Carbohydrate Processing: An Indispensable Platform of Pneumococcal Virulence

by

Edward Peter William Meier

B.Sc., (Honours) Biochemistry and Molecular Biology,
University of Northern British Columbia, 2017

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

MASTER OF SCIENCE

in the Department of Biochemistry and Microbiology

Supervisory Committee

Carbohydrate Processing: An Indispensable Platform of Pneumococcal Virulence

by

Edward Peter William Meier

B.Sc., (Honours) Biochemistry and Molecular Biology,
University of Northern British Columbia, 2017

Supervisory Committee Members

Dr. John E Burke, (Department of Biochemistry and Microbiology)
Supervisor

Dr. Alisdair Boraston, (Department of Biochemistry and Microbiology)
Departmental Member

Dr. Fraser Hof (Department of Chemistry)
Outside Member

Abstract

Carbohydrate processing is fundamental for *Streptococcus pneumoniae* (pneumococcus) colonization and it's also a formidable platform of its virulence. Many of the bacterium's key carbohydrate processing enzymes target different glycans produced by the human host. However, pneumococcus can also utilize a limited number of plant derived carbohydrates known as dietary saccharides. This thesis focuses on characterizing both structural and biochemical properties of four proteins involved in pneumococcal carbohydrate processing. One area of focus surrounds the pneumococcal *exo*-α1,2-mannosidase (SpGH92) which is essential for degrading host High-Mannose N-glycans (HMNGs) and is a major virulence determinant in models of sepsis and pneumonia. Another component of this thesis focuses on the virulence-associated dietary saccharide utilization locus that is required for raffinose family oligosaccharide (RFO) utilization.

The dietary saccharide utilization locus known as the *raf* locus belongs to the pneumococcal core genome, found in over 98% of clinical isolates. We hypothesize that if RFO utilization is important for pneumococcal infection, the enzymes responsible for RFO recognition and degradation will exhibit a strong affinity towards their putative RFO substrates. In this thesis work three crystal structures of the substrate binding protein RafE bound to the RFOs raffinose, stachyose and verbascose have been determined. In addition, this work has characterized several biochemical properties of substrate binding protein RafE; and the two glycoside hydrolases Aga and GtfA. Overall, it was found that the *raf* locus contains the biological machinery required for RFO utilization. However, all of the studied proteins show biochemical inefficiencies towards their putative targets.

On another note, the *S*. *pneumoniae* Glycoside Hydrolase family 92, or SpGH92, also belongs to the pneumococcal core genome, and it is the only known *exo*-α1,2-mannosidase encoded by *S. pneumoniae*. Due to its profound impact on pathogenesis, SpGH92 has high potential to be an alternative therapeutic target in future vaccine and drug trials. The work presented here describes a protocol utilizing Hydrogen-Deuterium-eXchange Mass-Spectrometry (HDX-MS) and Differential Scanning Fluorimetry (DSF) that effectively uncovers changes in both the heat stability and conformational dynamics of SpGH92 in response to different inhibitors binding. By utilizing a characterized α-mannosidase inhibitor, mannoimidazole, we demonstrate the

feasibility of detecting inhibitor binding interfaces to SpGH92, and compare the changes with a novel inhibitor. Overall, this body of work has uncovered novel structural insights into the virulence factor controlling the deconstruction of mammalian HMNGs and it has set the framework for future SpGH92 inhibitor screening assays using HDX-MS and DSF.

In summary, this thesis sheds light on the diverse roles of carbohydrate processing proteins belonging to the deadly pathobiont *Streptococcus pneumoniae*, thus contributing to the discovery and development of future therapeutics against pneumococcal infections.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ABC-transport	ATP binding cassette transporters
CAZymes	Carbohydrate active enzymes
C3	Complement component 3
GH13	Glycoside hydrolase family 13
GH36	Glycoside hydrolase family 36
GH92	Glycoside hydrolase family 92
HDX-MS	Hydrogen-Deuterium eXchange Mass-Spectrometry
ITC	Isothermal titration calorimetry
LTA	Lipoteichoic acid
NCSAPs	Non-classical surface-associated proteins
PCV13	13-valent pneumococcal conjugate vaccine
PCV7	7-valent conjugated pneumococcal vaccine
PHiD10	10-valent Pneumococcal-Haemophilus Influenzae-D surface protein
PPSV23	23-valent polysaccharide pneumococcal vaccine
PTS	Phosphotransferase transport system
PUL	Polysaccharide utilization locus
RGD	Arginine-glycine-aspartic acid integrin binding motif
SBP	Substrate binding protein
TIGR4	Streptococcus pneumoniae serotype 4
TLC	Thin-layer chromatography

Carbohydrates

AATGal	2-acetamido-4-amino-2,4,6-trideoxygalactose
F6P	Fructose-6-phosphate
FbP	Fructose-bisphosphate
Fru	Fructose
Gal	Galactose
GalNAc	N-acetyl-galactosamine
Glc	Glucose
Glc-1P	Glucose-1-phosphate
Glc-6P	Glucose-6-phosphate
GlcNAc	N-acetyl-glucosamine
HMNGs	High-mannose N-glycans
LacNAc	N-acetyl-lactosamine
Man	Mannose
MurNAc	N-acetyl-muramic acid
PCho	Phosphorylcholine
pNP	*para*-Nitrophenyl
RFOs	Raffinose family oligosaccharides
UDP-sugar	Uridine diphosphate - sugars

Acknowledgements

The path of graduate school, like so many aspects of science, is a collaborative effort that cannot be done without the support of others.

Foremost, I would like to thank my advisor, Professor John Burke, for his guidance, support and patience throughout my completing of this thesis. I am very grateful for your contagious enthusiasm and passion for research. It has been inspiring working with you, and I have learned so such in this short time.

I would also like to thank my committee members, Professors Alisdair Boraston and Fraser Hof, for their helpful discussions and assistance throughout this thesis. Cheers to both of you!

I would like to express my appreciation for all the members of the Burke, Boraston, Cameron, and other labs who have helped or supported me throughout this journey. Thank you, Brandon, for helping me learn the ropes of HDX-MS and for always being up for a round of tennis. Thank you, Ben and Andy, for all of your guidance in learning the ropes of X-ray crystallography, and of course for sharing some good brews and laughs at Fels. I would like to thank the Vocadlo lab for providing the (surprisingly, not so) novel inhibitor that was crucial to parts of this thesis.

I am extremely grateful for all of the fellow students and faculty that I am lucky to call my friends, thank you all for being a circle of support and laughter throughout this journey.

A special thank you to my friends and family, your support and encouragement gave me the strength to persevere through the hardest times. I would also like to thank my exceptionally furry family members Dexter and Tiger-lily for their infectious happiness, funny personalities, and constant cuddles.

Lastly, I would like to thank my caring partner, Maya, with your constant love and support anything is possible.

CHAPTER 1: Introduction

1.1 *Streptococcus pneumoniae*: Past, present and future

1.1.1 A brief history of pneumococcus

Historically, *Streptococcus pneumoniae* (the pneumococcus) was one of the leading causes of death around the world. Though, despite many advances it remains a significant cause of morbidity and mortality in children and elderly today. The progression of pneumococcal research is a fascinating tale filled with many amazing scientific breakthroughs[1]. One of the first published documents dates back to 1881 when the pathogenic potential of pneumococcus was first recognized by two independent microbiologists, Louis Pasteur in France; and George Sternberg in America[1,2]. In their early experiments, both researchers injected human saliva into healthy rabbits and subsequently recovered diplococci bacteria growing in the blood of the diseased animals. Approximately twenty years later Fred Neufeld, a German bacteriologist and physician, discovered different "types" of pneumococcus[3]. By using different bacterial isolates, Neufeld demonstrated that only specific rabbit antiserums were effective against select pneumococcal isolates. Furthermore, when he examined the bacterium under a light microscope, he observed that the bacteria would swell up and rupture, thus the reaction was named the Quellung reaction after the German word for swelling. Amazingly, these were the first of many serotyping experiments that are still used today and have led to the identification of over 100 unique serotypes to date. Knowing that isolates can be serologically distinct from one another, in 1928 the bacteriologist Frederick Griffith demonstrated the process of bacterial transformation[4]. Using non-virulent (unencapsulated or "rough" looking) strains of pneumococcus, and highly virulent (capsulated or "smooth") serotypes of pneumococcus, Griffith demonstrated that a non-virulent type, once incubated in the supernatant of a heat-killed virulent type, would adopt a similarly lethal phenotype while simultaneously changing its appearance from rough to smooth[4]. Still, arguably one of the greatest molecular biology discoveries involving pneumococcus occurred nearly sixteen years later in the year 1944. This is the year of the famous Avery-MacLeod-McCarty experiment that established the "hereditary material" in cells is DNA and not RNA or proteins,

thus breaking the seal on modern biology and establishing the central dogma of heredity[5]. Almost 10 years later this revolutionary discovery was driven home with the discovery of the double helix using X-ray crystallography, and consequently the world of structural biology was ignited[6].

1.1.2 Colonization, transmission and disease

Almost 140 years since stepping into the spotlight, pneumococcus has become one of the best studied bacterial pathobionts. Mountains of evidence undeniably demonstrates that pneumococcus is a part of the normal human nasopharyngeal microbiome[1,7,8]. Moreover, this data indicates that approximately 27-95% of children and up to 10% of adults are carriers depending on their geographical location. Luckily, in most colonized individuals pneumococcus will not cause disease. Instead, pneumococcus remains in an asymptomatic carriage state with the host, and these carriers are the principal reservoirs for transmission[7].

Transmission is known to occur from direct contact with micro-aerosol droplets produced by sneezing and coughing, the sharing of food items, or by person-person oral contact[7]. Likely owing to the increased levels of mucociliary flow and excretion, the rates of transmission increase in drier and colder months, and in conjunction with viral infections of the upper respiratory tract [9,10]. In addition to increased transmission, viral infections of the upper respiratory tract also increase the risk of developing pneumococcal disease.

In the small percentage of individuals who develop pneumococcal disease, the disease severity can range from a mild infection to a severe life-threatening illness. The infections occur when pneumococcus translocates to sterile tissues and organs of the body[9]. Pneumococcus is known to cause several non-invasive diseases including non-bacteremic pneumonia, acute sinusitis and otitis media[11]. These forms of disease are especially common in children where acute otitis media is one of the most common clinical presentations in the USA[12]. However, pneumococcus can also cause life-threatening invasive diseases including bacteremic pneumonia, meningitis, and sepsis (Figure 1-1). Generally, the population at high risk for developing invasive pneumococcal disease are those whom do not produce a robust T cell mediated immune response such as children (0-59 months), the elderly (65+) and immunocompromised individuals[13].

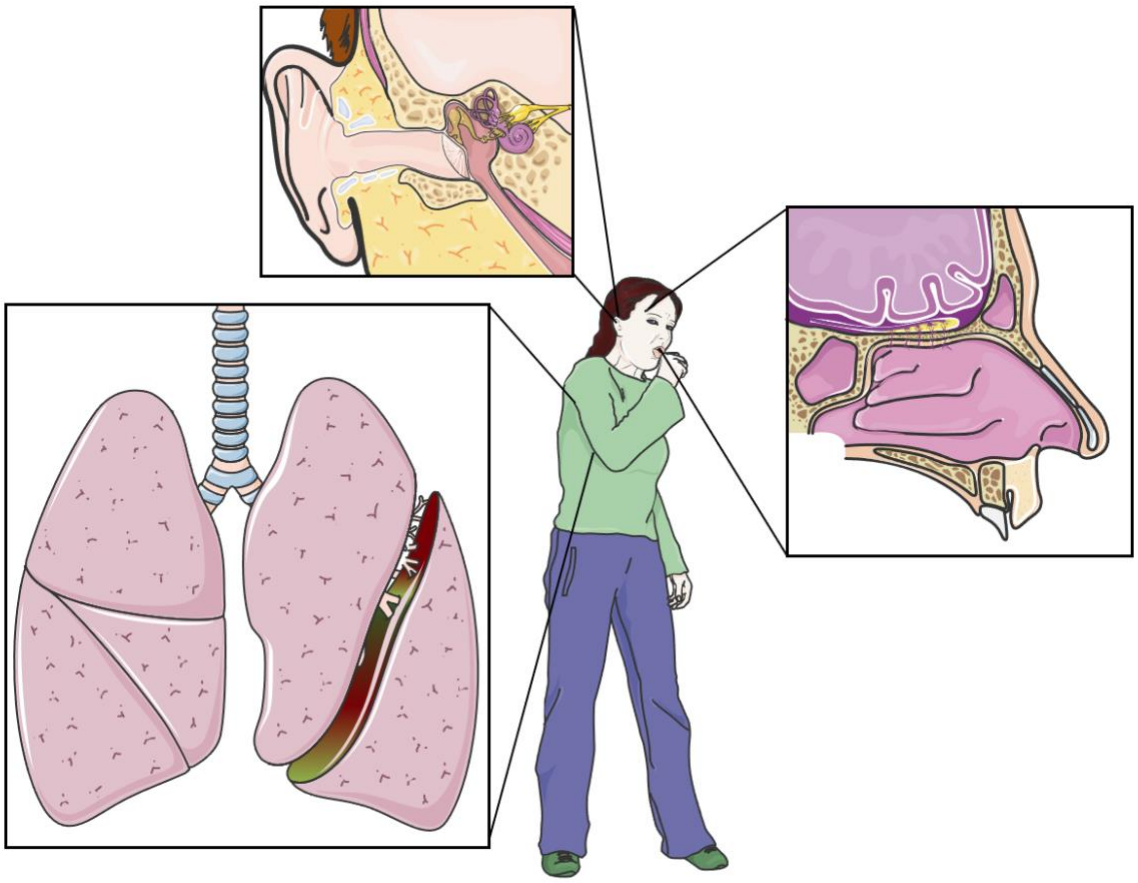


Figure 1-1. Pneumococcal disease. A staggering proportion of human nasal cavities are colonized by *S. pneumoniae* and although the majority of individuals will never experience disease, a small fraction will experience forms of invasive and non-invasive diseases. Forms of non-invasive disease includes acute inflammation in the nasal cavity (sinusitis), inner ear (otitis media) or lungs (pneumonia). On the other hand, invasive pneumococcal disease is a medical emergency and a leading cause of morbidity and mortality worldwide. Invasive disease occurs when pneumococcus triggers high levels of inflammation within sterile sites of the body such as the blood and organs, meninges, pleural fluid, joint fluid and pericardial fluid; additionally, meningitis can occur through retrograde axonal transport through olfactory neurons[14]. Images adapted from smart.servier.com (Creative Commons Attribution 3.0).

1.1.3 The global burden of pneumococcus

The socioeconomic cost associated with pneumococcal disease is massive. Global estimates indicate that prior to the 21st century, more than 10 million deaths arose annually due to invasive

pneumococcal disease[15]. It wasn't until the implementation of the first conjugated pneumococcal vaccine, at the beginning of the 21st century, that the number of casualties dropped between 1-2 million people annually[16]. However, even with the dramatic decline in recent years, young children from developing countries are still at high risk for succumbing to invasive disease. In Africa alone, between 1 and 4 million cases of pneumococcal pneumonia are estimated to occur annually, with roughly 800,000 child deaths caused by invasive disease[17]. Moreover, a large proportion of children who suffer from invasive disease do not present with pneumococcal pneumonia. Instead many children present with meningitis and sepsis which occurs upon passage through the nasal tissue or olfactory nerves[9,18]. In contrast, adults most commonly present with bacteremic pneumococcal pneumonia as the clinical presentation for pneumococcal disease[9,19].

Estimates show in Canada roughly 2,500 hospitalizations are attributed to invasive pneumococcal disease annually, often occurring in conjugation with a viral infection, and the average mortality rate is between 5.2% to 7.8% for those admitted[20]. In the USA approximately 150,000 hospitalizations result from pneumococcal infections annually, and the fatality rates are estimated to be between 5% to 7% according to the CDC[19]. Due to the implementation of several vaccines the rates of pneumococcal disease seen in Canada, the USA, as well as other countries across the world have decreased substantially since the beginning of the 21st century[21]. Still, the incidence of invasive disease involving non-vaccinated serotypes and the rates of antibiotic resistance in new isolates is continuing to rise at an alarming rate[22]. Despite best efforts, pneumococcus is still predicted to kill over half a million children under five years old each year, leaving countless others permanently effected[23–25]. Accordingly, numerous researcher groups worldwide are working to identify key pneumococcal virulence factors, such as cell wall components and extracellular proteins that may be included in the next generation of vaccine and drug developments.

1.1.4 Pneumococcal treatments and preventions

Several vaccines have been developed that target the pneumococcal capsule, thereby protecting against pneumococcal colonization and infection from specific serotypes[13,26]. The

pneumococcal capsular polysaccharide was first recognized as an immunogenic substance that elicits protection against infection in the early 20th century[27]. However, it was not until the mid to late 20th century that the first widely used pneumococcal vaccines, namely the 13-valent (PPSV13) and 23-valent (PPSV23) polysaccharide pneumococcal vaccines, were developed[28,29]. The PPSV23 was introduced in Canada in 1983, and it covered the twenty-three most common serotypes known to cause severe disease in the elderly and immunocompromised populations. Unfortunately, because the PPSV23 vaccine invokes a weak T cell response, it has a low immunogenic response in children. Thus, the first two vaccines created were largely ineffective in protecting this vulnerable age group[30]. Thankfully, between 2002 and 2006 the 7-valent pneumococcal conjugate vaccine (PCV7) was introduced in Canada. In the conjugate vaccines the pneumococcal capsule is covalently attached to an inactive protein, namely the diphtheria toxin for PCV7, and this results in a robust immune response thereby providing sufficient protection in children[31]. The implementation of PCV7 resulted in approximately a 90% decrease in the incidence of invasive disease in children under five[30,31]. However, a few years after its introduction the rates of pediatric disease began re-escalating with the emergence of several non-vaccine, multi-drug resistant strains[32]. This led to the release of the 10-valent Pneumococcal-*Haemophilus Influenzae*-D surface protein (PHiD10) vaccine that was introduced in 2010. The PHiD10 vaccine covered all of the PCV7 serotypes plus three emerging serotypes that were conjugated to the *H. influenzae* D protein[12]. Though, this vaccine was not widely implemented as less than a year after its release the 13-valent pneumococcal conjugate vaccine (PCV13) was introduced. PCV13 offers a strong protection against the PHiD10 serotypes, plus an additional three serotypes that were rising in disease incidence at the time of its release, and it is still used in most global vaccination programs today[16].

Nearly 10 years have passed since our last pneumococcal vaccine upgrade. However, there are currently several vaccines in clinical trials that are being prepared to defend against the next wave of emerging serotypes. Most notably are the pneumococcal conjugate vaccines PCV15 and PCV20 that are currently both in phase 3 clinical trials and include the emerging serotypes 22F and 33F which are both associated with multidrug resistance and together constitute nearly 20% of the invasive disease cases of children in the USA[11,33]. In addition, several inactivated whole-

cell and protein-based vaccines are in Phase 1 and 2 clinical trials that may confer a broad spectrum of protection against multiple pneumococcal antigens[34].

In addition to capsular replacement, pneumococcus readily acquires antibiotic resistance genes through transformation[30]. In 2017 the World Health Organization officially included *Streptococcus pneumoniae* as one of the 12 priority pathogens for which new antibiotics are required[35]. This comes to light as recent estimates show between 20% and 95.8% of new pneumococcal isolates contain one or more antibiotic resistance genes depending upon the geographic location[36,37]. Several non-vaccine serotypes are increasing in incidence and antibiotic resistance to penicillin (29%), chloramphenicol (5%), erythromycin (11%), cotrimoxazole (39%) and tetracycline (14%); additionally, almost 10% of non-vaccine serotype isolates are multi-drug resistant[38]. Put into the context of increasing antibiotic resistance and emerging serotypes associated with invasive disease, our current vaccine strategies may only provide a short-term solution in preventing pneumococcal disease. Therefore, recent efforts are directed towards discovering new proteins that are ubiquitous throughout the majority of pneumococcal strains that hold therapeutic potential. One such area of interest, that is the focus of this thesis, is the research field of carbohydrate processing.

