior to use. Synthetic type 2 A- or B-tetrasaccharide (Meloncelli and Lowary 2010) was titrated into CBM51-1, CBM51-2, and CBM51-1.2 (all >150 μM) under conditions maintained such that C-values were in excess of 5. Heat of dilution corrected data were fit to a one-site binding model to yield the association constant (K_a), change in enthalpy (ΔH), and stoichiometry (n). The change in Gibb's free energy (ΔG) and change in entropy (ΔS) were calculated from these values.

Antibody production and western blotting

Purified recombinant GH98 catalytic module from either TIGR4 or SP3-BS71 strains, Sp4GH98CM and Sp3GH98CM (Higgins, Whitworth et al. 2009), respectively, and FcsI from TIGR4, SpFcsI, were used to immunize rabbits for the production of polyclonal anti-sera against those proteins (ImmunoPrecise, Victoria, Canada).

S. pneumoniae TIGR4 or SP3-BS71 was streaked out on TSA plates supplemented with defibrinated sheep blood and incubated for 12 hours at 37°C in a candle jar. Cultures were started by inoculating 15-20 mL of AGCHY media supplemented with 1%

D-galactose and 1% L-fucose to induce fucose operon expression (Chan, O'Dwyer et al. 2003) and incubated for 8-9 hours in a candle jar at 37°C. Cultures were subsequently pelleted at 8000 RPM for 5 minutes. The *S. pneumoniae* TIGR4 bacterial pellet was resuspended in phosphate-buffered saline, pH 7.4 (PBS) with protease inhibitor (Roche) and lysed by sonication to yield the cellular fraction. Supernatant from the *S. pneumoniae* TIGR4 culture was saved for further trichloroacetic acid (TCA) precipitation steps to concentrate secreted proteins. Briefly, one part 100% TCA was added to 4 parts supernatant, left on ice for approximately 30 minutes, and centrifuged for 5 minutes at 14000 RPM. The precipitated pellet was washed twice with cold acetone, dried at room temperature, and resuspended in SDS loading buffer yielding the extracellular fraction. *S. pneumoniae* SP3-BS71 did not produce a condensed pellet like that of TIGR4. Instead, the bacteria formed a mucus-like consistency that was directly sonicated giving the cellular fraction. The *S. pneumoniae* SP3-BS71 supernatant was collected for an ethanol precipitation step. Briefly, 12 volumes of cold 95% ethanol was added to 1 volume of supernatant and kept at -20°C for two nights. Precipitant was centrifuged and the pellet washed with cold 95% ethanol then dried at room temperature and resuspended in SDS loading buffer.

Each fraction was run on a 10% SDS-polyacrylamide gel electrophoresis (PAGE), transferred to a nitrocellulose membrane (Whatman), and analysed by Western blot using a 1/7,000, 1/4,000, and 1/10,000 dilution of rabbit anti-serum raised against purified recombinant GH98 catalytic module from SP3-BS71 or TIGR4 strains or FcsI, respectively (Higgins, Whitworth et al. 2009). A 1/20,000 dilution of IRDye800 conjugated anti-rabbit IgG antibody raised in goat (Rockland 611-132-002) was used as a secondary antibody and imaged with a LICOR Odyssey System (LICOR Biosciences) at 800 nm.

Crystal structure determination

Crystals of uncomplexed CBM51-1.2 (23 mg/ml) were obtained in 17% polyethylene glycol 8K, 0.18 M magnesium acetate, 0.1M sodium cacodylate, pH 6.8. protein. Crystals of CBM51-1.2 (23 mg/ml) in complex with the blood group B tetrasaccharide were obtained by co-crystallization of the sugar at a final concentration of 5 mM in 17%

polyethylene glycol 3350, 0.2 M potassium bromide, 0.1 M sodium cacodylate, pH 6.2. I performed all crystallization experiments with the hanging drop vapour diffusion method at 18 °C. Crystals were cryoprotected in crystallization solution supplemented with 25% ethylene glycol and flash cooled at 113 K. Diffraction data were collected at either BL 9-2 (Stanford Synchrotron Radiation Laboratories) or CMCF1 (Canadian Light Source) and processed with MOSFLM (Powell 1999) and SCALA (1994) (Table 8).

The structure of uncomplexed CBM51-1.2 was solved by molecular replacement. Using the coordinates of CpGH98CBM51 (PDB code 2vnr; (Gregg, Finn et al. 2008)) as a search model PHASER (McCoy, Grosse-Kunstleve et al. 2007) was able to find four clear solutions corresponding to the two CBM51 modules present in two CBM51-1.2 tandems present in the asymmetric unit. This initial model was manually corrected and individual modules rearranged using COOT (Emsley and Cowtan 2004) such that two complete CBM51-1.2 polypeptide models could be built. This was iteratively corrected and refined with COOT and REFMAC (Murshudov, Vagin et al. 1997), respectively.

The structure of CBM51-1.2 complexed with BGB tetrasaccharide (obtained from Core H of the Consortium for Functional Glycomics) was solved by molecular replacement using MOLREP (Vagin and Teplyakov 2000) and the individual CBM51-1 and CBM51-2 modules from the uncomplexed structure as search models. This yielded clear solutions for two CBM51-1 modules and two CBM51-2 modules comprising the two molecules of CBM51-1.2 in the asymmetric unit. The resulting solutions were rearranged such that two complete models of the CBM51-1.2 polypeptide chain could be built. This model was corrected and refined with COOT and REFMAC, respectively.

In all cases, the addition of water molecules was performed in COOT with FINDWATERS and manually checked after refinement. Refinement procedures were monitored by flagging 5% of all observation as ‘free’ (Brunger 1992). Model validation was performed with SFCHECK (Vaguine, Richelle et al. 1999), PROCHECK (Laskowski, Moss et al. 1993), and MOLPROBITY (Chen, Arendall et al. 2010). All model statistics are shown in Table 8.

Table 8: Data collection and refinement statistics for CBM51-1.2

	CBM51-1.2	CBM51-1.2 + BGB
Data collection		
Beamline	SSRL BL 9.2	CLS CMCF1
Wavelength	0.97607 Å	0.92140 Å
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a, *b*, *c* (Å)	76.0, 90.6, 119.9	75.7, 78.5, 131.6
Resolution (Å)	40.00-2.10 (2.21-2.10)	30.00-2.35 (2.48-2.35)
R_{merge}	0.095 (0.383)	0.079 (0.406)
$I / \sigma I$	22.9 (7.3)	12.6 (4.0)
Completeness (%)	99.7 (99.7)	99.9 (99.9)
Redundancy	13.6 (13.3)	4.9 (4.9)
Refinement		
Resolution (Å)	2.10	2.35
No. reflections	46386	31713
R_{work} / R_{free}	0.21/.25	0.25/0.30
No. atoms		
Protein	2706 (A); 2702 (B)	2712 (A); 2712 (B)
Ligand	NA	47 (BGB 1); 47 (BGB 2)
Water	624	175
B-factors (Å^2)		
Protein	22.3 (A); 23.0 (B)	45.4 (A); 47.1 (B)
Ligand	NA	57.4 (BGB 1); 52.8 (BGB 2)
Water	34.4	42.0
R.m.s. deviations		
Bond lengths (Å)	0.006	0.006
Bond angles (°)	0.93	0.98
Ramachandran		
Preferred (%)	97.7	96.3
Allowed (%)	2.3	3.6
Disallowed (%)	0.0	0.1

Values in parentheses are for the highest-resolution shell.

Small-angle X-ray scattering (SAXS)

Solute samples for SAXS experiments were prepared from protein purified as above, flow through was collected during concentration of the protein samples from an Amicon stir cell concentrator for the solvent blanks. The experimental buffer used for Sp3GH98 was 20 mM Tris pH 8.0 and 150 mM NaCl. For Sp4GH98 the experimental buffer used was 20 mM Tris pH 8.0 and 1 mM DTT. The BSA reference, purchased from Sigma, was prepared in 20mM Tris pH 8.0 and 75 mM NaCl and aggregate was removed by gel filtration on a Sephacryl S-200 HR column from GE Healthcare. SAXS data was collected over a sample concentration ranging from 0.4-13.6 mg/ml depending on the protein. SAXS data were recorded on Beamline 4-2 (BL4-2) at the Stanford Synchrotron Radiation Laboratory (Smolsky, Liu et al. 2007) using a MarCCD165 detector. The sample to detector distance was 1.7 m with associated scattering vectors q (defined as $q = 4\pi/\lambda \sin\theta$, where 2θ is the scattering angle) in the range of $0.008\ Å^{-1} < q < 0.511\ Å^{-1}$ and the beam energy was 11 keV, corresponding to a wavelength of 1.127 Å. A data series of 10 frames with 3 second exposures were collected for each concentration sample or buffer. The program SasTool was used for integration of the 2-dimensional scattering patterns and reduction to a 1-dimensional scattering profile, normalizing for incident beam intensity, correcting for detector response and averaging over the entire data series. The buffer data was subtracted from the sample data using the program PRIMUS (Konarev, Volkov et al. 2003) from the ATSAS suite of software, resulting in a buffer subtracted scattering curve.

The Guinier approximation (Guinier and Fournet 1955), $\ln[I(q)]=\ln[I(0)]-q^2R_G^2/3$, where $I(q)$ is the scattered intensity and $I(0)$ is the forward scattered intensity extrapolated at zero angle, was used to determine the radius of gyration, Rg, and $I(0)$ by plotting data in the low q range as $\ln[I(q)]$ vs q^2. R_G is estimated from the slope and $I(0)$ from the intercept of the linear fit of this plot in the q-range $q \cdot R_G < 1.3$. At low angles the scattered intensities were well approximated by the Guinier approximation. All scattering curves were indicative of monomers in solution. The pair-density distance distribution function, P(r), which provides real space information about the distances between electrons in the scattering sample, was calculated by indirect Fourier transform

methods of the scattering intensity, *I(q)*, using the program GNOM (Svergun 1992). The molecular weights of Sp3GH98 and Sp4GH98 were also determined by using experimental data from a single SAXS curve on a relative scale and within a restricted q range as described by Fischer *et al* (Fischer, Neto et al. 2010).

Low resolution envelopes of the proteins were generated *ab initio* using the program DAMMIF (Franke and Svergun 2009). 20 independent DAMMIF reconstruction runs were done with no shape constraints introduced. The DAMAVER (Volkov and Svergun 2003) suite of programs was used to align the *ab initio* models, select for the most typical one, and produce an averaged model. The normal spatial discrepancy (NSD) values were computed using DAMAVER. The X-ray crystallographic atomic coordinates of the individual modules were manually fit to the DAMAVER generated envelopes using the program PyMol. CRYSOL (Svergun, Barberato et al. 1995) was used to evaluate the solution scattering for the atomic molecules, fit the SAXS experimental curve, and calculate the goodness of fit by minimizing the discrepancy (Chi-square value). The combined *ab initio* and rigid body fitting program BUNCH (Petoukhov and Svergun 2005) was used to generate models from the scattering data using the available atomic coordinates over a series of scattering datasets. For Sp3GH98 BUNCH was used for rigid body modeling, treating the four available crystallographic dimer conformations of CBM51-1.2 as rigid bodies alongside the Sp3GH98 catalytic module (PDB code 2WMI). BUNCH was also run with the Sp3GH98 catalytic module and each of the individual CBM51 modules treated as rigid bodies. BUNCH was also used for rigid body modeling on Sp4GH98 and each module was treated as an independent rigid bodies: Sp4GH98 catalytic module (PDB code 2WMF), CBM47-1 (PDB code 2J1R), CBM47-2, and CBM47-3 (PDB code 2J22). Since atomic coordinates are only available for CBM47-1 and CBM47-3, CBM47-3 was used as a substitute model for CBM47-2. SUBCOMB (Kozin and Svergun 2001) was used to superimpose the 3D structures generated onto the averaged DAMMIF model described above.

The SAXS data collected on Sp3GH98 was also analyzed by the EOM (Ensemble Optimization Method) (Bernado, Mylonas et al. 2007). Sp3GH98 was initially considered as having three rigid bodies: CBM51-1, CBM51-2 and the catalytic module. The program RanCh was used to generate a pool of 10 000 models where the conformations of the

flexible linker regions was varied randomly. Selection of the optimized ensemble was done using GAJOE and the default program parameters (1000 generations, 50 ensembles of theoretical curves to compare with experimental data, 20 curves per ensemble, 10 mutations per ensemble and 20 crossings per generation). This generated an ensemble where CBM51-1 was distributed almost randomly but CBM51-2 adopted three distinct conformational clusters relative to the catalytic module. An average position for CBM51-2 from each of the three clusters was selected, fixed in its position relative to the catalytic module and EOM re-performed as above using CBM51-2/catalytic module as one component and CBM51-1 as the second component. Thus, flexibility was now considered only between CBM51-1 and CBM51-2. RanCh then generated a pool of 10 000 models for each of these options and GAJOE was used to select the ensembles. CRYSOL was used to evaluate the solution scattering for each of the individual atomic molecules from EOM, fit the SAXS experimental curve, and calculate the goodness of fit by minimizing the discrepancy (Chi-square value).

3.4 Results and Discussion

Family 98 glycoside hydrolases are secreted by S. pneumoniae

Antibodies raised against either Sp3GH98 or Sp4GH98 catalytic modules were used to determine the cellular location of these enzymes in *S. pneumoniae* SP3-BS71 and TIGR4, respectively, by western blotting cellular fractions. Both GH98 enzymes were found solely in the protein fraction precipitated from culture medium that was previously cleared of bacterial cells (Figure 14). In contrast, fucose isomerase (SpFcsI), which was used as an intracellular control, was present only in the cellular fraction (Figure 14). Along with bioinformatics data predicting the presence of an N-terminal secretion sequence and absence of a classical cell wall attachment motif (not shown), this data gives strong evidence that the GH98 enzymes are fully secreted from the bacterium into the extracellular medium.

The N-terminal CBM51 of Sp3GH98 binds blood-group A and B antigens

The two N-terminal family 51 CBMs from Sp3GH98, CBM51-1 and CBM51-2, show roughly 40% amino acid sequence identity to the BGA and BGB antigen specific

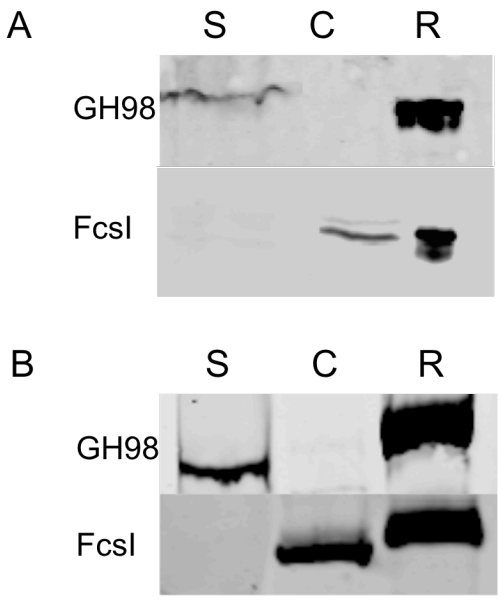


Figure 14: Cellular location of the GH98 enzymes.

The cellular localization of GH98 and FcsI *S. pneumoniae* SP3-BS71 (A) and *S. pneumoniae* TIGR4 (B) cultures. S indicates the secreted fraction obtained by precipitation of the culture medium, C indicates the total cellular fraction, and R indicates purified recombinant protein.

N-terminal CBM51 found on the family 98 glycoside hydrolase EABase from *C. perfringens* (Gregg, Finn et al. 2008). Furthermore, Sp3GH98 is known to have catalytic specificity for the BGA and BGB antigens indicating that the two CBMs from Sp3GH98 might be specific for the A and/or B blood group antigens. To test this hypothesis the gene fragments encoding separate CBM51-1 and CBM51-2 polypeptides were cloned and the polypeptides produced and purified. The purified polypeptides were then screened for binding specificity by glycan microarray analysis through Core H of the Consortium for Functional Glycomics. CBM51-1 showed strong binding to the arrays in a pattern consistent with specificity for the BGA and BGB antigen motifs (Figure 15). More specifically, this CBM bound only to the type 2 antigens with a Galβ1-4GlcNAc core and showed no capacity to bind to the type 1 antigens with a Galβ1-3GlcNAc core. Surprisingly, CBM51-2 showed no glycan binding capacity on the array (not shown).

To confirm the array results we further explored the ability of the CBM51 modules to bind BGA and BGB by isothermal titration calorimetry (Table 9). Consistent with the microarray analysis CBM51-1 bound well to the type 2 BGA and BGB with a 1:1 (carbohydrate:protein) stoichiometry whereas CBM51-2 showed no binding. The tandem construct of the these two CBMs, CBM51-1.2, also bound well to the antigens but with a 1:1 stoichiometry, rather than the 2:1 stoichiometry one would expect if both CBMs were functional. Thus the results clearly support the array results and indicate that, even in the context of the tandem, only CBM51-1 is functional.

Based on the ΔG values CBM51-1 binds the blood group A- and B-antigens with indistinguishable affinities. However, the ΔH and ΔS parameters differed significantly. Binding to BGA was associated with a more favourable change in enthalpy than binding to BGB; however, this was offset by a less favourable change in entropy. The exact source of this compensation is unknown but because BGA and BGB differ by only the C2-acetamido group on the blood group-A determining N-acetylgalactosamine it is very likely that the different thermodynamics of recognition arise from accommodation of this chemical moiety *versus* the O2-hydroxyl of the blood-group B determining galactose.

Like the CBM51 from EABase, CBM51-1 from Sp3GH98 is a blood-group A/B-antigen specific binding module, which is consistent with the specificity of the Sp3GH98 catalytic module. The inability of CBM51-2 to bind carbohydrate was unexpected given

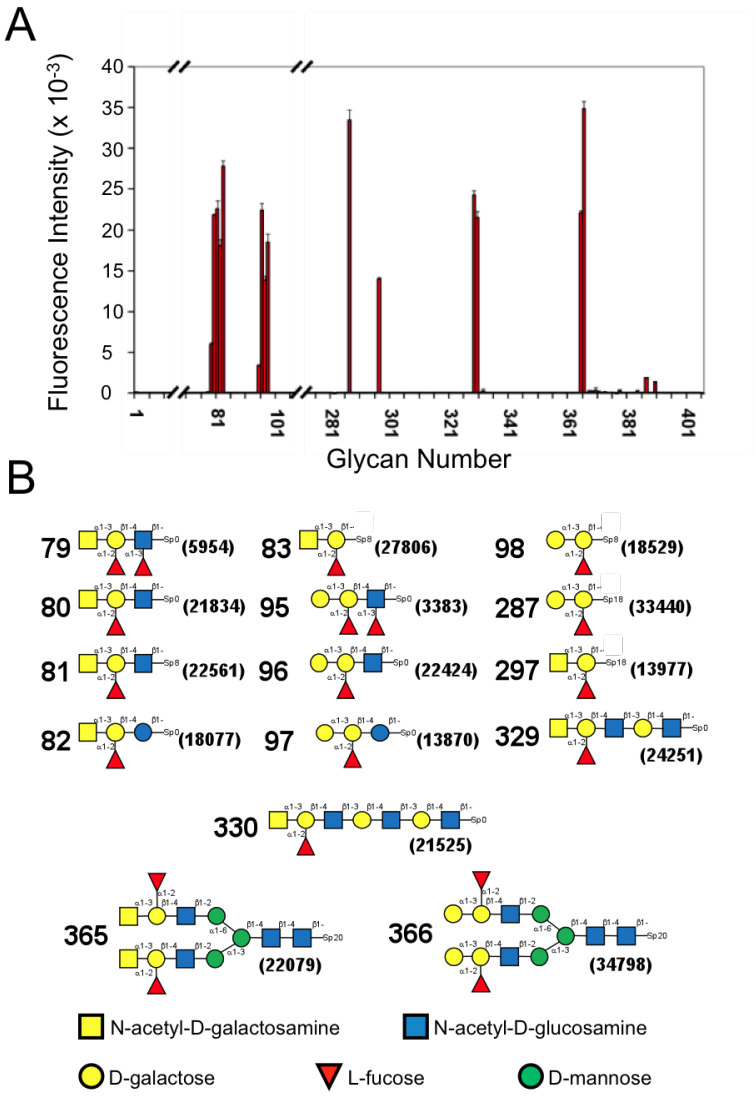


Figure 15: CBM51-1 and CBM51-2 glycan arrays analysis

Glycan array binding analysis of CBM51-1 showing the array results (A) and the schematic structures of glycans to which substantial binding was observed (B). The glycan array number of the oligosaccharide is given to the left of the schematic while the raw fluorescence intensity from the array analysis is given in parentheses to the right of the schematic.

Table 9: Isothermal titration calorimetry analysis of blood-group antigen binding by the CBM51 modules from Sp3GH98

Protein	Ligand	*n*	K_a x 10^4 (M^{-1})	ΔG (kcal/mole)	ΔH (kcal/mole)	ΔS (cal/mole/K)
CBM51-1	BGA	0.97 ± 0.01	6.63 ± 0.49	-6.38 ± 0.04	-14.77 ± 0.24	-27.5 ± 0.92
	BGB	1.11 ± 0.01	7.76 ± 0.21	-6.47 ± 0.01	-12.03 ± 0.34	-18.0 ± 1.20
CBM51-2	BGA	NB	NB	NB	NB	NB
	BGB	NB	NB	NB	NB	NB
CBM51-1.2	BGA	1.01 ± 0.01	5.58 ± 0.30	-6.30 ± 0.03	-14.93 ± 0.05	-28.3 ± 0.28
	BGB	1.18 ± 0.01	6.90 ± 0.48	-6.40 ± 0.04	-12.30 ± 0.16	-19.1 ± 0.55

NB, no binding

its high level (50%) of amino acid sequence identity with CBM51-1. The molecular basis of carbohydrate recognition by CBM51-1 and the lack of binding by CBM51-2 was explored by determination of the crystal structures of these CBMs.

CBM51 structure and flexibility

Crystals of the tandem CBM construct CBM51-1.2 were obtained in two different forms. The first was grown in the absence of a carbohydrate ligand while the second was acquired by co-crystallization with a large excess of the type 2 BGB tetrasaccharide. Both crystal forms contained two molecules of the tandem CBM in the asymmetric unit and each of the molecules could be completely modeled as contiguous polypeptide chains. Both modules, CBM51-1 and CBM51-2, adopt the 9-stranded β-sandwich fold previously described for family 51 carbohydrate binding modules (Figure 16A). The two tandem modules are separated by a roughly 20 amino acid polypeptide linker comprised of primarily random coil with a small α-helix (Figure 16A). In both the native and complexed structures the CBM51-1 module was found to bind a single metal ion, which on the basis of B-factor, bond lengths and coordination geometry, was modeled as calcium (Figure 16A). CBM51-2, however, was found to bind what was modeled as a sodium ion, but this was only present in the complexed structure and is thus likely bound quite weakly. The putative metal binding site of CBM51-2 in the native structure was either empty (molecule A) or the sidechain of D400 had collapsed into the space that could be occupied by a metal atom (molecule B).

A feature immediately evident in the four molecules of the tandem CBMs representing the asymmetric unit contents of the two different crystal forms is the inconsistent relative orientations of the two CBMs present in a single polypeptide. An overlap of the four different molecules using the CBM51-2 module as a reference reveals the CBM51-1 module to be in a different position in each molecule highlighting the flexibility of the linker separating the two modules (Figure 16B). However, the α-helix of the linker maintains its orientation relative to the CBM51-2 module indicating its comparatively static position and localizing the flexible linker region to amino acids 230-235 that separate the two CBM modules. The fortuitous crystallization of these four different

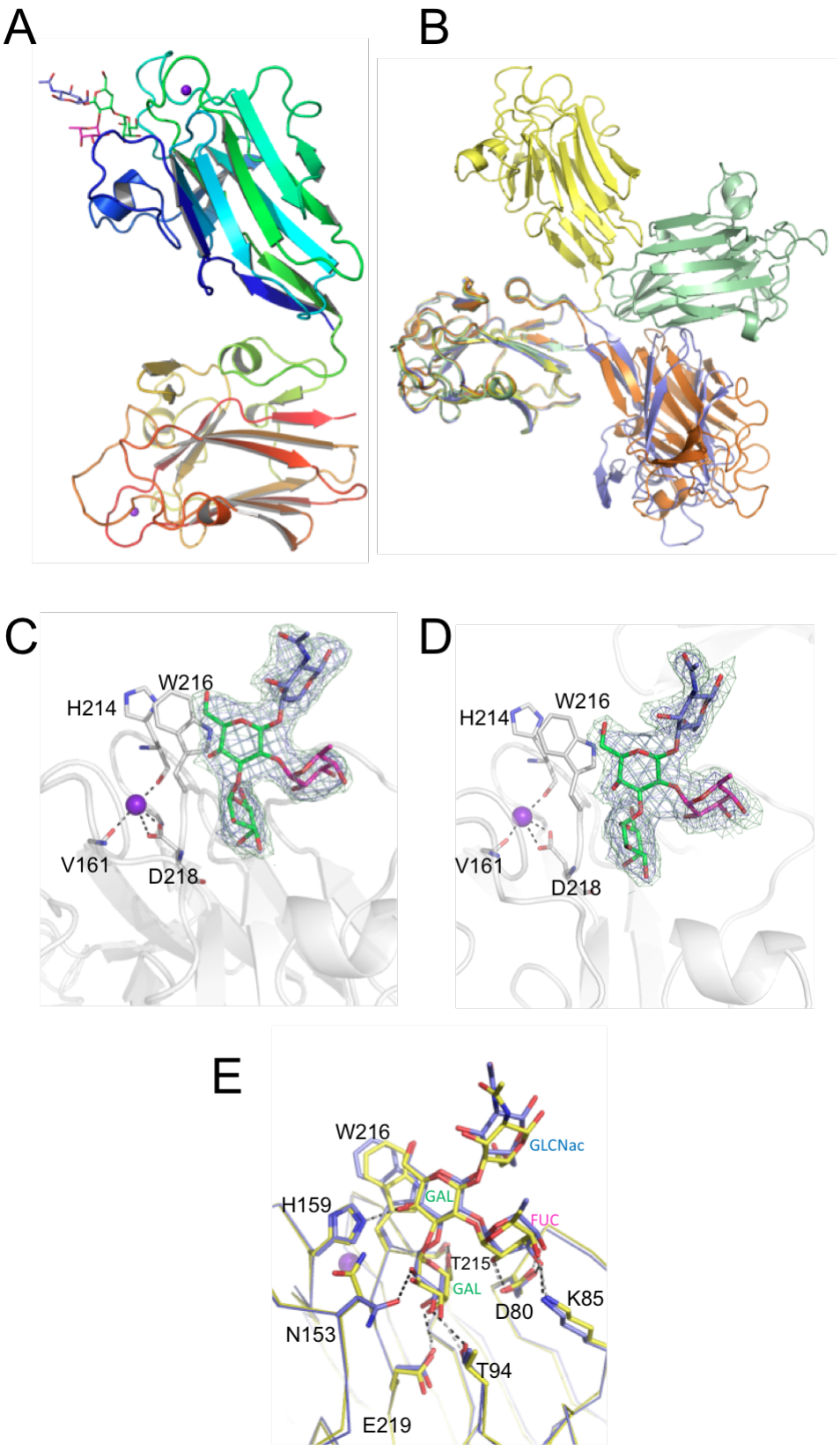

A
B
C
W216
H214
V161
D218
D
W216
H214
V161
D218
E
W216
GLCNac
H159
GAL
FUC
T215
GAL
K85
D80
N153
T94
E219

Figure 16: X-ray crystallographic analysis of the CBM51-1.2 tandem

(On previous page). (A) A cartoon representation of one of two molecules of CBM51-1.2 in the asymmetric unit of the BGB complex. Bound metal ions are shown as purple spheres. (B) An overlap of the four molecules of CBM51-1.2 found in the two asymmetric units of the complexed and uncomplexed forms of CBM51-1.2. The overlap was performed using only the Cα residues of the second CBM module, CBM51-2. The yellow and blue structures are from the complexed form of the protein while the green and orange were from the uncomplexed structure. Panels (C) and (D) show the bound carbohydrate in the binding site of CBM51-1 in the CBM51-1.2 tandem for molecule A (C) and molecule B (D) in the asymmetric unit. The sugar is shown in stick representation with maximum-likelihood (Murshudov, Vagin et al. 1997)/σ_A (Read 1986) weighted $2F_{obs} - F_{calc}$ (green mesh) and $F_{obs} - F_{calc}$ (blue mesh) electron density maps contoured at 0.23 electrons/Å^3 (1.0 σ) and 0.14 electrons/Å^3 (3.0 σ), respectively. The bound metal ions are shown as purple spheres and relevant sidechains as sticks. E) An overlap of the binding sites of CBM51-1 showing specific interactions for molecules A (yellow) and B (blue) in the asymmetric unit. Relevant sidechains are shown as sticks and hydrogen bonds as dashed lines. In panels A, C, and D the bound carbohydrate is shown in stick representation with galactose residues coloured green, N-acetylglucosamine coloured blue and fucose coloured magenta.

conformational snapshots of the CBM51-1.2 tandem hints at the possible structural heterogeneity of Sp3GH98.

The molecular basis of blood-group antigen binding

CBM51-1.2 co-crystallized with the BGB tetrasaccharide was indeed complexed with this sugar with clear electron density in the binding site of both CBM51-1 modules for each CBM tandem in the asymmetric unit (Figure 16C and 16D). This allowed clear modeling of the two sugar units, though the azido-ethyl tail of this synthetic sugar was too disordered to model. Consistent with the observed lack of sugar binding in solution, no electron density that could be interpreted as a carbohydrate was found associated with either CBM51-2 module in the asymmetric unit.

The binding site of CBM51-1 is primarily defined by a tryptophan residue, W216, that forms one side-wall of the binding site. The plane of the tryptophan indole ring lies approximately parallel to the plane of the C4-C5-C6 atoms in the central galactose residue of the carbohydrate ligand. The other side of the W216 indole ring packs against H214 in an interaction that is most likely relevant due to its stabilization of the tryptophan side chain orientation and also because of the role of the H214 backbone carbonyl in coordinating an adjacent calcium atom (Figure 16C and 16D). Thus, though not directly involved in carbohydrate recognition, the coordination of the calcium atom may aid in stabilizing the binding site conformation.

The conformations of the carbohydrate in the binding site of both CBM51-1 modules present in the asymmetric unit were essentially identical (Figure 16E). Only a small variation in the GlcNAc residue was observed, which is possibly due to its lack of direct interaction with the protein. The interaction of the sugar with the binding site is stabilized by a number of hydrogen bonds that involve the two galactose residues and the fucose residue (Figure 16E), providing absolute specificity for the branched carbohydrate antigen.

A notable difference in the two CBM51-1 modules present in the asymmetric unit is the conformation of N153, which only in one arrangement is able to hydrogen bond with the blood-group B determining terminal galactose residue (Figure 16E). Indeed, this particular observation raised questions of how CBM51-1 is able to recognize the blood-

group A antigen, which it clearly does based on the microarray and calorimetry studies. CBM51-1 possesses an extremely similar binding site to CpGH98CBM51, which also binds both blood-group antigens, with all of the residues involved in binding being identical or at least conserved in their properties (Figure 17A and 17E). This comparison reveals, however, that in the trapped conformation of CBM51-1 that the acetamido group of a blood-group A determining terminal GalNAc residue would clash with N153 (Figure 17A), and this would occur in either orientation of this side chain as it is seen in the CBM51-1 modules in the ligand bound structures. Considering the two CBM51-1 modules from the uncomplexed structure provides additional insight into this. This reveals that N153 in one of the uncomplexed CBM51-1 modules to match the conformation of this residue in one of the carbohydrate-bound forms of CBM51-1 (Figure 17B). In the other unbound form N153 is significantly displaced by a margin large enough to allow accommodation of the GalNAc acetamido group (light blue structure in Figure 17B). Thus, the ability of CBM51-1 to bind both blood group A and B antigens is likely facilitated by the mobility of both the N153 sidechain and slight displacement of the backbone of residues 152-154.

CBM51-1, CBM51-2 and CpGH98CBM51 display excellent conservation of residues in the binding site (Figure 5c and 5e). Structurally, the residues involved in hydrogen bonding are nearly absolutely conserved between CBM51-1 and CBM51-2 (Figure 17C). The major difference between the two CBMs leading to the inactivation of CBM51-2 becomes evident when examining the position of the binding site tryptophan. While W216 of CBM51-1 is well positioned to interact with the conserved galactose residue of the blood-group A and B antigens, W398 of CBM51-2 is found to be somewhat disordered in the crystal structure and present in different conformations in the individual monomers modeled in the asymmetric unit (Figure 17C and 17D). There is no evidence of W398 being appropriately positioned to interact with a carbohydrate ligand (Figure 17C and 17D). This difference in CBM51-2 is a result of a series of somewhat subtle structural changes that combine to result in functionally relevant rearrangements to the architecture of what would be the CBM51-2 binding site.