1.2 Carbohydrates processing: An indispensable platform of pneumococcal pathogenesis

The remainder of this thesis is focused on pneumococcal carbohydrate processing which is an incredibly important aspect of its survival and pathogenesis. Thus, the remainder of Chapter 1 will introduce key concepts relating to pneumococcal carbohydrate processing. In Chapter 2 of this thesis, I will describe the work relating to the biochemical characterization of several proteins within the virulence associated dietary saccharide processing locus known as the *raf* locus. Following this, in Chapter 3 I will describe a methodology that utilizes Hydrogen Deuterium eXchange Mass-Spectrometry (HDX-MS) to rapidly characterize the binding interfaces of inhibitors with a major virulence determinant known as SpGH92. Therefore, going onward from

here, an understanding of carbohydrates and their nomenclature will aid in understanding the information presented in this thesis.

1.2.1 Carbohydrate structures and nomenclature

Carbohydrates are extremely abundant organic molecules that range from rather simple monomers (monosaccharides) to complex oligo- and poly-saccharides[39]. The name carbohydrate originates from their general chemical composition, $C_n(H2O)_n$, that consists primarily of carbon (carbo-) and water (hydrates). Several distinguishing characteristics of carbohydrates that contribute to their uniqueness and heterogeneity include the relatively large number of structurally distinct monomer building blocks (3 to 9 carbon atoms chains) that form covalent linkages; whether the hydroxyl group stereochemistry at each carbon atom is axial or equatorial (Figure 1-2E); the range of functional group modifications occurring at the different hydroxyl groups (Figure 1-2CF); the ability to form glycosidic bonds with one or more carbohydrate and non-carbohydrate structures (Figure 1-2DEH-M); and the linkage type formed by the anomeric carbon (α or β) (Figure 1-2BD). The numbering system of carbohydrates is relative to their anomeric carbon which signifies position 1, or C1; likewise, hydroxyl groups are numbered according to their respective carbon atoms (Figure 1-2A). Carbohydrates lacking a glycosidic bond at their anomeric carbon can alternate between a linear and cyclic forms and have been shown to transfer between α- and β- conformations (Figure 1-2AB)[40].

1.2.2 Carbohydrate structures targeted by *S. pneumoniae*

Notably, enzymes that are active towards carbohydrates are often very selective towards the specific type of carbohydrate (galactose, glucose, fructose, etc.), the linkage type(s), the extent of branching and the total length. Thus, microbes isolated from different ecological niches are adapted to the particular carbohydrate structures present within their specific niches[41]. Consequently, many of the carbohydrate processing enzymes that are integral to pneumococcal colonization and pathogenesis are involved in human glycan degradation, though several plant sugars are also utilized. Many of the enzymes involved in carbohydrate processing are also

implicated in virulence and several direct examples are discussed in more detail in Chapters 1.5, 1.9, 2 and 3.

1.2.2.1 Host glycans

In humans, glycans are extremely diverse and abundant carbohydrate structures present on the surface of all cells. Glycans are covalently linked to lipids and proteins, and form structures generally referred to as glycoconjugates. Furthermore, glycan expression and composition are incredibly distinct between different taxonomic groups, and can even vary between organisms of the same species in response to different states of health, genetics, and age[42]. With regards to humans, a large proportion of extracellular proteins are glycosylated and these glycans serve important functional roles in protein recognition, adherence, signaling, protection, and stabilization[43]. With this in mind, perhaps it is unsurprising that carbohydrate processing plays an enormous role in pneumococcal pathogenesis. Several enzymes involved in degrading complex N-linked (Figure 1-2J), high-mannose N-linked glycans (Figure 1-2H), as well as O-linked glycans (Figure 1-2L) are essential for colonization and virulence, with evidence indicating a role in preventing complement mediated clearance[44–47].

1.2.2.2 O-linked glycans

O-linked glycosylation occurs between a carbohydrate moiety and a lipid or a hydroxyl group belonging to a serine or threonine residue. O-glycans commonly vary between a single monosaccharide to more than twenty diverse sugar additions (or up to 120,000 in the case of glycogen). The structures of O-linked glycans are extremely diverse[48–50]. In fact, several classes of O-linked glycosylations exist based upon the carbohydrate initiating glycosylation and this includes O-fucosylation, O-glucosylation, O-GalNAcylation, O-GlcNAcylation, O-mannosylation, and O-xylosylation[49]. An example of O-glycosylation that is related to pneumococcus colonization involves a group of heavily glycosylated proteins known as mucins. Mucins constitute an important component of our innate immune defense called mucus. However, pneumococcus can effectively degrade the glycans found in mucus resulting in nutrient acquisition and facilitating the exposure of host cell receptors[51].

1.2.2.3 N-linked glycans

N-linked glycosylation involves the covalent linkage of an N-acetyl-glucosamine (GlcNAc; Figure 1-2C) to a protein at an asparagine residue. In humans and other eukaryotes, all N-glycans begin with a GlcNAc-β1,4-asparagine linkage and share a common N-glycan core structure ($Man_3GlcNAc_2$) with varying branching patterns (Figure 1-2HJK). Importantly, pneumococcus is capable of degrading the N-glycan structures found throughout our bodies and several N-glycan degrading enzymes are vital for its colonization and pathogenesis[40,47,51,52]. An interesting area of research which has yet to be elucidated is how these enzymes are reducing complement mediated clearance during hematogenous spread[53]. Indeed, many human complement pathway proteins are glycosylated, and it is feasible that their glycans are a target of *S. pneumoniae*; however, this has yet to be shown directly. Notably, the N-glycans known as High-Mannose N-Glycans (Figure 1-2H) are a target of the virulence factor, SpGH92, that is the focus of Chapter 3.

1.2.2.4 Host dietary saccharides

In addition to host glycans, pneumococcus can breakdown and metabolize a limited number of plant-derived dietary saccharides. Several of the dietary saccharides utilized by pneumococcus include cellobiose, trehalose, sucrose, fructo-oligosaccharides, and raffinose family oligosaccharides (RFOs). Intriguingly, several dietary saccharide utilization pathways are implicated in pneumococcal pathogenesis. For example, cellobiose, a disaccharide comprised of β1,4-glucobiose, is a derivative of cellulose and is incorporated into plant cell walls. Interestingly, signature-tagged mutagenesis studies revealed components of the cellobiose transporter and the locus regulator (CelR) are required for lung infection[54]. Another dietary saccharide, known as fructo-oligosaccharides, are comprised of a sucrose (Glc-α1,2-Fru; Figure 1-2E) base with varying degrees of linear β2,1-linked fructose extensions stemming from the fructose subunit of a sucrose molecule ($[Fru\text{-}\beta\text{-}2,1]_n$-Fru-β2,1-Glc). Signature-tagged mutagenesis studies revealed the substrate binding protein FusA is important for non-invasive pneumococcal ear infections[55]. Furthermore, pneumococci also contain two distinct loci dedicated to sucrose uptake that are identified virulence factors. The *sus* (ABC-transporter) locus and the *scr* (PTS) locus are implicated in both lung and nasopharyngeal colonization, respectively[56].

Integral to this thesis work are the RFOs. RFOs comprise a sugar structure similar to fructo-oligosaccharides containing a sucrose molecule base. However, RFOs possess varying degrees of α1,6-galactosyl extensions emanating from the glucose subunit ($[Gal\text{-}\alpha1,6]_n\text{-}Glc\text{-}\alpha1,2\text{-}Fru$). In nature, RFOs are found in plants and seeds where they serve as a form of energy storage and for enduring cellular stresses like freezing and dessication[57,58]. The most abundant RFOs include the trisaccharide raffinose, the tetrasaccharide stachyose, and the pentasaccharide verbascose (Figure 1-2I). RFO utilization is important for *S. pneumoniae* pathogenesis as half of the genes encoded in the *raf* locus have been identified in signature-tagged mutagenesis and knockout studies as virulence factors[59,60]. Despite this the proteins involved in RFO utilization have not been biochemically characterized. Thus, the underlying mechanism of virulence is poorly understood and the general role of dietary saccharides during pathogenesis, beyond nutrient acquisition, is enigmatic.

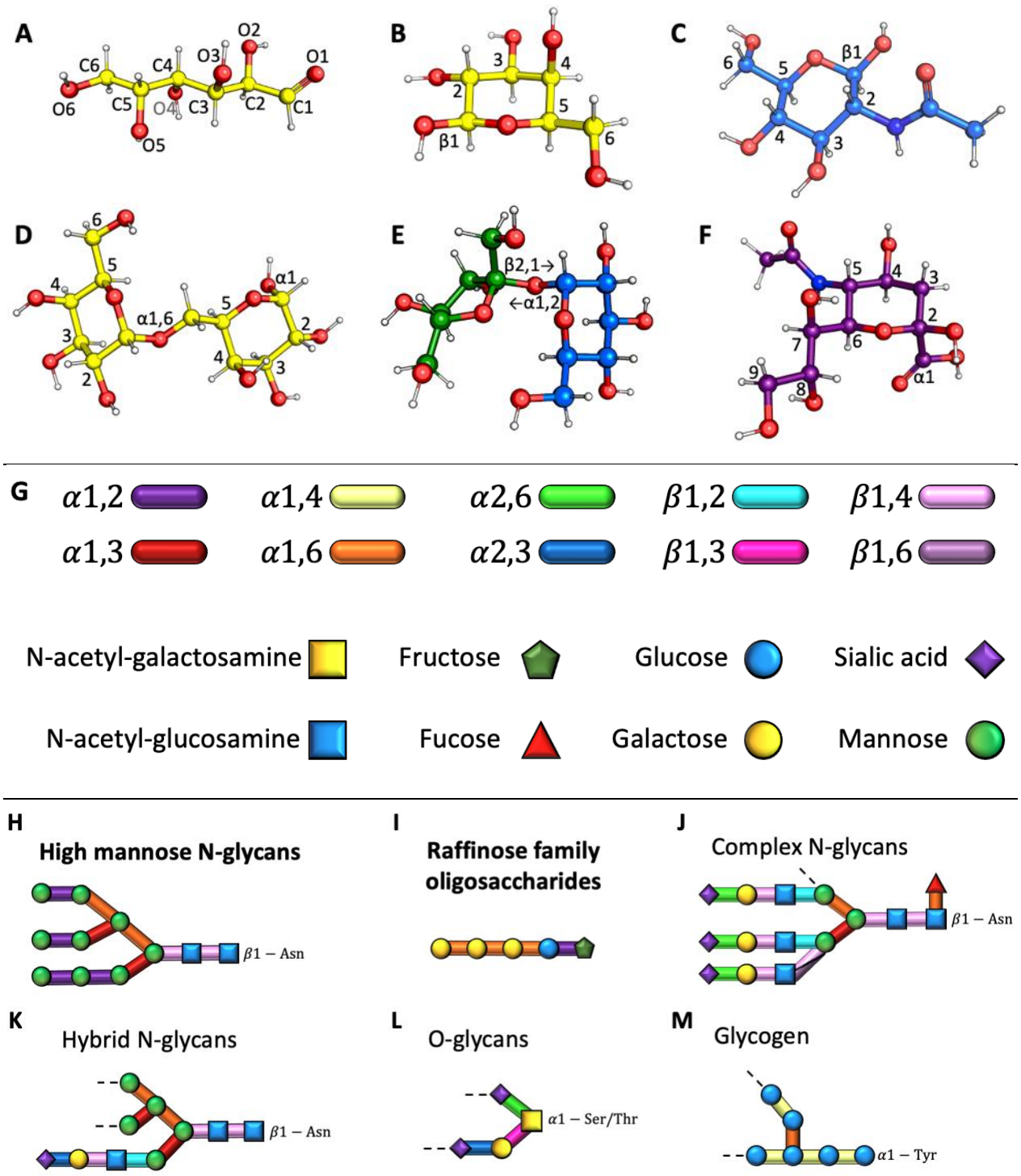


Figure 1-2. Carbohydrate structure, nomenclature, representations and complexity. Carbohydrates are numbered according to their anomeric carbon **(A)** acyclic galactose. **(B)** Cyclic galactose with a β-anomeric carbon as the hydroxyl group is in the equatorial position. **(C)** N-acetyl-glucosamine (GlcNAc); a common sugar modification is N-acetylation, as seen at the C-4

position of GlcNAc. **(D)** α1,6-galactobiose; carbohydrates can form glycosidic linkages between the anomeric carbon of the reducing end sugar and a hydroxyl group of another sugar, as shown with α1,6-galactobiose where the anomeric carbon's hydroxyl group is in the axial position. **(E)** Sucrose comprises a fructose (green) and a glucose (blue) residue, where the sugars are linked at their anomeric carbons, leaving no reducing-end. **(F)** Sialic acid is one of the most complex monosaccharides, containing several functional group modifications. Sialic acids are a common terminally linked sugar found on N- and O-linked glycans. **(G)** To represent sugars simplistically they can be represented using symbols and colored lines, where the colored lines depict specific linkage types, the different shapes represent distinct functional group modifications, and the different shape colors represent dissimilarities in stereochemical arrangement. **(H—M)** Pneumococcus is capable of depolymerizing several important host carbohydrate structures, and dietary saccharides. The bolded carbohydrates (**H** and **I**) represent the two carbohydrate structures which are the targets of the proteins discussed in this thesis.

1.3 Carbohydrate active enzymes

The modification of carbohydrates is fundamental to the survival of *S. pneumoniae*. This is true with regards to its growth, metabolism, replication, and also with regards to subverting the host's immune response[61]. Many of the important carbohydrate biosynthesis and degradation pathways have been illuminated and this includes those involved in the synthesis of the cell wall and capsule[26]; as well as those involved in the breakdown of host glycans and dietary saccharides[62].

Notably, many microbes including *S*. *pneumoniae* possess carbohydrate utilization systems that are organized into polysaccharide utilization loci (PULs)[41]. Often, PULs contain a regulator, a transport system, and one or more carbohydrate active enzymes. Thus, PULs are described as strictly regulated, colocalized gene clusters encoding the required enzymes and proteins required for the utilization of complex carbohydrates. A prime example related to this thesis is the *raf* locus, which contains eight colocalized genes and one detached gene that is also upregulated in the presence of the PULs target substrate (Figure 1-3). The *raf* locus contains two regulators

involved in tightly controlling gene expression; two glycoside hydrolases responsible for substrate depolymerization; an ABC transporter that is required for substrate internalization; and a seemingly out of place lipoteichoic acid ligase, required for lipoteichoic acid formation (discussed further in Chapter 1.5 and Chapter 2).

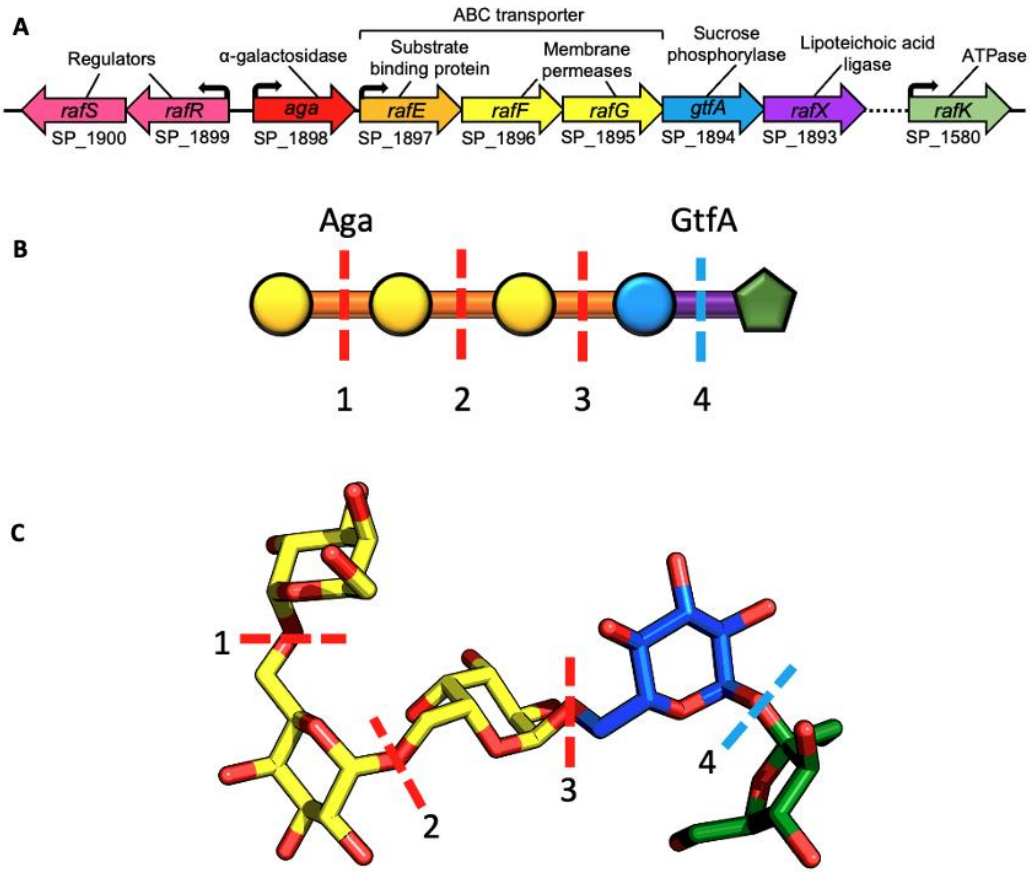


Figure 1-3. Polysaccharide utilization loci layout within microbial genomes. Genes integral to carbohydrate utilization are commonly clustered into polysaccharide utilization loci. **(A)** The pneumococcal machinery for RFO utilization is clustered together into the *raf* locus. The locus is strictly regulated by the repressor and activator (pink) that in the presence of RFOs increases expression of the *raf* locus genes. Internalization of the substrate is facilitated by the ABC transport system (orange, yellow, and green) and depolymerization results from the glycoside hydrolases (red and blue). **(B)** A cartoon symbol representation of verbascose, the RFO pentasaccharide, that once inside the cytoplasm is the α-galactosidase Aga (red) sequentially removes the terminal galactose residues. The resulting sucrose molecule is then cleaved by GtfA

(blue). **(C)** 3D-stick representation of verbascose, reiterating the bonds which are cleaved by both Aga and GtfA.

Over the years, numerous carbohydrate processing enzymes have been characterized and organized into distinct groups based upon their amino acid sequences and functions. The enzymes involved in both creating and destroying the glycosidic linkages holding carbohydrate structures together are referred to as **C**arbohydrate **a**ctive en**zymes** (CAZymes). CAZymes are fundamental instruments for dynamically assembling and degrading all known carbohydrate structures. Thus, CAZyme characterization is an incredibly important area of research. Hence the CAZy database (www.cazy.org) was created in 1991.

The CAZy database is a valuable online resource devoted to organizing CAZymes into different classes, families, subfamilies and clans[63]. The enzymes are first organized into classes based upon their general functions; for example, the class of glycosyltransferases are chiefly responsible for assembling the carbohydrate structures like those found attached to bacterial cell walls and human mucin proteins[50,64]. While those responsible for carbohydrate degradation are either classified into glycoside hydrolases, polysaccharide lyases or carbohydrate esterases with regards to their substrate targets and mechanisms. Notably, some glycoside hydrolases are also involved in carbohydrate biosynthesis, however they are not classified as glycosyltransferases as they are structurally more homologous to other glycoside hydrolases. Within each class, enzymes are assigned to a family based upon their amino acid sequences. The functions within each family can often be quite diverse, therefore enzyme families are further divided into subfamilies according to specific structural motifs and their enzymatic functions. Lastly, as overall 3D structures can be shared between very unrelated amino acid sequences, glycoside hydrolases are grouped into clans to indicate the common 3D structures seen amongst different families. As *Streptococcus pneumoniae* is an adept carbohydrate processor, it contains a repertoire of CAZymes from each forementioned class and examples within the pneumococcal genome are briefly discussed here.

1.3.1 Glycosyltransferases

The class of glycosyltransferases plays an important role in creating the majority of oligosaccharides and glycoconjugates in nature. Firstly, to form glycosidic linkages glycosyltransferases catalyze the transfer of an activated sugar donor, such as nucleotide diphospho-sugars, nucleotide monophospho-sugars or sugar phosphates, to a nucleophilic acceptor molecule which can be a sugar or non-sugar moiety. Glycosyltransferases can possess either retaining or inverting mechanisms relative to the configuration of the donor molecule. The type of mechanism is determined by the architecture of the active site that coordinates the location of the acceptor molecule in relation to the plane of the donor sugar[65]. With regards to the pneumococcal glycosyltransferases, many of its enzymes are involved in deactivating antibiotics, forming oligosaccharides in biofilms, and constructing cell wall repeating units like those in its capsule[65,66]. Currently there are 111 sequence based glycosyltransferases families on the CAZy database and this is the second largest class of studied CAZymes[67].