The first structural change leading to the misplacement of W398 is the position of H396. In CBM51-1 the analogous histidine, H214, whose placement is likely stabilized

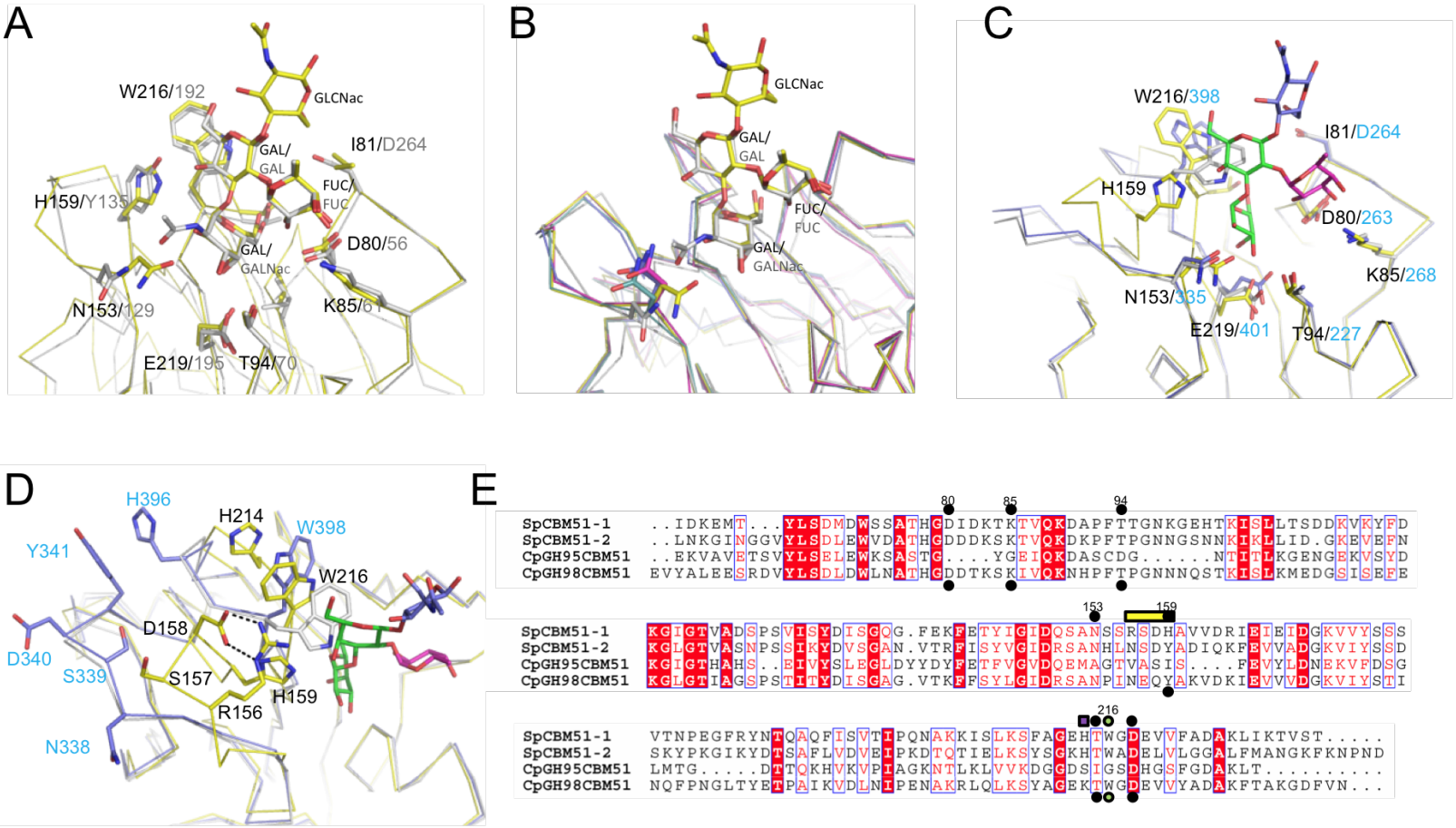

A
W216/192
GLCNac
GAL/
GAL
I81/D264
FUC/
FUC
H159/Y135
GAL/
GALNac
D80/56
N153/129
K85/61
E219/195
T94/70
B
GLCNac
GAL/
GAL
FUC/
FUC
GAL/
GALNac
C
W216/398
I81/D264
H159
D80/263
K85/268
N153/335
E219/401
T94/227
D
H396
H214
W398
Y341
W216
D158
D340
S339
S157
H159
R156
N338
E
80
85
94
SpCBM51-1 ..IDKEMT...YLSDMDWSSATHGDIDKTKTVQKDAPFTTGNKGEHTKISLLTSDDKVKYFD
SpCBM51-2 ..LNKGINGGVYLSDLEWVDATHGDDDKSKTVQKDKPFTPGNNGSNNKIKLLID.GKEVEFN
CpGH95CBM51 ..EKVAVETSVYLSELEWKSASTG....YGEIQKDASCDG.....NTITLKGENGEKVSYD
CpGH98CBM51 EVYALEESRDVYLSDLDWLNATHGDDTKSKIVQKNHPFTPGNNNQSTKISLKMEDGSISEFE
153
159
SpCBM51-1 KGIGTVADSPSVISYDISGQG.FEKFETYIGIDQSANSSRSDHAVVDRIEIEIDGKVVYSSS
SpCBM51-2 KGLGTVASNPSSIKYDVSGAN.VTRFISYVGIDRSANHLNSDYADIQKFEVVADGKVIYSSD
CpGH95CBM51 KGIGTHAHS..EIVYSLEGLDYYDYFETFVGVDQEMAGTVASIS....FEVYLDNEKVFDSG
CpGH98CBM51 KGLGTIAGSPSTITYDISGAG.VTKFFSYLGIDRSANPINEQYAKVDKIEVVVDGKVIYSTI
216
SpCBM51-1 VTNPEGFRYNTQAQFISVTIPQNAKKISLKSFAGEHTWGDEVVFADAKLIKTVST.....
SpCBM51-2 SKYPKGIKYDTSAFLVDVEIPKDTQTIELKSYSGKHTWADELVLGGALFMANGKFKNPND
CpGH95CBM51 LMTG.....DTTQKHVKVPIAGKNTLKLVVKDGGDSIGSDHGSFGDAKLT.........
CpGH98CBM51 NQFPNGLTYETPAIKVDLNIPENAKRLQLKSYAGEKTWGDEVVYADAKFTAKGDFVN...

Figure 17: Analysis of the CBM51 binding site

A) An overlap of the CBM51-1 binding site molecule B (yellow) with the binding site of CpGH98CBM51 (grey). The sugar and relevant sidechains involved in coordinating the carbohydrate are shown as sticks an reveal the conservation of the binding site. B) An overlap of the CBM51-1 binding site from all four molecules in the complexed and uncomplexed crystal forms with the binding site of CpGH98CBM51 (grey). Molecules A and B of the complexed crystal form are coloured blue and yellow, respectively, while molecules A and B of the uncomplexed form are colored magenta and light blue. N153 is shown in stick representation to illustrate its different position in the four different CBM51-1 modules found in the two crystal forms of CBM51-1.2. C) An overlap of the CBM51-1 binding site (yellow) from molecule B of the complexed form with the same region of CBM51-2 from molecules A (blue) and B (grey), also from the complexed form of the polypeptide. Conserved residues involved in hydrogen bonding in CBM51-1 are shown in stick representation, as is the tryptophan residue in the binding site that in CBM51-1 interacts with the sugar. D) An overlap as in panel C that illustrates changes in loop structure. Relevant residues are shown in stick representation. E) A comparison of the primary structure of known CBM51s. The pneumococcal CBM51s are preceded by Sp for clarity. Symbols above the alignment indicate residues involved in carbohydrate binding in CBM51-1 while those below are relevant to CpGH98CBM51. Black circles indicate residues involved in hydrogen bonding while green circles indicate aromatic sidechains that stack against the central galactose unit of the blood-group antigens. The purple square shows the histidine 214 and 396 in CBM51-1 and CBM51-2, respectively. The yellow rectangle indicates the loop of resides 338-341 that is displaced in CBM51-2. In all panels labels are coloured black for CBM51-1, blue for CBM51-2, and grey for CpGH98CBM51. In panels C) and D) galactose residues are coloured green, N-acetylglucosamine in blue, and fucose in magenta.

by the participation of its carbonyl group in calcium coordination, packs against the back of W216 maintaining the position of this sidechain. In contrast, the equivalent residue in CBM51-2, H396, and the loop housing it are displaced several angstroms and thus play no role in stabilizing the conformation of the tryptophan sidechain. This appears to result in the loss of the calcium binding site in CBM51-2. At first glance the amino acid sequence surrounding the key tryptophan residue appears to be well conserved between the two CBMs; however, it appears that a contributing factor to the altered conformation in CBM51-2 is the presence of an alanine (A399) residue immediately following W398 (Figure 17E). In CBM51-1 and CpGH98CBM51 this residue is a glycine (Figure 17E). These glycine residues adopt phi and psi angles that are not favored by alanine suggesting that the alanine substitution in CBM51-2 constrains the conformation of the preceding polypeptide chain resulting in the slightly displaced loop structure observed in CBM51-2.

The position of H396 in CBM51-2 is itself stabilized by its packing against the sidechain of Y341, which is so placed due to displacement of the loop containing residues 338-341 relative to the analogous loop in CBM51-1 (residues 156-159). In CBM51-1 this loop is pulled adjacent to the binding site and held in this conformation by hydrogen bonding between the sidechains of D158 and R156. This allows H159 to protrude in to the binding site and form a hydrogen bond with the carbohydrate. No such interaction occurs with this loop in CBM51-2, rather it adopts the conformation where Y341 interacts with H396. To complete the structural rearrangement of CBM51-2, the space vacated by the loop comprising residues 338-341 becomes occupied by the rearranged loop containing H396 and the sidechain of T397.

It is notable that, with the exception of slight variations in sidechain placement, the "binding" site of CBM51-2 is consistent for all of the CBM51-2 modules observed in the two molecules present in the asymmetric units of the complexed and uncomplexed forms (i.e. four CBM51-2 modules). This indicates that the conformation discussed is likely not an artifact of crystallization and represents the true conformation of the apparently inactive module. What is quite remarkable about the structural differences between CBM51-1 and CBM51-2 is that they could not be predicted on the basis of primary structures. The relevant regions of the modules are sufficiently similar in amino acid sequence to predict identical conformations and thus functions. This highlights the

subtlety of structural changes required to make quite dramatic alterations to the function of proteins.

A question that remains is the purpose for Sp3GH98 retaining a CBM51-2 module that is apparently rendered unable to bind carbohydrate. An option that cannot be unequivocally ruled out is that CBM51-2 can bind an as yet unidentified ligand. Considering that the glycan microarray contains a wide variety of commonly occurring mammalian glycans it seems improbable that CBM51-2 can bind an alternate mammalian carbohydrate; however, its ligand may be another macromolecule. Alternatively, CBM51-2 may simply act as a linker module separating the catalytic module and the functional CBM51-1. Indeed, this has been postulated in the past for CBM-like polypeptides (Boraston, Creagh et al. 2001; Notenboom, Boraston et al. 2001). The implication of CBM51-2 lacking carbohydrate-binding activity is that this tandem is very unlikely to have an increased affinity for clustered glycans through an avidity effect. In contrast, the three family 47 CBMs from Sp4GH98 rely on the multivalent binding imparted by the triplet for a high affinity interaction (Boraston, Wang et al. 2006).

It is now evident that Sp3GH98 and Sp4GH98 are deployed as secreted proteins. The catalytic modules of these enzymes impart the ability to degrade carbohydrate antigens presented on cell surfaces (Higgins, Whitworth et al. 2009) while the CBMs non-catalytically adhere the enzyme to the appropriate substrates for the enzymes. In this light, it is possible that the overall spatial organization of the modules influence how the adherent and catalytic functions of the proteins are coordinated at the surface of the host cell, as has been suggested for the *exo*-β-N-acetylglucosaminidase GH84C from *C. perfringens* (Ficko-Blean, Gregg et al. 2009). However, despite extensive attempts to address this question by X-ray crystallography, full-length Sp3GH98 and Sp4GH98 would not crystallize. To approach the structural analysis of full-length Sp3GH98 and Sp4GH98 small-angle X-ray scattering (SAXS) was chosen as a solution method that could, in combination with the available high-resolution X-ray crystal structures, yield insight into the overall solution conformation(s) of both of these proteins.

Table 10: Small angle X-ray scattering parameters

Enzyme	Molecular weight kDa	Number of amino acids	R_g (Guinier) Å	R_g (Gnom) Å	D_{max} Å	χ (Dammif)	Normalized spatial discrepancy
Sp3GH98	106.6	963					
3.45mg.ml-1	113.8		40.4±0.1	41.7±0.1	140	1.82±0.11	0.70±0.04
1.73mg.ml-1	108.8		39.4±0.1	40.7±0.1	130	1.62±0.03	0.66±0.02
0.86mg.ml-1	105.5		39.8±0.2	40.4±0.1	125	1.37±0.01	0.63±0.02
0.43mg.ml-1	103.7		39.0±0.3	40.1±0.1	125	1.49±0.01	0.75±0.03
Sp4GH98	114.4	1030					
4.89mg.ml-1	114.7		42.5±0.1	44.1±0.1	150	1.93±0.19	0.79±0.05
2.44mg.ml-1	114.9		43.8±0.1	43.7±0.1	140	1.91±0.03	0.70±0.03
1.22mg.ml-1	113.3		43.2±0.2	43.5±0.1	140	1.64±0.01	0.70±0.02
0.61mg.ml-1	114.7		45.2±0.4	42.7±0.1	130	1.79±0.01	0.72±0.01
BSA	66.4	607					
1.7mg.ml-1	65.0		29.3±0.1	29.1±0.1	90	N/A	N/A

The overall architecture of pneumococcal GH98 enzymes

Four different concentrations of both Sp3GH98 and Sp4GH98 were studied by SAXS analysis (see Table 8). In all cases the agreement between the concentrations for each protein was very good and the determined molecular weights were in good agreement with the expected molecular weights. The R_g of Sp3GH98 was ~40 Å and its D_{max} ~130 Å; the corresponding values for Sp4GH98 were a R_g of ~ ~43.5 Å and a D_{max} of ~140 Å, both values reflecting the larger size of this protein. Twenty *ab initio* models were generated with DAMMIF for each protein at all four concentrations. This yielded averaged χ-values (average of 20) in the range of 1.4-1.8 for Sp3GH98 and 1.6-1.9 for Sp4GH98. The normalized spatial discrepancies of the 20 models determined at each concentration were between 0.6 and 0.8, suggesting relatively uniform solution conformations for each protein.

The averaged *ab initio* form of Sp3GH98 shows an elbow-shape with a measured largest dimension of ~131 Å (Figure 18A), which was in keeping with the D_{max} determined from the scattering. The thin arm of the elbow presumably represents the CBMs while the thick arm is the catalytic module. To probe this hypothesis, rigid-body modeling with the program BUNCH was used to generate five different models. In all cases the catalytic module was used as an independent rigid body. In four of the rigid body modeling runs the four individual molecules of the intact CBM51-1.2 tandems observed in the crystal structures (2 molecules from each of the asymmetric units of the 2 different crystal forms) were used as rigid bodies. In the fifth model the two CBM51 modules were treated as individual rigid bodies. All five models had the general shape of the *ab initio* form obtained with DAMMIF but yielded χ_{crysol} values that ranged from quite good at 1.43 (model 5) to relatively poor at 3.12 (model 4; Figure 18B-F). The two models that displayed the lowest χ_{crysol} values, and thus best represented the scattering data, were model 5 generated using three independent modules (χ_{crysol} = 1.43) and model 1 generated with a CBM tandem from the crystal structure (χ_{crysol} = 1.55). Notably, these had reasonably similar relative arrangements of the CBMs and catalytic modules suggesting that of the five models these two are the most likely possible representations of the solution structure of Sp3GH98.

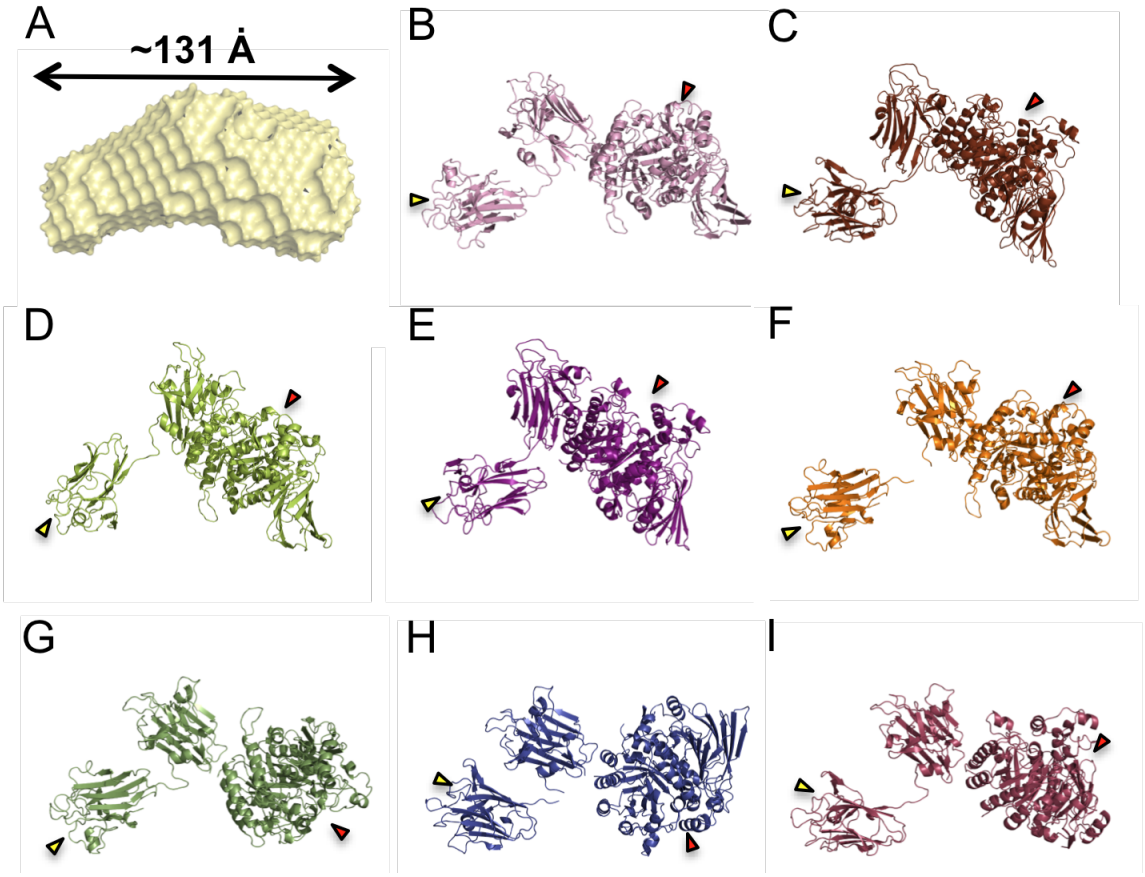


Figure 18: Small-angle X-ray scattering analysis of Sp3GH98

Small-angle X-ray scattering analysis of Sp3GH98 (at 0.4 mg/ml). A) *ab initio* model of Sp3GH98 created from the average of 20 models generated by DAMMIF. Panels A-F show five rigid body models generated using the program BUNCH. The details are as follows. B) Model 1 generated using the catalytic module (CM, the $(\alpha/\beta)_8$ domain; PDB code 2WMI) as a rigid body and the CBM51-1.2 tandem, molecule A from the complexed structure, as a second rigid body (χ_{crysol} = 1.55). C) Model 2 generated using the CM as a rigid body and the CBM51-1.2 tandem, molecule B from the complexed structure, as a second rigid body (χ_{crysol} = 2.32). D) Model 3 generated using the CM as a rigid body and the CBM51-1.2 tandem, molecule A from the uncomplexed structure, as a second rigid body (χ_{crysol} = 2.10). E) Model 4 generated using the CM as a rigid body and the CBM51-1.2 tandem, molecule B from the uncomplexed structure, as a second rigid body (χ_{crysol} = 3.11). F) Model 5 generated using the CM as a rigid body and the individual CBM51-1 and CBM51-2 modules as individual rigid bodies (χ_{crysol} = 1.43).

Panels G-I show individual models 6 (G), 7 (H), and 8 (I), selected as representatives of three separate EOM runs (see text). The χ_{crysol} values of these models are 1.37, 1.50, and 1.39, respectively. All of the models are shown in an orientation that is consistent with their fit into the shape shown in panel A. The red triangles indicate the position of the catalytic site while the yellow arrows indicate the binding site in CBM51-1.

The crystal structures of the CBM51 tandem suggested flexibility in the linker between the two CBMs, the potential for which was only partly accounted for by using the different CBM51 tandems in our application of rigid-body modeling. To further assess the possible contribution of inter-domain flexibility to the conformational heterogeneity of Sp3GH98 we used the ensemble optimization method (EOM). This was initially done using the 3 modules (2 CBMs and the catalytic module) as independent domains. The resulting ensemble of 20 models yielded an overall χ_{EOM} of 1.31 with the individual models having χ_{crysol} values ranging from 1.53-5.36 (not shown). The resulting ensemble was somewhat complicated and included what appeared to be a large number of poor individual models. Within the ensemble, however, was a useful pattern: the CBM51 adjacent to the catalytic module, CBM51-2, was found in three distinct clusters. An average position for CBM51-2 from each of the three clusters was selected, fixed in its position relative to the catalytic module and EOM re-performed treating only the linker between CBM51-1 and CBM51-2 as flexible (i.e. the catalytic module and CBM51-2 were treated as a single rigid body). The three separate ensembles of 20 models yielded χ_{EOM} values of 1.44 (χ_{crysol} range 1.34-1.69), 1.42 (χ_{crysol} range 1.36-1.89), and 1.44 (χ_{crysol} range 1.37-1.89). The overall χ_{EOM} values were thus not substantially different from EOM performed with three independent modules; however, individual models with poor χ_{crysol} values were not obtained by this method. Single models selected as approximate average representatives of each of the final three ensembles fit the *ab initio* shape very well and had a good χ_{crysol} values of 1.37 (model 6), 1.50 (model 7), and 1.39 (model 8) (Figure 18G-I). A comparison of these models reveals substantially different relative orientations of the CBMs and catalytic module (Figure 19). It is possible that these models and the ensembles that they represent indicate quite large flexibility in Sp3GH98 with the average solution conformation having the general "elbow-shaped" seen in the *ab initio* reconstruction. It is notable, however, that only one of representative models, model 7, matched the relative arrangement of the modules in the two preferred models, models 1 and 5, that were obtained by the rigid-body modeling with BUNCH (Figure 19). The convergence upon the one general conformation given by the two approaches to SAXS analysis suggests this is the most probable representation of Sp3GH98 in solution. This, and the observation that single models agree very well with

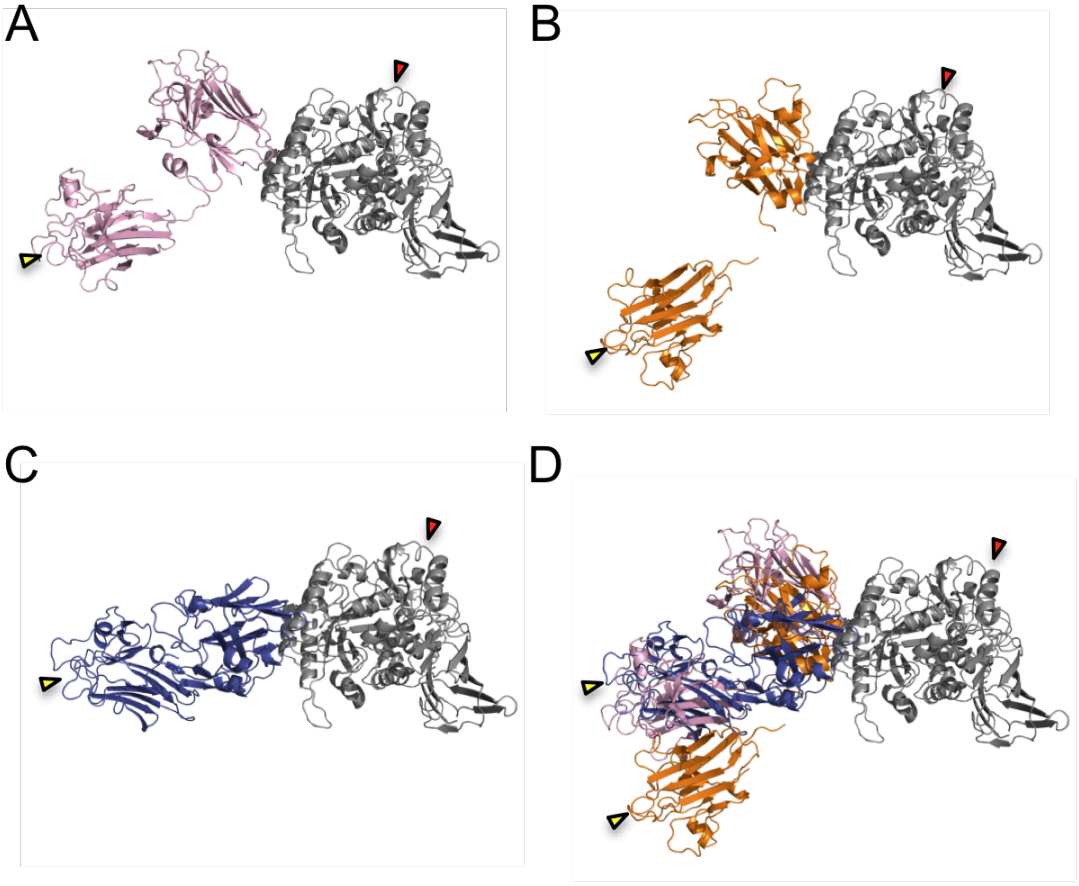


Figure 19: Comparison of Sp3GH98 SAXS models

Panels A – C show models 1, 5, and 7, respectively, from a consistent angle relative to the catalytic module, which is shown in grey. Panel D shows the composite of these panels (i.e. the three models overlapped by their catalytic modules). The red triangles indicate the position of the catalytic site while the yellow arrows indicate the binding site in CBM51-1.

the scattering data (χ_{crysol} values approaching 1), in turn suggests that the flexibility of intact Sp3GH98 is not profound and a single conformation likely predominates in solution. Despite the ability of the CBM tandems to adopt quite different conformations in the crystal structures it appears that not all of these conformations represent conformations that are preferred in solution or, alternatively, are not preferred in the context of the complete enzyme where the proximity of the catalytic module may influence the conformation of the neighbouring CBMs. This being said, we cannot completely discount the possibility that conformations similar to models 6 and 8 also represent alternative solution conformations.

The averaged *ab initio* form of Sp4GH98 had an overall largest dimension of ~140 Å, which was in keeping with the D_{max} measured from the scattering, and displayed an asymmetric bi-lobed shape (Figure 20A). The X-ray crystal structures of the Sp4GH98 catalytic module, CBM47-1 and CBM47-3 are available, which allowed these to be manually fit into the averaged DAMMIF form using the location of the N- and C-termini as restraints; no structure of CBM47-2 is available so CBM47-3, which has greater than 50% amino acid sequence identity to CBM47-2, was used as a model for CBM47-2 (Figure 20B). The resulting model yielded a χ_{crysol} value of 1.64. An alternative approach used BUNCH to rigid body model Sp4GH98 using the coordinates of the four modules as individually refined rigid bodies (Figure 20C). This gave a model that was remarkably similar to the manually derived model and had a χ_{crysol} value of 1.57. Both models possess a similarly positioned trio of CBMs in which the binding sites of the CBMs all face outwards from a common centre. Given the presence of linkers between the CBMs there is ambiguity in the relative positions of these modules. Indeed, if the general CBM positions relative to the catalytic module are considered 1, 2, and 3 (Figure 20D), then, given the constraint of linker length, the possible occupation of these positions by CBM47-1, CBM47-2, and CBM47-3, respectively, could be 1-2-3, 3-2-1, 1-3-2, or 3-1-2 (i.e. CBM47-1 could occupy position 1 or 3, CBM47-2 could occupy position 1,2 or 3 and CBM47-3 could occupy position 1,2, or 3). We cannot discriminate between these possibilities at this time. Regardless, the impact of the order is likely to be negligible as

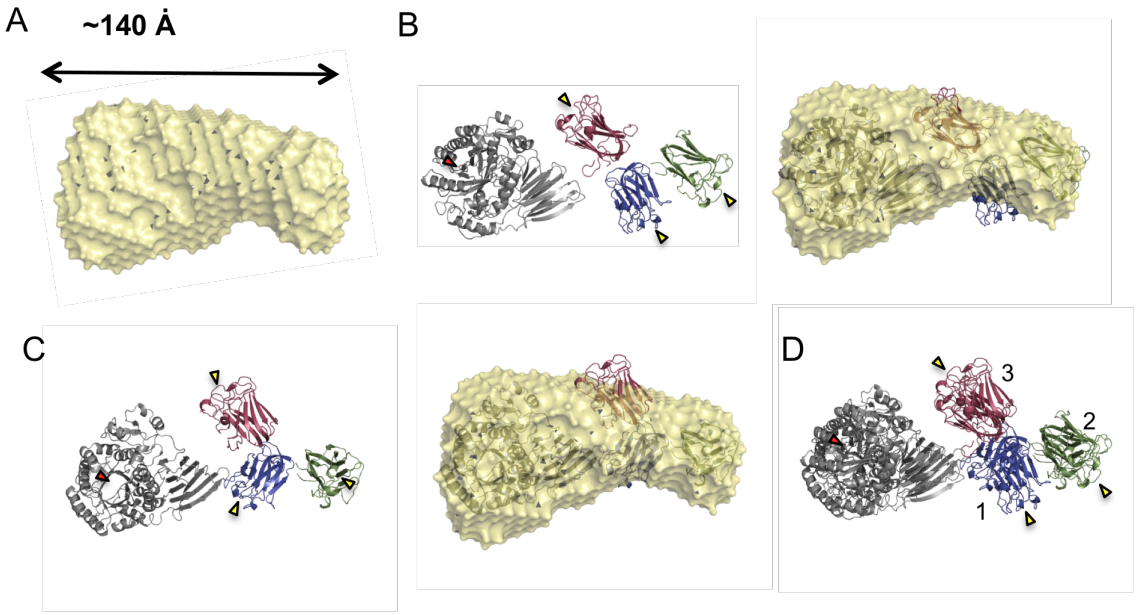


Figure 20: Small-angle X-ray scattering analysis of Sp4GH98

Small-angle X-ray scattering analysis of Sp4GH98 (at 0.6 mg/ml). A) *ab initio* model of Sp4GH98 created from the average of 20 models generated by DAMMIF. B) Left, model 1 generated by manually fitting the separate catalytic module (CM, the $(\alpha/\beta)_8$ domain; PDB code 2WMF), CBM47-1 (blue; PDB code 2J1R), CBM47-2 (purple, modeled using CBM47-3), and CBM47-3 (green; PDB code 2J22) coordinates into the shape shown in panel A (χ_{crysol} = 1.64). The overlap of the model with the form is shown at right. B) Left, model 2 generated by rigid body modeling with BUNCH using the CM, CBM47-1, CBM47-2, and CBM47-3 as a individual rigid bodies (χ_{crysol} = 1.57). The overlap of the model with the form is shown at right. D) An overlap of models 1 and 2 performed using SUPCOMB shows the similar placement of the modules in the two models. The red triangles indicate the position of the catalytic site while the yellow arrows indicate the binding sites in the CBM47 modules.

the individual CBMs all have high structural identity and are known to possess the same carbohydrate binding function (Boraston, Wang et al. 2006).

The locations in the primary structure of the CBMs relative to the catalytic modules in both Sp3GH98 and Sp4GH98 suggest different overall three-dimensional organizations. Indeed, the SAXS models of the full-length proteins confirm this and show quite different dispositions of the composite modules with the CBMs of Sp3GH98 present on the opposite side of the catalytic module in comparison to the CBMs of Sp4GH98 (Fig 21). A general similarity between Sp3GH98 and Sp4GH98, however, is the consistent orientation of the CBM binding sites away from the catalytic site. This suggests that CBM recognition of a carbohydrate receptor is not directly coupled to the action of the catalytic module on a carbohydrate, even though the CBMs possess identical specificities to the catalytic modules. The GH98 proteins are deployed as soluble proteins and thus the CBM cannot function to adhere the entire bacterium to carbohydrates. The CBMs therefore likely function to target the enzymes to regions of a cell surface that are dense in the specific receptor, either BGA, BGB or Lewis^{Y} antigen, thus concentrating the enzyme there allowing the catalytic module to graze up to ~140 Å away from the receptor site. This is very much in keeping with the generally accepted role of CBMs in glycoside hydrolases that are involved in the depolymerization of plant cell-wall polysaccharides (Boraston, Bolam et al. 2004). Also, as is frequently observed with plant cell wall specific enzymes, the CBMs of the GH98 enzymes have a specificity that is identical to that of the catalytic module, allowing the CBMs to target the enzyme to the appropriate substrate.

Pneumococcal carbohydrate-active enzymes are typically either covalently or non-covalently attached to the bacterial cell surface. Some of the best-characterized examples of such pneumococcal carbohydrate-active enzymes that are also considered as virulence factors are the N-glycan processing quartet of NanA, BgaA, StrH and EndoD and the glycogen degrading SpuA (Clarke, Platt et al. 1995; King, Hippe et al. 2006; Burnaugh, Frantz et al. 2008; Abbott, Macauley et al. 2009; Abbott, Higgins et al. 2010; Lammerts van Bueren, Ficko-Blean et al. 2011). Like most pneumococcal carbohydrate-active enzymes, including the GH98 enzymes, these five proteins are large and multimodular. However, due to their attachment to the cell wall, their activity on host glycans is limited

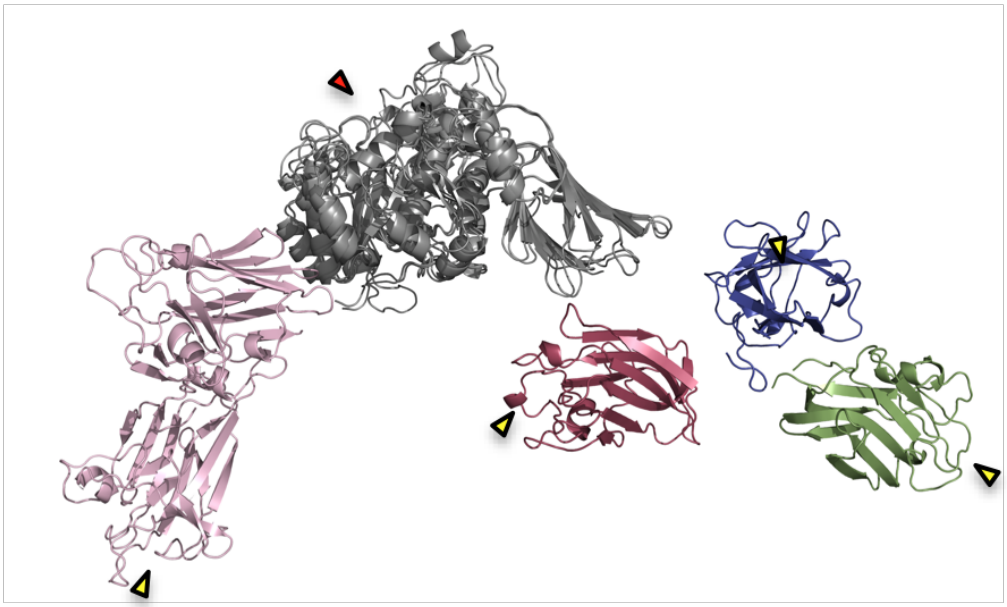


Figure 21: Comparison of full-length models of Sp3GH98 and Sp4GH98

Model 1 of Sp3GH98 and Model 2 of Sp4GH98 were overlapped via the catalytic domains (shown in grey). The CBM51 modules of Sp3GH98 are shown in pink while the CBM47 modules of Sp4GH98 are shown in purple, blue and green. The catalytic site is indicated by a red arrow while the binding sites of the CBMs are indicated by yellow triangles.

by the proximity of the bacterium to glycan bearing molecules. Based on the general dimensions of SpuA determined from a complete X-ray crystal structure we have also postulated that the presence a capsule around the bacterium, which is typically more than 1000 Å deep, would likely limit access of substrates to the cell-wall associated carbohydrate-active enzymes. This suggests that the capsular status of the bacterium influences when the activity of enzymes such as NanA, BgaA, StrH, EndoD and SpuA might exert their effect. In contrast, as secreted soluble extracellular proteins the activities of the GH98 enzymes are neither likely to be dependent on the capsular status of the bacterium nor necessarily on the proximity of the bacterium to host cells. Moreover, enzymes such as EndoD and SpuA, which have putative and proven CBMs, respectively, may also function as adhesins whereas the CBMs found in the GH98 enzymes are unlikely to perform this function but instead target the soluble enzymes to regions in the host that are dense in the target glycans. Thus, the GH98 enzymes differ substantially from the developing paradigm of pneumococcal processing enzymes.