1.3.2 Polysaccharide lyases

Polysaccharide lyases are another important class of CAZyme that breaks down uronic acid containing polysaccharides. Uronic acids are negatively charged sugars that possess a carboxylic acid functional group modification in their structures resulting from oxidation at the C6 position. Polysaccharide lyases are further distinguished by their β-elimination mechanism that produces an unsaturated hexene-uronic acid and a new reducing end strand upon cleavage[68,69]. Hyaluronate lyase (Hyl), is a pneumococcal polysaccharide lyases that plays an important role in pathogenesis by dismantling the extracellular matrix component, hyaluronan[70]. In humans, hyaluronan is an important polymer found in the extracellular matrix of cells and is composed of repeating glucuronic acid-β1,3-N-acetyl-galactosamine disaccharide units. Hyaluronan provides a highly viscous layer that acts as a diffusion barrier between cells; and by dismantling hyaluronan, pneumococcus increases both membrane permeability and surface exposure, thereby facilitating invasion[71]. Currently there are 40 sequence-based classes for polysaccharide lyases on the CAZy database.

1.3.3 Carbohydrate esterases

Another important class of CAZymes are the carbohydrate esterases. These enzymes catalyze the removal of ester-based functional group modifications including phosphorylcholine, O-acetyl and N-acetyl groups. Ester functional groups are important carbohydrate features that affect both the charge and 3D profile of a given sugar residue. Thus, carbohydrate esterases often facilitate the activities of glycoside hydrolases by removing ester-based functional groups that would otherwise prevent a glycoside hydrolase from cleaving the otherwise esterified carbohydrate. The mechanism of esterase cleavage is not as well described as other CAZymes. However, a Ser-His-Asp catalytic triad mechanism similar to those seen with serine proteases is believed to occur. In this reaction, the nucleophilic serine attacks the ester's carbonyl carbon, resulting in the release of the carbohydrate moiety and formation of a covalent bond with the ester group before a nucleophilic water subsequently displaces the covalent bond between the protein and ester-group[72]. A prime example of an important pneumococcal carbohydrate esterases is the teichoic acid phosphorylcholine esterase (Pce or CbpE). Pce is vital in cleaving the phosphodiester bonds linking phosphorylcholine moieties to the GalNAc residues in its cell surface anchored lipoteichoic and wall teichoic acid structures[73]. Additionally, the peptidoglycan N-acetylglucosamine deacetylase A (PgdA) esterase is a key virulence-factor in pneumococcus. PgdA functions by de-acetylating the GlcNAc residues within peptidoglycan, thereby reducing the effectiveness of the host lysozyme[74]. Currently, there are only 18 sequence-based carbohydrate esterases families on the CAZy database.

1.3.4 Glycoside hydrolases and associated modules

Glycoside hydrolases are arguably the most important class of CAZyme, and they will be the central CAZymes focused on in later chapters of this thesis. Glycoside hydrolases are the largest class of CAZymes and are found within all three domains of life, as well as in viruses. Glycoside hydrolases are responsible for the hydrolysis and rearrangement of glycosidic bonds between carbohydrates, or between carbohydrates and non-carbohydrate moieties. Many mechanisms for glycoside hydrolases exist, though they can generally be described as exo- or endo-acting in relation to where cleavage occurs. When cleavage takes place at either ends of a carbohydrate

structure the enzyme is termed exo-acting; conversely, when cleavage occurs somewhere within the middle of the structure, the enzyme is called endo-acting. Although multiple mechanisms have been described for glycoside hydrolases[75], the two most common are the retaining and inverting mechanisms described in Figure 1-4. Glycoside hydrolases are incredibly important CAZymes within the pneumococcal genome, and numerous glycoside hydrolases that are vital for the colonization and pathogenesis of *S. pneumoniae* are discussed in Chapter 1.5. There are currently 168 amino acid sequence-based families of glycoside hydrolases in the CAZy database.

Interestingly, many glycoside hydrolases also contain carbohydrate-binding modules, which are described as independently folding and functioning CAZyme associated modules. Notably, these modules possess no catalytic activity. Instead, they act as adhesins and potentiate the activity of their associated CAZymes by promoting a close interaction with the target substrate and catalytic domain[76]. It is common for glycoside hydrolases and other carbohydrate degrading enzymes to be attached to one or more carbohydrate-binding modules of the same or different family, thereby targeting different carbohydrates moieties. In summary, CAZymes and their associated modules are key components of pneumococcal pathogenesis. Thus, with its capacity to depolymerize a diverse collection of carbohydrate structures, pneumococcus has dedicated a large percentage of its transport systems to carbohydrate uptake.

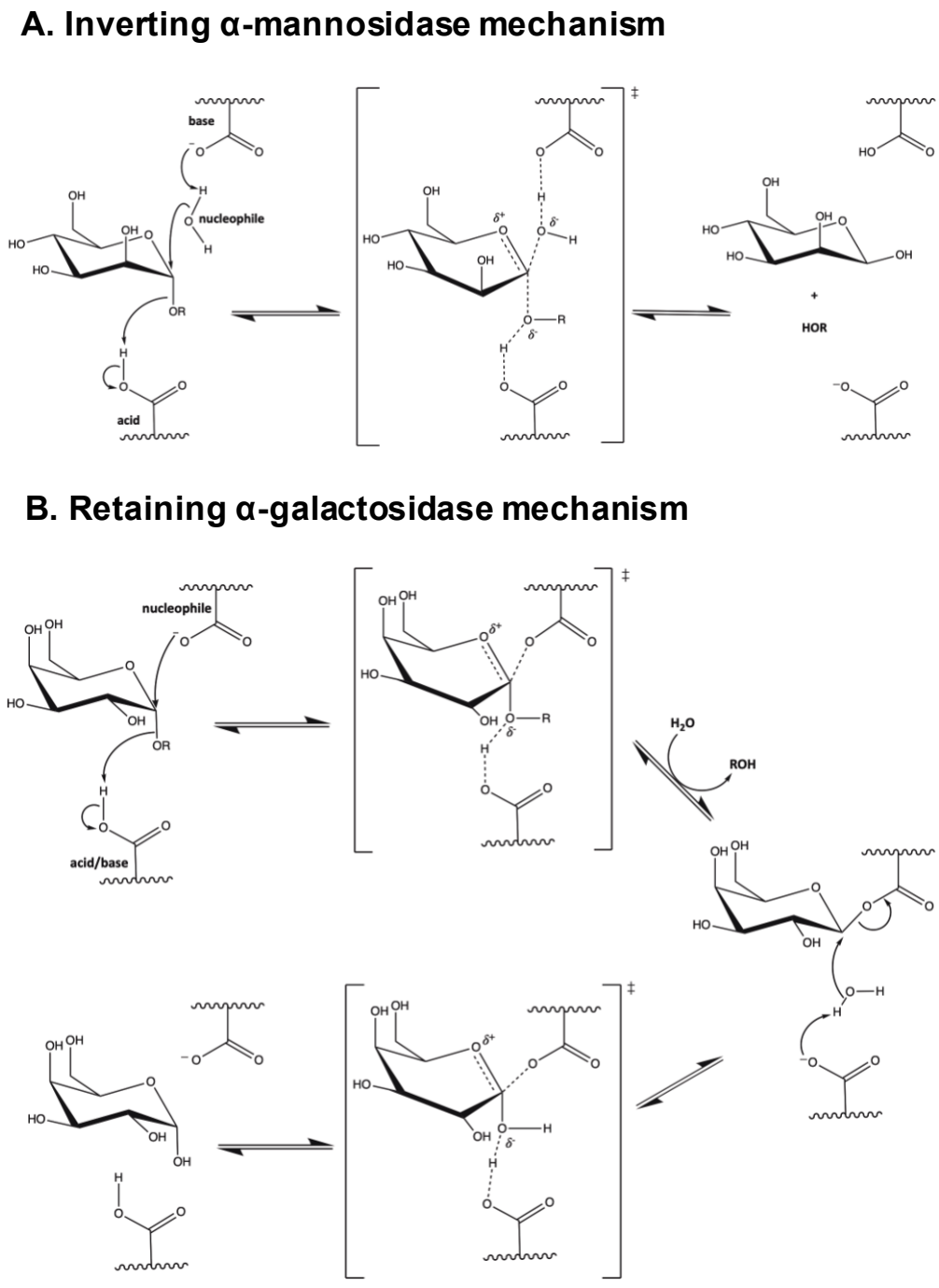


Figure 1-4. Common mechanisms of glycoside hydrolases. (A) The proposed one-step, single-displacement *inverting* mechanism performed by most inverting glycoside hydrolases[40]. In this example the saccharide is representative of an α-mannoside, the target of α-mannosidases like SpGH92 discussed in Chapter 3. **(B)** Representation of a two-step, double-displacement mechanism performed by most *retaining* glycoside hydrolases[77]. The saccharide in this

representation is an α-galactoside, the target of α-galactosidases such as the GH36 enzyme Aga discussed in Chapter 2. Notably, the mechanism of retaining phosphorylases, like GtfA (the GH13 discussed in Chapter 2), only differs by the nucleophilic water introduced after step 2 in **(B)** which is replaced with a nucleophilic phosphate ion[78]. Images were created using ChemDraw18.2.

1.4 Pneumococcal carbohydrate utilization

The pneumococcal genome contains approximately thirty different carbohydrate transporters[79,80]. These transporters primarily belong to one of two structurally and functionally distinct systems, the phosphotransferase transport system (PTS) and the ATP binding cassette (ABC) transport system. The PTS is a multicomponent sugar uptake systems which transport and simultaneously phosphorylates selective monosaccharides, disaccharides, and other sugar derivatives across the bacterial membrane (Figure 1-5)[81]. To perform its phosphorylation-transport function, PTS use the glycolytic intermediate phosphoenolpyruvate as both the phosphoryl donor and energy source. Several proteins are involved in the cascade of events that transfers the phosphoenolpyruvate phosphoryl group to the incoming sugars. The first two phosphotransferases in the pathway, creatively named enzyme I (EI) and heat-stable phosphocarrier protein (HPr), are shared among multiple different PTS systems. EI initiates the cascade by taking the phosphoryl group from the phosphoenolpyruvate, resulting in unphosphorylated pyruvate. Next the phosphoryl group is serially transferred from the cytosolic components EI, to heat-stable phosphocarrier protein (HPr), which then transfers it to enzyme II (EII). EII is composed of the three domains EIIA, EIIB, and EIIC which can be translated separately or as a polypeptide depending on the particular transporter[81]. The phosphate group is first transferred from HPr to EIIA and subsequently to EIIB, before being linked to the incoming sugar residue. Notably, EIIC is crucial for binding the sugar, though it does not appear to form a phosphointermediate[82]. The phosphorylation states of the multiple cytosolic components involved work as indicators of intracellular energy levels. Thus, the PTS is a tool for carbohydrate uptake, but also for gene regulation and metabolism[83].

Relative to PTS systems, ABC transporters can import much larger and more complex carbohydrates. ABC transporters consist of a substrate binding protein (SBP), two transmembrane permease domains and two nucleotide binding domains. The SBP recognizes extracellular carbohydrates or other cellular substrates and presents them to the transmembrane permeases. The two cytosolic ATPases are responsible for binding ATP inducing an "outward-facing" conformational change in the two transmembrane domains, thereby allowing the extracellular carbohydrate partial access into the cell. Consequently, upon ATP hydrolysis the conformation reverses back to the "inward-facing" state facilitating the substrates entry into the cytoplasm. Importantly, the substrate binding proteins encoded within a PUL is a strong indicators of the transporter systems substrate target[84]. Within pneumococcus many carbohydrate transport systems have been characterized that facilitate the uptake of several host-glycoconjugates and dietary saccharides.

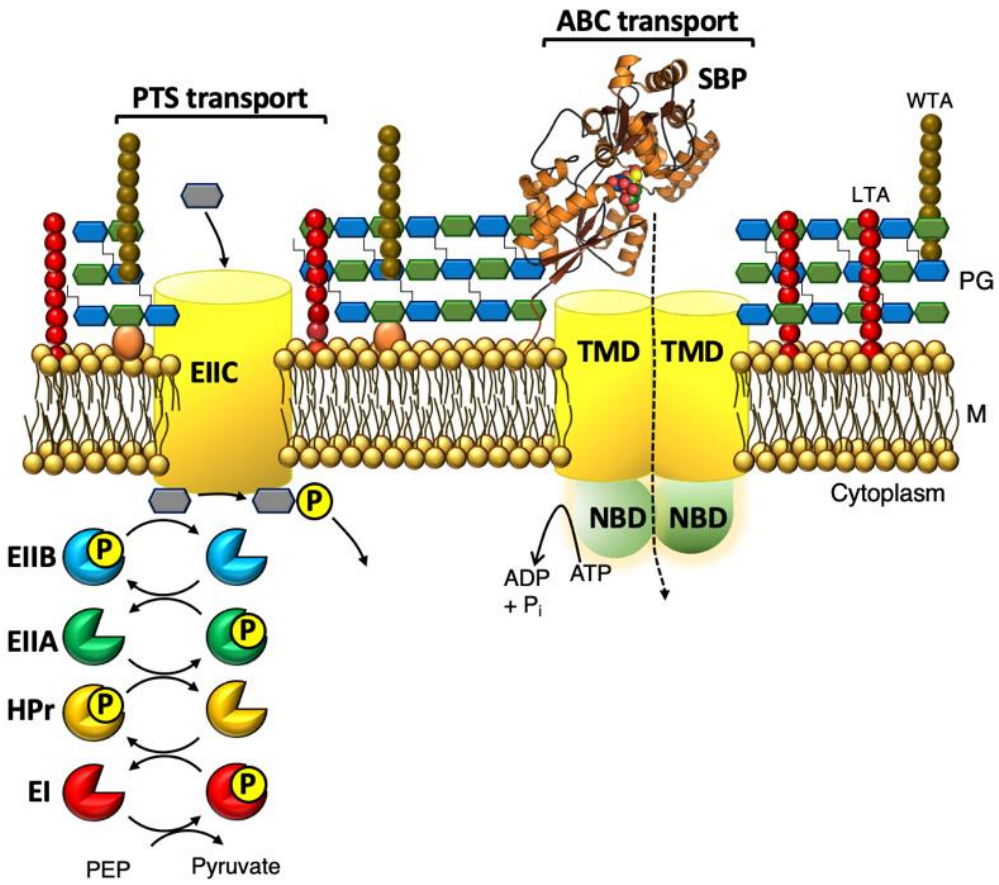


Figure 1-5. Carbohydrate transport in *S. pneumoniae*.

Figure 1-5. Carbohydrate transport in *S. pneumoniae* (continued). Representations of the two main importers of carbohydrates used by pneumococcus, the PTS and ABC transporter organizations. Abbreviation: PEP, phosphoenolpyruvate; SBP, substrate binding protein; TMD, transmembrane domain; NBD, nucleotide binding domain; LTA; lipoteichoic acid; WTA; wall teichoic acid; PG, peptidoglycan; M, membrane.

1.4.1 Carbohydrate metabolism

The ability to utilize multiple carbohydrates with varying compositions and oligomeric states is undeniably a critical survival advantage that increases nutrient acquisition. However, the capacity of carbohydrate utilization also plays an important role in environmental sensing and in modulating transcription[85–88]. Within *S. pneumoniae*, carbohydrates can enter into several different glycolytic pathways to fulfill the demand for ATP and cofactors such as NADH, FADH, pyruvate, and acetyl-CoA. Furthermore, the type of sugar being metabolized dictates what metabolic pathway it will enter, and consequently what impacts on transcriptional regulation will be incurred[86,89,90].

Glucose is one of the most important carbohydrate in pneumococcal polysaccharide synthesis, energy production and energy storage[86,87,91]. Gluco-oligosaccharides destined for catabolism are depolymerized and converted into glucose-6-phosphate (Glc-6P) where they are shuttled into the pentose phosphate pathway, glycolysis and ending in pyruvate metabolism[90]. Notably, the end-product of glucose metabolism is primarily lactate, pyruvate, ATP and NADH[92]. Alternatively, Glc-6P is converted to the biosynthetic precursors Glc-1P or fructose-6-phosphate (F6P). Glc-1P is an important precursor to glycogen and UDP-Glc (uridine diphosphate glucose) synthesis. Notably, UDP-Glc is the only known precursor for producing UDP-Gal in the absence of exogenous galactose and both sugars are critical components in multiple capsule serotypes[26].

The remaining cell wall and capsule components stem from the alteration of F6P[25,87,93]. F6P is a central intermediate of glycolysis, thus multiple sugars types are first converted to F6P before being fully metabolized. Importantly, F6P is also the link to N-acetylated sugar formation[90,94]. In a series of enzymatic reactions, F6P can be converted to UDP-GlcNAc which is an important component of peptidoglycan and several capsules types[26]. Additionally, UDP-GlcNAc can be

transformed into UDP-MurNAc, UDP-FucNAc, UDP-ManNAc, and UDP-GalNAc, thereby facilitating the production of many important cellular structures including teichoic acids, peptidoglycan, and capsular polysaccharides[26].

Lastly, one of the most influential carbohydrates affecting pneumococcal virulence is galactose[85,93]. Galactose uptake *in vitro* is relatively slower than most other sugar types[85], and utilization appears to be significantly regulated by the quorum sensing molecule, autoinducer-2[88]. Galactose is imported by multiple transporters including two PTS and two ABC transporters[79,85]. In PTS uptake, galactose import results in galactose-6-phosphate production which is fed into glycolysis through the tagatose-6-phosphate pathway. Alternatively, uptake through the Gal-specific or the *raf* locus ABC transporters results in unphosphorylated cytosolic galactose which subsequently enters into the Leloir pathway and is converted into UDP-Gal[93]. Surprisingly, galactose uptake results in a poorly described shift in transcriptional regulation and pyruvate metabolism[95,96]. Both pyruvate oxidase (*spxB*) and pyruvate formate lyase (*pfl*) are significantly upregulated in response to galactose[96,97]. Furthermore, their upregulation causes a shift in pyruvate metabolism away from lactate formation and towards mixed acid fermentation, thus leading to an increase in acetyl-CoA production[90,97–99]. It is likely that the increased acetyl-CoA production drives virulence through the production of N-acetylated carbohydrate structures that are essential for many cell wall components including peptidoglycan, teichoic acids and some capsular polysaccharide serotype repeating units.

1.5 *S. pneumoniae* cell surface carbohydrates

1.5.1 Cell wall peptidoglycan

One of the most important structural features of Gram positive bacteria like *S. pneumoniae* is the thick cell wall enveloping the cell membrane[100]. Peptidoglycan is a main component of the cell wall and its chemical structure comprises rows of repeating β1,4-linked N-acetyl-glucosamine (GlcNAc) and N-acetyl-muramic acid (MurNAc) residues. In addition, the rows of repeating -β1,4-GlcNAc-β1,4-MurNAc- are cross-linked together by short oligopeptides originating from the C3-lactic acid group of MurNAc[101]. The size of the oligopeptides range from tri- to penta-peptides, and predominantly form the core tripeptide structure β1,4-MurNAc(C3 lactic acid-L-Ala-D-Gln-L-

Lys) with L-Ser–L-Ala or L-Ala–L-Ala extending from the L-Lys residue within the tripeptide core[102]. Cross-linking occurs between the ϵ-amino group of a lysine residue in one oligopeptide with an L-Ser or L-Ala of an adjacent chain[101]. Ultimately, the extensive crosslinking between different rows of peptidoglycan forms a strong 3D mesh surrounding the pneumococcal plasma membrane providing strength to the membrane.

1.5.2 Cell surface teichoic acids

Another critical component of the pneumococcal cell surface are its cell wall and membrane anchored teichoic acids[101]. Teichoic acids are complex carbohydrate structures that are found in Gram positive microbes. In *S. pneumoniae* teichoic acids comprise five to seven repeating units of -(PCho-O6-)GalNAc-α1,3-(PCho-6O-)GalNAc-β1,1-ribitol-5-phosphate-6-Glc-β1,3-AATGalp-α1,4-[64,101]. Distinguishing features of pneumococcal teichoic acids include the rare amino-sugar, 2-acetamido-4-amino-2,4,6-trideoxygalactose (AATGal), that are found at the reducing end of each repeating unit. Additionally, pneumococcal teichoic acids contain multiple phosphorylcholine (PCho) functional groups attached to the GalNAc residues found at the non-reducing end of each repeating unit. Importantly, both wall and membrane bound teichoic acids are essential docking points for vital choline binding proteins that influence various aspects of pathogenesis including cell wall remodeling and host receptor binding (discussed in Chapter 1.6.3)[103]. Notably, teichoic acids are very antigenic, similar to lipopolysaccharide structures in Gram-negative bacteria, and need to be shielded by the capsular polysaccharide to avoid creating a robust immune response from the host[104]. Therefore, phosphorylcholine attachments and capsular polysaccharide production are both tightly regulated to control the balance between antigen exposure and key teichoic acid and choline binding protein interactions[105].

1.5.3 The capsular polysaccharide

The capsular polysaccharide is incredibly important for shielding antigenic structures and is undeniably a crucial virulence determinant in *S. pneumoniae*[106]. Capsular polysaccharides are formed from repeating units of relatively simple to extremely complex carbohydrate structures[26]. In total, there are over one hundred distinct capsular polysaccharide serotypes based upon poly

and monoclonal antibody reactivity[107,108]. Of the known serotypes approximately forty-eight share some degree of cross-reactivity due to structural similarities, thus the corresponding capsules belong to a shared serogroup. In nearly all serotypes the required machinery for capsule synthesis is located between the genes *dexB* and *aliA* within the pneumococcal genome[26]. The genes within this region encompass a diverse set of glycosyltransferases, glycoside hydrolases, epimerases, phosphatases, transposases, transposons and transporters[109]. The exception to this is serotype 37 which only uses a single gene (*tts*) that is located elsewhere in the genome for its entire capsule biosynthesis[110].

That being said, two mechanisms of capsular synthesis have been described. The Wzy-dependent mechanism and the synthase dependent mechanism, both of which operate in significantly different manners[110–113]. In Wzy-dependent capsule synthesis, the protein CpsE anchors the initiating sugar (Glc-1P) to a membrane-bound undecaprenol. This is then followed by an elongation step as numerous proteins contribute to building and processing the capsule's repeating unit. Once completed, the repeating units are flipped to the cell surface and ligated together by the Wzy polymerase. Conversely, in the synthase dependent mechanism only one protein is responsible for initiating, elongating and transporting the capsular polysaccharide across the membrane[66]. In general, most capsules are composed of repeating units comprising two to eight carbohydrate residues with varying degrees of N- and O-acetyl, phospho-glycerol and O-pyruvyl functional group modifications[26,114–116]. Furthermore, some capsule can grow as large as 400 nm in thickness depending on the serotype and strain[117].

The contribution of the capsule to pathogenesis is exemplified in models of pneumonia and sepsis where unencapsulated strains are unable to sustain hematogenous spread[106]. Notably, unencapsulated strains can cause persistent otitis media and conjunctivitis, possibly owing to biofilm formation and interactions with teichoic acid and cell wall components[118]. Conversely, in encapsulated strains the thick polysaccharide coat is known to benefit the bacterium in a multitude of ways. For example, within the nasopharyngeal mucus the negatively charged capsule repels negatively charged mucus components, thereby resisting mucosal entrapment[119]. Furthermore, as pneumococcus advances towards the epithelium and into the blood, the capsule functions to prevent complement mediated clearance[120]. The mechanism by

which that the capsule resists activation of the classical compliment pathway is thought to be mediated through preventing host antibodies and C-reactive proteins from recognizing subcapsular antigens like teichoic acids and peptidoglycan components. Similarly, many antigenic factors present on the pneumococcal surface like CBPs, wall anchored enzymes and membrane bound lipoproteins are shielded by the capsule from specific antibody recognition[120].

1.6 The "surface" of pneumococcal carbohydrate processing

A possible alternative route for treating and preventing invasive pneumococcal disease is aimed at targeting pneumococcal virulence factors and antigenic surface proteins[34]. In *S. pneumoniae*, approximately 4.3% of its open reading frames are predicted or characterized surface exposed proteins[80,121]. Surface proteins are usually identified by structural signatures like signal sequences and anchoring motifs, and several identified classes of surface proteins includes CBPs, sortase anchored proteins, and lipoproteins. In addition, some cell surface associated proteins lack the classical leader peptide and membrane anchoring motif, yet they are found to localize to the cell surface through a non-classical pathway. This non-classical pathway is not well understood and only a few key proteins have been identified. Interestingly, several cell surface proteins are involved in carbohydrate processing and pneumococcal pathogenesis.

1.6.1 Pneumococcal lipoproteins

Lipoproteins represent the largest class of cell surface proteins in *S. pneumoniae* and consequently comprise a large assortment of different functions[80,122–125]. Lipoproteins are defined by their N-terminal signal peptide and lipobox (LXXC) sequences that direct their export and covalent attachment to a membrane-bound diacylglycerol lipid. In *S. pneumoniae,* roughly half of the encoded lipoproteins are substrate binding proteins that operate in conjunction with various ABC transporter systems, facilitating nutrient acquisition and environmental sensing[80]. Substrate binding proteins target an array of different substrates ranging from simple to complex carbohydrates, metal ions, amino acids, peptides, and polyamines[80]. Pneumococcal carbohydrate substrate binding proteins have not been directly studied for their role in virulence.

However, several carbohydrate utilization ABC transport systems are important during pathogenesis including the RFO ABC transport system[54,79].

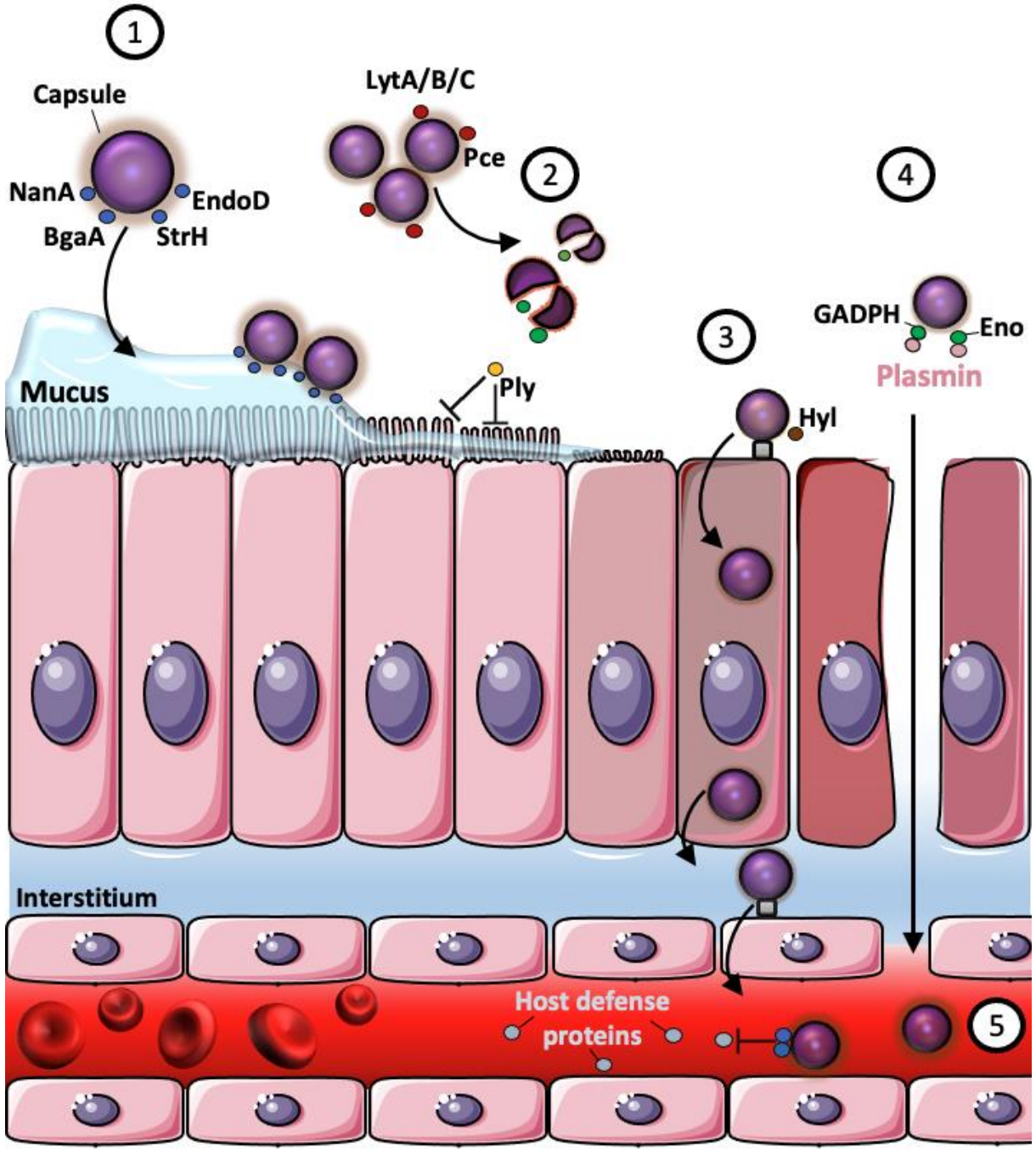


Figure 1-6. Pneumococcal colonization and pathogenesis depends on key CAZymes. The major reservoirs of pneumococcal colonization are the mucosal surfaces of the nasal cavity. **(1)** Pneumococcus relies upon several glycoside hydrolases (NanA, BgaA, StrH, EndoD) and

carbohydrate transporters for the exploitation of glycans found on host mucins[51,126]. **(2)** During the stage of mucus utilization, bacterial growth leads to cellular competence that results in fratricide by cell wall degrading enzymes like (Pce, LytA, LytB and LytC)[127]. Upon lysis, the release of intracellular components promotes virulence as several key virulence factors including Ply, Eno and GADPH are released into the environment and begin breaking down host's cells. **(3)** Further degradation of extracellular matrix glycans is achieved through the activity of CAZymes like Hyl. Subsequently, exposure of host's receptors enhances transmigration across the epithelium through receptor-mediated endocytosis[134]. The CBP CbpA, in addition to phosphorylcholine residues on pneumococcal teichoic acids bind to host polymeric immunoglobulin receptor and platelet-activating factor receptors resulting in endocytosis and transmigration across the host membrane[73,128,129]. **(4)** Pneumococcal "moonlighting" proteins like the glycolysis enzymes, Eno and GADPH, bind to host plasminogen thereby activating it to the serine protease plasmin. Plasmin dramatically alters the host's surface and facilitates dissemination by degrading connective tissue, extracellular matrix, and adhesion proteins[130]. **(5)** Once in the host circulatory system several glycoside hydrolases promote immune evasion through a poorly described mechanism. Notably, the glycoside hydrolase SpGH92 (discussed in Chapter 3) imparts the most profound contribution to hematogenous spread compared to all known pneumococcal CAZymes. Interestingly, multiple dietary saccharides are also implicated in pneumococcal pathogenesis; however, the role these saccharides play during nasal colonization and hematogenous spread is less understood. *Some images in this figure are adapted from smart.servier.com (Creative Commons Attribution 3.0).

1.6.2 Sortase-anchored proteins

In addition to lipoproteins, sortase-anchored proteins are an important multi-functional class of cell surface proteins. The proteins in this class all possess an N-terminal signal sequence and a C-terminal cell surface anchoring (LPXTG) motif that directs their attachment to the oligopeptide crosslinker regions in peptidoglycan[131]. Several of the sortase-anchored proteins in *S. pneumoniae* are important glycoside hydrolases including NanA, BgaA, StrH, EndoD and Hyl.

These CAZymes have important roles in colonization, cell adherence and immune evasion[47,53,62,126,132].

The glycoside hydrolases NanA, BgaA, StrH and EndoD function together to sequentially degrade complex carbohydrate structures found on various glycans including complex N-glycans (Figure 1-2J)[47,126]. Using the complex N-glycan structure as a model, the cascade of mucin glycan degradation begins with NanA, an exo-(α2,3-, α2,6-, α2,8-)-α-neuraminidase. NanA adheres to sialic acid containing glycans with its carbohydrate binding module and simultaneously cleave terminal sialic acid residues from host glycans. The removal of terminal sialic acids results in a terminal Gal-β1,4-GlcNAc- structure that is recognized by BgaA, an exo-β1,4-galactosidase. BgaA has two carbohydrate-binding modules that facilitate adhesion, while its catalytic domain cleaves Gal-β1,4-GlcNAc- creating a GlcNAc-β1,2-Man moiety that is subsequently cleaved by StrH. StrH is an exo-β-N-acetyl-glucosaminidase that removes the terminal N-acetylglucosamine residues ultimately leading to the exposure of terminal mannose residues in the N-glycan core structure ($Man_3GlcNAc_2$-Asn) that remains attached to the asparagine residue of the target glycoprotein[126]. The final step of N-glycan degradation involves EndoD, an endo-β-N-acetyl-glucosaminidase. EndoD cleaves between the two GlcNAc residues of $Man_3GlcNAc_2$ releasing $Man_3GlcNAc$ and producing a GlcNAc-Asp-glycoconjugate. Together these glycoside hydrolases are capable of deconstructing glycan structures fitting the carbohydrate type- and linkage-specificities of the forementioned enzymes so they can be subsequently imported and metabolized[47]. It is important to note that many glycans share similar structures to the “classic” N-glycan described here, thus glycans other than the classic N-glycan described above are likely targets of these enzymes[48].

1.6.3 Choline binding proteins

All known choline binding proteins share a common overall structure comprising an N-terminal variable region followed by a proline rich block and a C-terminal choline-binding domain. The N-terminal variable region functions of CBPs are extremely diverse, as are their roles in pneumococcal pathogenesis. Several key choline binding proteins that contribute to pneumococcal pathogenesis include the phosphorylcholine esterase (Pce or CbpE), N-acetyl-muramoyl-L-alanine amidase (LytA) and the N-acetyl-glucosaminidase (LytB)[133–135]. Pce and LytB

are CAZymes that modify teichoic acid and peptide structures thereby contributing to autolysis (Figure 1-6[2])[136]. Pce is a multifunctional phosphorylcholine esterase that removes phosphorylcholine moieties from teichoic acids, thereby decreasing the concentrations of surface CBPs[73]. In addition to its CAZyme activity Pce activates host plasminogen, a human extracellular matrix remodeling peptidase, and reduces neutrophil activation by inactivating host platelet activating factor[107,111]. LytA and LytB on the other hand function as autolysins that contribute to peptidoglycan turnover, resulting in autolysis and fratricide[138,139]. LytB is a recognized CAZyme as it cleaves between the GlcNAc-β1,4-MurNAc linkages in peptidoglycan. In contrast, LytA dedicates a large proportion of its N-terminal domain to binding the peptidoglycan structure, however it is not a CAZyme as it cleaves the oligopeptide crosslinkers in peptidoglycan. Importantly, autolysis triggers the release of many key virulence factors like pneumolysin (Ply), a pore-forming toxin that modulates host inflammation and diminishes ciliary beating (Figure 1-6[2])[10,140]. Moreover, fratricide is heavily implicated in pneumococcal horizontal gene transfer and transformation, thus contributing to antibiotic resistance and serotype replacement[127,138,139].

1.6.4 Non-classical surface-associate proteins

The least understood group of pneumococcal surface proteins are the non-classical surface-associated proteins (NCSAPs), which lack both a leader signal peptide as well as classical membrane-anchoring motif. The release of NCSAPs into the environment is possible through autolysis; however, some studies suggest accessory secretion systems and extracellular membrane-derived vesicles are underlying mechanisms driving the transport of these proteins outside of the cell[141,142]. The glycolytic pathway NCSAPs, enolase (Eno)[130] and glyceraldehyde-3-phosphate dehydrogenase (GAPDH)[143] are important enzymes in glycolysis. However, they also serve as potent virulence factors once released into the host environment (Figure 1-6[4]). Both Eno and GAPDH hijack host plasminogen which loosens the connection between neighboring epithelial cells leading to increased transmigration and systemic pneumococcal dissemination[103,143,144]. Notably, the virulence factor SpGH92 that is discussed in Chapter 3 is a suspected NCSAP.

1.7 Research objectives

In summary, I have described numerous carbohydrate active enzymes in *S. pneumoniae* that serve integral roles in pathogenesis. Hence, this information supports the notion that carbohydrate processing is a major virulence platform in *S. pneumoniae,* and it highlights the need for continued investment in this area of research as many important unanswered questions remain.

Objective 1: Biochemically characterizing the pneumococcal *raf* locus

One unanswered question that this thesis focuses on is what the functions and purpose of the dietary saccharide utilization loci are found within *S. pneumoniae*. Particularly, what functions do the *raf* locus' SBP and glycoside hydrolases perform, and how might they be contributing to the observed virulence phenotype?

The *raf* locus is recurrently associated with pneumococcal virulence. However, the literature surrounding its role and function are primarily genomic studies that investigate the phenotypic changes occurring in response to gene deletions. Consequently, the biochemical properties of the *raf* proteins have never been investigated directly. Thus, one goal of this thesis is to biochemically characterize the *raf* locus' SBP RafE, as well as the glycoside hydrolases Aga and GtfA to gain a better understanding of what roles they have *in vivo*. We hypothesize that if utilizing RFOs is important during pneumococcal infections, the proteins and enzymes responsible for the recognition and deconstruction of RFOs will have strong affinities towards their putative RFO substrates.

The specificity of the forementioned proteins towards their putative RFO substrates are examined using a mixture of biophysical and biochemical techniques. Using numerous fluorometric assays the substrate specificities of Aga and GtfA are examined. Subsequently, the kinetics of Aga are investigated using a coupled-fluorometric assay. Likewise, the specificity and affinity of the SBP RafE towards various RFOs is investigated using Isothermal Titration Calorimetry (ITC) and several crystal structures are determined using X-ray crystallography. The

binding pockets of RafE in complex with the different RFOs are studied and the key binding interactions that drive binding with very structurally diverse RFO substrates are identified.

Objective 2: Mapping inhibitor binding interfaces on SpGH92

Another important question that remains unanswered involves to the virulence factor SpGH92. We recently learned that SpGH92 confers a deadly virulence phenotype in models of pneumonia and sepsis. However, it is not clear if the virulence phenotype is associated with the enzymes activity, or if it a result of an unrelated interaction. To begin exploring this question, the availability of potent SpGH92-specific inhibitor(s) is vital. Notably, a novel inhibitor of SpGH92 with a K_i of 2 μM has been identified by the Vocadlo lab (unpublished). However, after attempting multiple crystal soaks, we have been unable to produce a crystal complex of SpGH92 with the bound inhibitor. We hypothesize that if the novel inhibitor (LIPS343) is binding to SpGH92, we will be able to elucidate the binding interface using HDX-MS.

Thus, the second goal of this thesis focuses on creating a protocol utilizing HDX-MS that can be used to accurately and unbiasedly determine inhibitor binding interfaces of inhibitor candidates with SpGH92. Utilizing a known α-mannosidase inhibitor (mannoimmidazole) and a novel inhibitor (LIPS343) an HDX-MS protocol is developed that can effectively identify inhibitor binding interfaces on SpGH92. Subsequently, the results of the HDX-MS analyses performed on the two inhibitors are compared and conclusions are drawn regarding the binding interface of LIPS343.

Overall, this thesis work furthers our understanding of carbohydrate active enzymes and carbohydrate processing in *S. pneumoniae*. The material covered in Objective 1 is published in the Journal of Biological Chemistry (JBC) and is discussed within Chapter 2 of this thesis. Likewise, the results of Objective 2 are discussed in further detail within Chapter 3 of this thesis. In Chapter 4 the conclusions and future directions of this work are discussed. Both Chapter 2 and Chapter 3 are presented in the form of a manuscript with an abstract, introduction, results, discussion and methods.

CHAPTER 2: The pneumococcal *raf* locus

Adapted from: Hobbs, J. K.[1], Meier, E. P. W.[1], Pluvinage, B.[1], Mey, M. A., & Boraston, A. B. (2019). Molecular analysis of an enigmatic Streptococcus pneumoniae virulence factor: The raffinose-family oligosaccharide utilization system. *Journal of Biological Chemistry*, 294(46), 17197–17208. https://doi.org/10.1074/jbc.RA119.010280.