Chapter 4: The fucose mutarotase from *S. pneumoniae*

Melanie A. Higgins[1] and Alisdair B. Boraston[1]

Adapted from Structure of the fucose mutarotase from *S. pneumoniae* in complex with L-fucose. Acta Crystallogr Sect F Struct Biol Cryst Commun. (2011) 67(Pt 12):1524-30.

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

Contributions: I performed the cloning, protein production and purification, crystallography, data collection and structure refinement.

4.1 Abstract

Streptococcus pneumoniae relies on a variety of carbohydrate utilization pathways for both colonization of its human host and full virulence during development of invasive disease states. One such pathway is the fucose utilization pathway, a component of which is the fucose mutarotase (SpFcsU), an enzyme that performs the interconversion between α- and β-L-fucose. This protein was crystallized and the 3-dimensional structure solved in complex with L-fucose. The structure shows a complex decameric quaternary organization with a high overall degree of structural identity with *Escherichia coli* FcsU (EcFcsU). Furthermore, the active site architecture of SpFcsU is nearly identical to that of EcFcsU. When considered in the context of the fucose utilization pathway found in *S. pneumoniae*, SpFcsU appears to link two halves of the pathway by enhancing the rate of conversion from the product of the final glycoside hydrolysis step, β-fucose, into the substrate for the fucose isomerase, α-fucose.

4.2 Introduction

Carbohydrate metabolism is a fundamental mode of energy acquisition and generation across all of the taxonomic kingdoms. The most well-established and central carbohydrate utilization pathway is glycolysis, which converts glucose into pyruvate through a number of intermediates. Other hexose sugars as well as pentose sugars have specific metabolic pathways that produce intermediates that feed into glycolysis (Figure 5). One such example is demonstrated by the utilization of fucose in *E. coli* whereby free L-fucose is converted into lactaldehyde and dihydroxyacetone phosphate (DHAP), the latter of which is a glycolytic intermediate (Chen, Zhu et al. 1987). The first biochemical modification in the pathway requires fucose isomerase (FcsI) to convert the α-anomer of fucose into its ketose form, fuculose. However, upstream of FcsI is a fucose mutarotase (FcsU), which is present to ensure that as the α-fucose is depleted by FcsI the comparatively slow rate of natural mutarotation between the α- and β-fucose anomers is enzymatically enhanced thus maintaining the pool of α-fucose substrate for FcsI (Seemann and Schulz 1997; Lee, Ryu et al. 2009). Following FcsI, the fuculose kinase (FcsK) then uses ATP to phosphorylate fuculose creating a fuculose 1-phosphate substrate (Heath and Ghalambor 1962) for fuculose 1-phosphate aldolase (FcsA), which performs the final cleavage to produce lactaldehyde and dihydroxyacetone phosphate (Ghalambor and Heath 1962).

S. pneumoniae also encodes for a putative fucose utilization operon; however it appears more complex than the *E. coli* operon and along with the fucose processing enzymes (FcsU, FcsI, FcsK, and FcsA), consists of two putative glycoside hyrdolases (GH98 and GH95) and a putative PTS (phosphoenolpyruvate: carbohydrate phosphotransferase system) carbohydrate transporter system (EIIA, EIIB, EIIC, and EIID). This fucose utilization pathway allows *S. pneumoniae* to harvest the Lewis^{Y} blood group antigen from host cells (Higgins, Whitworth et al. 2009) and a strain lacking the operon renders the bacterium nearly avirulent in a mouse model of pneumonia (Embry, Hinojosa et al. 2007), showing the importance of this pathway in pathogenesis. Surprisingly, *S. pneumoniae* is unable to use fucose as a sole carbon source (Chan, O'Dwyer et al. 2003) presenting questions about the function of this operon. With the exception of GH98,

details regarding the individual functions of enzymes comprising the *S. pneumoniae* fucose utilization pathway are presently lacking.

The predicted *S. pneumoniae* fucose mutarotase (SpFcsU) has an amino acid identity of 43% with *E. coli* FcsU (EcFcsU) and thus is thought to have activity interconverting the α- and β forms of L-fucose. Here, we report the 3-dimensional crystal structure of the fucose mutarotase from *S. pneumoniae* TIGR4 in complex with L-fucose and discuss a possible role with regards to the blood group antigen processing by *S. pneumoniae*.

5.2 Materials and Methods

Expression and Purification

The *fcsU* gene (SP_2165) was amplified from *S. pneumoniae* TIGR4 genomic DNA (American Type culture Collection BAA-334D) using primers FcsU-Fwd (GGCAGCCATAGTAAAACATATACCGAAA) and FcsU-Rev (GTGGTGCTCGAGTTATTGAACATTTTCTCTTTC) with engineered *NdeI* and *XhoI* restriction sites, respectively. This DNA fragment was subsequently cloned into pET 28a using these specific restriction sites to yield pFcsU; the sequence fidelity of which was confirmed by bidirectional DNA sequencing. The construct encoded the desired FcsU protein with a thrombin cleavable N-terminal polyhistidine tag.

Protein production and purification was carried out using previously described procedures (Higgins, Whitworth et al. 2009). Briefly, *E. coli* BL21 star (DE3) cells (Novagen) were transformed with the pFcsU plasmid. Cultures of the transformed *E. coli* harboring pFcsU were initially grown at 37°C in LB broth supplemented with 50 μg ml^{-1} kanamycin until they reached an absorbance at 600 nm of about 0.6. Gene expression was then induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside followed by incubation over night at 25°C with shaking. After centrifugation, the cells were ruptured via chemical lysis (Robb, Nano et al. 2010). Initial purification of the target protein was carried out by Ni^{2+} immobilized metal affinity chromatography (Amersham) followed by thrombin cleavage to remove the polyhistidine tag. After a final size exclusion chromatography step using a Sephacryl S-200 column (GE Healthcare), the purified protein was concentrated for crystallization screens.

Protein concentration was determined by measuring the absorbance at 280 nm (A_{280}) and using a calculated molar extinction coefficient of 0.017545 cm^{-1} μM^{-1} (Wilkins, Gasteiger et al. 1999).

Crystallization.

Crystallization was carried out using hanging drop vapour diffusion at 18°C and a protein concentration of 20 mg ml^{-1}. Optimization of initial crystallization conditions yielded the best crystals in 20% PEG 3350, 0.1 M HEPES pH 8.0, and 0.2 KCl. Crystals were soaked with crystallization solution supplemented with a molar excess of L-fucose to obtain a structure in complex with this sugar.

Data Collection, Structure Determination and Refinement.

Crystals of SpFcsU proved difficult to cryoprotect thus a protocol was used whereby powdered fucose was added directly to the crystallization drop, the drops were exposed to air until evaporation was sufficient to result in cryoprotecting conditions at which point the crystals were mounted straight onto the detector in a N_2 cryo-stream at 113 K. Diffraction data was collected with a Rigaku R-AXIS 4++ area detector coupled to a MM-002 X-ray generator with Osmic 'blue' optics and an Oxford Cryostream 700.

Using a pentamer of *Bacillus subtilis* RbsD (PDB1OGC) as a search model a molecular replacement solution comprising four of these pentamers was found with the program PHASER (Kim, Shin et al. 2003; McCoy, Grosse-Kunstleve et al. 2005). Manual correction of the twenty monomers in the asymmetric unit was performed using COOT (Emsley and Cowtan 2004) followed by refinement using REFMAC (Murshudov, Vagin et al. 1997). Finally, twenty β-L-fucose molecules were manually placed in each active site and water molecules found using COOT:FINDWATERS and manually inspected after refinement. Refinement procedures were monitored by flagging 5% of all observation as 'free' (Brunger 1992). Model validation was performed with RAMPAGE (Lovell, Davis et al. 2003) and MOLPROBITY (Chen, Arendall et al. 2010). Data processing and refinement statistics are given in Table 11. Coordinates and structure factors have been deposited in the Protein Data Bank with the PDB code of 4a34.

Table 11: Data collection and refinement statistics for SpFcsU

Data collection	
Beamline	Home source
Wavelength	
Space group	P21
Cell dimensions	
a, *b*, *c* (Å)	64.3, 144.1, 165.4
Resolution (Å)	20.00 – 2.50 (2.59-2.50)
R_{sym} or R_{merge}	0.087 (0.400)
$I / \sigma I$	8.4 (2.4)
Completeness (%)	99.7 (98.9)
Redundancy	4.0 (3.8)
Refinement	
Resolution (Å)	2.5
No. reflections	97486
R_{work} / R_{free}	0.225/0.292
No. atoms	
Protein	2801
Ligand	20
Water	849
B-factors (Å^2)	
Protein	31.6
Ligand	40.3
Water	31.7
R.m.s. deviations	
Bond lengths (Å)	0.0084
Bond angles (°)	1.20
Ramachandran*	
Preferrred (%)	96.1
Generously allowed (%)	3.1
Disallowed (%)	0.8

Values in parentheses are for highest-resolution shell.

5.3 Results and Discussion

S. pneumoniae fucose mutarotase structure

The refined structure of SpFcsU, determined to a resolution of 2.5 Å, revealed twenty monomeric SpFcsU subunits in the asymmetric unit that were arranged as two decamers, with each decamer comprising a dimer of pentamers. The SpFcsU monomer consists of a central β-sheet packed between three α-helices on one side of the β-sheet and two α-helices on the opposing side (Figure 22A). The pentamer of these subunits forms a ring-like structure in which the diameter of the inner and outer rings is approximately 40 and 90 Å, respectively (Figure 22B). The decameric structure, which has a height of approximately 56 Å, is formed by two pentamers that stack one on top of one another and interact via the same faces of the pentamer ring (Figure 22C). The pentamer rings, however, are rotated by about 36^{o} relative to one another resulting in a slightly offset association of the pentamer rings (Figure 22C).

The decameric quaternary structure of SpFcsU is stabilized by three molecular interfaces: (i), (ii), and (iii) (Figure 23A). The first and likely most significant, interface (i), occurs between adjacent monomers, referred to as monomers A and B, in the associated pentameric ring. The extensive interface of 985 A^2 is made by identical surfaces donated by each monomer, reflecting the two-fold symmetry that relates the two monomers, and is created by the pairing of the N-terminal α-helices (α1, residues 11-17) from each monomer while the arms created by residues 2-10 reach over to embrace the neighbouring monomer (Figure 23A and 23B). Four hydrogen bonds are formed at this interface with additional extensive van der Waals interactions, notably including the parallel stacking of the sidechain rings of His4 and Tyr59 from separate monomers (Figure 23B). Interface (ii) occurs between monomers B and C within the same pentamer ring. The interacting surface area of this interface is 442 A^2 and includes nine H-bonds (Figure 23A and 23C). Interface (iii) has an even smaller surface area at 258 A^2 and no polar interactions (Figure 23A and 23D). Notably, PISA predictions suggest that interfaces (ii) and (iii) are energetically insignificant and do not contribute to stabilization

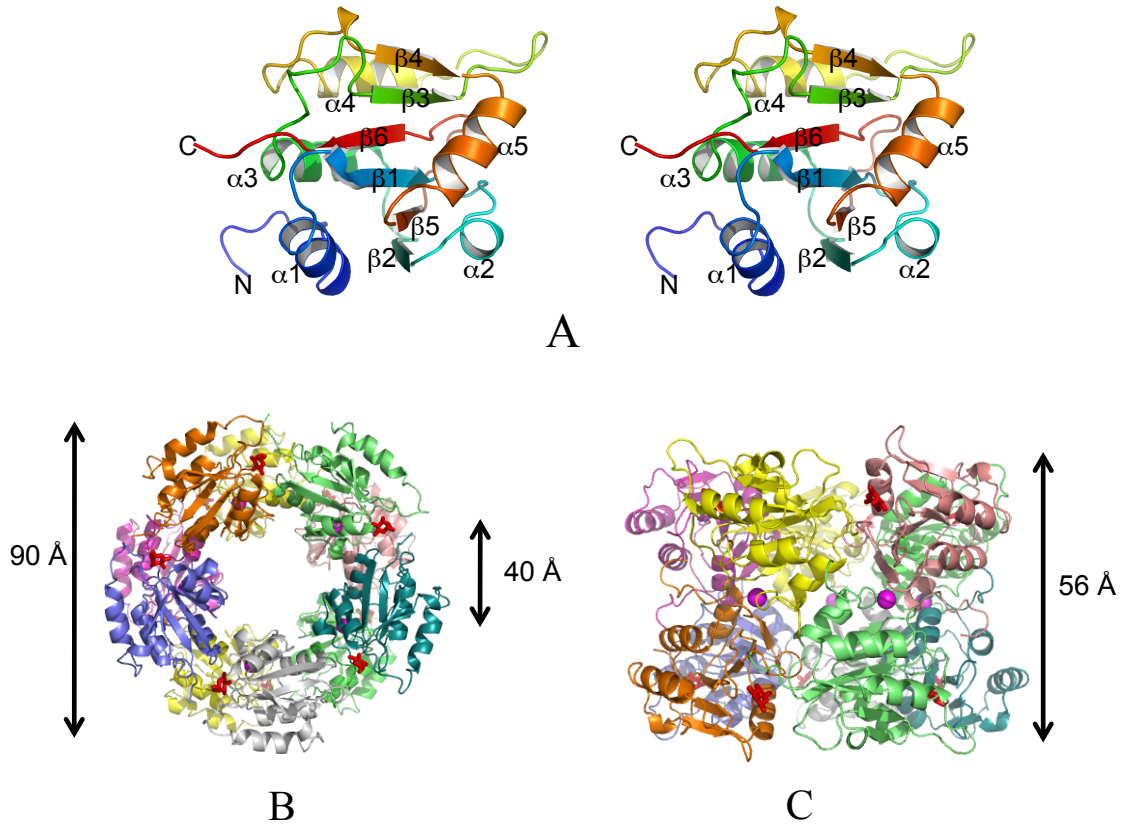


Figure 22: Structure of SpFcsU

(A) Stereo image of a cartoon representation of an SpFcsU monomer. (B) Top and (C) side views of the decamer organization of SpFcsU in complex with L-fucose. Each monomer is displayed in a different color with the potassium ions shown as magenta spheres and the fucose residues as red stick representations.

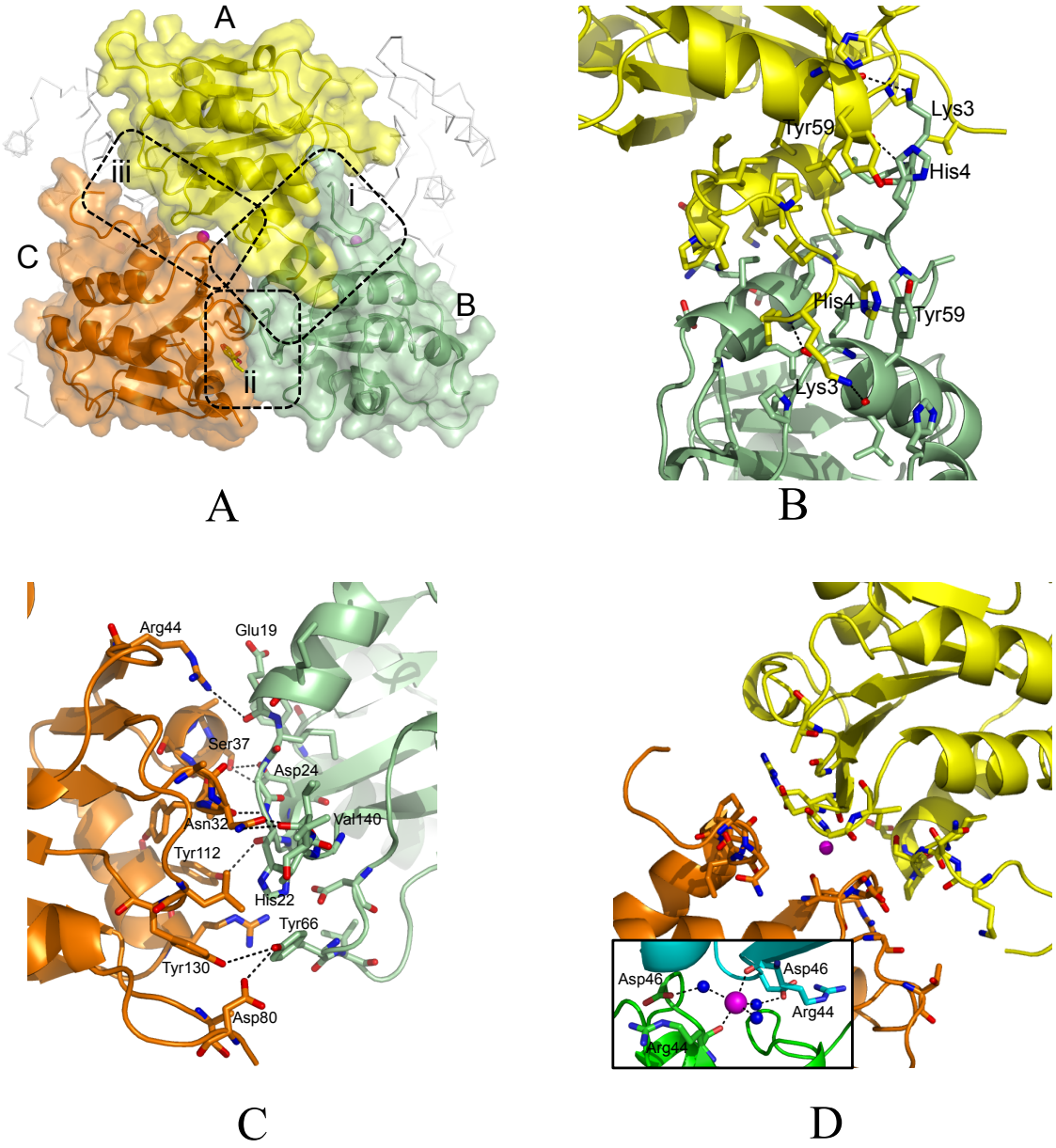


Figure 23: The interfaces of SpFcsU monomers contributing to decamer organization

(A) Side view of the decamer showing the three monomers involved in interfaces (i), (ii), and (iii). Monomer A is shown in yellow, B is shown in green, and C is shown in orange. Stick representation of amino acids involved in (B) interface (i), (C) interface (ii), and *(d)* interface (iii) with an inlay showing the coordination of the potassium ion with the main

chain oxygen of two Arg44 residues and three water molecules. Side chain nitrogens and oxygens are shown in blue and red, respectively, while potassium ions and water molecules are magenta and blue spheres, respectively.

of quaternary structure whereas interface (i) is predicted to have an essential role in complex formation. Thus, the decameric organization of FcsU is more accurately described as a pentamer of dimers where the dimers are of the A/B monomer type. These apparently stable dimers associate via interfaces (ii) and (iii), which together may provide sufficient driving force for the assembly of the decamer. However, in addition to the protein-protein interactions at the monomer interfaces, metal ions were also identified at each interface (iii) (Figure 23A and 23D). These ions were modeled as potassium on the basis of the presence of KCl in the crystallization conditions, coordination geometry, bond-lengths and analysis of the refined B-factors (average B-factor of 47, which was in accordance with neighbouring atoms). This potassium ion was in the same position in all cases and was penta-coordinated by the main chain oxygens of two Arg44 residues, one each from monomers A and C (Fig. 23D inset). Two Asp46 residues, again originating from the same A and C monomers, help coordinate two of the water molecules around the potassium ion. The coordination of this ion at interface (iii) likely aids in stabilizing decamer formation by improving the interaction between A/B type dimers at interface (iii).

Recognition of fucose

The crystals of SpFcsU were soaked in an excess of fucose and electron density that could be easily modeled as β-L-fucose was found in each of the active sites (Figure 22B and C). The active site resides at interface (ii) (Figure 23A) thus giving five active sites per pentamer ring and ten active sites in the entire decamer. The fucose monomer is coordinated by amino acid sidechains donated by both SpFcsU monomers that contribute to interface (ii), though the C-type monomer contributes the majority of the residues (Figure 24). Specificity appears to be conferred by hydrogen bonds from Asn132 and Asp30 to the axial hydroxyl groups of the fucose C2 and C3. The ring of Tyr130 packs against the apolar surface created by the plane of C3-C4-C5 in the fucose residue, selecting for the axial hydroxyl on C4. His22 (from the adjacent monomer) and Tyr112 hydrogen bond with O1 and O5, respectively, and also make up part of the catalytic machinery (see below).

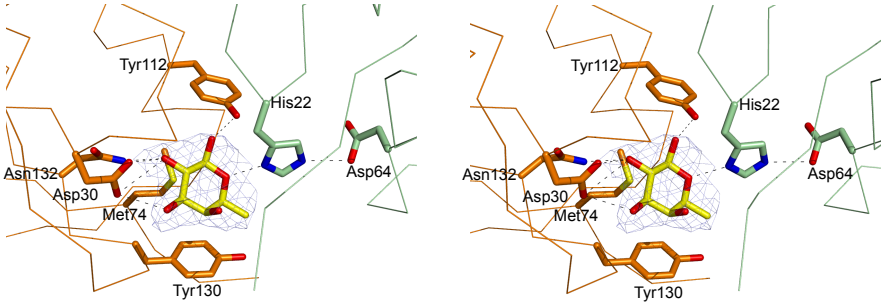


Figure 24: The active site architecture of SpFcsU

The active site of SpFcsU in a stereo image with the side chains of the catalytic and/or interacting residues from monomer B in green and from monomer C in orange while the fucose molecule is shown in yellow with the side chain oxygens, nitrogens, and sulfurs in red, blue, and yellow, respectively. The blue mesh shows the maximum-likelihood/σ_a-weighted F_{obs}-F_{calc} maps for fucose contoured at 0.25 electrons/Å^3.

Similarities Among Fucose Mutarotases

The SpFcsU monomer has amino acid identities of 43%, 50%, and 27% and root mean square deviations (RMSD) of 1.25 Å, 1.01 Å, and 1.72 Å with corresponding monomers from EcFcsU (PDB code 2WCV), MmFcsU (PDB code 2WCU), and BsRbsD (PDB code 1OGD), respectively (Figure 25A and 25B). Though all four of these structures are very similar, as revealed by the RMSD values, only the prokaryote structures share the same decamer organization (Figure 25C), while the mammalian fucose mutarotase (MmFcsU) is a homodimer (Kim, Shin et al. 2003; Lee, Ryu et al. 2009). Not surprisingly, the dimensions of the decamers formed by the prokaryotic enzymes are similar, as in general are many characteristics of the interfaces that result in the formation of the quaternary structures. A somewhat distinctive feature of SpFcsU, however, is the potassium ion found in interface (iii). Notably, there are no ions present in the EcFcsU structure, though the pocket that accommodates the potassium in SpFcsU is conserved in EcFcsU. The BsRbsD structure has well-ordered chloride ions, which are present at the equivalent of interface (i). The role of metal ions in the quaternary structure of these enzymes is unclear and, indeed, their observation in the structures of SpFcsU and BsRbsD may be artifacts of the crystallization. However, it is also plausible that these ions stabilize the quaternary structures in a manner that compensates for the slight species specific differences at the molecular interfaces of the protein monomers.

The residues in the catalytic site of SpFcsU are highly conserved with those of EcFcsU (Fig. 25A). The residues Arg108, Tyr112, His22', and Asp64' of SpFcsU overlap extremely well with those of EcFcsU (Figure 25D) (Kim, Shin et al. 2003), while these residues are also highly conserved among related family members. This high level of identity in the active site, along with the observed ability of SpFcsU to accommodate fucose in its active site, supports the contention that SpFcsU is indeed a fucose mutatrotase.

Biological Relevance of FcsU in S. pneumoniae

In prokaryotes, the assimilation of fucose is initiated by the fucose isomerase (FcsI) catalyzing the isomerization of fucose to fuculose, which is subsequently processed in

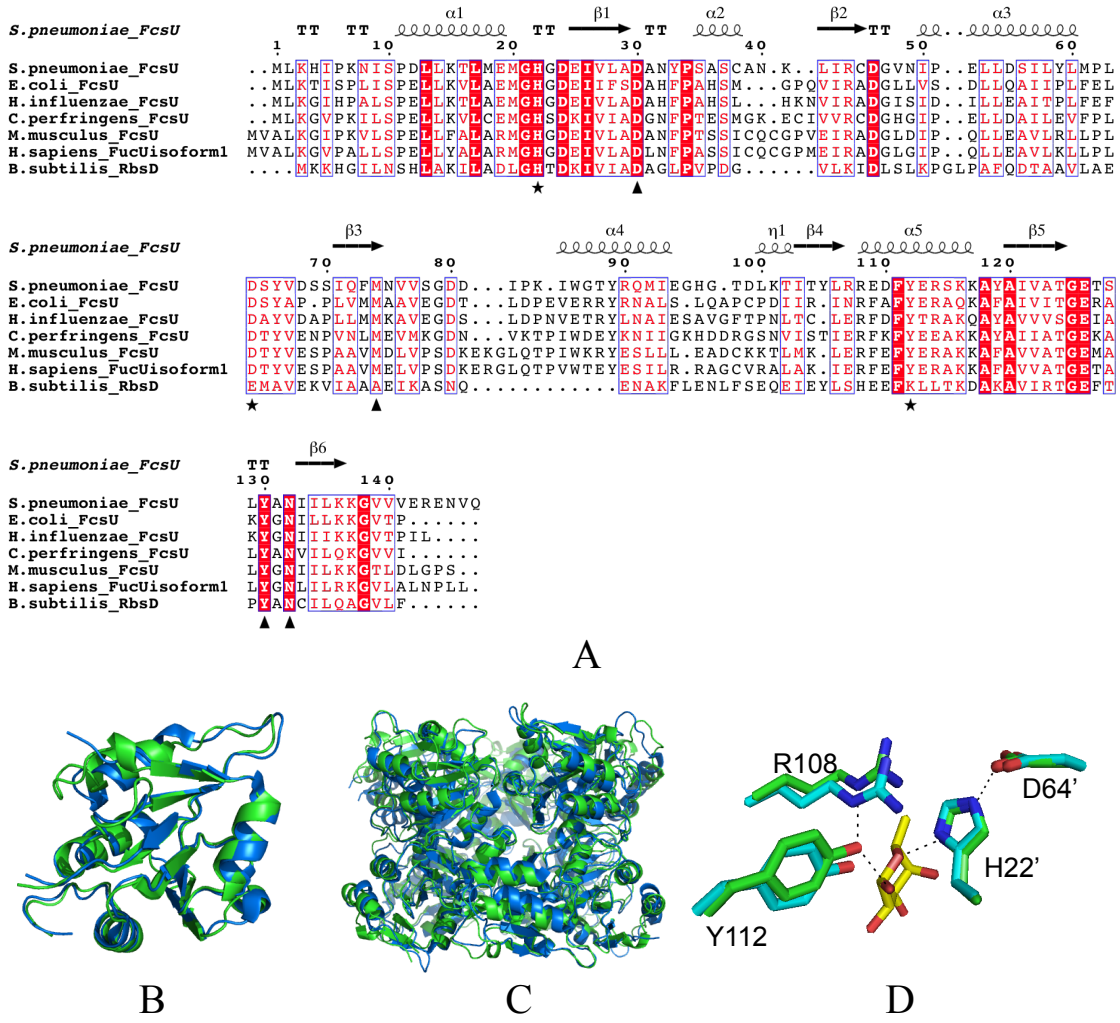


Figure 25: Comparison of SpFcsU

(A) Alignment of SpFcsU with other fucose mutarotases from both prokaryotes and eukaryotes and the paralog RbsD. Black stars and triangles denote the catalytic residues and fucose interacting residues, respectively. Cartoon representation overlay of the (B) monomers and (C) decamers of SpFcsU, in green, and EcFcsU, in blue. (D) A stick representation of the catalytic residues of SpFcsU and EcFcsU. Side chain oxygens and nitrogens are shown in red ad blue, respectively.

two additional steps to dihydroxyacetone phosphate and lactaldehyde. Like D-xylose isomerase, fucose isomerase is thought to be specific for the α-anomer of fucose (Seemann and Schulz 1997). In solution, however, the anomeric equilibrium of fucose favors β-L-fucose over α-L-fucose roughly 2.5:1 (Park, Ryu et al. 2007), providing a potential bottleneck for fucose metabolism. This limitation is compounded by inherently slow rates of spontaneous mutarotation, which, for example, can be on the order of minutes for glucose (Livingstone, Fanks et al. 1977). The specificity of FcsI for the less abundant anomer of fucose and the potential limits this imposes on flux through the fucose assimilation pathway provides a rationale for the incorporation of FcsU into fucose utilization pathways. As FcsI depletes the α-L-fucose the more abundant β-L-fucose is converted into α-L-fucose by FcsU at a rate more consistent with the needs of the downstream enzymes and speeding fucose conversion. Indeed, as an analogous example, it has been shown that an *E. coli* strain lacking the *YiiL* gene, which encodes an L-rhamnose mutarotase, when grown on a minimal media with a low concentration of rhamnose displays a longer lag phase and reduced maximal growth as compared to the wild type (Ryu, Kim et al. 2005).

The importance of FcsU to control the rate of fucose metabolism is clearly demonstrated in *S. pneumoniae*. Though this bacterium is unable to use exogenously supplied monosaccharide fucose as a sole carbon source (Chan, O'Dwyer et al. 2003; Embry, Hinojosa et al. 2007) it does possess a fucose utilization operon. Indeed, two types of mutually exclusive variations of fucose operons, called type 1 and type 2, are found in a strain dependent manner in *S. pneumoniae* (Higgins, Whitworth et al. 2009). A common feature of both operon-encoded pathways are the fucose processing enzymes FcsI, FcsK and FcsA. The pathways differ, however, in their type of carbohydrate transporter and, importantly, their complement of glycan processing enzymes that harvest fucose-containing glycans from host tissue and subsequently break them down into monosaccharide components.

Based on available genome sequences the majority of the *S. pneumoniae* strains have the type 1 fucose utilization pathway. This pathway is initiated by the action of an extracellular family 98 glycoside hydrolase that cleaves host Lewis^{Y} antigen yielding the H-disaccharide (Higgins, Whitworth et al. 2009). Once transported into the bacterial cell,

it is suggested that a putative intracellular GH95 hydrolyzes the H-disaccharide with the inverting mechanism demonstrated for the *Bifidobacterium bifidum* α-1,2-L-fucosidase homolog (Katayama, Sakuma et al. 2004). This ultimately releases the β-anomer of L-fucose. As the α-anomer is the supposed substrate for the fucose isomerase this pathway employs the SpFcsU, studied here, to speed the conversion of β-L-fucose to α-L-fucose and thus increasing the flux through the fucose conversion steps of the pathway (Figure 26). In contrast, the type 2 fucose utilization pathway is dedicated to harvesting and depolymerizing the blood group A- and B-antigens (Higgins, Whitworth et al. 2009). These sugars share similarities with the Lewis^{Y} antigen, including the α-1,2-fucose appended to a galactose. In this type 2 pathway, the fucose is predicted to be liberated from the blood group antigens by a family 29 glycoside hydrolase; previously characterized GH29 enzymes have shown to be retaining α-L-fucosidases (Berteau, McCort et al. 2002; Berteau, Bielicki et al. 2004) suggesting that the *S. pneumoniae* GH29 liberates α-L-fucose from its substrate (Figure 26). Unlike the type 1 pathway, therefore, the type 2 pathway releases α-L-fucose, which can be efficiently processed by FcsI without the need for SpFcsU.

FcsU is commonly found in pathogenic bacteria, such as *E. coli* and *C. perfringens*, and is often the first step in fucose utilization pathways where it converts excess free β-L-fucose to α-L-fucose and thus creates substrate for FcsI, which is the next enzyme in the pathway. In *S. pneumoniae*, however, the initiation of the fucose pathway is not with FcsU acting on free fucose; instead with a series of glycoside hydrolases releasing fucose from host complex carbohydrates (Higgins, Whitworth et al. 2009). One such glycoside hydrolase, GH95, releases β-L-fucose, and thus, SpFcsU acts as a link between glycoside hydrolase activity on host carbohydrates and bacterial carbohydrate metabolism.