[1] These authors contributed equally to this work.

Contributions: I performed the cloning, protein production, purification, ITC assays and analysis, crystallization, and structure refinement of the SBP RafE. I performed the cloning, protein production, purification and biochemical assays of GtfA. I performed the protein production, purification, TLC assays* and enzyme kinetics of Aga. *Mackenzie A. Mey produced the substrate specificity TLC analysis seen in Figure 2A.

2.1 ABSTRACT

Streptococcus pneumoniae (pneumococcus) is an obligate human pathobiont capable of utilizing a select number of dietary saccharides. The cluster of genes responsible utilizing raffinose family oligosaccharides (RFOs), better known as the *raf* locus, comprises several genes associated with pneumococcal virulence. However, no direct biochemical or structural information regarding the encoded proteins has been described in the literature. In this work, the crystal structures of the substrate binding protein RafE bound to the RFOs raffinose, stachyose and verbascose have been determined. Additionally, we have biochemically characterized the substrate binding protein RafE and the two glycoside hydrolases Aga and GtfA towards their putative substrates. In summary, this work has demonstrated that the *raf* locus contains the biological machinery required for RFOs utilization. However, the proteins involved appear biochemically inefficient. Thus, based on the information gathered here along with

existing knowledge pertaining to the *raf* locus, we suspect this locus may have an alternative role *in* vivo.

2.2 INTRODUCTION

A dense and highly specialized microflora populates the mucosal layer of the human nasal cavity and upper respiratory tract. Amongst the regular inhabitants is the pathobiont *Streptococcus pneumoniae* (pneumococcus). Pneumococcus transiently colonizes the upper airway in humans and can be isolated from as many as 8-10% of adults and 65-95% of children depending on the geographical location[7]. Although considered part of the normal microbiota, pneumococcus is the leading cause of community acquired bacterial pneumonia and a major cause of morbidity and death worldwide[21,107,145].

To survive in the host environment, pneumococcus relies heavily upon its adept carbohydrate processing abilities. In fact, the pneumococcal genome is reported to contain over 40 different carbohydrate degrading enzymes and over 30 carbohydrate transport systems, thereby facilitating the utilization of over 30 unique carbohydrates moieties. Many of the identified carbohydrate processing genes are known virulence factors that target glycans found on host glycoconjugates[61,62,80,146]. In addition, pneumococcus can utilize an array of plant derived oligosaccharides, known as dietary saccharides. The major dietary saccharides that pneumococcus can utilize include cellobiose, trehalose, sucrose, fructo-oligosaccharides, and raffinose family oligosaccharides (RFOs). Notably, several of the genes involved in utilizing these different dietary saccharides are classified as virulence factors from signature-tagged mutagenesis screens[54,147]. Indeed, this has been shown with the loci responsible for cellobiose, sucrose, fructo-oligosaccharides and RFO utilization. In some instances, the constituent gene products of the forementioned loci have been studied in detail, however their particular roles in pathogenesis remain enigmatic[54,56,148,149].

Despite the characterization of several virulence factors involved in dietary saccharide utilization, it remains unknown how the presence of these genes and their products are affecting pneumococcus pathogenesis. The abundance of dietary saccharides within the upper airway

mucus is expected to be quite low outside of the oral cavity. Furthermore, pneumococcus is completely capable of degrading the abundant N- and O-linked glycans within the host environment. Thus, prompting many researchers to question the roles of these dietary saccharide utilization loci *in vivo*. It is possible that pneumococci are encountering the dietary saccharides in the mucosa or saliva of the host, and that they are a source of nutrition. Alternatively, the dietary saccharides may resemble the extracellular structures on fellow microbes within the upper airway. However, this doesn't explain the phenomenon in which several genes proposed to function in dietary saccharide utilization also contribute to tissue adhesion and hematogenous spread[54,150]. Intriguingly, some dietary saccharides share homology with host glycans. For example, cellobiose (Glc-β1,4-Glc) can be derived from the glycosaminoglycan hyaluronan (glucuronic acid-β1,4-GlcNAc-β1,4-)[151]. Another possibility is that the dietary saccharide utilization systems are not critical for carbohydrate utilization. Instead, they may serve as environmental or quorum sensing receptors. Trappetti et al., recently demonstrated that the fructose phosphotransferase system, FruA (EIIABC^{Fru}; SPD_0773), is required for the transduction of autoinducer-2 mediated signaling and downstream transcriptional regulation that contributes to a hypervirulent phenotype[88]. Altogether, there is a growing body of evidence that is revealing many of the dietary saccharide utilization systems in *S. pneumoniae* are multi-functional; and that some of loci contain important virulence genes that possess activities that are distinct from their putative functions. That being said, almost every gene within the pneumococcal RFO dietary saccharide utilization locus has been previously identified as a virulence gene in signature-tagged mutagenesis screens and genetic direct mutagenesis studies. However, the proteins encoded have not been studied in significant biochemical or structural detail.

Within the pneumococcal genome the RFO utilization locus (the *raf* locus*)* is recognized as one of the core loci that is highly conserved and present in over 98% of clinical isolates[152]*.* The *raf* locus shares homology to the multiple-sugar metabolism (*msm*) locus identified within *Streptococcus mutans* and *Streptococcus suis*[153,154]; however, similarities are also present across many other streptococci and related bacteria. Within *S. mutans* and *S. suis*, the *msm* locus has been demonstrated to utilize an array of dietary sugars including melibiose, isomaltose,

isomaltotriose, stachyose and raffinose. In comparison, the *S. pneumoniae raf* locus is only known to utilize raffinose and stachyose[59,153]. The composition of the *raf* locus in *S. pneumoniae* comprises three promotors, eight local genes and one disconnected gene with its own promotor (Figure 2-1). The two transcriptional regulators which share a single promotor are the AraC-like transcriptional activator (*rafR*) and the biotin/lipoate ligase (BirA-like) transcriptional repressor (*rafS*). The family 36 glycoside hydrolase (GH36) α-galactosidase (*aga*) is located next to *rafR*, facing the opposite direction, and has its own promotor. Downstream to *aga*, the third and last promotor belonging directly to the *raf* locus is shared by five genes. Together, these genes encode an ABC transporter that comprises an extracellular substrate binding protein (SBP) *rafE*, and two membrane permeases (*rafF* and *rafG*). Downstream of *rafG* is the putative glycoside hydrolase family 13 (GH13) sucrose phosphorylase (*gtfA*) and the putative WzyC-like lipoteichoic acid ligase (*rafX*). Lastly, the gene encoding the Carbohydrate-Uptake Transporter-family 1 ATPase (*rafK*) is located elsewhere in the genome. Notably, the ATPase RafK powers several ABC transporters in *S. pneumoniae*[155].

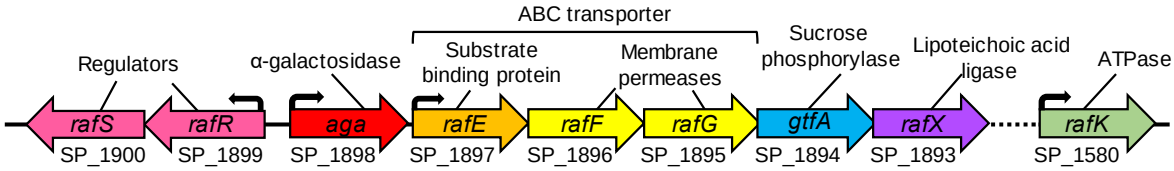


Figure 2-1. Organization of the *raf* locus within *S. pneumoniae*. The genetic organization of the RFO processing machinery found within the core genome of *S. pneumoniae*. The colored arrows indicate the direction that the corresponding gene is transcribed in, and the locus tags located underneath correspond to the TIGR4 genome open reading frames. Black arrows indicate the promotor sequences.

Remarkably, nearly every component of the *raf* locus is linked to pneumococcal virulence and multiple large scale transcriptomic studies have also shown the *raf* locus is upregulated in early-response to human macrophages and lung epithelial cells[156–158]. A total of three independent signature-tagged mutagenesis screens have identified *aga*, *rafF*, and *rafK* to be virulence genes in pneumococcal lung infection models[11,12,23]. In addition, mutants of *rafK* and *rafR* significantly

diminish the virulence phenotype of multiple pneumococcal serotypes and strains[60]. Perhaps the most telling gene of all within the *raf* locus that hints towards an alternative role is *rafX*. This gene encodes a putative lipoteichoic acid ligase that is seemingly unrelated to the rest of the *raf* locus. Heß et al., demonstrated that *rafX* is essential for lipoteichoic acid formation and that it serves an important role during pathogenesis[159,160]. The gene product of *rafX* is not well characterized aside from knock out studies which revealed Δ*rafX* mutants no longer produce lipoteichoic acid. However, *rafX* shares sequence identity with WzyC O-antigen polymerases found in Gram negative bacteria which are important for lipopolysaccharide assembly[161].

Taken together, the *raf* locus is an extremely enigmatic gene cluster involved in RFO utilization, lipoteichoic acid formation and host invasion[60,79,85,153,159,160]. To date, essentially all of the literature regarding the *raf* locus is centered around the phenotypic changes incurred by gene manipulations within *in vivo* and *in vitro* environments. A major obstacle preventing us from fully understanding the *raf* locus is the lack of information describing the biochemical functions of the *raf* proteins. In this study, we have elucidated the substrate specificities of several *raf* components including RafE, Aga and GtfA. Additionally, we have characterized the kinetic properties of Aga and RafE towards their identified substrate targets. We have also obtained several crystal structures of RafE bound to the three most prominent RFOs raffinose, stachyose and verbascose. Our results reveal that the proteins examined in this work are inefficient at RFO uptake and catabolism, at least with regards the conditions used in our assays. Thus, due to the low observed affinities and the high amino acid sequence conservation across the *raf* locus genes, we suspect this locus has an alternative role *in vivo* and the full scope of this locus has yet to be discovered.

2.3 RESULTS

2.3.1 Cloning, expression and purification of the *raf* locus glycoside hydrolases

The glycoside hydrolases Aga and GtfA represent a GH36 α-galactosidases and a putative GH13 sucrose phosphorylase, respectively. In order to define the biochemical properties of Aga and GtfA towards their putative substrates, we first obtained high-purity recombinant protein. Recombinant proteins containing an N-terminal 6x-histidine tag, were expressed and purified

from *Escherichia coli BL21* cells using immobilized Ni^{2+}-affinity chromatography (Figure 2-2; Aga); as well as both Ni^{2+}-affinity chromatography and size-exclusion chromatography (Figure 2-3; GtfA only).

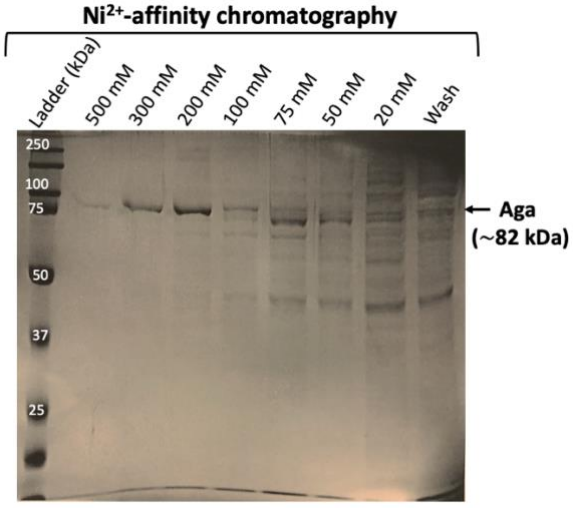


Figure 2-2. Purification of Aga. Aga was expressed in BL21 and purified using Ni^{2+}-affinity chromatography. The 200 mM and 300 mM imidazole fractions were pooled and concentrated for subsequent activity assays.

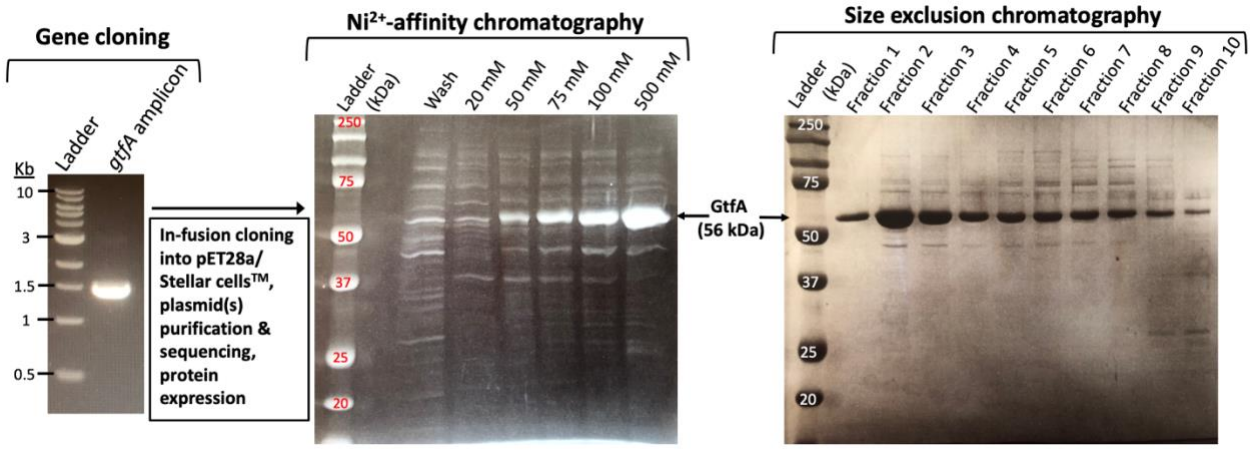


Figure 2-3. GtfA cloning, expression and purification. *gtfA* was cloned from the *S. pneumoniae* (TIGR4) genome and heterologously expressed and purified from *E. coli*. GtfA was purified by Ni^{2+}-affinity chromatography and S200-size exclusion chromatography.

2.3.1.1 Aga: An *exo*-α1,3-|α1,6-galactosidase

As Aga was previously demonstrated to be essential for RFO fermentation, and shares homology with GH36s, we were confident that it possessed α-galactosidase activity. Therefore, to confirm the presence of α-galactosidase activity and to explore other potential substrate targets, we tested Aga against a panel of synthetic α- and β-nitrophenyl monosaccharide analogues. Reactions were performed at 37°C and contained 1 mM substrate, 200 nM Aga and 50 mM phosphate buffer at pH 6.5. Activity was detected by measuring changes in absorbance at 405 nm which is produced upon hydrolysis of the *para*-nitrophenyl analogue. Using this approach, we found that Aga was only active towards *para*-nitrophenyl (pNP) α-galactosides (Gal-α-pNP).

Table 2-1. Substrate specificity of Aga towards synthetic pNP-carbohydrate analogues.

Substrate	Activity
para-Nitrophenyl α-galactoside	+
para-Nitrophenyl β-galactoside	-
ortho-Nitrophenyl β-galactoside	-
para-Nitrophenyl α-N-acetyl-galactosamine	-
para-Nitrophenyl β-N-acetyl- galactosamine	-
para-Nitrophenyl α-N-acetyl- glucosamine	-
para-Nitrophenyl β-N-acetyl- glucosamine	-
para-Nitrophenyl α-glucoside	-
para-Nitrophenyl α-mannoside	-
para-Nitrophenyl α-arabinoside	-
para-Nitrophenyl α-L-fucoside	-
para-Nitrophenyl β-xyloside	-

Activity assays were performed in triplicate, monitoring for release of *para*-Nitrophenyl at 405nm.

To investigate the substrate specificity of Aga further, we incubated the enzyme with a variety of biologically relevant α-galactoside substrates and probed for activity using Thin-Layer

Chromatography (TLC) (Figure 2-4A). As expected, Aga showed activity towards the RFOs, producing cleavage products matching the retention factors of galactose and sucrose (Figure 2-4A). Additionally, Aga cleaved the disaccharide constituent of the RFOs known as melibiose (Gal-α1,6-Glc) producing products matching the retention factors of glucose and galactose. In addition to RFOs and melibiose, Aga demonstrated activity towards α1,3-galactobiose and the linear type II B blood group antigen (Gal-α1,3-Gal-β1,4-GlcNAc) producing cleavage products matching the retention times of galactose and N-acetyl-lactosamine (LacNAc; Figure 2-4A). Conversely, our results revealed Aga cannot cleave β1,4-galactobiose, globotriaose (Gal-α1,4-Gal-β1,4-Glc), type II A blood group antigen (GalNAc-α1,3-(Fuc-α1,2-)Gal-β1,4-GlcNAc), or type II B blood group antigen (Gal-α1,3-(Fuc-α1,2-)Gal-β1,4-GlcNAc). Collectively, our analyses demonstrated that Aga could cleave α1,6- and α1,3- terminal galactose substrates. However, Aga was unable to cleave α1,6- or α1,3- substrates containing branching next to the terminal α-galactoside.

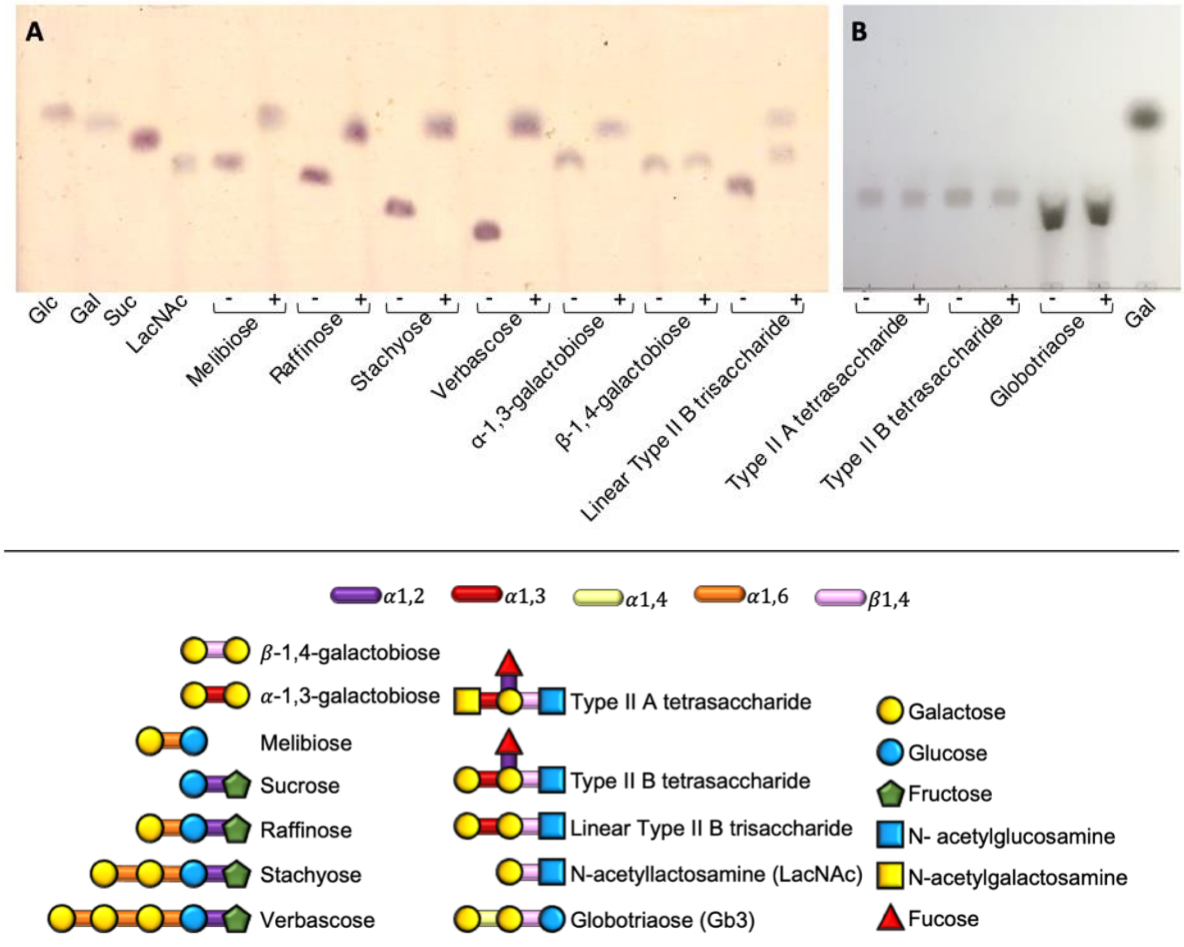


Figure 2-4. Aga is active towards unbranched terminal α1,3- and α1,6-galactosides. (**A**) Aga was incubated overnight with the specified substrates prior to TLC separation (produced by Mackenzie Mey). (**B**) Aga was incubated with the specified mammalian glycans overnight prior to TLC separation. A (+) sign indicates the presence of Aga, and a (-) signs indicates a control lane. Standards are abbreviated: Glc, glucose; Gal, galactose; Suc, sucrose; LacNAc, N-acetyl-lactosamine.