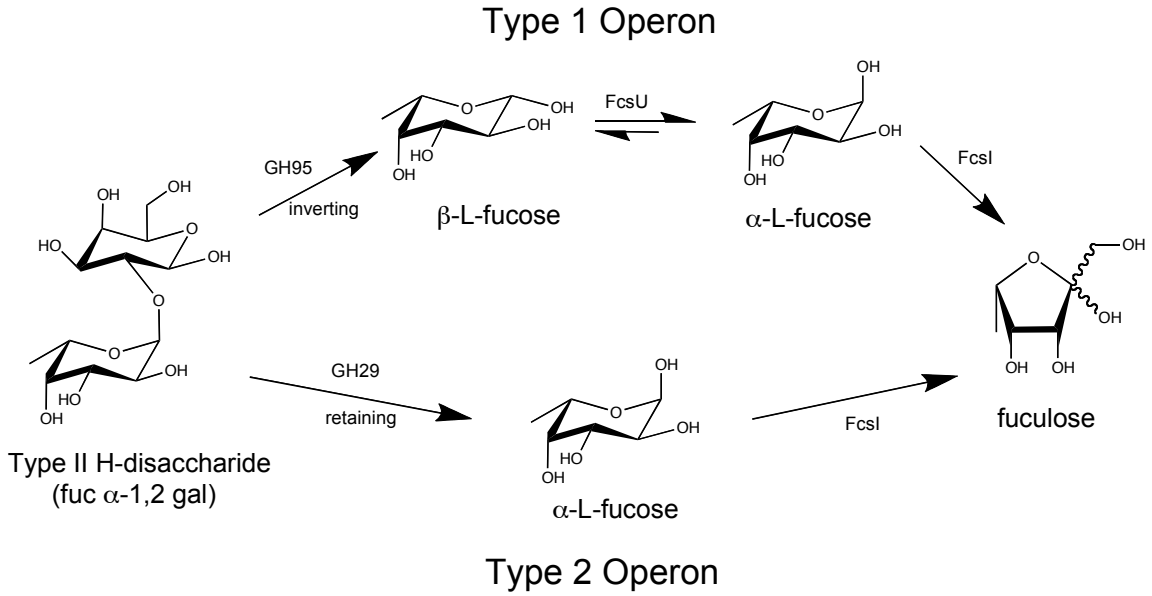


Figure 26: Biological role of SpFcsU

Schematic of the link between glycan degradation and fucose processing in *S. pneumoniae* by both putative fucose metabolic pathways, type 1 and type 2 operons. The type I operon harvests the Lewis^{Y} blood group antigen and encodes for a GH95 and SpFcsU to release fucose for further processing by FcsI while the type 2 operon harvests the A/B-blood group antigens and encodes for a GH29 that does not contain SpFcsU for further fucose processing. Inverting and retaining refer to the catalytic mechanism of the glycoside hydrolases (Figure 4).

Chapter 5: Pneumococcal fucose processing enzymes

Melanie A. Higgins[1], Nahida El Warry[1], Candace Marsters[1], Robert D. Burke[1] and Alisdair B. Boraston[1]

Microscopy data is adapted from Differential recognition and hydrolysis of host carbohydrate antigens by Streptococcus pneumoniae family 98 glycoside hydrolases. J Biol Chem. (2009) 284(38):26161-73. The rest of the data is unpublished.

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

Contributions: I performed the cloning and protein production and purification of all the *S. pneumoniae* proteins, the fucose processing enzyme assay, and the SpFcsK crystallization and structure refinement. I also created all of the *S. pneumoniae* mutant strains and performed the localization experiments and internalization assays. Candace Marsters performed the cloning and protein production and purification of the *E. coli* fucose processing proteins. The microscopy experiments were performed and analysed by Nahida El Warry and Robert Burke.

5.1 Abstract

Fucose metabolism pathways are present in many bacterial species and typically consist of the central fucose processing enzymes, fucose isomerase (FcsI), fuculose kinase (FcsK), and fuculose 1-phosphate aldolase (FcsA). Fucose initially undergoes isomerization by FcsI producing fuculose, which is then phosphorylated by FcsK. Finally, FcsA cleaves the fuculose 1-phosphate product into lactaldehyde and dihydroxyacetone phosphate (DHAP), which can be incorporated into central metabolism allowing the bacterium to use fucose as an energy source. However, *S. pneumoniae* cannot use exogenously supplied fucose as a sole energy source even though it contains the necessary fucose processing enzymes. Interestingly, enzyme assays along with structural data support that the *S. pneumoniae* fucose processing enzymes can act in a sequential manner to ultimately produce DHAP. Also similar to *E. coli* fucose metabolism, fuculose 1-phosphate appears to act as an inducing molecule for expression of the fucose operon. Lastly, the extracellular GH98 found in both types of fucose operon acts to remove di- or trisaccharides from histo-blood group antigens presented on lung epithelial cells and preliminary data shows that fucose utilization in *S. pneumoniae*

may be involved in host internalization suggests a novel role for fucose metabolism in pneumococcal pathogenesis.

5.2 Introduction

Fucose metabolism was first studied in the 1950's when Green and Cohen initially observed the isomerization of fucose in *E. coli* lysates (Green and Cohen 1956). This led to further research by Heath and colleagues on other enzymes involved in processing fucose ultimately revealing a fucose metabolism pathway. First, fucose is converted into fuculose by fucose isomerase (FcsI) (Green and Cohen 1956). Fuculose kinase (FcsK) then uses ATP to phosphorylate fuculose creating fuculose 1-phosphate (Heath and Ghalambor 1962), which acts as a substrate for fuculose 1-phosphate aldolase (FcsA) producing lactaldehyde and dihydroxyacetone phosphate (DHAP) (Ghalambor and Heath 1962), the latter a glycolysis intermediate (Figure 27). Upon sequencing of many bacterial genomes, it was discovered that these fucose processing enzymes (FcsI, FcsK, and FcsA) are usually part of an operon dedicated to metabolizing fucose that also includes a fucose permease (FcsP) that transports fucose from the environment into the cell, a fucose mutarotase (FcsU) that speeds conversion of the more abundant β-L-fucose into the α-L-fucose substrate of FcsI, and finally a fucose regulator (FcsR) that controls expression of the entire operon. Recently, structures have been solved for some of these enzymes in the operon (FcsU, FcsI, and FcsA) (Dreyer and Schulz 1996; Seemann and Schulz 1997; Grueninger and Schulz 2006; Lee, Ryu et al. 2009), giving greater insight into their mechanisms of catalysis and providing a better understanding of bacterial fucose metabolism.

The genome of *Streptococcus pneumoniae* encodes for two different strain-dependent putative fucose utilization operons; however, they appear more complex than the operon frequently found in other bacteria. Both operons contain different glycans processing enzymes, including extracellular GH98 enzymes, carbohydrate transport systems, a fucose regulator, and putative fucose processing enzymes, FcsU, FcsI, FcsK, and FcsA (see Section 1.5). However, unlike other fucose pathways which seem to have a strict

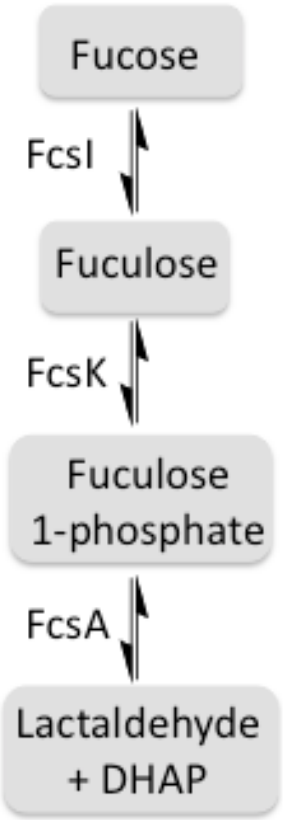


Figure 27: Schematic of the fucose processing pathway

General fucose metabolic pathway where fucose is converted to lactaldehyde and DHAP.

metabolic function, this pathway plays an important role in pneumococcal pathogenesis; an *S. pneumoniae* strain lacking the fucose operon renders the bacterium completely avirulent in a mouse model (Embry, Hinojosa et al. 2007). Surprisingly, *S. pneumoniae* is unable to use fucose as a sole carbon source suggesting a novel function for this operon.

Here, we investigate the activities of the pneumococcal fucose processing enzymes to verify that *S. pneumoniae* contains a similar fucose processing pathway to *E. coli*. Also, we examined the role of fucose utilization in *S. pneumoniae* pathogenesis. *In situ* studies determined the activities of the extracellular GH98 enzymes from both fucose operons on specific host complex carbohydrates. Furthermore, we explored the possibility of novel host internalization function for fucose metabolism in *S. pneumoniae* pathogenesis.

5.3 Material and Methods

Expression and purification

fcsI, fcsK, and *fcsA* gene fragments were amplified from *S. pneumoniae* TIGR4 genomic DNA (American Type culture Collection BAA-334D) using specific primers with engineered *NdeI* or *NheI* and *XhoI* or *HindIII* restriction sites in the forward and reverse primer, respectively. *fcsI, fcsK,* and *fcsA* gene fragments were also amplified from *E. coli* DH5α colony PCR also with specific primers engineered with *NheI* and *XhoI* *r*estriction sites. A full list of the primers is given in Table 12. The DNA fragments were subsequently cloned into pET 28a(+) (Novagen) using the specific restriction sites followed by bidirectional sequencing to confirm sequence fidelity.

Protein production and purification was carried out as previously described (Higgins, Whitworth et al. 2009). Briefly, the pET28a(+) plasmids containing the fucose processing gene fragments were transformed and expressed in *E. coli* BL21 star (DE3) cells (Novagen). Cultures were initially grown at 37°C in Luria-Bertani broth supplemented with 50 μg ml^{-1} kanamycin until they reached an absorbance at 600 nm of about 0.6, induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside, and then incubated at 37°C for 4-6 hours or at 16°C over night. After centrifugation, a French pressure cell or chemical lysis methods were used to rupture the cells (Robb, Nano et al. 2010). The target proteins were initially purified by Ni^{2+} immobilized metal

Table 12: Primer list for the fucose processing recombinant proteins

Primer	DNA sequence 5'-3'
SpFcsI-F	5’ CATATGGCTAGCATTCAACATCCACGTATT
SpFcsI-R	5’ GGCCGCAAGCTTTTATTTATGTAGTGGCCCCAA
SpFcsK-F	5’ GGCAGCCATATGGAAGCGGTTTTAGCAATA
SpFcsK-R	5’ GTGGTGCTCGAGTTAATTCTTTTGGGAGTAATA
SpFcsA-F	5’ CATATGGCTAGCTCAGATGTAAAACAAGAA
SpFcsA-R	5’ GTGGTGCTCGAGTTATTTTCTCTGACCGTATGT
EcFcsI-F	5’ CATATGGCTAGCTTACCGAAAATTGGTATCCGCCCG
EcFcsI-R	5’ GTGGTGCTCGAGTTAACGCTTGTACAACGGACCGTAGTT
EcFcsK-F	5’ CATATGGCTAGCTTATCCGGCTATATTGCAGGAGCG
EcFcsK-R	5’ GTGGTGCTCGAGTTACACTTCCTCTATAAATTVAGGTTC
EcFcsA-F	5’ CATATGGCTAGCGAACGAAATAAACTTGCTCGTCAG
EcFcsK-R	5’ GTGGTGCTCGAGTTACTCTTCAATTCGTAACCCATAGGT

Underlined DNA sequences show the restriction endonuclease sites.

affinity chromatography (Amersham), followed by a size exclusion chromatography step using a Sephacryl S-200 column (GE Healthcare). The purified target proteins, with intact polyhistidine tags, were concentrated for further studies.

Protein concentrations were determined by measuring the absorbance at 280nm (A_{280}) and using calculated molar extinction coefficients (0.088725, 0.059945, and 0.020525 cm^{-1} μM^{-1} for SpFcsI, SpFcsK, SpFcsA, respectively, and 0.1308, 0.08839, 0.02168 cm^{-1} μM^{-1} for EcFcsI, EcFcsK, and EcFcsA, respectively) (Gasteiger, Gattiker et al. 2003).

The cloning, protein production and purification methods for the catalytic modules from Sp3- and Sp4-GH98 (Sp3GH98CM and Sp4GH98CM) can be found in Section 2.3 (Higgins, Whitworth et al. 2009).

Fucose processing enzyme assay

A multi-enzyme assay was performed with specific fucose processing enzymes (FcsI, FcsK, and FcsA) from either *S. pneumoniae* or *E. coli*. The reactions contained 100 mM Tris pH 7.5, 1 mM $MgCl_2$, 1 mM $MnCl_2$, 1.5 mM L-fucose (Sigma), 5 mM ATP (Calbiochem), 120 μM NADH (Sigma), 1.18 U μl^{-1} α-glycerophosphate dehydrogenase (Sigma), and 40 μg ml^{-1} of the appropriate recombinant enzymes. Reactions were measured at 339 nm every 10 seconds by a Carey Bio UV-Visible Spectrophotometer to monitor the NADH depletion. Data points were plotted on a graph to obtain a slope of μM NADH depleted per minute. Assuming that NADH depletion is equivalent to NAD^+ production, ultimately the amount of NAD^+ produced was calculated.

Dynamic light scattering (DLS)

DLS experiments were performed for SpFcsI and SpFcsA to estimate their molecular weights in solution. Recombinant protein was purified as described above with the following exceptions. The polyhistidine tag was cleaved off of SpFcsI using thrombin treatment prior to size exclusion chromatography with the Sephacryl S-200 column (GE Biosciences) equilibrate in buffer containing 20 mM Tris pH 7.0 and 150 mM NaCl. For SpFcsA, size exclusion chromatography was also with the Sephacryl S-200 column (GE Biosciences) but was equilibrated in 20 mM Tris pH 7.0, 300 mM NaCl, and 5 mM

dithiothreitol (DTT). Two individual 1 mL fractions from the size exclusion chromatography of each SpFcsI and SpFcsA were collected, concentrated to 8-10 mg ml^{-1}, centrifuged for 30 minutes, and filtered before DLS analysis, which was carried out on the DynaPro^{TM} Plate Reader at 25 °C. Size and size distributions were calculated by Dynamics® 7 software using the method of Cumulants or Regularization (Koppel 1972).

Crystallization data collection, structure determination and refinement

Crystallization of SpFcsK was carried out using the hanging drop vapour diffusion method at 18°C. Preliminary screening using the Hampton Crystal Screen 1 and 2, with a protein concentration of 20 mg ml^{-1}, yielded initial crystals in a single condition. After optimization of this crystallization condition the best crystals for SpFcsK grew in 1.9 M Ammonium Sulfate, 0.1 M sodium citrate pH 6.8, 10% PEG 400.

SpFcsK crystals were flash frozen in crystallization solution supplemented with 15% (v/v) ethylene glycol. Diffraction data was collected with a Rigaku R-AXIS 4++ area detector coupled to a MM-002 X-ray generator with Osmic 'blue' optics and an Oxford Cryostream 700.

Molecular replacement using PHASER (McCoy, Grosse-Kunstleve et al. 2005) with L-rhamnulokinase from *E. coli* (PDB 2CGL) as a search model determined a clear solution for one SpFcsK molecule per asymmetric unit. Manual modeling was completed using COOT (Emsley and Cowtan 2004) followed by refinement using REFMAC (Murshudov, Vagin et al. 1997). Water molecules were found using COOT:FINDWATERS and manually inspected after refinement. Refinement procedures were monitored by flagging 5% of all observation as 'free' (Lovell, Davis et al. 2003). Model validation was performed by RAMPAGE (Lovell, Davis et al. 2003) and MOLPROBITY (Chen, Arendall et al. 2010). Data statistics can be found in Table 13.

S. pneumoniae transformation

S. pneumoniae transformation was carried out as previously described (Bricker and Camilli 1999). Briefly, an initial culture was grown in Todd Hewitt Broth supplemented with 0.5% yeast extract (THY) for 9-10 hours to an approximate OD of 600 nm of 0.9 in

Table 13: Data collection and refinement statistics for SpFcsK.

Data collection	
Beamline	Home source
Wavelength	1.514
Space group	4_32_12
Cell dimensions	
a, *b*, *c* (Å)	60.16, 60.16, 282.32
Resolution (Å)	19.904-2.2 (2.32-2.20)
R_{sym} or R_{merge}	0.057 (0.259)
$I / \sigma I$	16.1 (3.4)
Completeness (%)	98.7 (94.3)
Redundancy	4.5 (3.0)
Refinement	
Resolution (Å)	2.2
No. reflections	25685
R_{work} / R_{free}	0.241/0.300
No. atoms	
Protein	3537
SO4	10
CIT	13
Water	222
B-factors (Å^2)	
Protein	8
SO4	31
CIT	46
Water	11
R.m.s. deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.4
Ramachandran	
Preferrred (%)	97.4
Generously allowed (%)	2.4
Disallowed (%)	0.2

Values given in parentheses represent the highest resolution shell.

a candle jar at 37ºC. This starter culture was diluted 1/50 into pre-induction growth medium (THY, 0.5% (w/v) glycine, 11 mM HCl) and grown at 37 ºC in a candle jar until an OD of 600 nm of 0.04 was reached. The media was then supplemented in order with 10 mM NaOH, 0.2% (w/v) BSA, 1 mM $CaCl_2$, and 100 ng ml^{-1} Competence Stimulating Peptide 2 (CSP-2) (Pozzi, Masala et al. 1996; Bricker and Camilli 1999) and incubated for 14 minutes in a candle jar at 37ºC. At this time, the specific PCR product was added to the media, incubated for 1 hour at 37 ºC in a candle jar, then diluted 1/4 with THY broth and incubated for another 2 hours. The transformations were finally plated on sheep blood agar supplemented with the appropriate antibiotic for selection and incubated overnight in a candle jar at 37 ºC. Antibiotic concentrations were 200 μg ml^{-1} for kanamycin and 150 μg ml^{-1} for streptomycin.

S. pneumoniae mutagenesis

Deletion mutants of the fucose processing and *gh98* genes from the fucose utilization operon of *S. pneumoniae* TIGR4 were created using the JANUS cassette (Sung, Li et al. 2001). All of the *S. pneumoniae* strains and primers are listed in Table 14. This method exploits a recessive Sm^R allele allowing for gene replacement or deletion through negative selection. A single base change, Lys 56 (AAA) to Thr (ACA) in the *rpsl41* gene gives rise to the recessive Sm^R strain *RpsL41*(Claverys, Roger et al. 1980; Salles, Creancier et al. 1992). This mutation in the parent *S. pneumoniae* TIGR4 strain was created using an adapted 'megaprimer' PCR method to introduce this single base change (Barik 1996). However, instead of ligating the final PCR product containing the K56T mutation into a plasmid, it was directly transformed into *S. pneumoniae* TIGR4 and subsequent RpsL41 clones were selected for on TSA plates supplemented with streptomycin. Further *S. pneumoniae* deletion strains were then constructed from this RpsL41 strain.

A PCR based method was used to create the first JANUS construct (Figure 28). Flanking regions surrounding the target gene were amplified by PCR whereby the downstream flanking region (F1) contained a *BamHI* restriction site while the upstream flanking (F2) region an *ApaI* site and the JANUS cassette contains both *ApaI* and *BamHI* digested with *BamHI*, ligated together then amplified by PCR using F1-F and DAM406

Table 14: Strains and primers used to create the *S. pneumoniae* mutant strains

Strain	Description
TIGR4	ATCC BAA-334
RpsL41	TIGR4 derivitave r*psL41*; Sm^R, Kan^S
Δ*gh98*-Janus	*rpsL41* with gh*98::kan-rpsL*$^+$; Sm^S, Kan^R
Δ*gh98*	Δ*gh98*; Sm^R, Kan^S
Δ*fcsI*-Janus	*rpsL41* with *fcsI::kan-rpsL*$^+$; Sm^S, Kan^R
Δ*fcsI*	Δ*fcsI*; Sm^R, Kan^S
Δ*fcsK*-Janus	*rpsL41* with *fcsK::kan-rpsL*$^+$; Sm^S, Kan^R
Δ*fcsK*	Δ*fcsK*; Sm^R, Kan^S
Δ*fcsA*-Janus	*rpsL41* with *fcsA::kan-rpsL*$^+$; Sm^S, Kan^R
Δ*fcsA*	Δ*fcsA*; Sm^R, Kan^S
Δ*cps*-Janus	*rpsL41* with *cps::kan-rpsL*$^+$; Sm^S, Kan^R
Δ*cps*	Δ*cps*; Sm^R, Kan^S
Δ*cps*Δ*gh98*-Janus	Δ*cps* with *gh98::kan-rpsL*$^+$; Sm^S, Kan^R
Δcps	Δ*cps*Δ*gh98*; Sm^R, Kan^S
Primer	**DNA sequence**
RpsL41-F	5' TCTTGACAAGCAAGGGAAAA
Rps41L-R	5' CGATTGAAGCAGCAGTACCA
RpsL41K56T-F	5' ACAATGACACCTACAAAACCTAACTCA
RpsL41K56T-R	5' TGAGTTAGGTTTTGTAGGAGACATTGT
DAM351	5' CTAGGGCCCCTTTCCTTATGCTTTTGGAC
DAM406	5' TCTATGCCTATTCCAGAGGAAATGGAT
ΔGH98-F1-F	5' CACGGCTTAAAACAAGGTCTTCACGGT
ΔGH98-F1-R	5' ACGAGGATCCATTTTAAAAATTATCACCCAGGCTACCGTA
ΔGH98-F2-F	5' AGCAGGGCCCCTAAAAATTTAACTCAATAATTTTTTC
ΔGH98-F2-R	5' CTATCATCTACAATTGATAATCAAATT
ΔFcsI-F1-F	5' CGAGGAAGATATTCCTTGATAAGCTCA
ΔFcsI-F1-R	5' ACGAGGATCCAAACTTGTTTATATAGGAGGTGAACTT
ΔFcsI-F2-F	5' AGCAGGGCCCTTACTTATAAGTATAAACCTCTACCTC
ΔFcsI-F2-R	5' AATAAGGTAGCTACTCAAAGTTCAACC
ΔFcsK-F1-F	5' CCATTTCCATTAAAGTCTTCAATAAAT
ΔFcsK-F1-R	5' ACGAGGATCCAAAAGATTGAGAGTTTGTAAATTTGCC
ΔFcsK-F2-F	5' AGCAGGGCCCTTTTCTTCTCTCTTCGTCCTTGATTAA
ΔFcsK-F2-R	5' CCAGTCTTGTAAAATTTAGAACTATCA
ΔFcsA-F1-F	5' AACATTAAACAAACTATTTTGAGCATT
ΔFcsA-F1-R	5' ACGAGGATCCGGGAGGATATTAATGTTAAAACATATA
ΔFcsA-F2-F	5' AGCAGGGCCCTTAATTCTTTTGGGAGTAATAACTCTC
ΔFcsA-F2-R	5' TATTTGAATTGGATTCATCTTTACTTC
Δcps-F1-F	5' GCAAGAATTGTCTATGGTCTTCCAGTT
Δcps-F1-R	5' ACGAGGATCCTTATAGTAATTCCACACAGAAAGCATC
Δcps-F2-F	5' AGCAGGGCCCAATTGTATATTTTTAATAAATCAAAAC
Δcps-F2-R	5' AAACTAGGAAGGGCAATTGAATACAGC

Underlined DNA sequences show the restriction endonuclease sites.

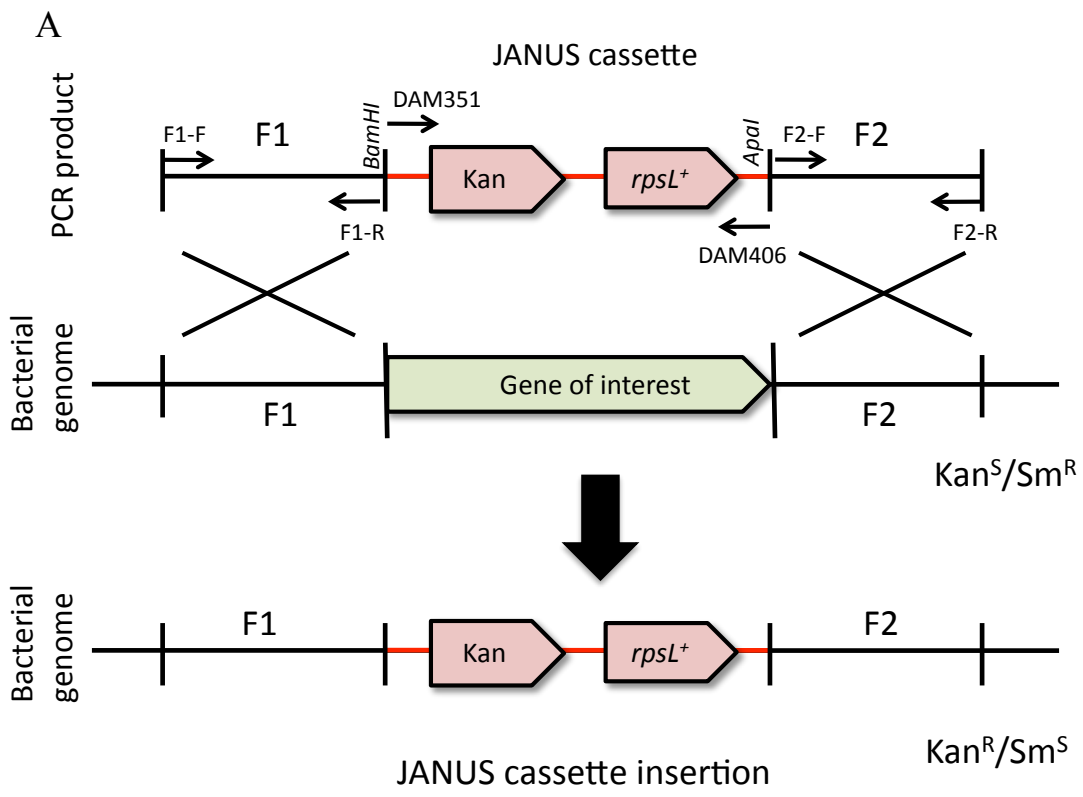
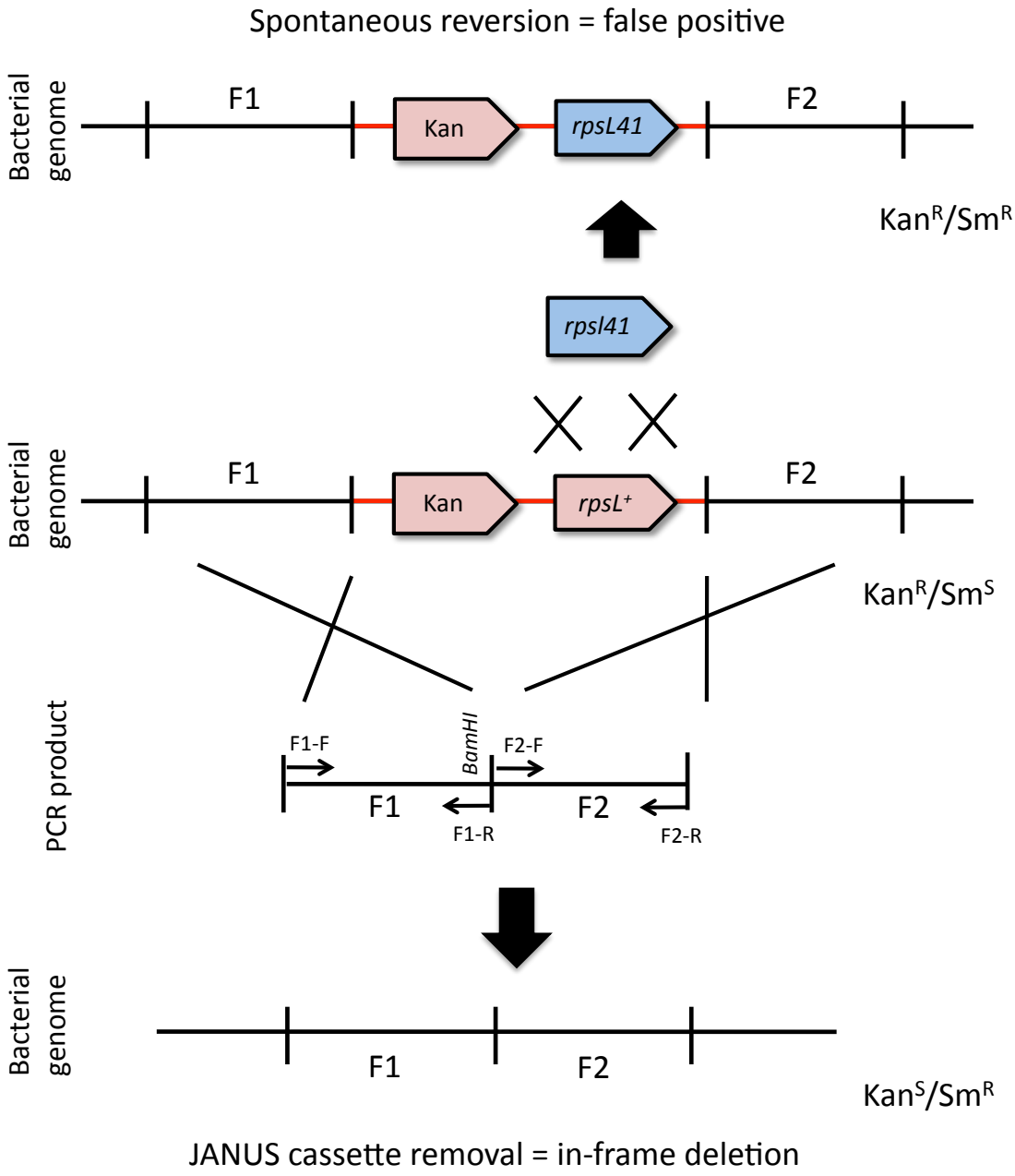


Figure 28: Schematic of *S. pneumoniae* mutagenesis

Schematic of *S. pneumoniae* mutagenesis to construct deletion strains using the JANUS cassette. Shown are the primers and restriction sites used to make the transformed PCR constructs (A) The first transformation step to create the JANUS strain (ie/ Δ*fcsI*-JANUS). (B) (On the next page) The second transformation step has two fates, a false positive or a proper deletion strain. F1 and F2 are flank 1 and 2, respectively. Primers are labelled F1-F and F1-R etc. (the DNA sequences for each deletion construct are found in Table 14).

Figure 28B – Legend on previous page.



sites at opposing ends (Figure 28A) (using DAM406 and DAM351 primers (Sung, Li et al. 2001; Trzcinski, Thompson et al. 2003)). F1 and the JANUS cassette were both (Trzcinski, Thompson et al. 2003) primers producing an F1-JANUS fusion product. F1-JANUS and F2 were then digested with *ApaI* and ligated together, then subsequently amplified by PCR, using F1-F and F2-R primers, yielding a final construct that was transformed into RpsL41. Strains containing the JANUS cassette (which are Kan^R and Sm^S due to the kanamycin resistant cassette and $rpsl^+$ gene, respectively) were selected for on TSA plates supplemented with kanamycin and subsequent PCR reactions tested for the correct location of JANUS cassette insertion.

A similar PCR-ligation protocol was employed to generate the final construct necessary to make the *S. pneumoniae* deletion mutants (Figure 28B). This method used the previous amplified F1 region along with a second F2 region that was amplified with an *BamHI* restriction site. After *BamHI* digestion, both F1 and F2 were ligated and further amplified with the F1-F and F2-R primers producing a DNA sequence that lacks the gene target. This product was transformed into the appropriate JANUS cassette containing strain (ie/ *fcsI::kan-rpsL*$^+$) and strains now lacking the JANUS cassette with the specific gene deletion (Kan^S/Sm^R) were selected on TSA plates supplemented with streptomycin. However, spontaneous recombination of the *RpsL41* (Sm^R) and $rpsl^+$ (Sm^S) alleles can produce false-positive clones (Kan^R/Sm^R) (Sung, Li et al. 2001) so further diagnostics were necessary. Initial colonies from the transformation were re-streaked on TSA plates supplemented with either kanamycin or streptomycin. Colonies that grew on both TSA plates were designated as false-positives due to their Kan^R/Sm^R phenotypes. On the other hand, colonies that grew on TSA plates with streptomycin but not on plates with kanamycin indicated the removal of the JANUS cassette due to their Kan^S/Sm^R phenotype. These clones were then screened by PCR to validate the correct gene deletion.

Using this method in-frame unmarked deletions were created in the RpsL41 strain producing Δ*gh98*, Δ*fcsI*, Δ*fcsK*, and Δ*fcsA* strains for the *S. pneumoniae* fractionation experiments. However, a Δ*gh98* mutant in a TIGR4 strain lacking its capsular polysaccharide was used for the internalization assays. First, a TIGR4 strain lacking the *cps* locus, which encodes for the capsular polysaccharide biosynthesis operon, was

created from the RpsL41 strain using the same deletion method described above. The Δ*cps* strain was then used to generate a strain lacking both the capsular polysaccharide and *gh98* (Δ*cps*Δ*gh98*).

S. pneumoniae fractionation and Western Blots

Fractionation and Western blot experiments were performed on the *S. pneumoniae* Δ*gh98*, Δ*fcsI*, Δ*fcsK*, and Δ*fcsA* strains as described in Section 3.3 (Higgins, Ficko-Blean et al. 2011).

Immunofluorescence microscopy

A549 human lung adenocarcinoma epithelial cells (ATCC CCL-185) were cultured in DMEM:F12, containing 10% FBS, 2 mM L-Glutamine, 1 mM sodium pyruvate and passaged into 8-well slide chambers. Cells were rinsed in PBS and fixed in 4% paraformaldehyde in PBS (15 min), rinsed, and blocked in 5% lamb serum in PBS Tween-20 (15 min, room temperature). Cells were treated with 2.5 mg ml^{-1} enzyme (Sp3GH98CM and/or Sp4GH98CM), diluted in 20 mM Tris-HCl, pH 8.0, with 1 mM DTT for Sp4GH98CM or 250 mM NaCl for the Sp3GH98CM (overnight, room temperature). Cells were rinsed three times in PBS before application of primary antibodies. Antibodies used were anti-$Lewis^Y$ antigen (Abcam, ab23911-100) and anti – A/B antigen (Abcam, ab24223). Primary antibodies were diluted 1:50 ($Lewis^Y$) and 1:400 (A/B) in 5% lamb serum in PBS and applied to the fixed cells overnight (4^oC). Specimens were rinsed three times in PBS and Alexa 568 conjugated goat anti-mouse IgG secondary antibody (Molecular Probes) diluted 1:800 in PBS was applied. After incubation (1-2h), cells were rinsed with PBS and counter stained with Hoechst 33342 (1 ng ml^{-1}, Sigma). Slides were mounted with coverslips, examined with a Leica DM-6000 epifluorescence microscope, and images captured with a Hamamatsu Orca widefield camera controlled with Openlab software (4.04). Overall contrast and brightness were adjusted and images cropped and assembled using Photoshop (CS2). For quantification, fluorescence intensity per unit area was determined within a 5 by 5 μm sample area placed over the cytoplasm of cells chosen at random in monochrome images using

ImageJ (1.40g). Statistical analysis (one-way ANOVA, Bonferroni's multiple comparison test) was done with Graph Pad Prism (4.03).

Internalization assays

Internalization assays were performed on the Δ*cps* and Δ*cps*Δ*gh98* strains and the methods were adapted from a previously described protocol (Hammerschmidt, Wolff et al. 2005). All assays were performed in 24-well plates at 37°C in the presence of 5% CO_2. A confluent monolayer of A549 human lung epithelial cells (2 x 10^5 per well) were incubated with TNFα (Sigma T0157) for 2 hours. Once the bacteria reached an OD of 600 nm between 0.1-0.2, 4 x 10^5 bacteria resuspended in DMEM:F12 media containing 1% FBS and 1 mM sodium pyruvate were added to the A549 cells, spun at 1000 RPM for 5 minutes, and incubated for 3 hours. To determine the number of intracellular surviving bacteria, the A549 cells were then treated with 200 ug ml^{-1} of gentamicin to kill any extracellular bacteria. Cells were then washed with PBS, lysed with 1% saponin, and the whole contents plated on a TSA plate. After the TSA plates were incubated at 37 °C for at least 24 hours, *S. pneumoniae* colonies were counted to yield the intracellular surviving bacteria per well. Each experiment was performed on at least 3 different occasions with 8 replicates each. Statistical analysis (two-tailed unpaired *t*-test) was calculated with Graph Pad Prism (5.03).

5.4 Results and Discussion

Fucose processing enzyme reactions

The activities of these fucose processing enzymes from *S. pneumoniae* were examined using a multi-enzyme assay. FcsI, FcsK, and FcsA were combined with glycerol phosphate dehydrogenase (GPDH), which requires NADH as an electron donor to convert DHAP to glycerol 3-phosphate. Upon the addition of the fucose, the generation of DHAP by the fucose processing enzymes will be observed by monitoring GPDH activity. GPDH will act on DHAP with NADH to produce glycerol 3-phosphate and NAD^+, and the depletion of NADH can be measured at 339 nm at different time points. Thus, a reaction rate for fucose processing can be determined since NADH depletion is equivalent to NAD^+ production ultimately validating the predicted fucose processing activities.

The *E. coli* fucose processing enzymes were used as a positive control showing that the assay conditions allow for the formation of DHAP and also for GPDH activity. An enzyme assay containing EcFcsI, EcFcsK, and EcFcsA gave a reaction rate of 0.924 μM NAD^+ min^{-1}, while SpFcsI, SpFcsK, and SpFcsA yielded 0.487 μM NAD^+ min^{-1} (Figure 29). In contrast, there was no NAD^+ production observed in the absence of all or one of the fucose processing enzymes showing that all three enzymes are necessary for the reaction to proceed (Figure 29). The fucose processing enzymes from *E. coli* were also substituted with *S. pneumoniae* enzymes in a stepwise manner to determine the compatibility of these two fucose pathways. All combinations of *E. coli* and *S. pneumoniae* fucose processing enzymes produced NAD^+; however, some arrangements generated significantly different rates of reactions when compared to others (Figure 29). The most rapid formation of NAD^+ per minute occurred with the assays containing SpFcsI and EcFcsK yielding 13.5 μM NAD^+ min^{-1} for SpFcsI, EcFcsK, and EcFcsA and 10.8 μM NAD^+ min^{-1} for SpFcsI, EcFcsK, and SpFcsA. In contrast the assays containing EcFcsI and SpFcsK yielded the slowest formation of NAD^+ per minute, with 0.134 μM NAD^+ min^{-1} for EcFcsI, SpFcsK, and EcFcsA and 0.203 μM NAD^+ min^{-1} for EcFcsI, SpFcsK, and SpFcsA.

From this data, it appears that the *S. pneumoniae* fucose processing enzymes act in the equivalent manner to the *E. coli* enzymes, producing a final DHAP product. However, the reactions' rates differ depending on the combinations of *S. pneumoniae* and *E. coli* enzymes used. The highest rates are seen with reactions containing SpFcsI and EcFcsK. These observations imply that SpFcsI and EcFcsK have a greater activity than EcFcsI and SpFcsK, respectively. Interestingly though, the reactions containing EcFcsI and SpFcsK have significantly lower rates than the standard reactions suggesting that the low activity of EcFcsI offsets the high activity of EcFcsK and conversely the high activity of SpFcsI balances the low activity of SpFcsK so that the overall fucose processing of *S. pneumoniae* and *E. coli* have similar reaction rates. Also of note, there was no significant variation in reactions rates between fucose processing combinations that only differed in the source of FcsA, suggesting the FcsA enzymes from *S. pneumoniae* and *E. coli* have similar activities. Further structural information for the SpFcsI, SpFcsK, and SpFcsA was obtained to verify fucose processing functions and possibly distinguish discrepancies

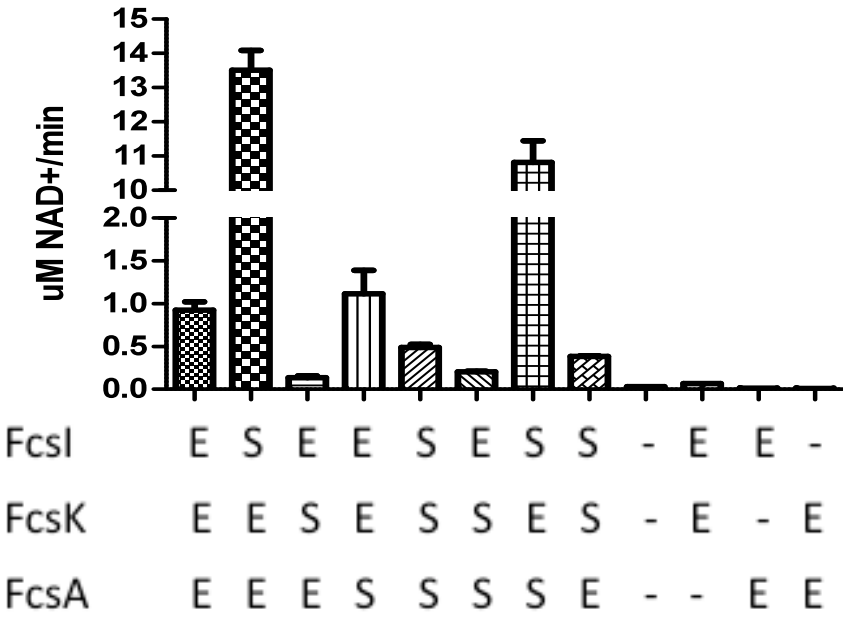


Figure 29: Multi-enzyme fucose processing

Fucose processing enzymes were assayed and NAD^+ formation monitored. E refers to the fucose processing enzyme from *E. coli* and S from *S. pneumoniae*.

between the activities of the *E. coli* and *S. pneumoniae* fucose processing enzymes by understanding the molecular basis of their actions.

SpFcsI quaternary organization and 3-dimensional structure

Belonging to the family of isomerases, FcsI is classified as a ketol isomerase and requires a Mn^{2+} or Co^{2+} cofactor to convert fucose to fuculose. The *E. coli* FcsI (EcFcsI) also has activity for D-arabinose, L-fucose lacking the 6-methyl group, producing D-ribulose (Oliver and Mortlock 1971) and thus, occasionally fucose isomerases have been called arabinose isomerases. One such example is the arabinose isomerase from *Bacillus pallidus* (BpAraI), which also has dual activity on arabinose and fucose (Oliver and Mortlock 1971; Chen, Zhu et al. 1987).

The 3-dimensional structures of EcFcsI and more recently BpAraI have both been solved in complex with L-fucitol, an alcohol derivative of fucose, allowing for identification of the enzyme's catalytic site (Seemann and Schulz 1997; Takeda, Yoshida et al. 2010). Unfortunately, diffraction quality crystals were not obtained for SpFcsI; however, since there are known structures of homologous FcsI enzymes, SpFcsI was modeled using a Protein Homology/analogY Recognition Engine V 2.0 (Phyre^2) (Kelley and Sternberg 2009). The secondary structure of SpFcsI can be broken up into 3 domains all of which consist of a β-structure surrounded by α-helices. Domains 1 and 2 are comprised of parallel β-sheets and domain 3 an anti-parallel β-barrel (Figure 30A), which is identical to other known fucose isomerase structures (Figure 30B).

EcFcsI (PDB code 1FUI) and BpAraI (PDB code PDB3A9T), which have a 60 and 69% amino acid identity to SpFcsI, respectively, are found as a hexamer, or more specifically a dimer of trimers. Thus, DLS was used to establish the molecular weight of SpFcsI, which in turn allows for an estimate of quaternary organization. Purification via size exclusion chromatography yielded one predominant peak at which two fractions were taken for DLS analysis (Figure 31A). From a hydrodynamic radius of 5.10 nm a molecular weight of 220 kDa was determined for the SpFcsI recombinant protein still containing a polyhistidine tag (Figure 31B). Thus, with an estimated molecular weight of 66 kDa (Gasteiger, Gattiker et al. 2003) SpFcsI is found as a trimer in these conditions, which is not consistent with the FcsI enzymes from *E. coli* and *B. pallidus*. Furthermore,

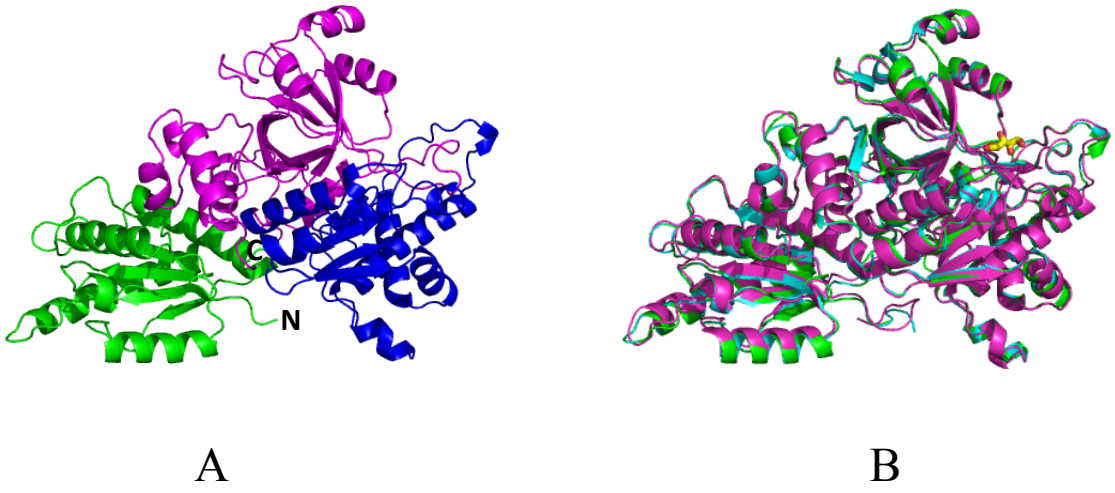


Figure 30: 3-dimensional homology model of SpFcsI

(A) Cartoon representation of SpFcsI showing domain 1 in green, domain 2 in blue and domain 3 in magenta. (B) Cartoon overlay of FcsI structures. SpFcsI is shown in green, EcFcsI in cyan, and BpAraI in magenta. Fucose is shown in a yellow stick representation with oxygen molecules in red.

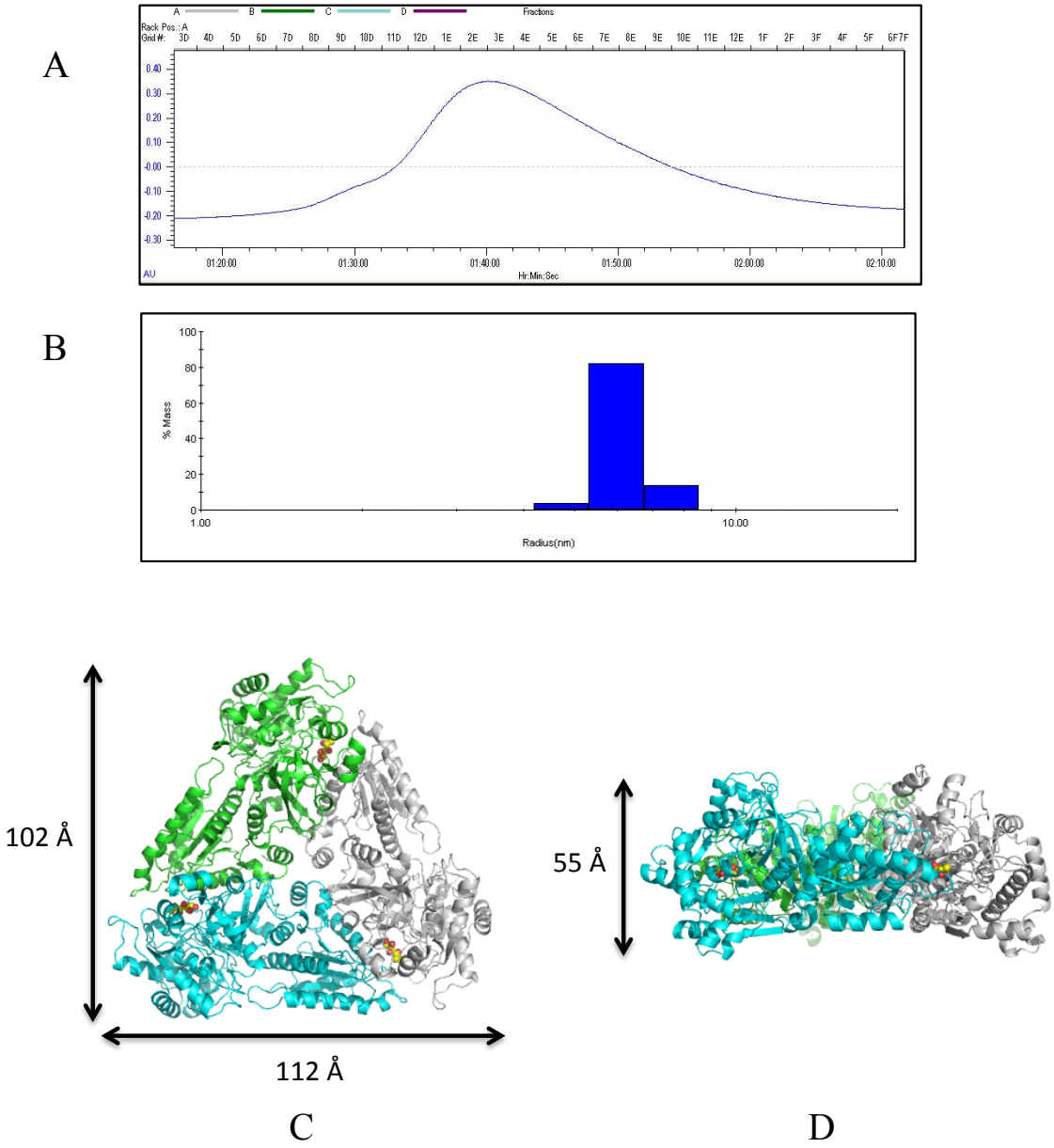


Figure 31: Quaternary organization of SpFcsI

(A) Size exclusion chromatography and (B) DLS profiles for SpFcsI with an intact polyhistidine tag. Top (C) and side (D) views of the hexamer organization of SpFcsI. Each monomer is shown in a different color and fucose molecules are yellow spheres with oxygen in red.

with the polyhistidine tag removed to ensure it was not affecting quaternary organization, a similar size exclusion chromatography profile was observed and a hydrodynamic radius of 5.07 nm yielding a calculated molecular weight of 203 kDa was determined by DLS (data not shown). The estimated molecular weight of SpFcsI without the polyhistidine tag is 61 kDa confirming a trimer organization. Lastly, DLS experiments were conducted with higher concentrations of SpFcsI (15-25 mg ml^{-1}) also gave molecular weights consistent with trimer organizations (data not shown) suggesting that higher concentrations of recombinant enzyme do not induce changes in quaternary organization.

Since other fucose isomerases are found to have a hexamer organization, or a dimer of trimers (Patrick and Lee 1969; Wallace, Eiserling et al. 1978; Seemann and Schulz 1997; Takeda, Yoshida et al. 2010), it is possible that SpFcsI forms a similar trimer as the homologue enzymes. Thus, SpFcsI was modeled as a trimer using one trimer unit from EcFcsI. The face of the SpFcsI trimer has dimensions of 102 Å and 112 Å (Figure 31C) while the height of the trimer is 55 Å (Figure 31D).

The catalytic residues between the fucose isomerases are conserved (Figure 31C and D) and are located between two monomers within the same trimer for EcFcsI and BpAraI. This highlights the importance of trimer formation for enzyme catalysis, providing further evidence that SpFcsI should have a similar trimer formation as homologous FcsI enzymes but may have lost the ability to dimerize forming a hexamer unit. The active site of SpFcsI is extremely similar to both EcFcsI and BpAraI with identical contributing residues, mostly from domains 2 and 3 of one monomer (Gln300, Glu335, Asp359, Arg361, Asn524, and His525) along with Arg16 from an adjacent monomer (Figure 32 and 33).

SpFcsI possesses the ability to functionally replace EcFcsI in the fucose processing enzyme reactions, strongly suggesting that SpFcsI is a bona fide fucose isomerase in the fucose metabolic pathway. SpFcsI appears to have greater activity than EcFcsI although the active site architecture remains almost identical between the SpFcsI model structure and the EcFcsI crystal structure, providing no rationale behind this observation. Since SpFcsI was modeled based on available FcsI structures, which includes EcFcsI,

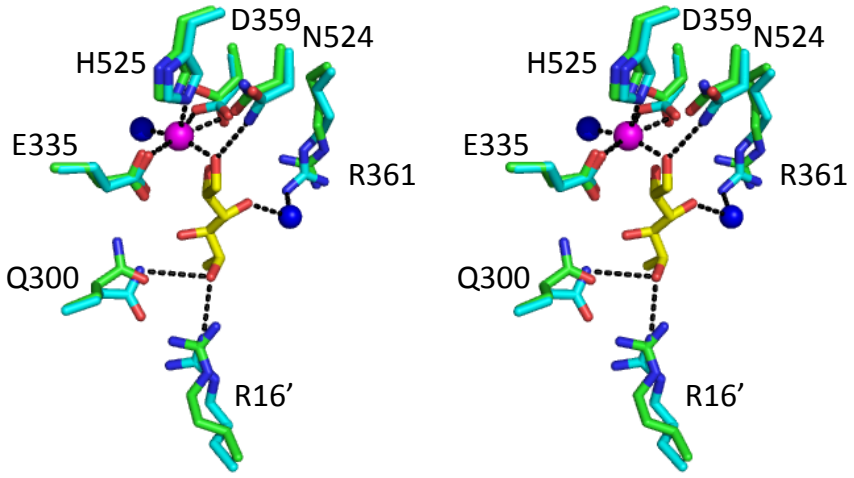


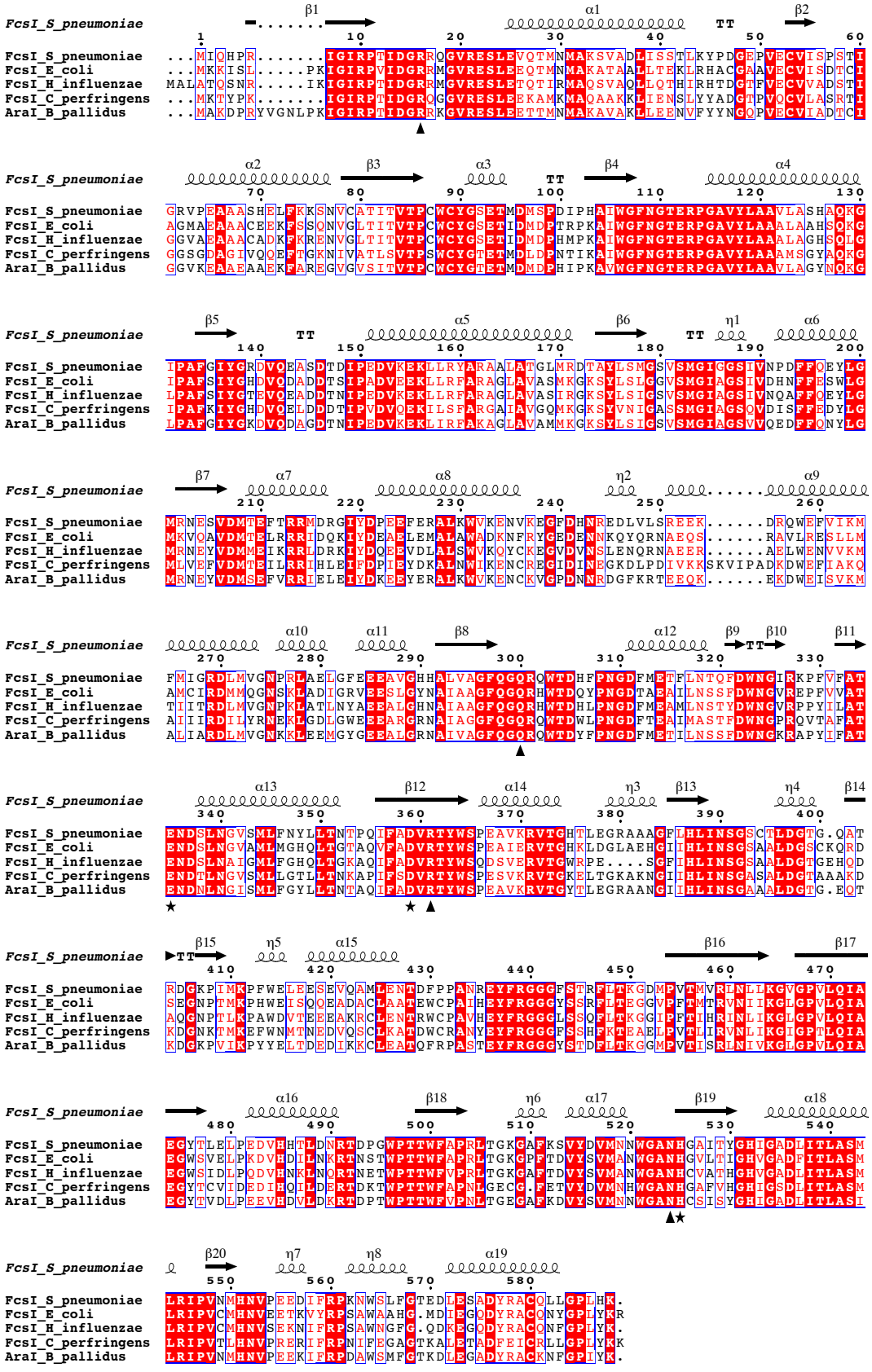
Figure 32: The active site architecture of SpFcsI

Stereo image of an overlay of the active site of SpFcsI, in green, and EcFcsI, in cyan. Fucose is shown in yellow, oxygens in red, and nitrogens in blue. Water molecules are blue spheres while the Mn^{2+} ion is a magenta sphere.

Figure 33: FcsI comparison

(On next page). Alignment of FcsI enzymes from different organisms. Black stars and triangles represent catalytic and substrate interacting residues, respectively.

Figure 33 (Legend on previous page)



differences in the surrounding sequence could alter the active site architecture, which cannot be predicted with *in silico* modeling. Alternately, this phenomenon could be an artefact of the assay conditions and thus further biochemical and structural studies are required to determine the molecular basis of SpFcsI activity.

SpFcsK structure

Fucuolse kinase (FcsK) is the second step in fucose metabolism converting the product of FcsI, fuculose, to the substrate of FcsA, fuculose 1-phosphate (Figure 27). It is found in the hexokinase-hsp70-actin superfamily and although this family has diverse overall sequence similarity, the structural similarity is remarkably high (Hurley 1996).

Although the 3-dimensional structure of a related rhamnulose kinase from *E. coli* (EcRhuK) has been solved (Grueninger and Schulz 2006), there is currently no structural information available for a fuculose kinase.

The apo-structure of FcsK from *S. pneumoniae* (SpFcsK) was solved to 2.2 Å by molecular replacement using the EcRhuK structure as a search model (PDB code 2CGK) (Grueninger and Schulz 2006). SpFcsK has an r.m.s.d. of 1.719 Å to EcRhuK and is found as a monomer containing a cleft, approximately 27 Å deep, that separates two large domains both consisting of characteristic five-stranded β-sheets (Figure 34A).

A notable feature of enzymes in the hexokinase-hsp70-actin superfamily is the existence of opened and closed conformations. The enzymes undergo an induced fit conformational change upon substrate binding where one domain rotates with respect to the other domain approximately 10° around the hinge region, altering the enzyme from an open to a closed conformation (Bork, Sander et al. 1992; Grueninger and Schulz 2006). This attribute is especially prominent for hexokinases, like RhuK, where the addition of rhamnulose or fructose and adenosine di-phospate (ADP) corresponds to a domain rotation along with specific active site residue displacements (Grueninger and Schulz 2006; Grueninger and Schulz 2007). The sugar molecule sits deep in the cleft, undergoing only slight shifts in few of the contributing residues. However, the ATP-binding site, at the top of the cleft, undergoes the majority of the conformational changes with substantial movement from certain α-helices.

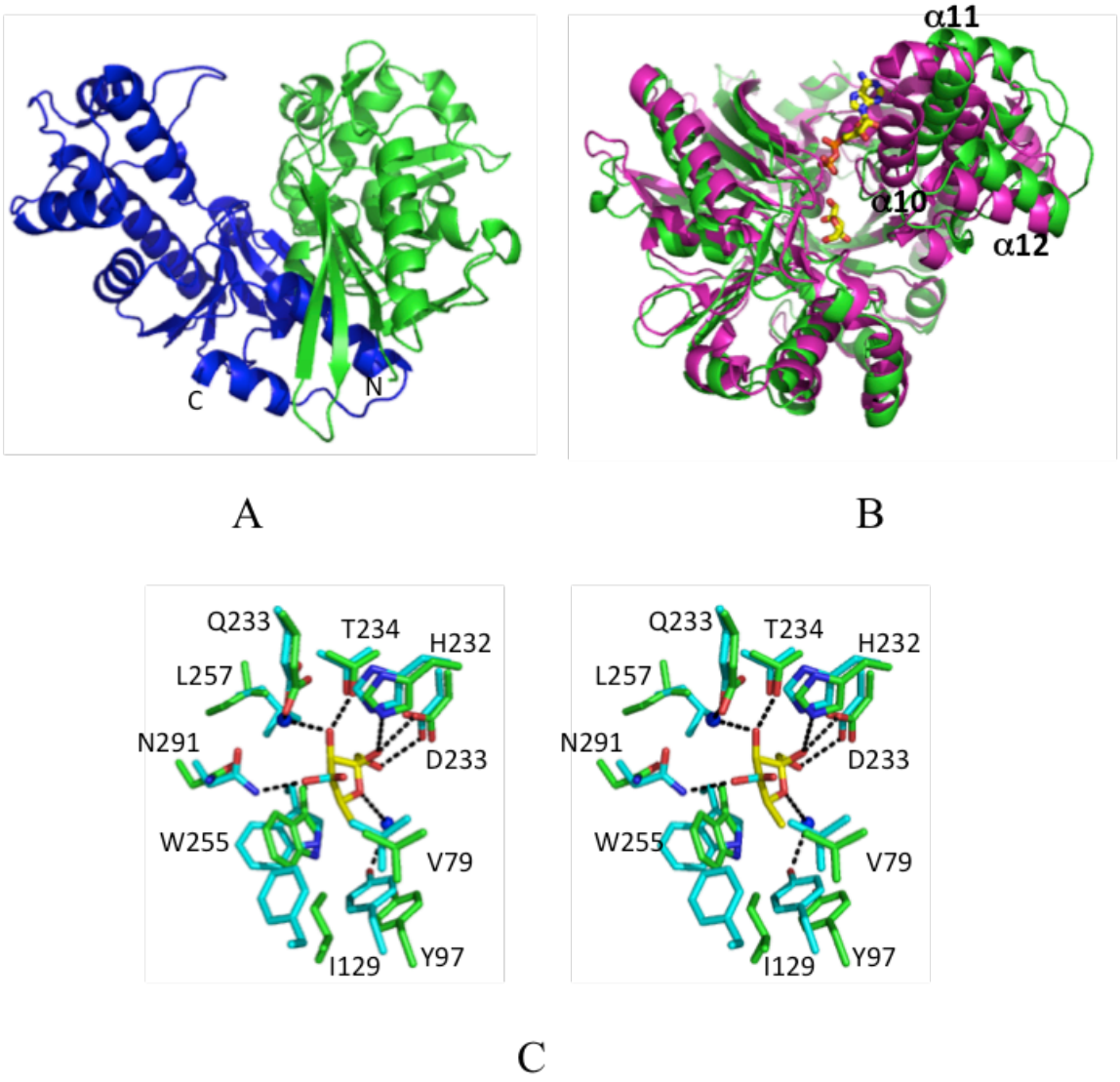


Figure 34: Overall structure and active site architecture of SpFcsK

(A) Cartoon representation of SpFcsK showing domain 1 in green and domain 2 in blue. (B) Cartoon overlay, showing the open conformation of SpFcsK in green and the closed conformation of EcRhuK in magenta. The ADP and fructose substrates are yellow stick representations with oxygen in red, nitrogen in blue, and phosphorus in orange. (C) Stereo image of an overlay of the sugar binding site of SpFcsK, in green, and EcRhuK, in cyan. Fucose is shown in yellow with the underlying rhamnose in cyan and water molecules are blue spheres.

SpFcsK was crystallized in the absence of substrate and thus is in the open conformation, which is clearly seen in an overlay with the closed conformation of EcRhuK found in the complex structure (Figure 34B). There is a distinct shift in the EcRhuK α-helices 10, 11, and 12 moving them closer to the substrate to form the closed conformation. Since the sugar-binding site of EcRhuK shows no major displacements between the open and closed conformations, fuculose was modeled into the active site of SpFcsK (Figure 34C). The residues of SpFcsK that would interact with fuculose (His232, Asp233, and Thr234) are identical to the rhamnulose interacting residues of EcRhuK. One main difference is the contribution from Asn296 from EcRhuK, which H-bonds with the equitorial hydroxyl group from C4 in fructose. However, the fuculose C4 hydroxyl is in the axial position and thus does not require contribution from the equivalent Asn291 from SpFcsK, although, there does not seem to be any surrounding residues that could interact with an axial C4 hydroxyl group. There are also non-polar contacts of SpFcsK with fuculose, Val79, Tyr97, Ile129, Ile255, and Leu257, which are all identical to EcRhuK with the exception of Ile129, which is Phe134 in RhuK.

Not surprising is the similarity in sugar binding site residues between SpFcsK and its homologue RhuK as their primary substrates, fuculose and rhamnulose, respectively, only differ by the configuration at the C4 position. Both FcsK and RhuK, from *E. coli*, has been shown to have activity on other 2-ketoses, albeit with less activity than their primary substrates (Heath and Ghalambor 1962; Grueninger and Schulz 2007) indicating that the sugar binding site is not particularly specific for fuculose or rhamnulose, respectively. It was postulated that for EcRhuK this lack of specificity is due to an abundance of water molecules covering the sugar binding site (Grueninger and Schulz 2006); however, the configuration at positions C2 and C3 prove the most important for EcRhuK specificity (Grueninger and Schulz 2007). EcRhuK has 56% activity for L-fuculose (Chiu and Feingold 1964); however, EcFcsK appears inactive on L-rhamulose (Heath and Ghalambor 1962) despite the homologous SpFcsK having an extremely similar active site architecture to EcRhuK. When taking a closer look at the C4 position of fuculose, in the SpFcsK structure there doesn't appear to be any surrounding residues that could confer specificity to fucuolse over rhamnulose. However, since SpFcsK is predicted to undergo a conformational change upon substrate binding, it is possible that

there are unique local induced fit changes that are absent with EcRhuK, which could lead to fuculose specificity. Further structural studies on SpFcsK could provide insight into the difference of specificity between these two enzymes that could not be explained from the apo-SpFcsK structure.

The ATP binding site is located at the top of the cleft and is involved in the majority of the conformational changes that occur upon substrate binding and thus ADP from the RhuK complex structure was not modeled into the apo-SpFcsK structure. However, through an alignment, SpFcsK does not seem to have much similarity in the ATP binding site of RhuK (Figure 35). Nonetheless, SpFcsK contains the five common sequence motifs present in the active sites of enzymes from hexokinase-hsp70-actin superfamily (Bork, Sander et al. 1992).

SpFcsA quaternary organization and 3-dimensional structure

FcsA is a class II aldolase and requires a divalent cation, such as zinc, to stabilize reaction intermediates, which has clearly been shown by *E. coli* FcsA (EcFcsA) structural studies (Dreyer and Schulz 1996; Dreyer and Schulz 1996; Joerger, Gosse et al. 2000; Joerger, Mueller-Dieckmann et al. 2000). Further structural information for FcsA from *Aquifex aeolicus* (AaFcsA), *Bacteroides thetaiotaomicron* (BtFcsA), and *Thermus thermophilus* (TtFcsA) has highlighted the molecular basis of FcsA activity (Jeyakanthan, Taka et al. 2005).