2.3.1.2 GtfA: An inefficient sucrose phosphorylase

To test the putative function of GtfA towards sucrose, we probed for sucrose phosphorylase activity in a variety of different buffers including a PBS buffer (50 mM Na_2HPO_4-KH_2PO_4 pH 7.4, 150 mM NaCl), a Tris buffer (20 mM Tris pH 8.0, 150 mM NaCl), and four phosphate buffers (50 mM Na_2HPO_4-KH_2PO_4 pH 6.6; 50 mM Na_2HPO_4-KH_2PO_4 pH 7.2; 100 mM Na_2HPO_4-KH_2PO_4 pH 6.6; and 100 mM Na_2HPO_4-KH_2PO_4 pH 7.2). For each reaction, 31 µM GtfA was incubated with sucrose overnight at 37°C and inspected for cleavage products the following morning using TLC. Following separation, we found that GtfA had the highest activity in 100 mM phosphate buffer at pH 6.6, and in all of the buffers lacking phosphate there was no detectable enzymatic activity. The products of GtfA cleavage matched the retention factors of the fructose, glucose and glucose-1-phosphate (Glc-1P) standards indicating GtfA possessed partial phosphorylase activity. Notably, only trace activity could be detected after using relatively large concentrations of enzyme (31 µM; Chapter 2.6.4) in an overnight incubation. In comparison, the *S. mutans* GtfA sucrose phosphorylase (84% sequence identity) showed distinct cleavage products after 10 minutes of incubation and this increased significantly after two hours[162]. To demonstrate the activity of GtfA in the context of RFO depolymerization with Aga (Figure 2-3A), we incubated both enzymes with RFOs overnight at 37°C and probed for activity the following morning. Our assay revealed that GtfA can cleave sucrose once released by Aga and no activity was apparent on full length RFOs.

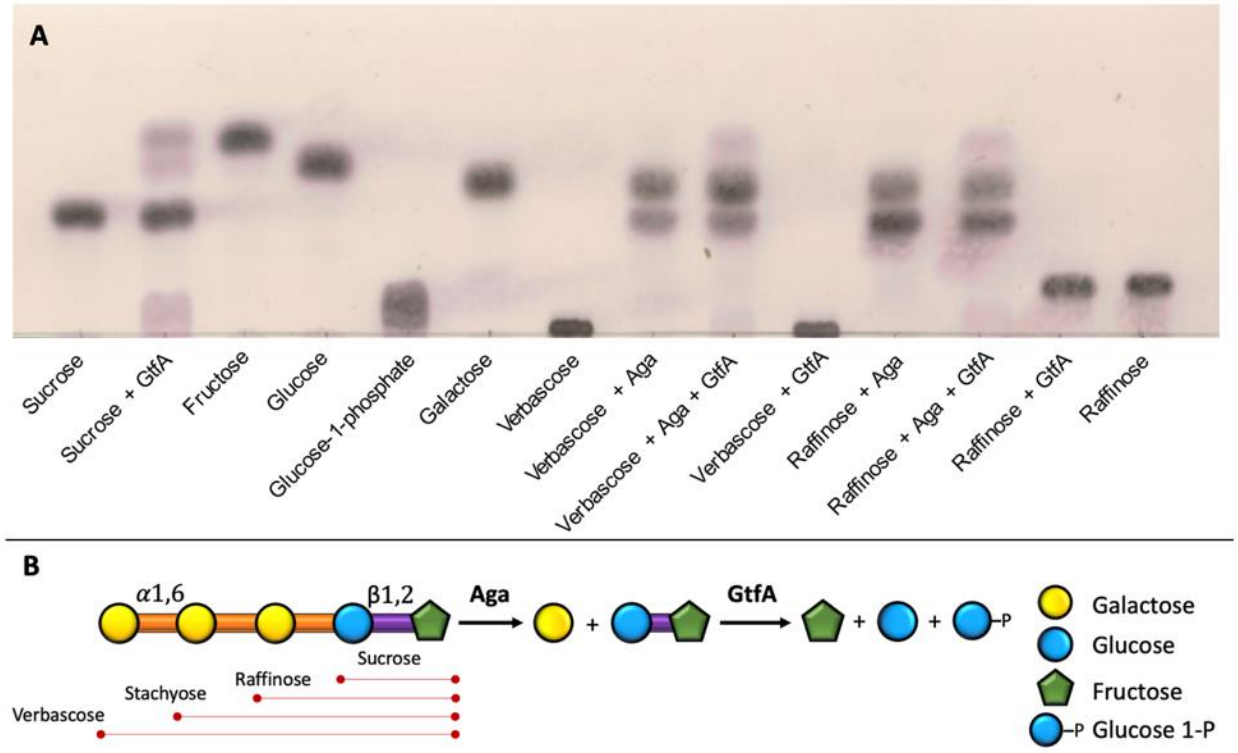


Figure 2-5. Aga and GtfA concertedly depolymerize the RFOs. (**A**) Analysis of the concerted activity of Aga and GtfA towards the RFOs verbascose and raffinose using TLC analysis. (**B**) A cartoon representation of RFO depolymerization by Aga and GtfA.

2.3.1.3 The kinetic parameters of Aga

Our pNP- and TLC assays revealed Aga is selective towards α1,3- and α1,6-galactoside linkages. Thus, we wanted to know what substrates Aga had a higher affinity towards, so we could gain further insight into its natural targets *in vivo*. To do this we analyzed the kinetics of Aga towards pNP-α-galactoside, α1,3-galactobiose, melibiose, raffinose and stachyose. For all of the substrates except pNP-α-galactoside, which we could directly measure at 405 nm, we used the Megazyme galactose detection kit (Figure 2-6).

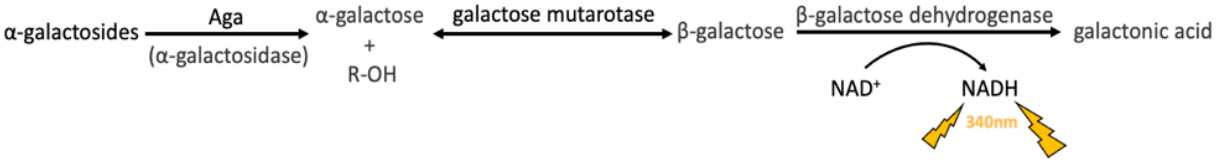


Figure 2-6. Aga kinetics assay layout. Kinetic constants of Aga were measured by quantifying the

number of α-galactose monosaccharides produced in a stopped assay format. Heat killed reaction, once cooled, were mixed with components from the L-arabinose/D-galactose assay kit (Megazyme). Detection of galactose production through the Megazyme assay kit was achieved by the reaction depicted above through the increase in absorbance at 340 nm.

Overall, we found Aga had the highest affinity and catalytic efficiency toward the synthetic pNP-α-galactoside substrate. However, with regarding biologically relevant substrates, the highest affinity was seen towards α1,3-galactobiose relative to all of the α1,6-galactoside substrates (Table 2-2). In addition, we found that the turnover rates were highest towards melibiose and stachyose and the catalytic efficiencies were highest towards melibiose and α1,3-galactobiose. In summary, our kinetic assays suggested that Aga is more efficient at cleaving shorter α-galactosides, with a preference towards α1,3-linkages.

Table 2-2. Kinetic constants of Aga towards terminal α-galactoside substrates.

Substrate		**K_M (mM) ± SEM**	**k_{cat} (s^{-1}) ± SEM**	**k_{cat}/K_M (s^{-1} mM^{-1}) ± SEM**
pNP-α-galactoside		0.43 ± 0.02	20.1 ± 0.9	46.7 ± 8.1
α1,3-galactobiose	(Gal-α1,3-Gal)	2.2 ± 0.2	76.2 ± 2.3	34.6 ± 3.1
Melibiose	(Gal-α1,6-Glc)	4.9 ± 0.4	162.0 ± 4.6	33.1 ± 2.9
Raffinose	(Gal-α1,6-Glc-α1,2-Fru)	5.8 ± 0.3	81.1 ± 1.3	14.0 ± 0.7
Stachyose	(Gal-α1,6-Gal-α1,6-Glc-α1,2-Fru)	11.0 ± 0.9	155.0 ± 4.3	14.1 ± 1.2

Values are representative of triplicates.

2.3.2 RafE: Cloning, expression and purification

To begin investigating the binding affinity and specificity of RafE, the *rafE* gene was cloned from the pneumococcal genome without the N-terminal signal sequence and lipid anchoring motif. Instead, an N-terminal 6x-histidine tag was added to facilitate purification. RafE was expressed in *E. coli* (BL21) and purified using immobilized metal affinity chromatography with a nickel resin (Figure 2-7). The protein was then concentrated and subsequently purified on a Superdex S100 column in very high purity due to an interaction between the protein and the

carbohydrate moieties that constitute bead resin. To analyze the purity and oligomeric state of RafE, dynamic light scattering was performed using a DynaPro plate reader (Wyatt) with approximately 1 mg/mL RafE isolated from size exclusion chromatography (Figure 2-7) in 20 mM Tris pH 8.0, 150 mM NaCl. The calculated polydispersity was below 10% and the predicted molecular weight was approximately 55 kDa suggesting RafE purified as a monomer.

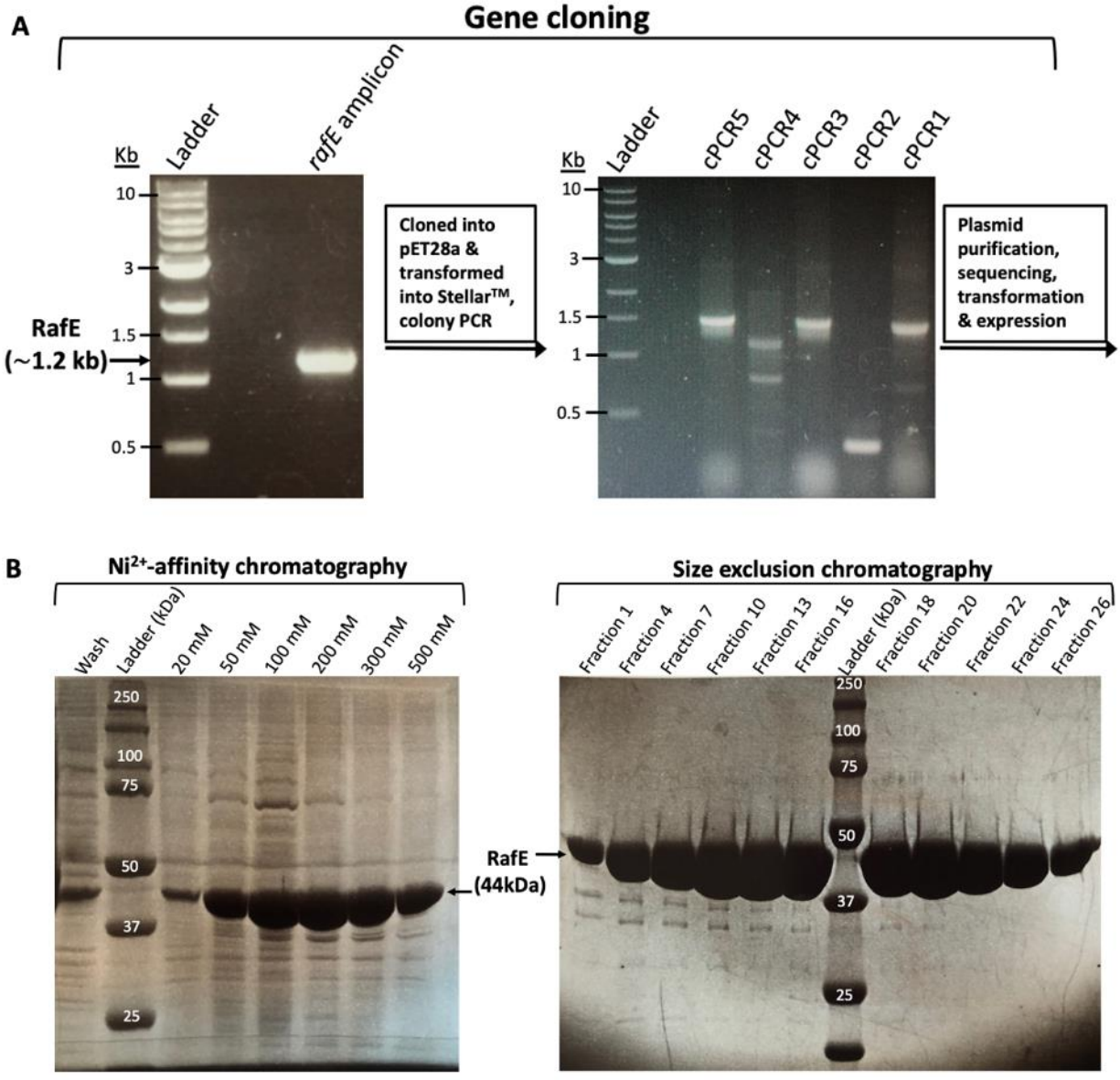


Figure 2-7. Cloning, expression, and purification of the substrate binding protein RafE. (A) RafE was cloned from the *S. pneumoniae* (TIGR4) genome and heterologously expressed and purified

from *E. coli* (BL21). **(B)** The purifications steps shown are typical for RafE when purified using Ni^{2+}-affinity chromatography followed by size exclusion chromatography on a Superdex S100 column. Notably, an interaction between RafE and the S100 column was beneficial for attaining a high purity RafE elution (RafE elution volume = ∼130 mL, void peak = ∼40 mL).

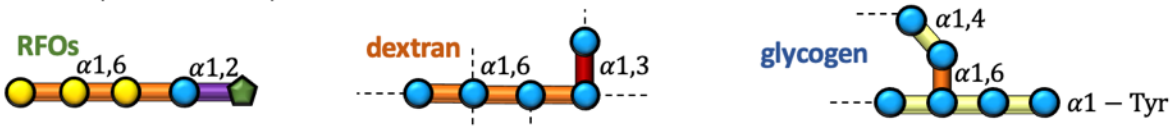
2.3.3 RafE: A promiscuous substrate binding protein

Using freshly purified protein with ligands that were precisely weighed and dissolved in isotonic buffer we analyzed the binding properties of RafE using a Malvern VP-ITC instrument. We initially investigated the affinity of RafE towards the RFOs followed by the RFO constituents (galactose, melibiose, and sucrose). Our initial results revealed RafE bound the RFOs with a relatively low affinity, measuring K_d values between 199 µM for stachyose to 479 µM for raffinose (Table 2-3). This was a surprising result as the average SBP K_d for its target substrate is around 1 µM[163]. Furthermore, any interactions with the RFO constituents were undetectable. Although we tested against a small carbohydrate panel, our results showed that RafE only bound oligosaccharides three residues or more in length. Therefore, we wanted to know if RafE bound other trisaccharides that it may encounter in the upper airway. Thus, we examined the binding affinity towards isomaltotriose, a derivative of dextran which is a common microbial extracellular polysaccharide; panose, a derivative of glycogen; and globotriaose, a globo-series glycolipid that is found predominantly on erythrocytes[164]. We found that RafE had a comparable K_d towards isomaltotriose (427 µM) as it did for raffinose (479 µM) and that it had the highest affinity for panose (80 µM) relative to all the substrates tested. Interestingly panose only differs from isomaltotriose at the reducing end α1,4-linkage, implying that the subsite three α1,6-glycosides are not the preferred substrates of the *raf* transporter.

Table 2-3. Equilibrium dissociation constants of RafE determined by ITC.

	Ligand	Structure	K_d (μM) ± SD
RFOs	Raffinose	(Gal-α-1,6-Glc-α-1,2-Fru)	479 ± 35.2
RFOs	Stachyose	(Gal-α-1,6-Gal-α-1,6-Glc-α-1,2-Fru)	199 ± 23.0
RFOs	Verbascose	(Gal-α-1,6-Gal-α-1,6-Gal-α-1,6-Glc-α-1,2-Fru)	369 ± 34.2
dextran	Isomaltotriose	(Glc-α-1,6-Glc-α-1,6-Glc)	427 ± 20.1
glycogen	Panose	(Glc-α-1,6-Glc-α-1,4-Glc)	80.2 ± 8.12
	Melibiose	(Gal-α-1,6-Glc)	Undetectable
	Galactobiose	(Gal-α-1,3-Gal)	Undetectable
	Sucrose	(Glc-α-1,2-Fru)	Undetectable
	Galactose		Undetectable

*Values are representative of triplicates



2.3.4 RafE: Crystal formation and optimization

Given the ability of RafE to accommodate multiple oligosaccharides of different lengths, sugar compositions, and linkage types we wanted to investigate the molecular basis of ligand binding using X-ray crystallography. To do this we screened over seven hundred different crystallization conditions using eight crystallization screens (Hampton, MCSG and in-house PEG/ion) containing RafE and the RFOs. We found that the RafE complex crystallized in several conditions including 0.2 M $(NH_4)F$, 20% (w/v) PEG 3350 (MCSG-I); 0.2 M $(NH_4)Cl$ pH 6.3, 20% (w/v) PEG 3350 (MCSG-I); 0.1 M citric acid: NaOH pH 3.5, 25% (w/v) PEG 3350 (MCSG-III); and 0.1 M Tris pH 8.0, 0.2 mM CsCl, 20% (w/v) PEG 3350 (House PEG). After examining the diffraction quality of several crystals (Figure 2-8) it was determined the best condition were 0.1 M Tris pH 8.0, 0.2 mM CsCl, 20% (w/v) PEG 3350. We then attempted crystal optimization using multiple additive screens. Though, this only resulted in higher quality crystals for the verbascose complex which improved slightly with the addition of 0.1 M $MnCl_2$.

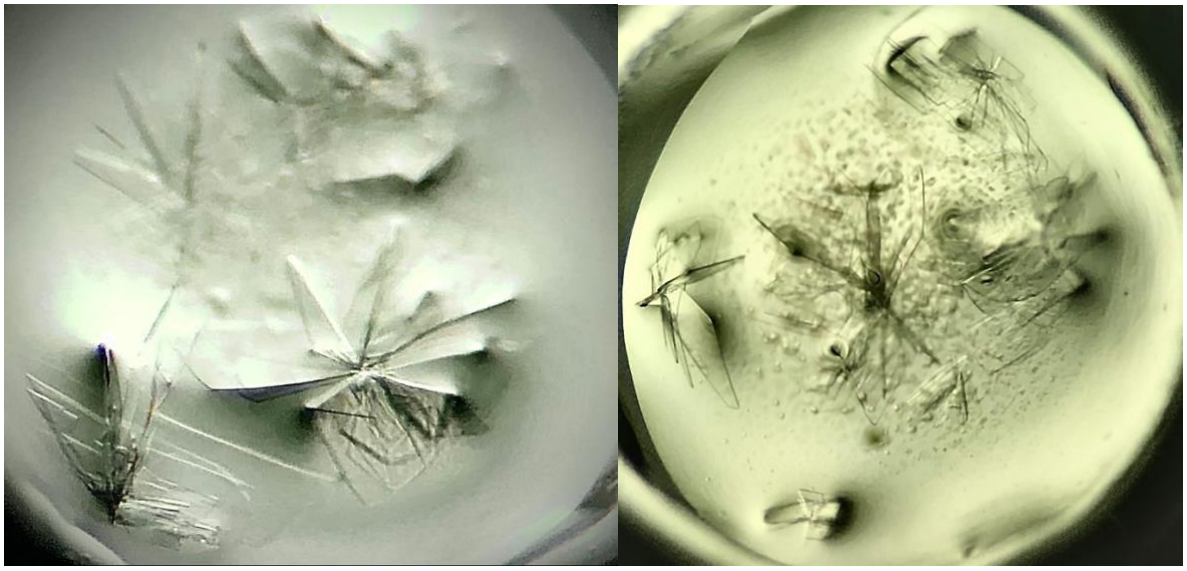


Figure 2-8. Crystals of RafE in complex with the RFOs. Large plate-like crystals of RafE bound to the RFOs were developed in 0.1 M Tris pH 8.0, 0.2 mM CsCl, 20% (w/v) PEG 3350 using the sitting-drop method. Notably, crystals of the verbascose complex were optimized by the addition of 0.1 M $MnCl_2$.