SpFcsA has a 41, 21, 20, 21%, amino acid identity to EcFcsA, AaFcsA, BtFcsA, and TtFcsA, respectively. Since SpFcsA did not yield diffraction quality crystals and there are FcsA homologue structures available, a model of SpFcsA was generated with Phyre^2 (Kelley and Sternberg 2009). SpFcsA consists of a 6-stranded β-sheet flanked by two α-helices to one side and four α-helices on the other side (Figure 36A), which is identical to other known FcsA structures (Figure 36B).

Since homologous FcsA enzymes are found as a multimer, the molecular weight and subsequent quaternary organization for SpFcsA was determined using DLS analysis. Two fractions found in the single peak produced from size exclusion chromatography were used for DLS analysis (Figure 37A and B). A hydrodynamic radius of 4.90 nm gives SpFcsA a molecular weight of 108 kDa in these conditions. Thus, with an estimated

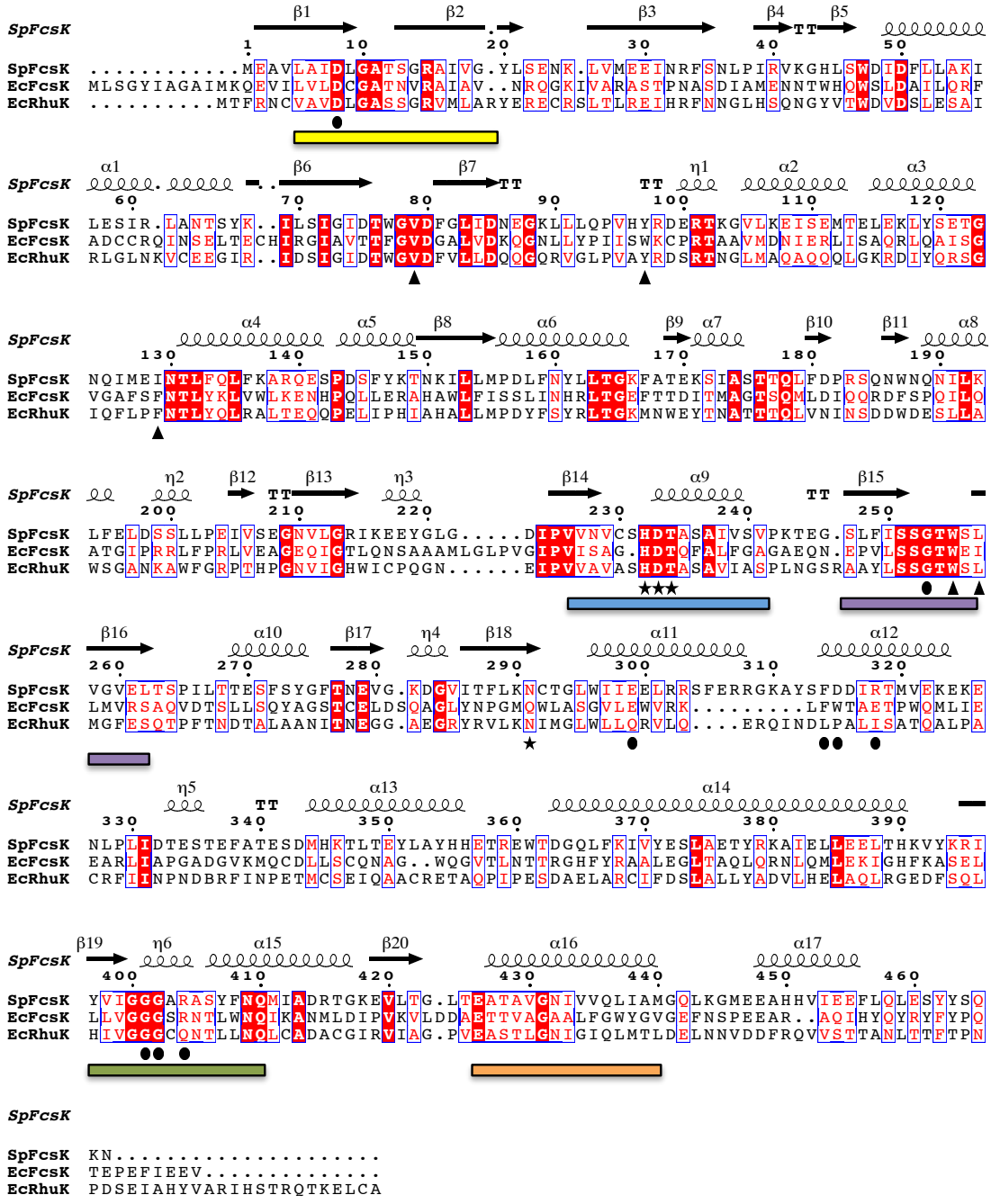
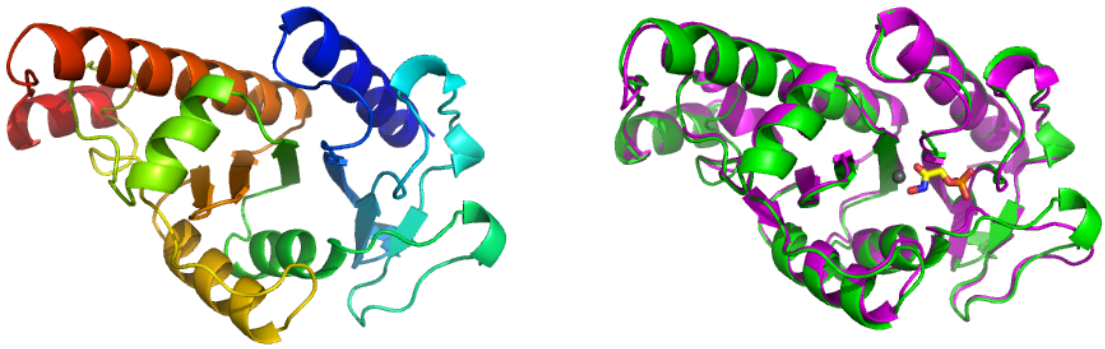


Figure 35: FcsK comparison

Alignment of SpFcsK, EcFcsK and EcRhuK enzymes. Black stars and triangles represent catalytic and sugar interacting residues, respectively while the black circles show the ATP-interacting residues. Coloured boxes represent the five common sequence motifs present in the active sites of enzymes from hexokinase-hsp70-actin superfamily (Bork, Sander et al. 1992).



A B

Figure 36: 3-dimensional homology model of SpFcsA

(A) Cartoon representation of SpFcsA in rainbow color. (B) Cartoon overlay of SpFcsA, in green, and EcFcsA, in magenta. PGH is shown as a yellow stick representation with oxygen molecules in red, nitrogen in blue, and phosphorus in orange.

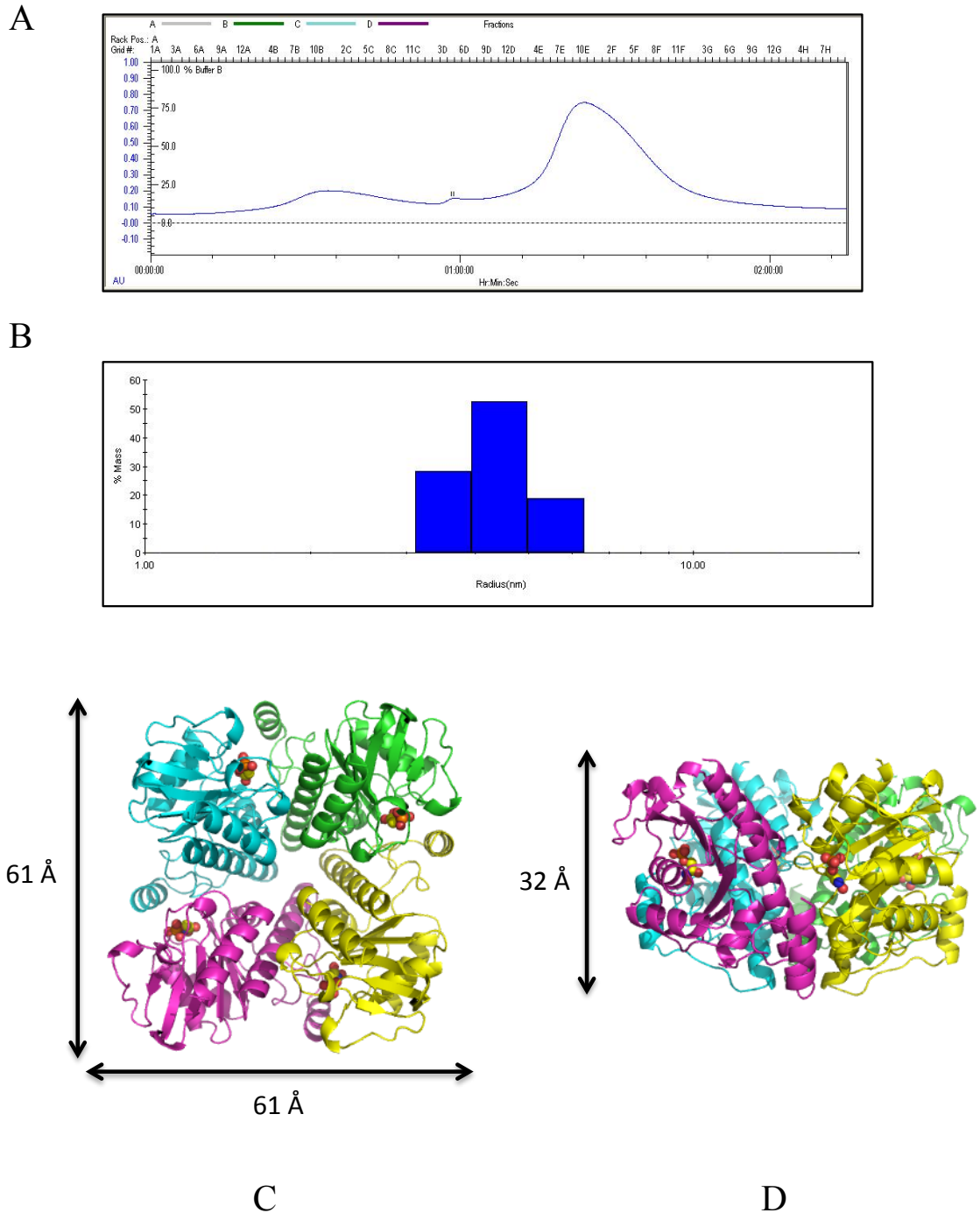


Figure 37: Quaternary organization of SpFcsA

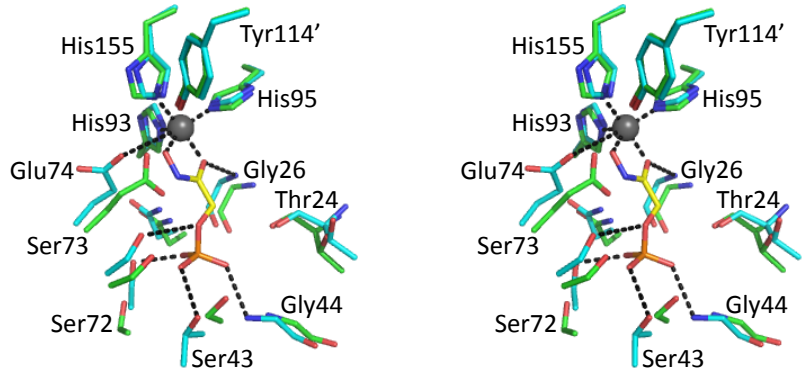
(A) Size exclusion chromatography and (B) DLS profiles for SpFcsA. Top (C) and side (D) views of the tetramer organization of SpFcsA. Monomers are shown in different colors and PGH as spheres.

molecular weight of 26 kDa (Gasteiger, Gattiker et al. 2003) SpFcsA appears to be found as a tetramer in solution, which is consistent with FcsA homologues (Dreyer and Schulz 1996; Kroemer, Merkel et al. 2003; Jeyakanthan, Taka et al. 2005). Using the EcFcsA structure (PDB code 3FUA), the tetramer of SpFcsA was modeled, forming a cube-like structure consisting of a square face with edges of 61 Å (Figure 37C) and a height of 32 Å (Figure 37D).

Extensive structural data has been collected on EcFcsA in complex with phosphoglycolohydroxamate (PGH), a transition state analog of DHAP, giving insight into critical catalytic residues and the mechanism for catalysis (Dreyer and Schulz 1996; Joerger, Gosse et al. 2000; Joerger, Mueller-Dieckmann et al. 2000). PGH was modeled into the active site of SpFcsA, revealing a very similar environment as EcFcsA (Figure 38A and B). The catalytic residues are all identical to EcFcsA, where one monomer contributes the majority of the residues (Glu74, His93, His95, His155) with Tyr114 coming from an adjacent monomer. The same monomer that contributes most of the catalytic residues also contributes all of residues making up the phosphate pocket (Thr24, Gly26, Asn27, Ser43, Gly44, Ser72, and Ser73). There is, however, one minor substitution in the phosphate pocket; Ser43 in SpFcsA has replaced Thr43 in EcFcsA. The main structural differences between the active sites of SpFcsA and EcFcsA are the orientation of the side chain of Glu74 and a slight shift in resides found in the phosphate pocket. These differences, however, could be an artefact of the *in silico* modeling or could be a result of the induced-fit nature of FcsA when it binds its substrate, which has clearly been shown in EcFcsA (Dreyer and Schulz 1996).

SpFcsA has high sequence identity to other known FcsA enzymes, which allows an *in silico* model to be built. Along with an identical active site architecture to EcFcsA, predicted from the modeled structure, SpFcsA also posses the ability to maintain DHAP production at a similar reaction rate when substituted for EcFcsA in the *E. coli* fucose processing reactions. All together, SpFcsA appears to have the ability to function as the last step in fucose metabolism as a fuculose 1-phosphate aldolase.

A



B

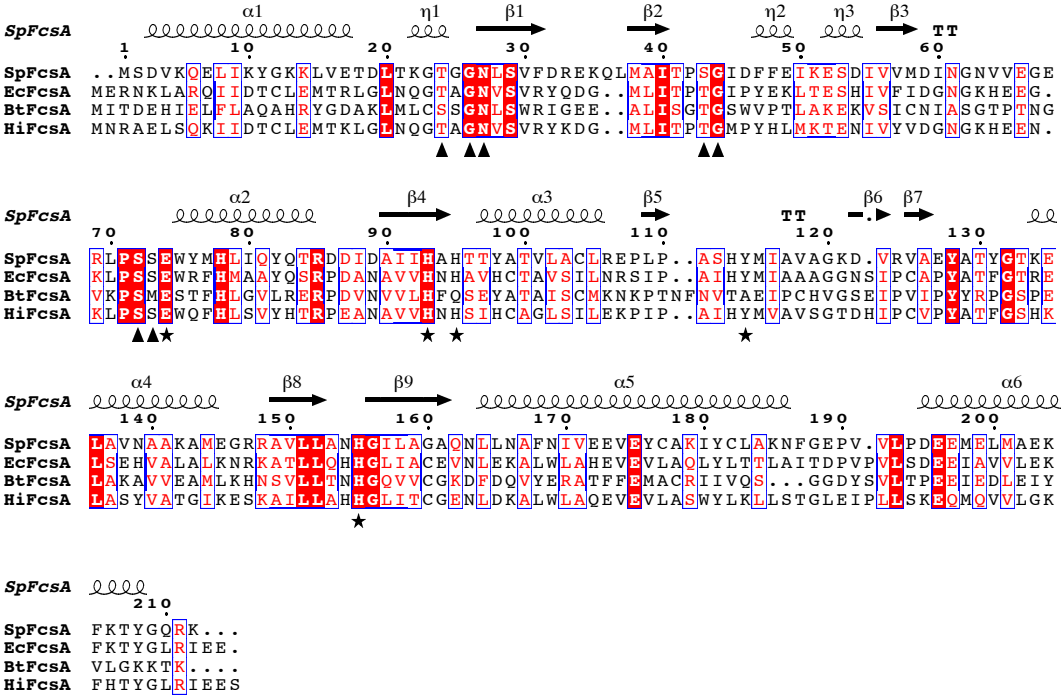


Figure 38: SpFcsA active site architecture

(A) Stereo image of an overlay of the active site of SpFcsA, in green, with EcFcsA, in cyan. PGH is shown in yellow, with oxygen molecules in red, nitrogen in blue, and phosphorus in orange. The Zn^{2+} ion is shown as a grey sphere. (B) Alignment of FcsA enzymes from different organisms. Black stars and triangles represent catalytic and substrate interacting residues, respectively.

Regulation of the fucose utilization operon

Regulation of the fucose operon occurs by a specific fucose regulator found in the operon. In *S. pneumoniae*, the fucose operon is tightly regulated and is repressed by glucose or sucrose and induced by fucose (Chan, O'Dwyer et al. 2003). To determine whether one of the carbohydrate intermediates of the fucose processing pathway acts as the inducer molecule, *S. pneumoniae* strains deficient in each of the fucose processing components were created. Subsequent Western blots probing for FcsI and GH98 were conducted on Δ*gh98*, Δ*fcsI*, Δ*fcsK*, and Δ*fcsA* to observe operon expression.

When fucose was added to the growth media, both FcsI and GH98 were detected for the wild-type *S. pneumoniae* TIGR4 strain (Figure 39). The Δ*gh98* strain showed no GH98 production but did show FcsI production showing that this gene deletion method does not produce any downstream effects since *fcsI* is the last gene in the operon (Figure 39). Both Δ*fcsI* and Δ*fcsK* strains did not produce FcsI or GH98 while Δ*fcsA* produced both FcsI and GH98. This suggests that the product of FcsK, fuculose 1-phosphate, acts as an inducer molecule for fucose operon expression, which is consistent with what was found for *E. coli* fucose metabolism (Bartkus and Mortlock 1986).

In situ activity of GH98 enzymes

Fucose utilization has been identified as a critical component for pneumococcal virulence and disease in the TIGR4 strain (Hava and Camilli 2002; Embry, Hinojosa et al. 2007). However, the bacterium is unable to use fucose as a carbon source (Chan, O'Dwyer et al. 2003) even though it possesses active fucose processing enzymes. Since the regulator of the fucose operon is fuculose 1-phosphate, the fucose processing enzymes are crucial for induction of the operon, but that may be the extent of their role in pathogenesis. Thus, the glycan processing activities of the extracellular glycoside hydrolases may be the critical components for virulence. Previous knowledge of the extracellular Sp4GH98 and Sp3GH98 enzymes (from the type I and type II operons, respectively) has shown *in vitro* activity on differing Type II histo-blood group antigens (Higgins, Abbott et al. 2009; Higgins, Ficko-Blean et al. 2011); however, it has not yet been established if these proteins have activity on cell-surface presented substrates.

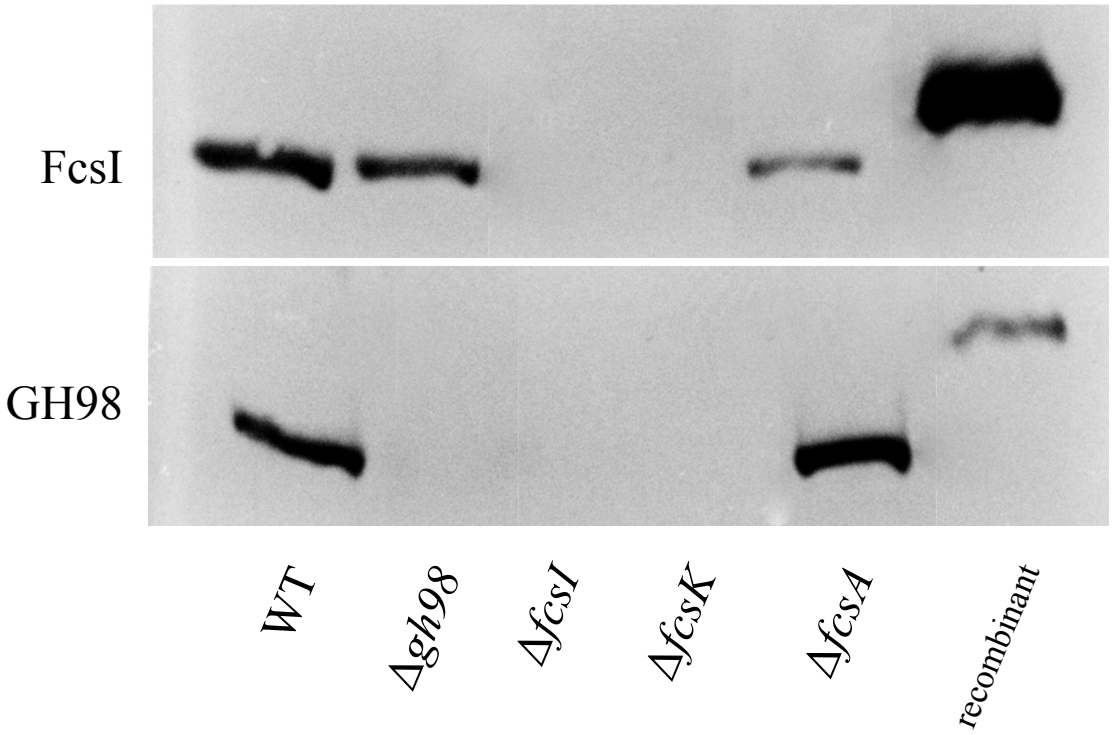


Figure 39: Regulation of fucose operon expression from *S. pneumoniae*

Western blot of fractionation experiments with *S. pneumoniae* TIGR4 and strains deficient in specific fucose utilization genes. The blot probed for FcsI are the cell lysates since FcsI is intracellular while the blot probed for Sp4GH98 are the supernatants since GH98 is secreted (Section 3.4 or (Higgins, Ficko-Blean et al. 2011)).

Given that the environmental niche targeted by *S. pneumoniae* is the lung we approached this question by using a type II alveolar cell line (A549 cells), which is commonly used to assess pneumococcal adherence and invasiveness. Thus, it was presumed that treatment of the cell line with one or both enzymes should remove the carbohydrate antigen from the cells in a manner consistent with specificity of these enzymes. Cells not treated with enzymes and probed with the anti-Lewis^Y antibody had small foci of fluorescence scattered over the surface of the cells (Figure 40A and 40I). Notably, cells treated with Sp4GH98CM and probed with the anti-Lewis^Y antibody had a significant reduction in the number and intensity of the immunoreactive foci, consistent with the Sp4GH98CM-catalyzed destruction of Lewis^Y antigen on the cells (Figure 40B and 40I). In contrast, when cells were treated with Sp3GH98CM and probed with the anti-Lewis^Y antibody the fluorescent foci were smaller and more dispersed, but the mean fluorescence was not significantly reduced (Figure 40C and I). When cells were digested with both enzymes, fluorescence after detection with the anti-Lewis^Y antibody was again reduced significantly but not more so than when treated with Sp4GH98CM alone (Figure 40D and I). In a parallel set of experiments, in which cells were probed after enzyme treatment with an antibody that recognizes both the A- and B-blood group antigens, only pre-treatment of the cells with Sp3GH98CM alone or with both enzymes resulted in significant decreases in fluorescence intensity (Figures 40E-H and I). The processing of cell surface A- and B-blood group antigens by Sp3GH98 is also consistent with studies showing that the *C. perfringens* E-ABase processes these antigens on erythrocytes (Anderson, Ashida et al. 2005). These results clearly indicate that both *S. pneumoniae* Sp4GH98CM and Sp3GH98CM are capable of removing carbohydrate antigens from a model lung cell-line and with activity in keeping with their *in vitro* specificity.

The role of Fucose Utilization in pneumococcal internalization to epithelial cells

Since the GH98 enzymes from both types of *S. pneumoniae* fucose utilization operon act on histo-blood group antigens presented on the surface of host lung epithelial cells, it was hypothesized that they may play a role in adhesion and internalization to the model lung cell-line. Thus, initial internalization assays were performed with an *S. pneumoniae*

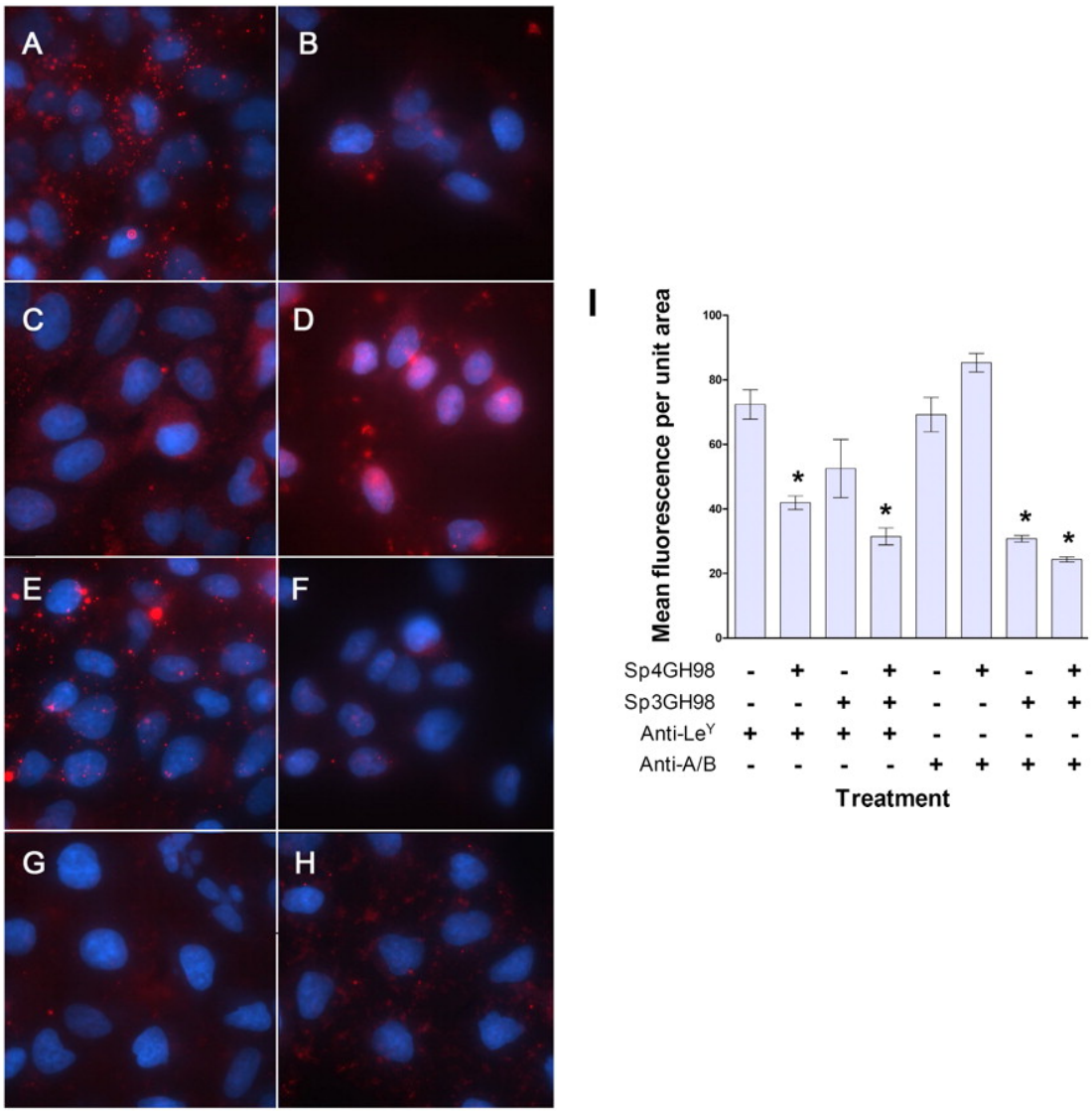


Figure 40: GH98 enzymes remove carbohydrate antigens from a model cell-line

Epifluorescent images of A549 cells treated with Sp4GH98 and Sp3GH98 enzymes then prepared with antibodies to the Lewis^{Y} antigen or the A/B antigens (red) and counter stained with DAPI (blue). A – D: cells were probed with anti-Lewis^{Y} antibody after no treatment (A); Sp4GH98 treatment (B); Sp3GH98 treatment (C) or treatment with both enzymes (D). E-H: cells were probed with anti-A/B antibody after no treatment (E); Sp4GH98 treatment (F); Sp3GH98 treatment (G) or treatment with both enzymes (H). Bar = 5 μm. (I) Quantification of fluorescence of A549 cells treated as in panels A-H

(treatment is indicated below the graph). Bars represent the mean (± standard error of the mean) fluorescence per unit area for a sampling of 5 randomly selected regions of cytoplasm. Bars marked with * are significantly different from cells not treated with an enzyme as judged by one-way ANOVA.

TIGR4 strain and a strain deficient in GH98. It has been shown that the amount of capsular polysaccharide can influence the bacterium's ability to adhere and invade host cells (Hammerschmidt, Wolff et al. 2005). Since there is considerable intrastrain capsule variation within an *S. pneumoniae* population, these experiments were conducted with strains lacking capsular polysaccharide to assure that any phenotype differences seen between the wild-type and Δ*gh98* are due to the absence of *gh98* and not potential capsule variation.

When quantifying bacteria inside of host lung epithelial cells, there was a significant decrease, of about 2/3, in the amount of bacteria that were internalized for the Δ*gh98* strain when compared to the wild-type (p value = 0.0130) (Figure 41). This could mean one of two things; either the bacteria were not entering the host cell or they were not surviving in the host cell. Since in TIGR4, GH98 cleaves the Lewis^{Y} antigen from the surface of host epithelial cells and blood group antigens are not found inside host cells, it seems unlikely that GH98 has an intracellular survival function. Thus, it is predicted that GH98 somehow aids in the internalization of pneumococci without playing a role in adhesion to host cells. Clearly, this is a preliminary study and more extensive experiments are required to validate this hypothesis and gain a better understanding of the mechanism by which *S. pneumoniae* uses enzymes from the fucose utilization operon for internalization by a host cell.

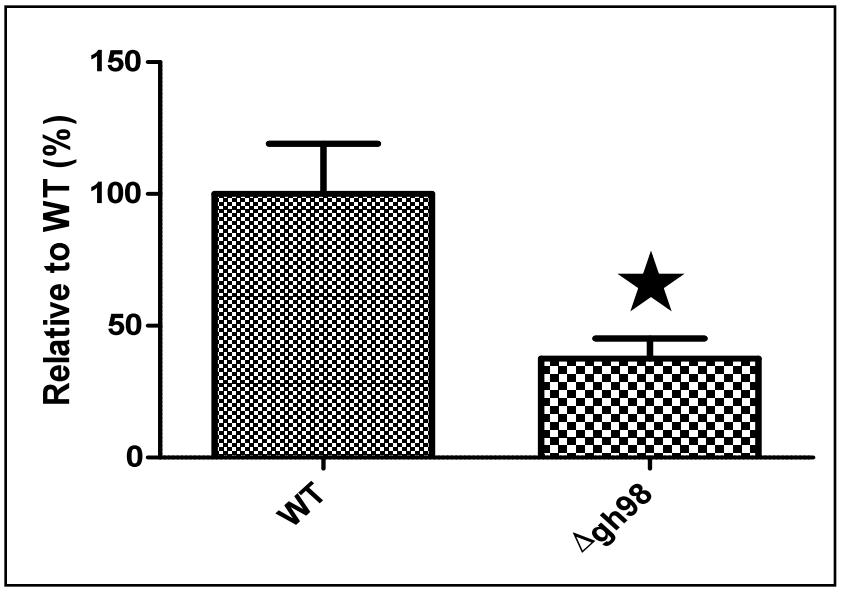


Figure 41: Δ*gh98* has reduced host internalization compared to WT *S. pneumoniae*

Internalization assay for *S. pneumoniae* unencapsulated TIGR4 strain (Δ*cps*) which is referred to as WT and an unencapsulated TIGR4 strain lacking GH98 (Δ*cps*Δ*gh98*) referred to as Δ*gh98*. The star represents a significant difference between host internalization of the WT and Δ*gh98* strains as judge by an unpaired *t*-test (p-value = 0.0130).

Chapter 6: Discussion

The exploitation of host glycans is a recurring theme in the establishment and maintenance of infections caused by microbial pathogens. A number of bacterial and viral pathogens are known to utilize human glycans as receptors for adherence as a precursor to establishing infection(Bergsten, Samuelsson et al. 2004; Odenbreit 2005; Cao, Lou et al. 2007). Many pathogens can also degrade or process host glycans as a part of the infection and/or dissemination process. For example, glycan processing and destruction is thought to be a key process in the severity and rapidity of disease onset in infections caused by *C. perfringens*. In this organism, a neuraminidase that removes terminal sialic acid residues potentiates the lethal cytolytic activity of the α-toxin (phospholipase) facilitating the massive tissue destruction that can be caused by this bacterium (Flores-Diaz, Alape-Giron et al. 2005). Perhaps the most prominent example of neuraminidase activity in a pathogen is that of the Influenza virus neuraminidase, whose activity is critical to the dissemination of the virus in the respiratory tract. Furthermore, this neuraminidase has been targeted by a number of inhibitors used to reduce the severity of Influenza virus infections (von Itzstein, Wu et al. 1993; von Itzstein 2008; Hayden 2009). In fact, the action of the Influenza virus neuraminidase has been associated with the lethal synergy between the virus and *S. pneumoniae* (McCullers 2006) and has been discussed in Section 1.2.4.

S. pneumoniae clearly relies on its capacity to process and/or degrade glycans during its infection process (described in Section 1.5.1) (Hava and Camilli 2002; Burnaugh, Frantz et al. 2008; Lammerts van Bueren, Ficko-Blean et al. 2011; Limoli, Sladek et al. 2011). The research presented in this dissertation has uncovered a novel process for pneumococcal glycan degradation that is essential component for disease. *S. pneumoniae* has adapted its fucose utilization operon to harvest and process fucosylated glycans from histo-blood group antigens. Potential roles and implications for fucose utilization in pneumococcal pathogenesis are discussed below.

Possible mechanisms for fucose utilization in pneumococcal pathogenesis.