2.3.5 RafE X-ray diffraction: Data collection and processing

Notably, it was found that the RFO which RafE had the highest affinity towards was stachyose thus it was the first model to be pursued and completed. The steps leading to model determining were similar between all structures and in this section are examples from the stachyose model building process. First, crystals were diffracted on the "Home beam X-ray source" (1.540 Å) located in the University of Victoria's Biochemistry and Microbiology department. Using crystals of RafE in complex with stachyose diffraction data could be collected to a resolution of 2.35 Å with an acceptable mosaicity of 0.52° (Figure 2-9). Autoindexing in *HKL-2000* revealed that the crystals belonged to the orthorhombic space group P222.

Satisfied with the diffraction quality of the crystals, we collected our full data set and integrated, scaled and merged the diffraction images using the *HKL-2000* software[165]. Data processing revealed the unit cells to have the cell dimensions *a*= 49.28 Å, *b*= 118.84 Å, *c*= 146.01 Å and α=β=γ= 90°, belonging to the multiple screw axis containing primitive orthorhombic space group $P2_12_12_1$. The completeness was high with 99.8% in our high-resolution shell and 99.9%

overall. Additionally, the redundancy was above 3.0 in both the overall- and high-resolutions bins. The I/σI was 20.2 and 6.7 for the low- and high-resolutions shells, respectively. Although a value above 2.0 in the high-resolution shell suggests that the crystal could diffract farther thereby increasing the maximum resolution, our data had to be truncated before this to preserve the quality of our data set[166]. The R_{merge} of the low- and high-resolution shells were 0.084 and 0.181, respectively. Notably, the increased R_{merge} in the high-resolution shell could be attributed to radiation damage or from smearing caused by the mosaicity in our crystal[166]. Given the size of our unit cell and the molecular weight of RafE, the Matthews coefficient was estimated to contain a solvent content of 46.56% (V_M = 2.3 Å^3/Da) with two molecules in the asymmetric unit. All data collection statistics were recorded in Table 2-4.

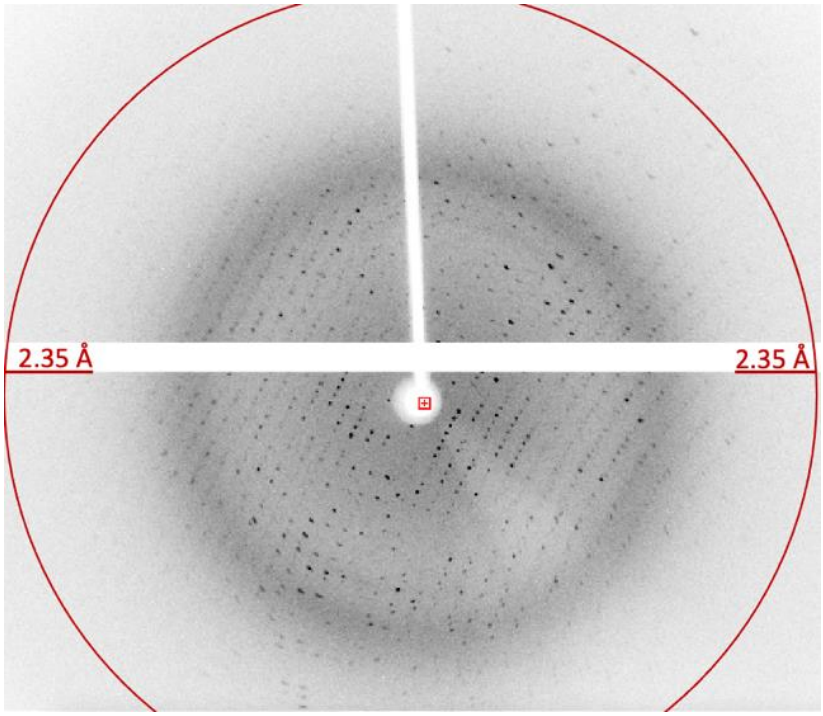


Figure 2-9. Diffraction image of the RafE-stachyose complex showing the 2.35 Å detection edge.

2.3.6 RafE: Molecular replacement, model building and refinement

We used molecular replacement with an existing structure of RafE in complex with raffinose (PDB: 2I58_ChainA)[167] as our template model with *PHASER*[168] to solve the phase problem, searching for two molecules in the asymmetric unit . The output ensemble (structure model)

from *PHASER* went through iterative cycles of manual building in *COOT*[169] and restrained refinement in *RAFMAC5*[170]. A geometric weighting term of 0.1-0.3 was used through the refinement process until model quality statistics were optimized.

Following backbone and sidechain refinement a clear region of unmodelled electron density corresponding to stachyose was seen in the RafE binding pocket (Figure 2-10B). Ligand building began with adding the expected monosaccharides into the regions of missing densities and subsequently linking them together through modifications within the corresponding pdb file. The model was then improved through several more rounds of refinement. Lastly, water molecules were added using *FINDWATERS* in *COOT* and remaining densities corresponding to solvent molecules were incorporated.

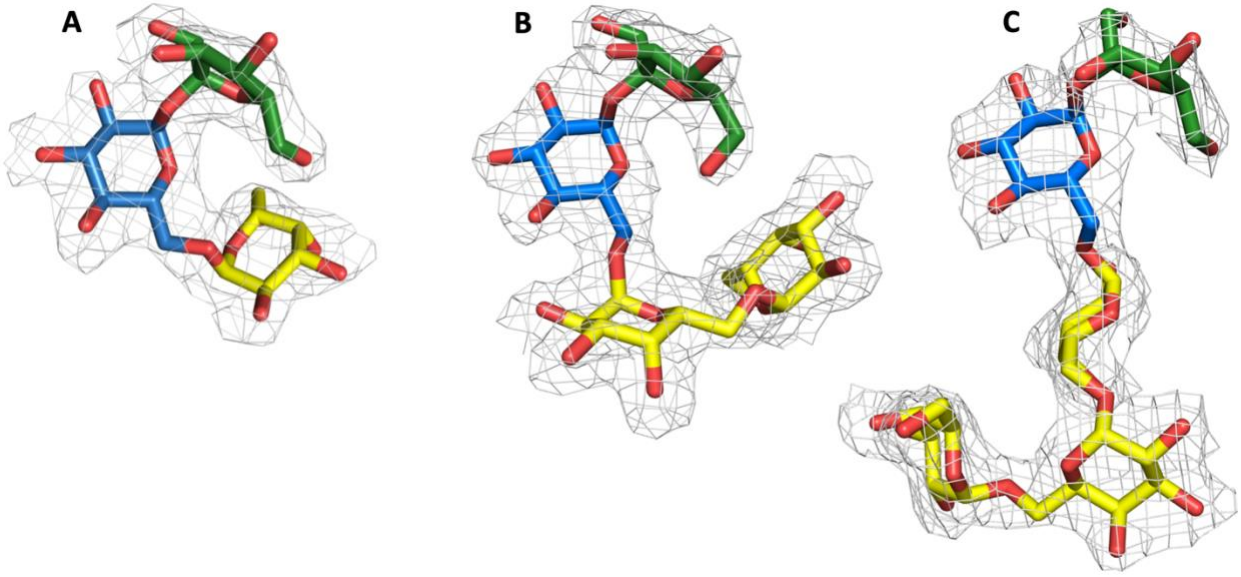


Figure 2-10. Models of the RafE binding pocket contain unambiguous density for each RFO. Stick representations of (**A**) raffinose, (**B**) stachyose, (**C**) verbascose within their electron densities shown as grey mesh. The electron densities represent maximum likelihood/σ_a-weighted F_o-F_c omit maps, contoured to 2.0σ. In all RFOs galactose (yellow) is α1,6-linked to another galactose or to glucose (blue) and glucose is α1,2-linked to fructose (green).

Final model validation statistics were collected after all refinement processes were completed using *PHENIX.validation* (Table 2-4)[171]. The final R_{work} and R_{free} values 0.22 and 0.24 were good

given the resolution of 2.35 Å. Furthermore, the root-mean-square deviations of the bond lengths and bond angles were very low, 0.006 Å and bond angles of 1.11° respectively, indicative of the tight geometric restraints used during *REFMAC5* refinement. A Ramachandran plot indicated the conformations of 97.8% amino acids were favorable while the remaining 2.2% were in allowed regions (Figure 2-11).

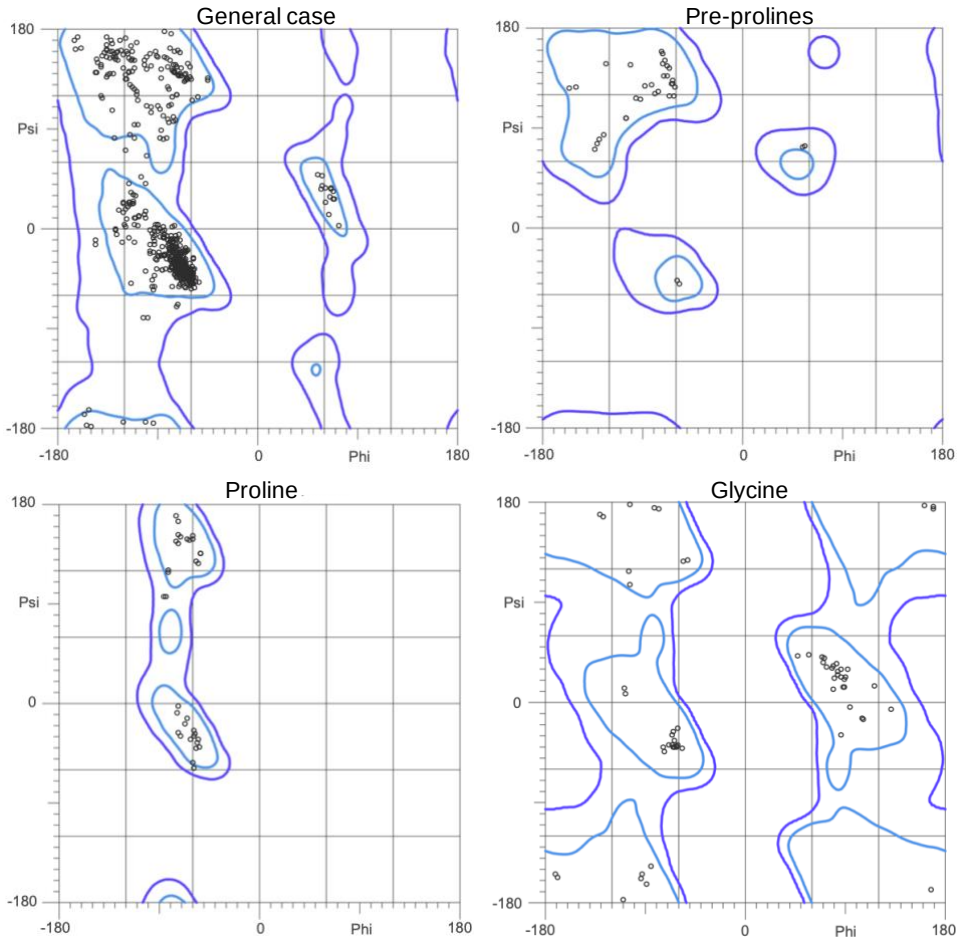


Figure 2-11. Ramachandran plots of the different amino acid classes. Ramachandran plots of the RafE-stachyose complex were generated using *MOLPROBITY*server[172] (PDB: 6PRG). Light blue contours designate favorable regions and dark blue contours designate allowed regions. Ramachandran statistics are as follows: favored = 97.8%, allowed = 2.2%, outliers = 0%.

The final model of stachyose bound RafE (PDB: 6PRG) was deposited in the RCSB Protein Data Bank (www.rcsb.org) and subsequent model building of RafE in complex with raffinose and

verbascose was performed similarly. Notably, the Chain A monomer from the RafE-stachyose model was used as the template for molecular replacement in both the raffinose (PDB: 6PQL) and verbascose (PDB: 6PRE) models.

2.3.7 RafE: Global structure analysis

The overall structure of RafE was found to be characteristic of the Cluster D-I SBPs (Figure 2-12), which are reported to bind both sugars and polyamines[173,174]. In general, the structure of RafE was found to comprise two α/β domains linked together by three loops that form the hinge region. The smaller N-terminal domain encompassed a five β-sheet core surrounded by five α-helices; and the larger C-terminal domain comprised a six β-sheet core with a total of nine α-helices. Like other SBPs, the binding pocket of RafE was located between the two α/β domains and was up to 19 Å deep. The base of the binding pocket contained three tryptophan residues that formed a hydrophobic platform facilitating substrate binding. Additionally, our crystal constructs revealed RafE adopts an open binding pocket when bound to the RFOs, thereby allowing the reducing end of longer sugars like the pentasaccharide verbascose to hang partially outside of the binding pocket. Notably, a sequence analysis (myhits.sib.swiss) of RafE indicated a putative arginine-glycine-aspartic acid (RGD) integrin binding motif that upon inspection appear to be surface localized. However, the potential binding interaction between RafE and integrin was not in the scope of this study.

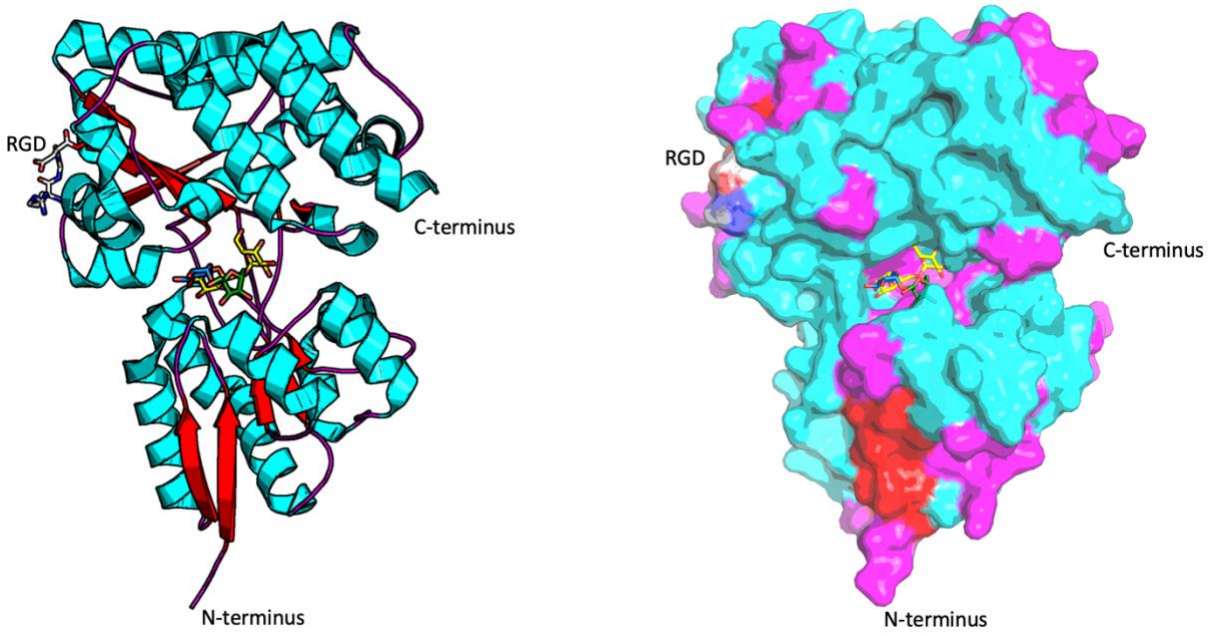


Figure 2-12. Domain architecture of RafE in complex with stachyose. The secondary structure of RafE is representative of Cluster D-I SBPs. RafE is shown here in a cartoon and surface model representation in complex with stachyose. α-helices are shown in blue, β-sheets are in red, and loops are purple. The RGD motif is shown as grey sticks.

2.3.8 RafE: Substrate recognition subsites

Overlaying all three RafE complexes revealed that regardless of the RFOs total length, a conserved binding site persisted at the non-reducing end of the sugars which we designated as subsite 1 (Figure 2-13). The binding interface of subsite 1 contained several hydrogen bonds between the RafE sidechains and the galactose residues O2 (Q95), O3 (Q378, E384), O4 (N193, Q378, E384), and O6 (N193) hydroxyl groups. Additionally, W376 appeared to form a platform with the galactose residues C2-C1-O1 atoms (Figure 2-13). Above the axial O4 hydroxyl group, a roughly 5 Å space exists indicating there would be ample room for an equatorial O4 hydroxyl like those present in the gluco-oligosaccharides panose and isomaltotriose.

A similar set of binding interactions was observed in subsite 2, with a mix of hydrophobic and polar interactions (Figure 2-14). Notably, when occupied by stachyose or verbascose, subsite 2 of RafE contained a galactose residue, but when bound to raffinose this site was occupied by glucose (Figure 2-14A). Though, the coordination between W274 appeared similar for both types

of sugar residues with an interaction between W274 and the O5-C1-C2 plane of the galactose/glucose residue. Hydrogen bonding interactions were apparent between RafE and the galactose residues O2 (E44), O3 (K42, E44) and O4 (K42, D308) hydroxyl groups. Similarly, the glucose residues O2 (E44), O3 (K42, E44), O4 (Y93, D308) hydroxyl groups formed nearly identical interactions (Figure 2-14).

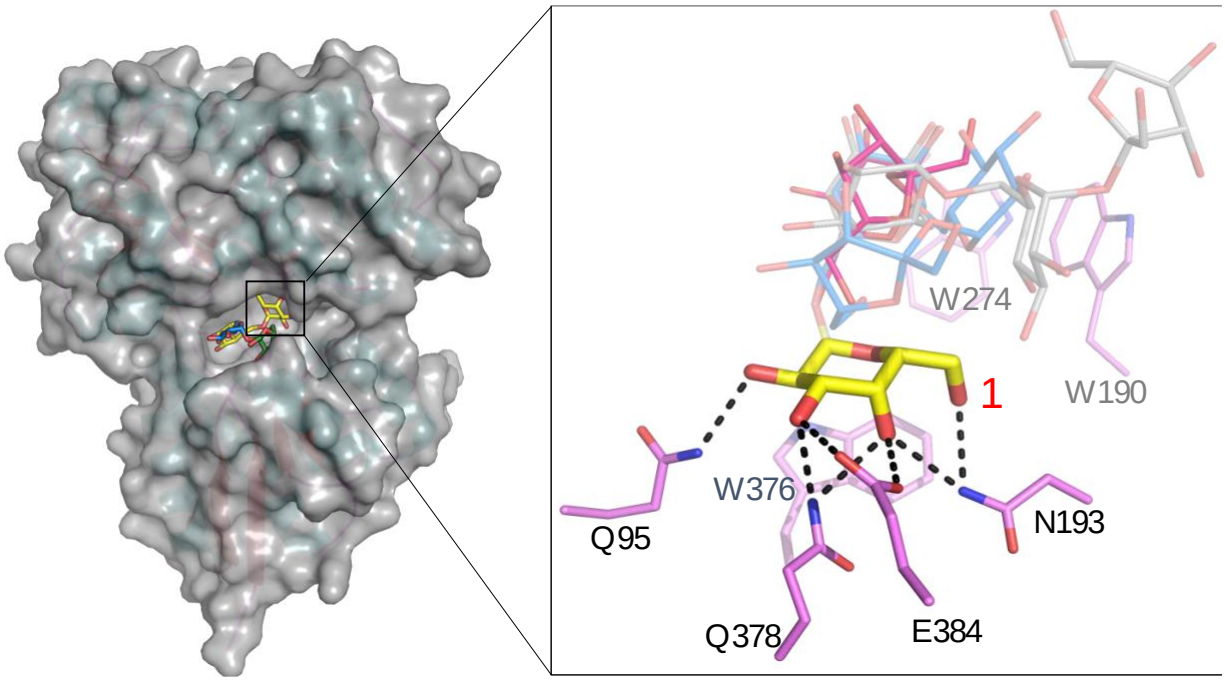


Figure 2-13. The conserved subsite 1 RFO binding interface. The RafE binding pocket overlayed with raffinose (pink), stachyose (blue), and verbascose (gray). Coordination of the subsite 1 non-reducing end galactose residue (yellow) is conserved between all three RFOs. Hydrogen bonding (black dashes) between galactose and RafE side chains (magenta) are between 2.2-3.2 Å apart.