It has been documented in two separate studies that fucose utilization is important for pneumococcal pathogenesis (Hava and Camilli 2002; Embry, Hinojosa et al. 2007); however the specific mechanism for virulence is yet to be determined. Complex carbohydrate catabolism has been linked to three aspects of pathogenesis in *Streptococci*: deglycosylation of host immune proteins, nutrient acquisition, and bacterial adhesion and host internalization (Shelburne, Davenport et al. 2008). It is unlikely that pneumococcal fucose utilization is involved in deglycosylation of immune proteins since histo-blood group antigens, which initiate the fucose pathway, are not commonly found on immune proteins. A possible role in nutrient acquisition is also unlikely since *S. pneumoniae* cannot use fucose as a carbon source; however, this option will be discussed below. Through internalization assays, we have shown a promising role for fucose utilization in host internalization of pneumococci (Chapter 5), although the mechanism is unknown.

Through signature-tagged mutagenesis studies, the individual GH98, EIIA, EIIC and FcsK components were identified as virulence factors suggesting that the entire operon may be involved in pathogenesis. However, expression of the fucose operon is tightly regulated and is induced by fucose, more specifically the product of FcsK fuculose 1-phosphate (Chapter 5 and (Chan, O'Dwyer et al. 2003)). As a result, the components of the fucose pathway necessary to produce fuculose 1-phosphate should be required for pathogenesis. Although, these processing enzymes may only be needed to induce expression of the extracellular GH enzymes, whose activities may be responsible for host internalization.

There are two types of fucose operons in *S. pneumoniae* that are initiated by two different GH98 enzymes that act on varying histo-blood group antigens presented on host cells. The type 1 operon contains the Sp4GH98 that cleaves the type 2 Lewis^{Y} antigen while the type 2 operon consists of Sp3GH98 that depolymerises the type 2 A/B blood group antigens. Cleavage of these antigens results in different carbohydrate structures remaining on the host cell surface; Sp4GH98 generates a terminal Fucα1-3GlcNAcβ1-3Gal on the host cell surface while Sp3GH98 creates GlcNAcβ1-3Gal (Figure 42). Interestingly, the addition of GlcNAcβ1-3Gal to an adhesion assay inhibits pneumococcal adhesion to epithelial cells in a concentration dependent manner and is

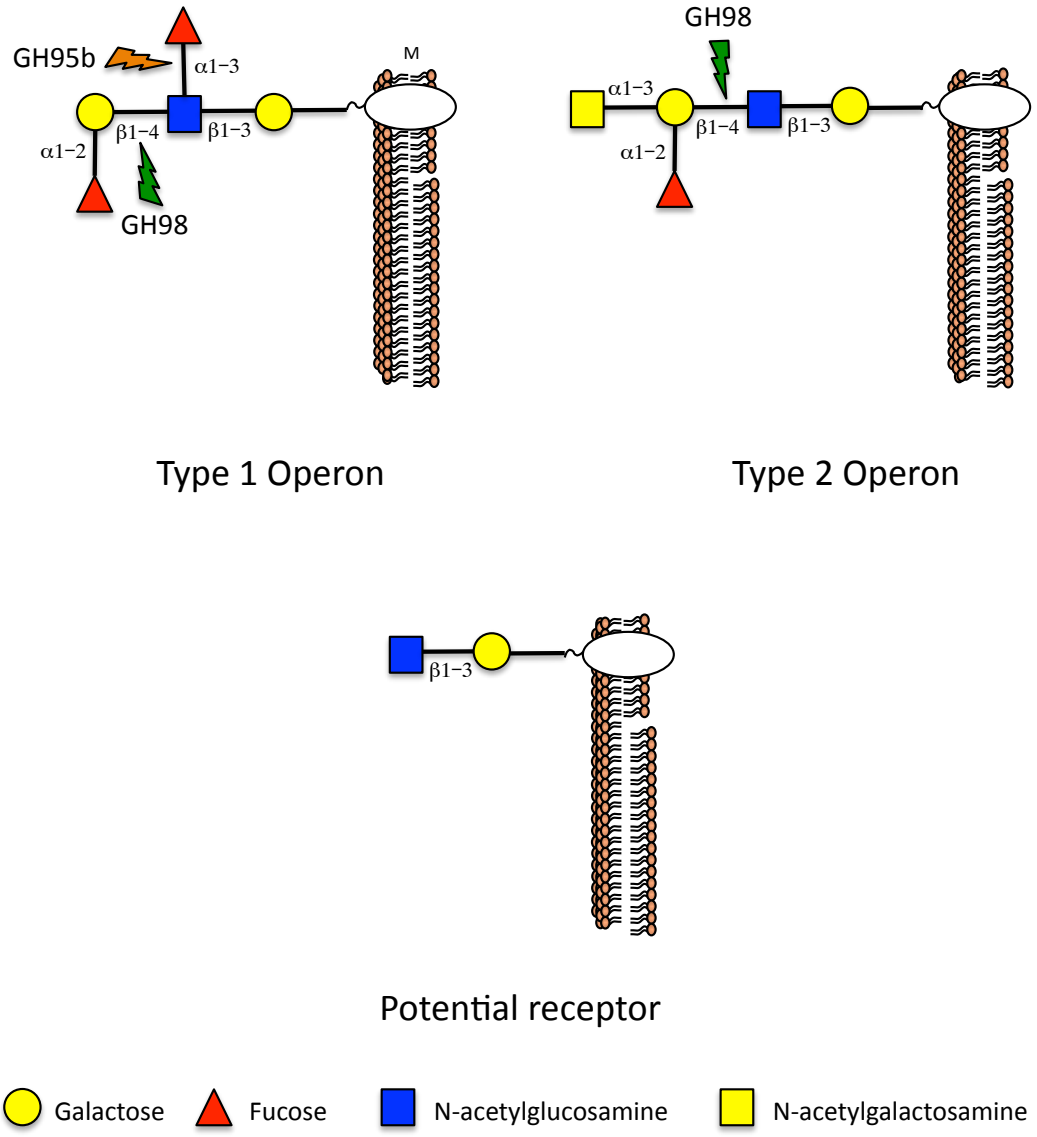


Figure 42: Schematic of glycan degradation for fucose utilization in *S. pneumoniae*

Both GH98 enzymes have been shown to act on the Galβ1-4GlcNAc linkage of their respective histo-blood group antigens. The GH95 activity is predicted to remove the terminal fucose reside for the type 1 operon to uncover a potential host receptor.

consequently thought to act as a host receptor for pneumococcal adhesion (Andersson, Dahmen et al. 1983). Therefore, cleavage of the A/B blood group antigens by Sp3GH98 would uncover the GlcNAcβ1-3Gal receptor for adherence to host cells for subsequent internalization. For pneumococcal strains with the type 1 fucose operon, an additional enzyme would be required to remove the terminal fucose molecule to generate the GlcNAcβ1-3Gal receptor. The gene SP_1654, which is not in the *fcs* locus, encodes for a putative α-fucosidase (GH95b) that has the potential to remove the terminal fucose residue from Fucα1-3GlcNAcβ1-3Gal resulting in a GlcNAcβ1-3Gal receptor (Figure 42).

Notably, there is a putative α-fucosidase enzyme (GH95a) present in the fucose operon that is predicted to act on the H-disaccharide in the bacterium cytosol releasing fucose for further processing (Figure 12). Since both GH95a and GH95b do not contain a secretion signal, either enzyme has the potential to be non-classically secreted for the removal of the terminal fucose histo-blood group antigen presented on host cells. Unfortunately, neither of these enzymes have been enzymatically characterized or have established 3-dimensional structures. When structural models of the GH95a and GH95b were created using Phyre^2 (Kelley and Sternberg 2009), the putative active sites were identified using alignments with the 2'-fucosyllatose complex structure of the homologous GH95 from *B. bifidum* (BbGH95) (Nagae, Tsuchiya et al. 2007). The GH95a model shows a closed active site with a loop region that appears to block additional subsites allowing for an extended substrate in the active site (Figure 43A and 43C). Thus, GH95a could not accommodate the extra sugars involved in attaching the substrate to the host cell surface and likely has its original predicted function to cleave a soluble H-disaccharide releasing fucose for further processing. Alternatively, the GH95b model appears to have a very open active site that could accommodate the additional sugars found on the remaining host surface-associated carbohydrate (Figure 43B). Also, it is thought that the α-fucosidase activity of GH95 would occur following Sp4GH98 hydrolysis since Sp4GH98 is not active on the H-trisaccharide and requires an intact Lewis^Y antigen for activity (Higgins, Whitworth et al. 2009).

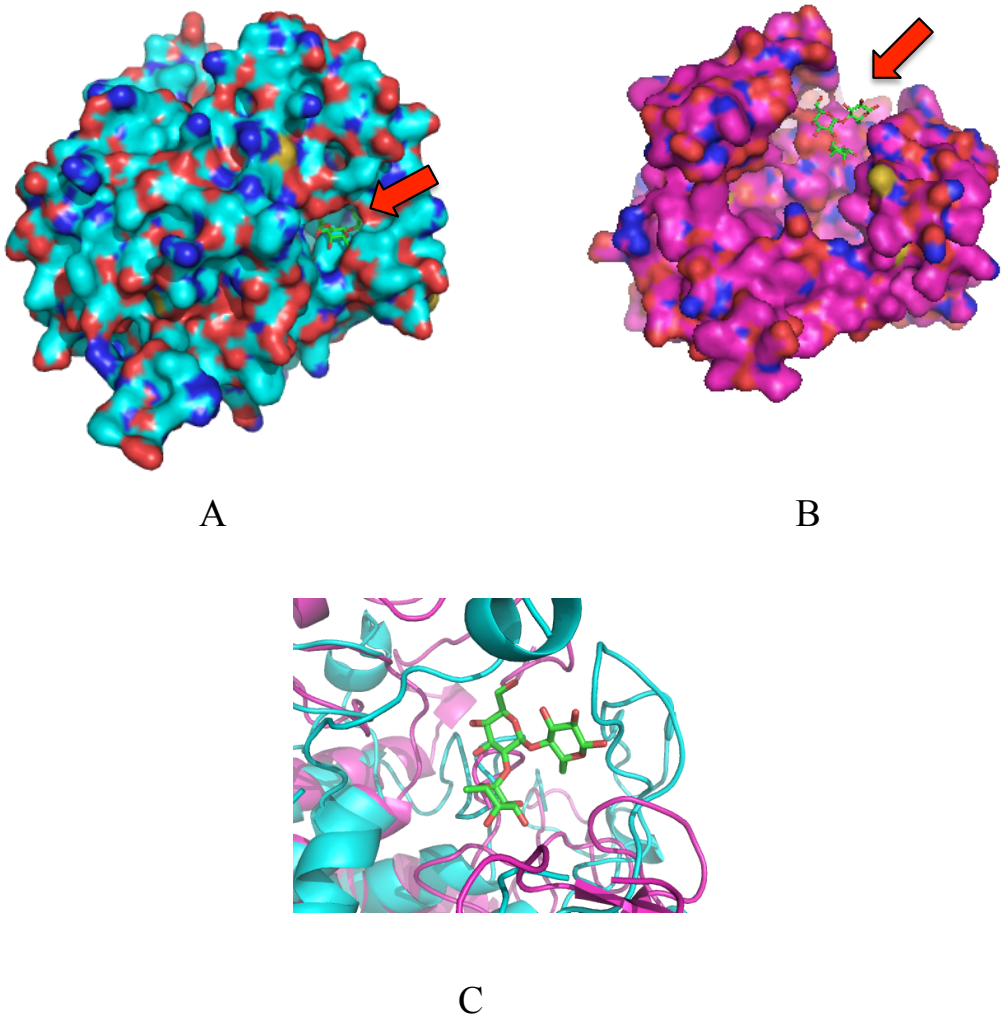


Figure 43: 3-dimensional models of GH95a and GH95b

Surface representations of (A) GH95a and (B) GH95b. The red arrow is pointing to the substrate of BbGH95 (Fucα1-2Galβ1-4Glc) that is modeled into the active site of each enzyme. (C) A closer look at the active site of GH95a (cyan) and GH95b (magenta). The carbohydrate substrate is shown in green with oxygen molecules in red.

Structure-function studies are required to initially confirm the predicted activities of GH95a and GH95b on the H-disaccharide and Fucα1-3GlcNAcβ1-3Gal, respectively. Furthermore, determining the molecular basis of the interactions could provide valuable information on the differences of each enzymes specificities. Also, if the GH98 enzymes are in fact involved in uncovering a receptor, it would be expected that all *S. pneumoniae* strain lacking *gh98* would have reduced adhesion and internalization to host cells. Additional internalization and adhesion assays could also provide evidence that GH95b is also involved uncovering the same receptor for strains with the type 1 operon. Also, restoring the internalization and predicted adhesion phenotype by the addition of recombinant Sp3GH98 to Δ*gh98* (from the TIGR4 strain) could suggest that that these enzymes are required to uncap a terminal carbohydrate structure essential for host cell internalization and that this receptor structure is identical for all strains of *S. pneumoniae*.

Another theory is that one of the fucose processing metabolites act as a signalling molecule to stimulate expression of other factors involved in internalization. Basal level expression of the fucose operon produces a low quantity of GH98 enzymes that are secreted to cleave host histo-blood group antigens. As the bacterium moves closer to the host cell it senses the H-disaccharide or A/B-trisaccharide GH98 products which are transported into the bacterium and subsequently metabolized. However, one of the fucose processing intermediates, such as fuculose or fuculose 1-phosphate, could act as a signalling molecule to induce expression of other factors required for bacterial adhesion and host internalization. To test for a hypothesis where fucose processing enzymes as the critical step in fucose utilization for adhesion and internalization, the addition of fucose or the H-disaccharide should restore Δ*gh98* (from the TIGR4 strain) internalization by host epithelial cells.

Pneumococcal fucose processing and nutrient acquisition

It is surprising with the presence of the same fucose processing enzymes that allow *E. coli* to metabolize fucose for energy that *S. pneumoniae* cannot use exogenous fucose as a sole carbon source. A major difference in the fucose operons of *E. coli* and *S. pneumoniae* is the transport system; *E. coli* harbours a single fucose permease (FcsP), while *S. pneumoniae* TIGR4 uses a more complex PTS transport system (EIIA, EIIB,

EIIC, EIID). The predicted model for the fucose utilization pathway in *S. pneumoniae* TIGR4 shows fucose entering the bacterium as the H-disaccharide via the PTS-transport system (Higgins, Whitworth et al. 2009) (Figure 12), instead of free fucose via the fucose permease as in *E. coli*. However, the addition of fucose, as a monosaccharide, is required for expression of the fucose operon in *S. pneumoniae* (Chan, O'Dwyer et al. 2003) and thus fucose appears to enter the bacterium. However, since FcsP is absent in *S. pneumoniae*, it is possible that fucose is entering the bacterium non-specifically through a different sugar permease, such as the galactose permease, and consequently is not entering the bacterium in abundance. Hence, enough fucose enters the bacterium to induce expression of the fucose operon, although not enough for the bacterium to sustain growth. Therefore, the substitution of FcsP for the PTS-transport system has rendered the bacterium unable to use exogenous fucose as a sole carbon source. It is predicted that fucose enters *S. pneumoniae* in the form of the H-disaccharide or the A/B-trisaccharides depending if the strain carries the type 1 or type 2 fucose utilization operons, respectively (Figure 12). Thus, the bacterium could possibly use these di- or tri-saccharides as a carbon source and thus a role in energy acquisition cannot be ruled out.

The ability of a bacterium to use fucose as a carbon source has been linked to providing a competitive advantage for both *Campylobacter jejuni* and *Bacteroides thetaiotamicron* in the gut (Salyers and Pajeau 1989; Muraoka and Zhang 2011; Stahl, Friis et al. 2011). Similarly, it is possible that pneumococcal fucose utilization could be involved in bacterial fitness using the histo-blood group antigens as a fucose source. Further investigation into this possibility is required and could illustrate a multi-functional role for fucose utilization similar to other pneumococcal virulence factors.

Pneumococcal fucose utilization and implications for blood type specific infections

The substrate selectivity of the two GH98 enzymes and the pathways that they represent has a major implication in blood type specificity. The type 2 ABO blood group antigens are all based on a Fucα1-2Galβ1-4GlcNAc (H-trisaccharide) type 2 core. This core structure, when acted upon by the FUT3 fucosyl transferase, forms the Lewis^{Y} antigen that is found among people of all ABO blood types (Marionneau, Cailleau-Thomas et al. 2001). Indeed, the more frequent presence among the sequenced

pneumococcal strains of the operon tuned to processing the Lewis^Y antigen, suggests that the Lewis^Y antigen is a more common target than the A/B-blood group antigens, which are the likely targets of the type 2 or Sp3GH98-containing operon. Assuming that fucose utilization is a conserved virulence factor among all pneumococcal strains, the implication of this research is that strains having a fucose utilization pathway tuned to activity on the Lewis^Y antigen would likely have potential for virulence in hosts of any blood type. In contrast, the specificity of the alternative operon for type 2 A/B-blood group antigens predicts that these strain should be less virulent in blood-type O hosts. Overall, this apparent relationship suggests that mismatching of the pneumococcal fucose utilization operon with the type of antigen expressed by the host, for example a strain with a Sp3GH98-type enzyme in a blood type O(H) individual, may play a role in the promotion of an invasive infection and/or the ability of the bacterium to establish asymptomatic colonization of the host.

Fucose utilization as a novel drug target

The rapid increase in bacterial resistance to antibiotic treatments has sparked the development of alternative approaches to managing infections. One such approach is to target specific bacterial virulence mechanisms, such as adhesion or internalization, to slow down or stop the bacterium from becoming pathogenic allowing the host immune system to contain the bacterium and thus reducing the chance of infection. These anti-infective therapies can have advantages over traditional antibiotic treatments, most significantly reducing bacterial selective pressures for resistance. More specifically, with carbohydrate-based inhibitors for bacterial lectins, it is believed that a mutation in the lectin rendering the inhibitor ineffective would also cause reduced or absent binding between the lectin and its natural host ligand making the selective mutants avirulent (Mulvey, Kitov et al. 2001). Also, unlike antibiotics, anti-infective treatments do not affect the delicate balance of the non-pathogenic bacteria species found in the host's natural flora. A major set back of anti-infective therapies is their inability to be effective for patients with reduced immune functions, such as people with cancer or HIV/AIDS. However, anti-infective therapies can be combined with conventional antibiotics to increase efficacy of bacterial killing. For example, bacteria are less sensitive to antibiotic

killing when attached to host tissue, part of a biofilm, or internalized in a host cell and thus a inhibitor to a component involved in these mechanisms would aid in antibiotic efficiency (Zopf and Roth 1996).

Sp4GH98 is an excellent candidate to design an inhibitor for a novel anti-infective therapy, as it is an established virulence factor in the *S. pneumoniae* TIGR4 strain and may play a role in host internalization (Section 5.4 and (Hava and Camilli 2002; Embry, Hinojosa et al. 2007)). As an extracellular enzyme, GH98 appears to initiate the pneumococcal fucose utilization pathway and therefore inhibiting its activity will essentially hinder the entire pathway. Also, its secreted nature provides easy access for the inhibitor to bind the enzyme so the inhibitor would not have to be able enter the bacterium. Furthermore, Sp4GH98 does not have a human homologues so a synthetic inhibitor for it is not expected to affect host functions.

There are two different GH98 enzymes found in *S. pneumoniae* depending on the type of fucose utilization operon present in the specific strain. Assuming the type 2 operon is also important for pneumococcal virulence, synthetic inhibitors to both Sp4GH98 and Sp3GH98 would be required to cover all pneumococcal strains. These inhibitors could act as a prophylactic for individuals with high susceptibility to *S. pneumoniae* infections. Although more likely, these inhibitors would be used in combination with conventional antibiotics. This approach has been shown to be effective for *S. pneumoniae* treatments using anti-viral drugs as anti-infective agents. One class of anti-viral agents, such as the popular drug Tamiflu®, are neuraminidase inhibitors. It is well established that the viral neuraminidase contributes to secondary pneumococcal infections (Peltola, Murti et al. 2005). More recently, it has been shown that treatment with an anti-viral agent, specific for the viral neuraminidase, upon an influenza infection can not only improve antibiotic efficiency to treat a subsequent pneumococcal infection but also increases survival in persons at high risk of complications and mortality (McCullers 2004). These kind of combination therapies could provide significantly more effective pneumococcal treatments and certainly warrants more intensive examination.

Conclusion

S. pneumoniae contains two strain-dependent varieties of fucose utilization operons that consist of glycan degradation enzymes, carbohydrate transport proteins, and fucose processing enzymes. In the TIGR4 strain, the fucose operon is required for the bacterium to cause infection illustrating its importance in pathogenesis (Hava and Camilli 2002; Embry, Hinojosa et al. 2007). This research focused on characterizing the extracellular GH98 enzymes from both operons revealing that the GH98 from the type 1 operon deglycosylates the type 2 Lewis^{Y} antigen while the GH98 from the type 2 operon cleaves the type 2 A/B-blood group antigens. Both GH98 enzymes act on their respective blood group antigens presented on lung epithelial cells and thus is thought to initiate the pneumococcal fucose pathway (See Figure 12 for schematics of both fucose pathways). Structural and biochemical data on the fucose processing enzymes (FcsU, FcsI, FcsK, and FcsA) have further corroborated the proposed fucose pathways. However, although the fucose processing enzymes from *S. pneumoniae* have been shown to convert fucose to DHAP and lactaldehyde for energy generation, the bacterium is unable to use fucose as a sole carbon source suggesting a novel function for this operon. Preliminary data hints at a role in host internalization indicating that further investigation into fucose utilization could demonstrate a novel mechanism for pneumococcal pathogenesis.

Bibliography

(1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallogr D Biol Crystallogr **50**(Pt 5): 760-763.

Abbott, D. W., M. A. Higgins, et al. (2010). "The molecular basis of glycogen breakdown and transport in Streptococcus pneumoniae." Molecular Microbiology **77**(1): 183-199.

Abbott, D. W., M. S. Macauley, et al. (2009). "Streptococcus pneumoniae endohexosaminidase D, structural and mechanistic insight into substrate-assisted catalysis in family 85 glycoside hydrolases." J Biol Chem **284**(17): 11676-11689.

Agarwal, V., T. M. Asmat, et al. (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**(46): 35615-35623.

Anderson, K. M., H. Ashida, et al. (2005). "A clostridial endo-beta-galactosidase that cleaves both blood group A and B glycotopes: the first member of a new glycoside hydrolase family, GH98." J Biol Chem **280**(9): 7720-7728.

Andersson, B., J. Dahmen, et al. (1983). "Identification of an active disaccharide unit of a glycoconjugate receptor for pneumococci attaching to human pharyngeal epithelial cells." J Exp Med **158**(2): 559-570.

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

Bartkus, J. M. and R. P. Mortlock (1986). "Isolation of a mutation resulting in constitutive synthesis of L-fucose catabolic enzymes." J Bacteriol **165**(3): 710-714.

Bergmann, S. and S. Hammerschmidt (2006). "Versatility of pneumococcal surface proteins." Microbiology **152**(Pt 2): 295-303.

Bergmann, S., M. Rohde, et al. (2001). "alpha-Enolase of Streptococcus pneumoniae is a plasmin(ogen)-binding protein displayed on the bacterial cell surface." Mol Microbiol **40**(6): 1273-1287.

Bergsten, G., M. Samuelsson, et al. (2004). "PapG-dependent adherence breaks mucosal inertia and triggers the innate host response." J Infect Dis **189**(9): 1734-1742.

Berkley, J. A., B. S. Lowe, et al. (2005). "Bacteremia among children admitted to a rural hospital in Kenya." N Engl J Med **352**(1): 39-47.

Bernado, P., E. Mylonas, et al. (2007). "Structural characterization of flexible proteins using small-angle X-ray scattering." Journal of the American Chemical Society **129**(17): 5656-5664.

Berry, A. M. and J. C. Paton (1996). "Sequence heterogeneity of PsaA, a 37-kilodalton putative adhesin essential for virulence of Streptococcus pneumoniae." Infect Immun **64**(12): 5255-5262.

Berteau, O., J. Bielicki, et al. (2004). "Alpha-L-fucosidases: exoglycosidases with unusual transglycosylation properties." Biochemistry **43**(24): 7881-7891.

Berteau, O., I. McCort, et al. (2002). "Characterization of a new alpha-L-fucosidase isolated from the marine mollusk Pecten maximus that catalyzes the hydrolysis of alpha-L-fucose from algal fucoidan (Ascophyllum nodosum)." Glycobiology **12**(4): 273-282.

Binet, M. R., M. N. Rager, et al. (1998). "Fructose and mannose metabolism in Aeromonas hydrophila: identification of transport systems and catabolic pathways." Microbiology **144 (Pt 4)**: 1113-1121.

Black, S., H. Shinefield, et al. (2000). "Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. Northern California Kaiser Permanente Vaccine Study Center Group." Pediatr Infect Dis J **19**(3): 187-195.

Blake, A. W., L. McCartney, et al. (2006). "Understanding the biological rationale for the diversity of cellulose-directed carbohydrate-binding modules in prokaryotic enzymes." J Biol Chem **281**(39): 29321-29329.

Bogaert, D., R. De Groot, et al. (2004). "Streptococcus pneumoniae colonisation: the key to pneumococcal disease." Lancet Infect Dis **4**(3): 144-154.

Bolam, D. N., A. Ciruela, et al. (1998). "Pseudomonas cellulose-binding domains mediate their effects by increasing enzyme substrate proximity." Biochem J **331 (Pt 3)**: 775-781.

Bongaerts, R. J., H. P. Heinz, et al. (2000). "Antigenicity, expression, and molecular characterization of surface-located pullulanase of Streptococcus pneumoniae." Infect Immun **68**(12): 7141-7143.

Boraston, A. B., D. N. Bolam, et al. (2004). "Carbohydrate-binding modules: fine-tuning polysaccharide recognition." Biochem J **382**(Pt 3): 769-781.

Boraston, A. B., A. L. Creagh, et al. (2001). "Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from Thermotoga maritima xylanase 10A." Biochemistry **40**(21): 6240-6247.

Boraston, A. B., D. Wang, et al. (2006). "Blood group antigen recognition by a Streptococcus pneumoniae virulence factor." J Biol Chem **281**(46): 35263-35271.

Bork, P., C. Sander, et al. (1992). "An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins." Proc Natl Acad Sci U S A **89**(16): 7290-7294.

Bricker, A. L. and A. Camilli (1999). "Transformation of a type 4 encapsulated strain of Streptococcus pneumoniae." FEMS Microbiol Lett **172**(2): 131-135.

Bridy-Pappas, A. E., M. B. Margolis, et al. (2005). "Streptococcus pneumoniae: description of the pathogen, disease epidemiology, treatment, and prevention." Pharmacotherapy **25**(9): 1193-1212.

Brown, J. S., T. Hussell, et al. (2002). "The classical pathway is the dominant complement pathway required for innate immunity to Streptococcus pneumoniae infection in mice." Proc Natl Acad Sci U S A **99**(26): 16969-16974.

Brunger, A. T. (1992). "Free R value: a novel statistical quantity for assessing the accuracy of crystal structures." Nature **355**(6359): 472-475.

Burnaugh, A. M., L. J. Frantz, et al. (2008). "Growth of Streptococcus pneumoniae on human glycoconjugates is dependent upon the sequential activity of bacterial exoglycosidases." J Bacteriol **190**(1): 221-230.

Caines, M. E., H. Zhu, et al. (2008). "The structural basis for T-antigen hydrolysis by Streptococcus pneumoniae: a target for structure-based vaccine design." J Biol Chem **283**(46): 31279-31283.

Calcutt, M. J., H. Y. Hsieh, et al. (2002). "Identification, molecular cloning and expression of an alpha-N-acetylgalactosaminidase gene from Clostridium perfringens." FEMS Microbiol Lett **214**(1): 77-80.

Cantarel, B. L., P. M. Coutinho, et al. (2009). "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics." Nucleic Acids Res **37**(Database issue): D233-238.

Cao, S., Z. Lou, et al. (2007). "Structural basis for the recognition of blood group trisaccharides by norovirus." J Virol **81**(11): 5949-5957.

Chan, P. F., K. M. O'Dwyer, et al. (2003). "Characterization of a novel fucose-regulated promoter (PfcsK) suitable for gene essentiality and antibacterial mode-of-action studies in Streptococcus pneumoniae." J Bacteriol **185**(6): 2051-2058.

Chen, V. B., W. B. Arendall, 3rd, et al. (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." Acta Crystallogr D Biol Crystallogr **66**(Pt 1): 12-21.

Chen, Y. M., Y. Zhu, et al. (1987). "The organization of the fuc regulon specifying L-fucose dissimilation in Escherichia coli K12 as determined by gene cloning." Mol Gen Genet **210**(2): 331-337.

Chiu, T. H. and D. S. Feingold (1964). "The Purification and Properties of L-Rhamnulokinase." Biochim Biophys Acta **92**: 489-497.

Clarke, V. A., N. Platt, et al. (1995). "Cloning and expression of the beta-N-acetylglucosaminidase gene from Streptococcus pneumoniae. Generation of truncated enzymes with modified aglycon specificity." J Biol Chem **270**(15): 8805-8814.

Claverys, J. P., M. Roger, et al. (1980). "Excision and repair of mismatched base pairs in transformation of Streptococcus pneumoniae." Mol Gen Genet **178**(1): 191-201.

Cowtan, K. D. and K. Y. Zhang (1999). "Density modification for macromolecular phase improvement." Prog Biophys Mol Biol **72**(3): 245-270.

Cron, L. E., H. J. Bootsma, et al. (2009). "Surface-associated lipoprotein PpmA of Streptococcus pneumoniae is involved in colonization in a strain-specific manner." Microbiology **155**(Pt 7): 2401-2410.

Dalia, A. B., A. J. Standish, et al. (2010). "Three surface exoglycosidases from Streptococcus pneumoniae, NanA, BgaA, and StrH, promote resistance to opsonophagocytic killing by human neutrophils." Infect Immun **78**(5): 2108-2116.

Daoud, Z., M. Kourani, et al. (2011). "Resistance of Streptococcus pneumoniae isolated from Lebanese patients between 2005 and 2009." Rev Esp Quimioter **24**(2): 84-90.

Davies, G. and B. Henrissat (1995). "Structures and mechanisms of glycosyl hydrolases." Structure **3**(9): 853-859.

Davies, G. and B. Henrissat (1995). "Structures and Mechanisms of Glycosyl Hydrolases." Structure **3**(9): 853-859.

Davies, G. J., K. S. Wilson, et al. (1997). "Nomenclature for sugar-binding subsites in glycosyl hydrolases." Biochem J **321 (Pt 2)**: 557-559.

Dillard, J. P., M. W. Vandersea, et al. (1995). "Characterization of the cassette containing genes for type 3 capsular polysaccharide biosynthesis in Streptococcus pneumoniae." J Exp Med **181**(3): 973-983.

Dintilhac, A., G. Alloing, et al. (1997). "Competence and virulence of Streptococcus pneumoniae: Adc and PsaA mutants exhibit a requirement for Zn and Mn resulting from inactivation of putative ABC metal permeases." Mol Microbiol **25**(4): 727-739.

Dreyer, M. K. and G. E. Schulz (1996). "Catalytic mechanism of the metal-dependent fuculose aldolase from Escherichia coli as derived from the structure." J Mol Biol **259**(3): 458-466.

Dreyer, M. K. and G. E. Schulz (1996). "Refined high-resolution structure of the metal-ion dependent L-fuculose-1-phosphate aldolase (class II) from Escherichia coli." Acta Crystallogr D Biol Crystallogr **52**(Pt 6): 1082-1091.

Drouillard, S., H. Driguez, et al. (2006). "Large-scale synthesis of H-antigen oligosaccharides by expressing Helicobacter pylori alpha1,2-fucosyltransferase in metabolically engineered Escherichia coli cells." Angew Chem Int Ed Engl **45**(11): 1778-1780.

Embry, A., E. Hinojosa, et al. (2007). "Regions of Diversity 8, 9 and 13 contribute to Streptococcus pneumoniae virulence." BMC Microbiol **7**: 80.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-2132.

Erni, B. and B. Zanolari (1985). "The mannose-permease of the bacterial phosphotransferase system. Gene cloning and purification of the enzyme IIMan/IIIMan complex of Escherichia coli." J Biol Chem **260**(29): 15495-15503.

Evans, G. and G. Bricogne (2002). "Triiodide derivatization and combinatorial counter-ion replacement: two methods for enhancing phasing signal using laboratory Cu

Kalpha X-ray equipment." Acta Crystallogr D Biol Crystallogr **58**(Pt 6 Pt 2): 976-991.

Ezer, A., E. Matalon, et al. (2008). "Cell surface enzyme attachment is mediated by family 37 carbohydrate-binding modules, unique to Ruminococcus albus." J Bacteriol **190**(24): 8220-8222.

Fenoll, A., L. Aguilar, et al. (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.

Ficko-Blean, E. and A. B. Boraston (2005). "Cloning, recombinant production, crystallization and preliminary X-ray diffraction studies of a family 84 glycoside hydrolase from Clostridium perfringens." Acta Crystallogr Sect F Struct Biol Cryst Commun **61**(Pt 9): 834-836.

Ficko-Blean, E. and A. B. Boraston (2006). "The interaction of a carbohydrate-binding module from a Clostridium perfringens N-acetyl-beta-hexosaminidase with its carbohydrate receptor." J Biol Chem **281**(49): 37748-37757.

Ficko-Blean, E., K. J. Gregg, et al. (2009). "Portrait of an enzyme, a complete structural analysis of a multimodular {beta}-N-acetylglucosaminidase from Clostridium perfringens." J Biol Chem **284**(15): 9876-9884.

Fiore, A. E., O. S. Levine, et al. (1999). "Effectiveness of pneumococcal polysaccharide vaccine for preschool-age children with chronic disease." Emerg Infect Dis **5**(6): 828-831.

Fischer, H., M. D. Neto, et al. (2010). "Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale." Journal of Applied Crystallography **43**: 101-109.

Flores-Diaz, M., A. Alape-Giron, et al. (2005). "A cellular deficiency of gangliosides causes hypersensitivity to Clostridium perfringens phospholipase C." J Biol Chem **280**(29): 26680-26689.

Fraenkel, D. G. (1968). "The phosphoenolpyruvate-initiated pathway of fructose metabolism in Escherichia coli." J Biol Chem **243**(24): 6458-6463.

Franke, D. and D. I. Svergun (2009). "DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering." Journal of Applied Crystallography **42**: 342-346.

Frey, P. A. (1996). "The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose." FASEB J **10**(4): 461-470.

Ganter, C., A. Bock, et al. (1988). "Production of thermostable, recombinant a-glactosidase suitable for raffinose elimination from sugar beet syrup." Journal of Biotechnology **8**(4): 301-310.

Garcia-Alles, L. F., A. Zahn, et al. (2002). "Sugar recognition by the glucose and mannose permeases of Escherichia coli. Steady-state kinetics and inhibition studies." Biochemistry **41**(31): 10077-10086.

Gardy, J. L., M. R. Laird, et al. (2005). "PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis." Bioinformatics **21**(5): 617-623.

Gasteiger, E., A. Gattiker, et al. (2003). "ExPASy: The proteomics server for in-depth protein knowledge and analysis." Nucleic Acids Res **31**(13): 3784-3788.

Gentile, A. and V. Bazan (2011). "Prevention of pneumococcal disease through vaccination." Vaccine **29 Suppl 3**: C15-25.

Ghalambor, M. A. and E. C. Heath (1962). "The metabolism of L-fucose. II. The enzymatic cleavage of L-fuculose 1-phosphate." J Biol Chem **237**: 2427-2433.

Giardina, T., A. P. Gunning, et al. (2001). "Both binding sites of the starch-binding domain of Aspergillus niger glucoamylase are essential for inducing a conformational change in amylose." J Mol Biol **313**(5): 1149-1159.

Gilbert, H. J. (2007). "Cellulosomes: microbial nanomachines that display plasticity in quaternary structure." Mol Microbiol **63**(6): 1568-1576.

Green, M. and S. S. Cohen (1956). "Enzymatic conversion of L-fucose to L-fuculose." J Biol Chem **219**(2): 557-568.

Gregg, K. J., R. Finn, et al. (2008). "Divergent modes of glycan recognition by a new family of carbohydrate-binding modules." J Biol Chem **283**(18): 12604-12613.

Grueninger, D. and G. E. Schulz (2006). "Structure and reaction mechanism of L-rhamnulose kinase from Escherichia coli." J Mol Biol **359**(3): 787-797.

Grueninger, D. and G. E. Schulz (2007). "Substrate spectrum of L-rhamnulose kinase related to models derived from two ternary complex structures." FEBS Lett **581**(16): 3127-3130.

Gschwind, R. M., G. Gemmecker, et al. (1997). "Secondary structure of the IIB domain of the Escherichia coli mannose transporter, a new fold in the class of alpha/beta twisted open-sheet structures." FEBS Lett **404**(1): 45-50.

Guillen, D., S. Sanchez, et al. (2010). "Carbohydrate-binding domains: multiplicity of biological roles." Appl Microbiol Biotechnol **85**(5): 1241-1249.

Guinier, A. and G. Fournet (1955). Small-angle scattering of X-rays. New York,, Wiley.

Hammerschmidt, S., S. Wolff, et al. (2005). "Illustration of pneumococcal polysaccharide capsule during adherence and invasion of epithelial cells." Infect Immun **73**(8): 4653-4667.

Hava, D. L. and A. Camilli (2002). "Large-scale identification of serotype 4 Streptococcus pneumoniae virulence factors." Mol Microbiol **45**(5): 1389-1406.

Hayden, F. (2009). "Developing new antiviral agents for influenza treatment: what does the future hold?" Clin Infect Dis **48 Suppl 1**: S3-13.

Heath, E. C. and M. A. Ghalambor (1962). "The metabolism of L-fucose. I. The purification and properties of L-fuculose kinase." J Biol Chem **237**: 2423-2426.

Helenius, A. and M. Aebi (2001). "Intracellular functions of N-linked glycans." Science **291**(5512): 2364-2369.

Henry, S., R. Oriol, et al. (1995). "Lewis histo-blood group system and associated secretory phenotypes." Vox Sang **69**(3): 166-182.

Hicks, L. A., Y. W. Chien, et al. (2011). "Outpatient antibiotic prescribing and nonsusceptible Streptococcus pneumoniae in the United States, 1996-2003." Clin Infect Dis **53**(7): 631-639.

Higgins, M. A., D. W. Abbott, et al. (2009). "Blood group antigen recognition by a solute-binding protein from a serotype 3 strain of Streptococcus pneumoniae." J Mol Biol **388**(2): 299-309.

Higgins, M. A., E. Ficko-Blean, et al. (2011). "The overall architecture and receptor binding of pneumococcal carbohydrate-antigen-hydrolyzing enzymes." J Mol Biol **411**(5): 1017-1036.

Higgins, M. A., G. E. Whitworth, et al. (2009). "Differential recognition and hydrolysis of host carbohydrate antigens by Streptococcus pneumoniae family 98 glycoside hydrolases." J Biol Chem **284**(38): 26161-26173.

Holden, H. M., I. Rayment, et al. (2003). "Structure and function of enzymes of the Leloir pathway for galactose metabolism." J Biol Chem **278**(45): 43885-43888.

Holmes, A. R., R. McNab, et al. (2001). "The pavA gene of Streptococcus pneumoniae encodes a fibronectin-binding protein that is essential for virulence." Mol Microbiol **41**(6): 1395-1408.

Hurley, J. H. (1996). "The sugar kinase/heat shock protein 70/actin superfamily: implications of conserved structure for mechanism." Annu Rev Biophys Biomol Struct **25**: 137-162.

Iyer, R. and A. Camilli (2007). "Sucrose metabolism contributes to in vivo fitness of Streptococcus pneumoniae." Mol Microbiol **66**(1): 1-13.

Jensen, S. O. and P. R. Reeves (1998). "Domain organisation in phosphomannose isomerases (types I and II)." Biochim Biophys Acta **1382**(1): 5-7.

Jeong, J. K., O. Kwon, et al. (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**(9): 3011-3023.

Jeyakanthan, J., J. Taka, et al. (2005). "Purification, crystallization and preliminary X-ray crystallographic study of the L-fuculose-1-phosphate aldolase (FucA) from Thermus thermophilus HB8." Acta Crystallogr Sect F Struct Biol Cryst Commun **61**(Pt 12): 1075-1077.

Joerger, A. C., C. Gosse, et al. (2000). "Catalytic action of fuculose 1-phosphate aldolase (class II) as derived from structure-directed mutagenesis." Biochemistry **39**(20): 6033-6041.

Joerger, A. C., C. Mueller-Dieckmann, et al. (2000). "Structures of l-fuculose-1-phosphate aldolase mutants outlining motions during catalysis." J Mol Biol **303**(4): 531-543.

Johansen, F. E. and C. S. Kaetzel (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**(6): 598-602.

Johnson, S. E., J. K. Dykes, et al. (2002). "Inhibition of pneumococcal carriage in mice by subcutaneous immunization with peptides from the common surface protein pneumococcal surface adhesin a." J Infect Dis **185**(4): 489-496.

Johnston, J. W., D. E. Briles, et al. (2006). "Mn2+-dependent regulation of multiple genes in Streptococcus pneumoniae through PsaR and the resultant impact on virulence." Infect Immun **74**(2): 1171-1180.

Johnston, J. W., L. E. Myers, et al. (2004). "Lipoprotein PsaA in virulence of Streptococcus pneumoniae: surface accessibility and role in protection from superoxide." Infect Immun **72**(10): 5858-5867.

Kadioglu, A., H. Brewin, et al. (2010). "Pneumococcal protein PavA is important for nasopharyngeal carriage and development of sepsis." Mol Oral Microbiol **25**(1): 50-60.

Kaetzel, C. S., J. K. Robinson, et al. (1991). "The polymeric immunoglobulin receptor (secretory component) mediates transport of immune complexes across epithelial cells: a local defense function for IgA." Proc Natl Acad Sci U S A **88**(19): 8796-8800.

Katayama, T., A. Sakuma, et al. (2004). "Molecular cloning and characterization of Bifidobacterium bifidum 1,2-alpha-L-fucosidase (AfcA), a novel inverting glycosidase (glycoside hydrolase family 95)." J Bacteriol **186**(15): 4885-4893.

Kelley, L. A. and M. J. Sternberg (2009). "Protein structure prediction on the Web: a case study using the Phyre server." Nat Protoc **4**(3): 363-371.

Kim, J. O. and J. N. Weiser (1998). "Association of intrastrain phase variation in quantity of capsular polysaccharide and teichoic acid with the virulence of Streptococcus pneumoniae." J Infect Dis **177**(2): 368-377.

Kim, M. S., J. Shin, et al. (2003). "Crystal structures of RbsD leading to the identification of cytoplasmic sugar-binding proteins with a novel folding architecture." J Biol Chem **278**(30): 28173-28180.

King, S. J., K. R. Hippe, et al. (2006). "Deglycosylation of human glycoconjugates by the sequential activities of exoglycosidases expressed by Streptococcus pneumoniae." Mol Microbiol **59**(3): 961-974.

Konarev, P. V., V. V. Volkov, et al. (2003). "PRIMUS: a Windows PC-based system for small-angle scattering data analysis." Journal of Applied Crystallography **36**: 1277-1282.

Koppel, D. E. (1972). "Analysis of Macromolecular Polydispersity in Intensity Correlation Spectroscopy - Method of Cumulants." Journal of Chemical Physics **57**(11): 4814-&.

Koshland, D. E. (1953). "Stereochemistry and the Mechanism of Enzymatic Reactions." Biological Reviews of the Cambridge Philosophical Society **28**(4): 416-436.

Kozin, M. B. and D. I. Svergun (2001). "Automated matching of high- and low-resolution structural models." Journal of Applied Crystallography **34**: 33-41.

Krissinel, E. and K. Henrick (2004). "Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2256-2268.

Kroemer, M., I. Merkel, et al. (2003). "Structure and catalytic mechanism of L-rhamnulose-1-phosphate aldolase." Biochemistry **42**(36): 10560-10568.

Kyaw, M. H., R. Lynfield, et al. (2006). "Effect of introduction of the pneumococcal conjugate vaccine on drug-resistant Streptococcus pneumoniae." N Engl J Med **354**(14): 1455-1463.

Lacks, S. (1966). "Integration efficiency and genetic recombination in pneumococcal transformation." Genetics **53**(1): 207-235.

Lammerts van Bueren, A., E. Ficko-Blean, et al. (2011). "The conformation and function of a multimodular glycogen-degrading pneumococcal virulence factor." Structure **19**(5): 640-651.

Landsteiner, K. (1900). "Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe." Zentralblatt Bakteriologie **27**: 357-362.

Laskowski, R. A., D. S. Moss, et al. (1993). "Main-chain bond lengths and bond angles in protein structures." J Mol Biol **231**(4): 1049-1067.

Lee, K. H., K. S. Ryu, et al. (2009). "Crystal structures and enzyme mechanisms of a dual fucose mutarotase/ribose pyranase." J Mol Biol **391**(1): 178-191.

Lillelund, V. H., H. H. Jensen, et al. (2002). "Recent developments of transition-state analogue glycosidase inhibitors of non-natural product origin." Chem Rev **102**(2): 515-553.

Limoli, D. H., J. A. Sladek, et al. (2011). "BgaA acts as an adhesin to mediate attachment of some pneumococcal strains to human epithelial cells." Microbiology **157**(Pt 8): 2369-2381.

Liu, Q. P., G. Sulzenbacher, et al. (2007). "Bacterial glycosidases for the production of universal red blood cells." Nat Biotechnol **25**(4): 454-464.

Livingstone, G., F. Fanks, et al. (1977). "The effects of aqueous solvent structure on the mutarotation kinetics of glucose." Journal of Solution Chemistry **6**(3): 203-216.

Lovell, S. C., I. W. Davis, et al. (2003). "Structure validation by Calpha geometry: phi,psi and Cbeta deviation." Proteins **50**(3): 437-450.

Maeda, S., C. Sugita, et al. (2006). "Latent nitrate transport activity of a novel sulfate permease-like protein of the cyanobacterium Synechococcus elongatus." J Biol Chem **281**(9): 5869-5876.

Manco, S., F. Hernon, et al. (2006). "Pneumococcal neuraminidases A and B both have essential roles during infection of the respiratory tract and sepsis." Infect Immun **74**(7): 4014-4020.

Marionneau, S., A. Cailleau-Thomas, et al. (2001). "ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world." Biochimie **83**(7): 565-573.

Marra, A., S. Lawson, et al. (2002). "In vivo characterization of the psa genes from Streptococcus pneumoniae in multiple models of infection." Microbiology **148**(Pt 5): 1483-1491.

Marriott, H. M., T. J. Mitchell, et al. (2008). "Pneumolysin: a double-edged sword during the host-pathogen interaction." Curr Mol Med **8**(6): 497-509.

Mccarter, J. D. and S. G. Withers (1994). "Mechanisms of Enzymatic Glycoside Hydrolysis." Current Opinion in Structural Biology **4**(6): 885-892.

McCoy, A. J., R. W. Grosse-Kunstleve, et al. (2007). "Phaser crystallographic software." J Appl Crystallogr **40**(Pt 4): 658-674.

McCoy, A. J., R. W. Grosse-Kunstleve, et al. (2005). "Likelihood-enhanced fast translation functions." Acta Crystallogr D Biol Crystallogr **61**(Pt 4): 458-464.

McCullers, J. A. (2004). "Effect of antiviral treatment on the outcome of secondary bacterial pneumonia after influenza." J Infect Dis **190**(3): 519-526.

McCullers, J. A. (2006). "Insights into the interaction between influenza virus and pneumococcus." Clin Microbiol Rev **19**(3): 571-582.

Meloncelli, P. J. and T. L. Lowary (2010). "Synthesis of ABO histo-blood group type I and II antigens." Carbohydr Res **345**(16): 2305-2322.

Miller, E., N. J. Andrews, et al. (2011). "Effectiveness of the new serotypes in the 13-valent pneumococcal conjugate vaccine." Vaccine **29**(49): 9127-9131.

Mollicone, R., I. Reguigne, et al. (1994). "Molecular basis for Lewis alpha(1,3/1,4)-fucosyltransferase gene deficiency (FUT3) found in Lewis-negative Indonesian pedigrees." J Biol Chem **269**(33): 20987-20994.

Mostov, K. E. (1994). "Transepithelial transport of immunoglobulins." Annu Rev Immunol **12**: 63-84.

Mulvey, G., P. I. Kitov, et al. (2001). "Glycan mimicry as a basis for novel anti-infective drugs." Biochimie **83**(8): 841-847.

Muraoka, W. T. and Q. Zhang (2011). "Phenotypic and genotypic evidence for L-fucose utilization by Campylobacter jejuni." J Bacteriol **193**(5): 1065-1075.

Murshudov, G. N., A. A. Vagin, et al. (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallogr D Biol Crystallogr **53**(Pt 3): 240-255.

Nagae, M., A. Tsuchiya, et al. (2007). "Structural basis of the catalytic reaction mechanism of novel 1,2-alpha-L-fucosidase from Bifidobacterium bifidum." J Biol Chem **282**(25): 18497-18509.

Notenboom, V., A. B. Boraston, et al. (2001). "Crystal structures of the family 9 carbohydrate-binding module from Thermotoga maritima xylanase 10A in native and ligand-bound forms." Biochemistry **40**(21): 6248-6256.

Nunn, R. S., Z. Markovic-Housley, et al. (1996). "Structure of the IIA domain of the mannose transporter from Escherichia coli at 1.7 angstroms resolution." J Mol Biol **259**(3): 502-511.

O'Brien, K. L., L. J. Wolfson, et al. (2009). "Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates." Lancet **374**(9693): 893-902.

Odenbreit, S. (2005). "Adherence properties of Helicobacter pylori: impact on pathogenesis and adaptation to the host." Int J Med Microbiol **295**(5): 317-324.

Oliver, E. J. and R. P. Mortlock (1971). "Metabolism of D-arabinose by Aerobacter aerogenes: purification of the isomerase." J Bacteriol **108**(1): 293-299.

Orihuela, C. J., J. N. Radin, et al. (2004). "Microarray analysis of pneumococcal gene expression during invasive disease." Infect Immun **72**(10): 5582-5596.

Oteo, J., E. Lazaro, et al. (2004). "Trends in antimicrobial resistance in 1,968 invasive Streptococcus pneumoniae strains isolated in Spanish hospitals (2001 to 2003): decreasing penicillin resistance in children's isolates." J Clin Microbiol **42**(12): 5571-5577.

Park, D., K. S. Ryu, et al. (2007). "Characterization and role of fucose mutarotase in mammalian cells." Glycobiology **17**(9): 955-962.

Patni, N. J. and J. K. Alexander (1971). "Catabolism of fructose and mannitol in Clostridium thermocellum: presence of phosphoenolpyruvate: fructose phosphotransferase, fructose 1-phosphate kinase, phosphoenolpyruvate: mannitol phosphotransferase, and mannitol 1-phosphate dehydrogenase in cell extracts." J Bacteriol **105**(1): 226-231.

Patrick, J. W. and N. Lee (1969). "Subunit structure of L-arabinose isomerase from Escherichia coli." J Biol Chem **244**(16): 4277-4283.

Peer, A., S. P. Smith, et al. (2009). "Noncellulosomal cohesin- and dockerin-like modules in the three domains of life." FEMS Microbiol Lett **291**(1): 1-16.

Peltola, V. T., K. G. Murti, et al. (2005). "Influenza virus neuraminidase contributes to secondary bacterial pneumonia." J Infect Dis **192**(2): 249-257.

Perrakis, A., R. Morris, et al. (1999). "Automated protein model building combined with iterative structure refinement." Nat Struct Biol **6**(5): 458-463.

Petoukhov, M. V. and D. I. Svergun (2005). "Global rigid body modeling of macromolecular complexes against small-angle scattering data." Biophysical Journal **89**(2): 1237-1250.

Pettigrew, M. M., K. P. Fennie, et al. (2006). "Variation in the presence of neuraminidase genes among Streptococcus pneumoniae isolates with identical sequence types." Infect Immun **74**(6): 3360-3365.

Pflugrath, J. W. (1999). "The finer things in X-ray diffraction data collection." Acta Crystallogr D Biol Crystallogr **55**(Pt 10): 1718-1725.

Pilishvili, T., C. Lexau, et al. (2010). "Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine." J Infect Dis **201**(1): 32-41.

Pluvinage, B., M. A. Higgins, et al. (2011). "Inhibition of the Pneumococcal Virulence Factor StrH and Molecular Insights into N-Glycan Recognition and Hydrolysis." Structure **19**(11): 1603-1614.

Polissi, A., A. Pontiggia, et al. (1998). "Large-scale identification of virulence genes from Streptococcus pneumoniae." Infect Immun **66**(12): 5620-5629.

Powell, H. R. (1999). "The Rossmann Fourier autoindexing algorithm in MOSFLM." Acta Crystallogr D Biol Crystallogr **55**(Pt 10): 1690-1695.

Pozzi, G., L. Masala, et al. (1996). "Competence for genetic transformation in encapsulated strains of Streptococcus pneumoniae: two allelic variants of the peptide pheromone." J Bacteriol **178**(20): 6087-6090.

Pracht, D., C. Elm, et al. (2005). "PavA of Streptococcus pneumoniae modulates adherence, invasion, and meningeal inflammation." Infect Immun **73**(5): 2680-2689.

Ramos-Sevillano, E., M. Moscoso, et al. (2011). "Nasopharyngeal colonization and invasive disease are enhanced by the cell wall hydrolases LytB and LytC of Streptococcus pneumoniae." PLoS One **6**(8): e23626.

Ravn, V. and E. Dabelsteen (2000). "Tissue distribution of histo-blood group antigens." APMIS **108**(1): 1-28.

Read, R. J. (1986). "Improved Fourier Coefficients for Maps Using Phases from Partial Structures with Errors." Acta Crystallographica Section A **42**: 140-149.

Rigden, D. J. (2005). "Analysis of glycoside hydrolase family 98: catalytic machinery, mechanism and a novel putative carbohydrate binding module." FEBS Lett **579**(25): 5466-5472.

Robb, C. S., F. E. Nano, et al. (2010). "Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of intracellular growth locus E (IglE) protein from Francisella tularensis subsp. novicida." Acta Crystallogr Sect F Struct Biol Cryst Commun **66**(Pt 12): 1596-1598.

Robinson, K. A., W. Baughman, et al. (2001). "Epidemiology of invasive Streptococcus pneumoniae infections in the United States, 1995-1998: Opportunities for prevention in the conjugate vaccine era." JAMA **285**(13): 1729-1735.

Rosen, J. B., A. R. Thomas, et al. (2011). "Geographic variation in invasive pneumococcal disease following pneumococcal conjugate vaccine introduction in the United States." Clin Infect Dis **53**(2): 137-143.

Rosenow, C., P. Ryan, et al. (1997). "Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of Streptococcus pneumoniae." Mol Microbiol **25**(5): 819-829.

Rosey, E. L. and G. C. Stewart (1992). "Nucleotide and deduced amino acid sequences of the lacR, lacABCD, and lacFE genes encoding the repressor, tagatose 6-phosphate gene cluster, and sugar-specific phosphotransferase system components of the lactose operon of Streptococcus mutans." J Bacteriol **174**(19): 6159-6170.

Ryu, K. S., C. Kim, et al. (2004). "NMR application probes a novel and ubiquitous family of enzymes that alter monosaccharide configuration." J Biol Chem **279**(24): 25544-25548.

Ryu, K. S., J. I. Kim, et al. (2005). "Structural insights into the monosaccharide specificity of Escherichia coli rhamnose mutarotase." J Mol Biol **349**(1): 153-162.

Salles, C., L. Creancier, et al. (1992). "The high level streptomycin resistance gene from Streptococcus pneumoniae is a homologue of the ribosomal protein S12 gene from Escherichia coli." Nucleic Acids Res **20**(22): 6103.

Salyers, A. A. and M. Pajeau (1989). "Competitiveness of different polysaccharide utilization mutants of Bacteroides thetaiotaomicron in the intestinal tracts of germfree mice." Appl Environ Microbiol **55**(10): 2572-2578.

Schneider, T. R. and G. M. Sheldrick (2002). "Substructure solution with SHELXD." Acta Crystallogr D Biol Crystallogr **58**(Pt 10 Pt 2): 1772-1779.

Seemann, J. E. and G. E. Schulz (1997). "Structure and mechanism of L-fucose isomerase from Escherichia coli." J Mol Biol **273**(1): 256-268.

Shaikh, F. A. and S. G. Withers (2008). "Teaching old enzymes new tricks: engineering and evolution of glycosidases and glycosyl transferases for improved glycoside synthesis." Biochem Cell Biol **86**(2): 169-177.

Shelburne, S. A., M. T. Davenport, et al. (2008). "The role of complex carbohydrate catabolism in the pathogenesis of invasive streptococci." Trends Microbiol **16**(7): 318-325.

Shouval, D. S., D. Greenberg, et al. (2006). "Site-specific disease potential of individual Streptococcus pneumoniae serotypes in pediatric invasive disease, acute otitis media and acute conjunctivitis." Pediatr Infect Dis J **25**(7): 602-607.

Singleton, R. J., T. W. Hennessy, et al. (2007). "Invasive pneumococcal disease caused by nonvaccine serotypes among alaska native children with high levels of 7-valent pneumococcal conjugate vaccine coverage." JAMA **297**(16): 1784-1792.

Skaper, S. D., A. Leon, et al. (1989). "Ganglioside function in the development and repair of the nervous system. From basic science to clinical application." Mol Neurobiol **3**(3): 173-199.

Smolsky, I. L., P. Liu, et al. (2007). "Biological small-angle x-ray scattering facility at the Stanford synchrotron radiation laboratory." Journal of Applied Crystallography **40**: S453-S458.

Song, X. M., W. Connor, et al. (2008). "Microarray analysis of Streptococcus pneumoniae gene expression changes to human lung epithelial cells." Can J Microbiol **54**(3): 189-200.

Southall, S. M., P. J. Simpson, et al. (1999). "The starch-binding domain from glucoamylase disrupts the structure of starch." FEBS Lett **447**(1): 58-60.

Stahl, M., L. M. Friis, et al. (2011). "L-fucose utilization provides Campylobacter jejuni with a competitive advantage." Proc Natl Acad Sci U S A **108**(17): 7194-7199.

Stapleton, A., E. Nudelman, et al. (1992). "Binding of uropathogenic Escherichia coli R45 to glycolipids extracted from vaginal epithelial cells is dependent on histo-blood group secretor status." J Clin Invest **90**(3): 965-972.

Sulzenbacher, G., C. Bignon, et al. (2004). "Crystal structure of Thermotoga maritima alpha-L-fucosidase. Insights into the catalytic mechanism and the molecular basis for fucosidosis." J Biol Chem **279**(13): 13119-13128.

Sung, C. K., H. Li, et al. (2001). "An rpsL cassette, janus, for gene replacement through negative selection in Streptococcus pneumoniae." Appl Environ Microbiol **67**(11): 5190-5196.

Svergun, D., C. Barberato, et al. (1995). "CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates." Journal of Applied Crystallography **28**: 768-773.

Svergun, D. I. (1992). "Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria." Journal of Applied Crystallography **25**: 495-503.

Takeda, K., H. Yoshida, et al. (2010). "X-ray structures of Bacillus pallidus d-arabinose isomerase and its complex with l-fucitol." Biochim Biophys Acta **1804**(6): 1359-1368.

Tong, H. H., L. E. Blue, et al. (2000). "Evaluation of the virulence of a Streptococcus pneumoniae neuraminidase-deficient mutant in nasopharyngeal colonization and development of otitis media in the chinchilla model." Infect Immun **68**(2): 921-924.

Trouw, L. A. and M. R. Daha (2011). "Role of complement in innate immunity and host defense." Immunol Lett **138**(1): 35-37.

Trzcinski, K., C. M. Thompson, et al. (2003). "Construction of otherwise isogenic serotype 6B, 7F, 14, and 19F capsular variants of Streptococcus pneumoniae strain TIGR4." Appl Environ Microbiol **69**(12): 7364-7370.

Tseng, H. J., A. G. McEwan, et al. (2002). "Virulence of Streptococcus pneumoniae: PsaA mutants are hypersensitive to oxidative stress." Infect Immun **70**(3): 1635-1639.

Tyx, R. E., H. Roche-Hakansson, et al. (2011). "Role of dihydrolipoamide dehydrogenase in regulation of raffinose transport in Streptococcus pneumoniae." J Bacteriol **193**(14): 3512-3524.

Vagin, A. and A. Teplyakov (2000). "An approach to multi-copy search in molecular replacement." Acta Crystallogr D Biol Crystallogr **56**(Pt 12): 1622-1624.

Vaguine, A. A., J. Richelle, et al. (1999). "SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model." Acta Crystallogr D Biol Crystallogr **55**(Pt 1): 191-205.

Van den Steen, P., P. M. Rudd, et al. (1998). "Concepts and principles of O-linked glycosylation." Crit Rev Biochem Mol Biol **33**(3): 151-208.

van der Pol, W., G. Vidarsson, et al. (2000). "Pneumococcal capsular polysaccharide-specific IgA triggers efficient neutrophil effector functions via FcalphaRI (CD89)." J Infect Dis **182**(4): 1139-1145.

Varrot, A., V. L. Yip, et al. (2005). "NAD+ and metal-ion dependent hydrolysis by family 4 glycosidases: structural insight into specificity for phospho-beta-D-glucosides." J Mol Biol **346**(2): 423-435.

Vasella, A., G. J. Davies, et al. (2002). "Glycosidase mechanisms." Curr Opin Chem Biol **6**(5): 619-629.

Vocadlo, D. J. and G. J. Davies (2008). "Mechanistic insights into glycosidase chemistry." Curr Opin Chem Biol **12**(5): 539-555.

Volkov, V. V. and D. I. Svergun (2003). "Uniqueness of ab initio shape determination in small-angle scattering." Journal of Applied Crystallography **36**: 860-864.

von Itzstein, M. (2008). "Disease-associated carbohydrate-recognising proteins and structure-based inhibitor design." Curr Opin Struct Biol **18**(5): 558-566.

von Itzstein, M., W. Y. Wu, et al. (1993). "Rational design of potent sialidase-based inhibitors of influenza virus replication." Nature **363**(6428): 418-423.

Wallace, L. J., F. A. Eiserling, et al. (1978). "The shape of L-arabinose isomerase from Escherichia coli." J Biol Chem **253**(10): 3717-3720.

Weinberger, D. M., K. Trzcinski, et al. (2009). "Pneumococcal capsular polysaccharide structure predicts serotype prevalence." PLoS Pathog **5**(6): e1000476.

Whitney, C. G., M. M. Farley, et al. (2000). "Increasing prevalence of multidrug-resistant Streptococcus pneumoniae in the United States." N Engl J Med **343**(26): 1917-1924.

Wilkins, M. R., E. Gasteiger, et al. (1999). "Protein identification and analysis tools in the ExPASy server." Methods Mol Biol **112**: 531-552.

Xue, L., K. Yao, et al. (2010). "Serotype distribution and antimicrobial resistance of Streptococcus pneumoniae isolates that cause invasive disease among Chinese children." Clin Infect Dis **50**(5): 741-744.

Yip, V. L., A. Varrot, et al. (2004). "An unusual mechanism of glycoside hydrolysis involving redox and elimination steps by a family 4 beta-glycosidase from

Thermotoga maritima." Journal of the American Chemical Society **126**(27): 8354-8355.

Zhang, J. R., K. E. Mostov, et al. (2000). "The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells." Cell **102**(6): 827-837.

Zhu, Y. and E. C. Lin (1989). "L-1,2-propanediol exits more rapidly than L-lactaldehyde from Escherichia coli." J Bacteriol **171**(2): 862-867.

Zopf, D. and S. Roth (1996). "Oligosaccharide anti-infective agents." Lancet **347**(9007): 1017-1021.

Appendix: Components of the pneumococcal fucose utilization operons

Table 15: The regulatory protein and fucose processing enzymes of the fucose utilization operons.

Protein	Operon Type	Locus Tag (TIGR4/SP3-BS71)	Homolog (% identity)	Cellular location[2]	Putative function	Reference[1]
FcsR	1 and 2	Sp_2168/ CGSSp3BS71 _10443	*S. mutans* LacR (29%)	Cytoplasm	Negative regulator	(Rosey and Stewart 1992)
FucK	1 and 2	Sp_2167/ CGSSp3BS71 _10438	*E. coli* RhaB (31%)	Cytoplasm	Fuculose kinase	(Grueninger and Schulz 2006)
FucA	1 and 2	Sp_2166/ CGSSp3BS71 _10398	*E. coli* FucA (41%)	Cytoplasm	Fuculose phosphate aldolase/epi merase	(Dreyer and Schulz 1996)
FucU	1	Sp_2165	*E. coli* FucU (45%)	Cytoplasm	Fucose mutarotase	(Ryu, Kim et al. 2004)
FucI	1 and 2	Sp_2158/ CGSSp3BS71 _10393	*E. coli* FucI (60%)	Cytoplasm	Fucose isomerase	(Green and Cohen 1956)

[1] Literature reference for properties of the homolog

[2] Predictions performed by PSORT (Gardy, Laird et al. 2005)

Table 16: The carbohydrate transport proteins of the fucose utilization operons

Protein	Operon Type	Locus Tag (TIGR4/SP3-BS71)	Homolog (% identity)	Cellular location[2]	Putative function	Reference[1]
EIIA	1	Sp_2164	*E. coli* EIIA-mannose (27%)	Cytoplasm	PTS transport soluble component	(Nunn, Markovic-Housley et al. 1996)
EIIB	1	Sp_2163	*E. coli* EIIB-mannose (33%)	Cytoplasm	PTS transport soluble component	(Gschwind, Gemmecker et al. 1997)
EIIC	1	Sp_2162	*E. coli* EIIC-mannose (32%)	Membrane	PTS transport membrane component	(Erni and Zanolari 1985)
EIID	1	Sp_2161	*E. coli* EIID-mannose (40%)	Membrane	PTS transport membrane component	(Erni and Zanolari 1985)
PBP	2	CGSSp3BS71_10433	*O. iheyensis* ABC sugar transport sugar-binding protein (24%)	Membrane /cell wall	Periplasmic binding protein	(Higgins, Abbott et al. 2009)
ABC-a	2	CGSSp3BS71_10428	*S. elongatus* LacF (26%)	Membrane	ABC Transport Permease	(Maeda, Sugita et al. 2006)
ABC-b	2	CGSSp3BS71_10423	*S. elongatus* LacG (29%)	Membrane	ABC Transport Permease	(Maeda, Sugita et al. 2006)

[1] Literature reference for properties of the homolog

[2] Predictions performed by PSORT (Gardy, Laird et al. 2005)

Table 17: The glycoside hydrolases of the fucose utilization operons

Protein	Operon Type	Locus Tag (TIGR4/SP3-BS71)	Homolog (% identity)	Cellular location[2]	Putative function	Reference[1]
GH95	1	Sp_2160	*B. bifidum* α-fucosidase (26%)	Cytoplasm	α-fucosidase	(Katayama, Sakuma et al. 2004)
GH98	1	Sp_2159	*C. perfringens* endo-β-galactosidase (34%)	Secreted	Endo-β-galactosidase	(Anderson, Ashida et al. 2005)
GH29	2	CGSSp3BS71_10418	*Thermotoga maritima MSB8* α-L-fucosidase (21%)	Cytoplasm	α-L-fucosidase	(Sulzenbacher, Bignon et al. 2004)
GH36A	2	CGSSp3BS71_10413	*C. perfringens* α-N-acetylgalactosaminidase (58%)	Cytoplasm	α-N-acetylgalactosaminidase	(Calcutt, Hsieh et al. 2002)
GH36B	2	CGSSp3BS71_10408	*G. stearothermophilus* α-galactosidase (46%)	Cytoplasm	α-galactosidase	(Ganter, Bock et al. 1988)
GH98	2	CGSSp3BS71_10403	*C. perfringens* A/B-antigen cleaving endo-beta-galactosidase (60%)	Secreted	A/B-antigen cleaving β-galactosidase	(Anderson, Ashida et al. 2005)

[1] Literature reference for properties of the homologue

[2] Predictions performed by PSORT (Gardy, Laird et al. 2005)