The binding interactions presented in subsite 3 were much more diverse due to the presence of fructose residue when bound to raffinose, a glucose residue when bound to stachyose, and a galactose residue when bound to verbascose (Figure 2-14). The hydrophobic platform formed by W190 was only fully utilized by stachyose as both raffinose and verbascose formed very minor contacts (Figure 2-14). The hydrogen bonding interactions were limited in raffinose, as the O3 and O4 of fructose only formed hydrogen bonds with the C_α carbonyl and $O\epsilon$ of Q41, respectively. Regarding verbascose, the subsite 3 galactose residue was flipped away from W190, and instead

made hydrogen bonds between O1 (K43), O2 (G74), O3 (Q41), and O4 (Y93). Notably, the glucose and fructose residues of verbascose did not appear to form any hydrogen bonds with RafE side chains (Figure 2-14C). Stachyose had the highest affinity out of all the RFOs and upon examining the binding interface at subsite 3 revealed limited hydrogen bonds formed between K43 at hydroxyls O3 and O4. However, stachyose formed extensive hydrogen bonding at subsite 4 between its fructose residues O1 (N72), O3 (G74), O4 (Q41) hydroxyl groups. Notably, stachyose was the only RFO identified that forms a direct hydrogen bond with N72 and a tight interface with W190 (Figure 2-14B).

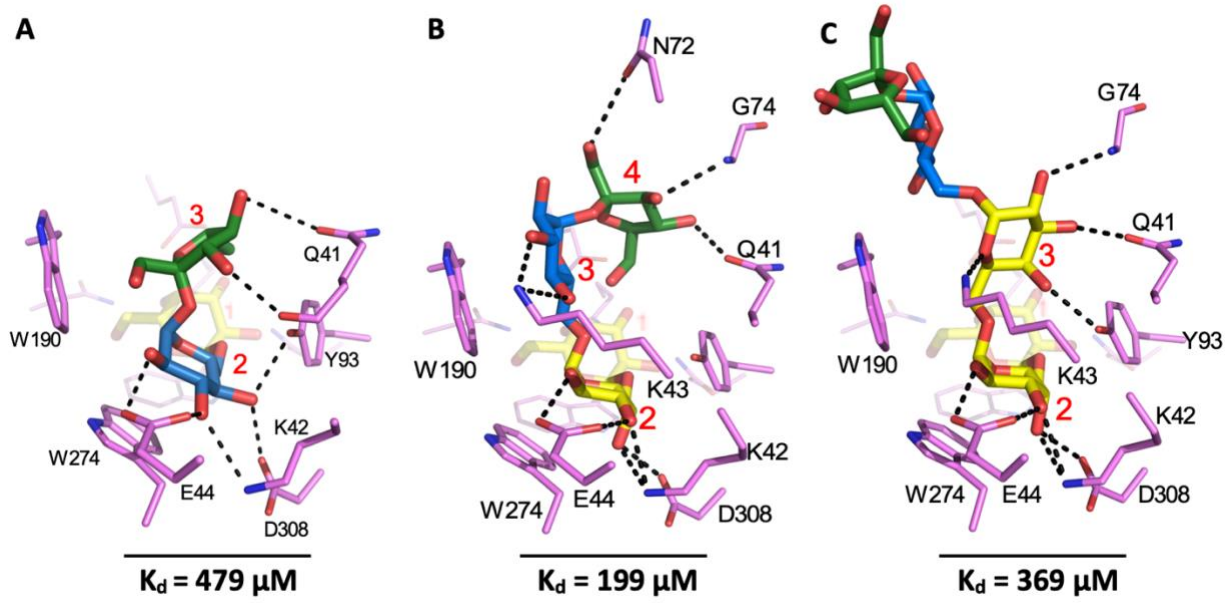


Figure 2-14. The adaptability of RafE subsites 2 to 4. Subsites 2-4 within the binding pocket of RafE can accommodates a multitude of carbohydrates as seen in the binding pocket interactions with (**A**) raffinose, (**B**) stachyose and (**C**) verbascose. Hydrogen bonding interactions (2.2-3.2 Å) between RafE side chains (magenta) and fructose (green), glucose (blue) and galactose (yellow) are indicated by black dashes. Experimentally calculated K_d values are indicated below each RFO.

2.3.9: A structural comparison of the RafE and BlG16BP binding pockets

A structural comparison search using the DALI[175] server indicated the raffinose- and panose-binding SBP BlG16BP (Z-score 39.3; 23% sequence identity; PDB 4ZZE) from *Bifidobacterium animalis* subsp*. lactis* Bl-04 was the highest match. Similar to RafE, BlG16BP was found to have a higher affinity towards panose relative to raffinose (17.5 µM and 27 µM, respectively)[176]. In addition, both SBPs share a similar ligand stacking platform comprised of three aromatic residues denoting subsites 1 to 3 (Figure 2-15). Though the platform created by BIG16BP differs with two less bulky aromatic residues at subsite 1 (F392) and subsite 2 (Y291). Despite the higher affinity seen with BIG16BP towards panose and raffinose, RafE has nearly 2-fold more hydrogen bonding interactions in subsite 1. Furthermore, there were no conserved sidechains observed within this binding subsite (Figure 2-15). Overall, the two enzymes share approximately 50% conservation in the residues involved in hydrogen bonding coordination.

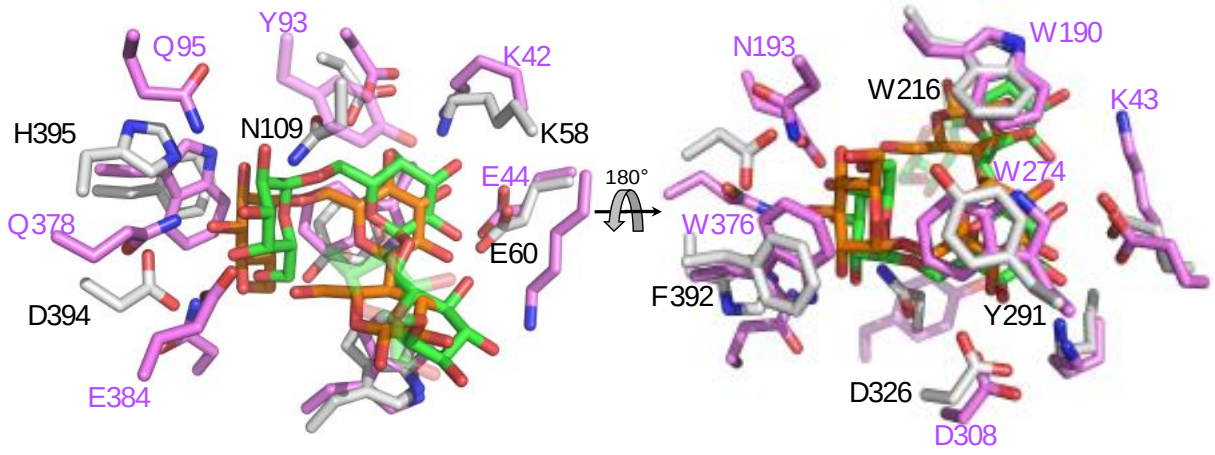


Figure 2-15. Binding pocket overlay of RafE and BlG16BP. The binding pockets of RafE (magenta) bound to stachyose (green), and the *B. animalis* SBP BlG16BP (light gray) bound to panose (orange). Residue numbering is shown in pink for RafE and black for BlG16BP.

2.4 DISCUSSION

Prior to this work, the biochemical functions of Aga, GtfA, and RafE had only been inferred through crude protein extracts and deletion mutations[153]. The goal of this study was to

biochemically characterize the three proteins associated with extracellular RFO recognition and subsequent depolymerization. Here we identified several biochemical properties pertaining to RafE, Aga and GtfA. Furthermore, we have constructed three structure complexes of RafE bound to the most prominent RFOs raffinose, stachyose and verbascose. In summary, the results shown here provide direct biochemical evidence that the *raf* locus within the *S. pneumoniae* genome contains the biological machinery required to recognize and depolymerize RFOs. Though, we have demonstrated the proteins are relatively inefficient in their expected roles.

2.4.1 RafE targets and determinants of recognition

One goal of this study was to characterize the binding specificity and affinity of RafE. Using ITC, we found that RafE binds to a diverse range of substrates including panose (Glu-α1,6-Glu-α1,4-Glu), isomaltotriose (Glu-α1,6-Glu-α1,6-Glu) and the RFOs (Gal_n-α1,6-Glu-α1,2-Fru). A similar range in substrate binding is also present in the multiple sugar metabolism loci of *S. mutans* and *S. suis*, which utilize the RFOs, melibiose, and gluco-oligosaccharides[154,177]. However, unlike its relatives the pneumococcal SBP does not bind to the disaccharide melibiose; and with respects to the substrates tested here RafE appears to discriminate against sugars that are less than three residues in length (Table 2-3). The unexpected preference towards panose was an interesting discovery. However, we were unable to produce crystals complexes with a resolution limit below 3 Å in a timely manner. Thus, a model of RafE in complex with panose is not included in this work.

By extrapolating from the structure of BIG16BP bound to panose overlayed in the binding site of RafE (Figure 2-15) it appears that panose would form an addition hydrogen bond between the subsite 1 H388 residue and the equatorial O4 hydroxyl group present in gluco-oligosaccharides like panose. Additionally, within the RafE subsite 2 position panose would likely resemble that of the raffinose complexes' subsite 2 binding interface which also contains a glucose residue at that position (Figure 2-14A). Thus, an additional hydrogen bond would be expected to occur between Y93 and O4 of the glucose residue in panose. In subsite 3 we could expect additional hydrogen bonds between the O6 hydroxyl group of panose and E384 of RafE (Figure 2-15). Additionally, we might see a hydrogen bond between the O2 of panose and N72 of RafE. However, this interaction

is highly dependent upon the loss of fructose at subsite 4 seen in stachyose (Figure 2-14B) permitting more flexibility in the N-terminal lobe of the binding pocket thereby allowing the N72 residue to move approximately 2.2 Å closer to the glucose residue creating a hydrogen bonding distance of around 3.0 Å. Notably, an induced fit caused upon panose binding may be leading to the high affinity for panose relative to the RFOs, and we would not be able to account for this by simply comparing alignments. In summary, our findings suggest that RafE has evolved towards utilizing panose, or substrates resembling, relative to RFO substrates. While panose bound RafE with 3- to 6-fold stronger affinity than any of the RFOs (Table 2-3), it still had a relatively weak K_d that is roughly 80-fold higher than the average SBP K_d value of 1 μM[163].

In addition to panose, we also found that RafE bound to the gluco-oligosaccharide isomaltotriose with a K_d of 427 μM. Considering isomaltotriose and panose only differ by the linkage type at the reducing end, it can be rationalized that the α1,6-linkage is not the preferred substrate linkage at this position. This is likely due to the increased length and flexibility of α/β1,6-linkages which is evident in the verbascose structure (Figure 2-14C) as the subsite 3 sugar residue "twists" away from W190. Thus, it can be proposed that the subsite 3 reducing end of isomaltotriose twists away from W190 in a similar fashion as verbascose, reducing the hydrophobic interactions while also losing the hydrogen bonding interaction between the subsite 3 O4 hydroxyl group and Y93 due to the presence of an equatorial hydroxyl group in glucose.

2.4.2 Rationalizing the apparent binding affinities

RafE bound all of the substrates tested with relatively low affinity (80 μM – 479 μM). This suggests that we did not identify the main target of this transport system and we should absolutely increase the carbohydrate panel tested in follow-up experiments. However, an interesting alternative for RFO utilization, that is not focused on nutrient acquisition, is that RafE may have evolved to become a low-affinity SBP towards the RFOs and other oligosaccharides in the environment. Once the substrates are internalized, they act as potent cytosolic transcriptional regulators. This idea is supported by several studies which have shown that RFOs uptake in *S. pneumoniae* profoundly upregulates transcription of the *raf* locus and the shared ATPase RafK[153]. Marion et al., also demonstrated that *rafE* gene transcripts, and in theory the *raf*

locus, are remarkably upregulated (>1000-fold increase from base level) when pneumococcus is grown in raffinose relative to sialic acid and maltotetraose[155]. Comparatively, when pneumococcus was grown in malto-oligosaccharides or sialic acid, the respective sialic acid and maltotetraose SBPs were only upregulated a fraction of the change seen with raffinose (40- to 70-fold increases, respectively) suggesting RFOs may be potent transcriptional regulators. Furthermore, Tan et al., demonstrated that when *S. suis* was grown in 1% w/v raffinose or glycogen, the homolog of *rafK* (*msmK*; 79.3% sequence identity) was substantially upregulated compared to several other sugar types[154]; however, this has yet to be shown in *S. pneumoniae*. The same authors later revealed MsmK is essential for upregulating several stress response genes in *S. suis*[178], thus MsmK appears to have regulator properties in addition to ATPase activity. Altogether, these findings provoke the idea that RFOs and other choice oligosaccharides may not be utilized with a high affinity; however, the small amounts that are imported have a profound impact on transcriptional regulation.

The current understanding of RFO utilization demonstrates that upon RFO uptake, a robust increase of the *raf* locus transcription occurs[153,155]. Though, the effects of utilizing the remaining RafE targets on pneumococcal gene regulation is unknown. Given the ability of RafE to target multiple oligosaccharides, and the observed effects of RFO utilization on the transcriptional regulation of the *raf* locus, understanding the targets of RafE may help elucidate the mysterious roles associated with the *raf* locus. That being said, future studies could include additional ITC assays using a larger carbohydrate panel to expand our understanding of what sugars RafE may be targeting *in vivo*. As panose bound with the highest affinity, continuing to search for a high affinity targets of the *raf* locus could include additional panose-series gluco-oligosaccharides including tetra- and penta-saccharide derivatives of glycogen. As a follow up to this, using the identified substrates as a sole carbon source for pneumococcal growth, we could perform multiple RT-qPCR transcriptomic analyses focusing on identifying changes in the expression of select genes within the *raf* locus, like lipoteichoic acid ligase (*rafX*); as well as genes located outside of the *raf* locus such as *rafK* and the stress response genes identified by Tan et al[178].

To date, no available literature has explored the link between *rafX* mediated teichoic acid synthesis, and RFOs utilization. As RFO utilization leads to a profound increase in *raf* expression,

and teichoic acids serve essential roles in multiple aspects of pneumococcal biology, it is important to fully understand the key drivers regulating its production. Therefore, we could explore the effects of RFO utilization on the formation of teichoic acids in both the bacterial cell wall and bound to its lipid membrane. Utilizing mass spectrometry or immunoblotting, we could quantify and contrast the differences in teichoic acid production when pneumococcus is grown in different carbohydrate sources that are targets of RafE.

2.4.3 Aga substrate specificity and kinetics

A second goal of this study was to identify the substrate specificity and kinetics of the α-galactosidase Aga. Structurally, GH36 enzymes embody a large N-terminal distorted β-sandwich domain connected to a central α/β TIM-barrel domain that is linked to a small β-sandwich C-terminal domain. GH36 enzymes are reported to contain several activities including α-galactosidase, α-N-acetyl-galactosaminidase, raffinose/stachyose synthase, transgalactosylase and kinase activity[58,179–181]. The sequence of Aga relative to other GH36 family members places it into subgroup I which is reported to contain members with α-galactosidase and transgalactosylase activities[182].

Previous studies have identified *aga* is essential for pneumococcal RFO fermentation and α-galactosidase activity[153]. Additionally, *aga* was identified in two independent signature-tagged mutagenesis screens as a virulence gene in pneumococcal pneumonia[54,148]. This work demonstrates Aga is an exo-(α1,6-, α1,3-)-galactosidase that cannot tolerate branching adjacent to the non-reducing end galactose residue like those present in type II blood group antigens (Figure 2-4). The kinetic assays performed in this work reveal Aga has a slightly higher K_M (~2- to 20-fold) relative to other GH36 enzymes, thus indicating that it may be slightly inefficient at depolymerizing the examined substrates[181,183]. Additionally, our kinetic assays indicate Aga has a higher catalytic efficiency towards shorter oligosaccharides containing α1,3-galactosyl linkages relative to α1,6-linkages.

Within humans, the galactose residues found in various glycans are predominantly β1,4-galactosides and α1,3-galactosides . β1,4-galactosides are found in the abundant lactose linkages found in various N- and O-linked glycans, while the only known terminal α1,3-galactosyl linkages

are in the type II B blood group antigen[42,49]. According to our TLC assays Aga is unable to cleave the branched type II B blood group antigen, presumably because of the α1,2-fucose attachment adjacent to the terminal galactose (Figure 2-4). Furthermore, the only known pneumococcal α1,2-fucosidase is incapable of removing fucose residues on type II B blood group antigens[184]. Consequently, unless an α-fucosidase capable of generating linear type II B blood group antigen is identified, our results only support the role of Aga in degrading dietary and microbial carbohydrate structures. As Aga is highly conserved amongst different *S. pneumoniae* serotypes and strains (average >98% sequence identity; KEGG[185]) it likely serves an important role within the bacterium. However, how this is related to α1,3- and α1,6-galactosidase activity is currently unknown.

2.4.4 GtfA: The inefficient sucrose phosphorylase

The third goal of this study was to identify the activity of the GH13, GtfA. The GH13 family is one of the most functionally diverse families of GHs, containing over thirty unique functions and forty-three subfamilies. Despite the wealth of functionality, all GH13s embrace a common three domain structure[186]. The largest component, domain A, encompasses the central α/β TIM-barrel which houses the catalytic site. Domain B comprises an extended arm region which reaches outwards from the TIM-barrel to varying degrees and is involved in substrate recognition and coordination. Whereas domain C comprises a β-sheet or β-barrel structure at the C-terminus, and there are no known functions of this domain. The amino acid sequence of the GtfA from *S.* pneumoniae places it nearest to the GH13_18 subfamily. GH13_18 comprises sucrose phosphorylases, sucrose-6'-phosphorylases, glucosyl-glycerol phosphorylases and glucosyl-glycerate phosphorylases[187].

Overall, GtfA is the least understood component of the *raf* locus. The only literature relating to GtfA has revealed that it is non-essential for growth on raffinose[153], and otherwise there is no other literature regarding this enzyme. In our experiments, we found that GtfA had trace activity towards sucrose when assayed under similar conditions as the *S. mutans* GtfA (84% sequence identity), suggesting the two enzymes may have divergent functions[162]. Thus, the scarcity of sucrose phosphorylase activity witnessed in our TLC assays may result from a lack of appropriate

acceptor-doner molecules used within our analyses. To determine if GtfA has a more efficient activity than what we have uncovered, future studies could include a wider range of acceptor and donor molecules, using a similar TLC approach.

2.4.5 A putative model for the highly conserved pneumococcal *raf* locus

Collectively, using the data gathered in this portion of the thesis, together with current knowledge of the pneumococcal *raf* locus, we can build a model for the *raf* locus. In this model, the SBP RafE binds to extracellular carbohydrates like the RFOs, and subsequently presents them to the membrane permeases RafF and RafG. Once bound to the membrane permeases, the ATPase RafK drives the conformational changes within the ABC transporter's transmembrane domains, thereby facilitating substrate internalization. Within the cytoplasm the RFOs, or perhaps their mono- and disaccharide constituents, can interact with cytosolic proteins including transcriptional regulators that results in the upregulation of the *raf* locus[153,155,188]. The RFOs are depolymerized by Aga forming galactose and sucrose molecules. Subsequently, GtfA cleaves sucrose into fructose, glucose and glucose-1-phosphate. The resulting monosaccharides are then further metabolized through the Leloir pathway, glycolysis, and pentose phosphate pathway; or they are repurposed towards various biosynthetic pathways such as capsule, peptidoglycan and teichoic acid synthesis (Figure 2-16)[87,93,96,189].

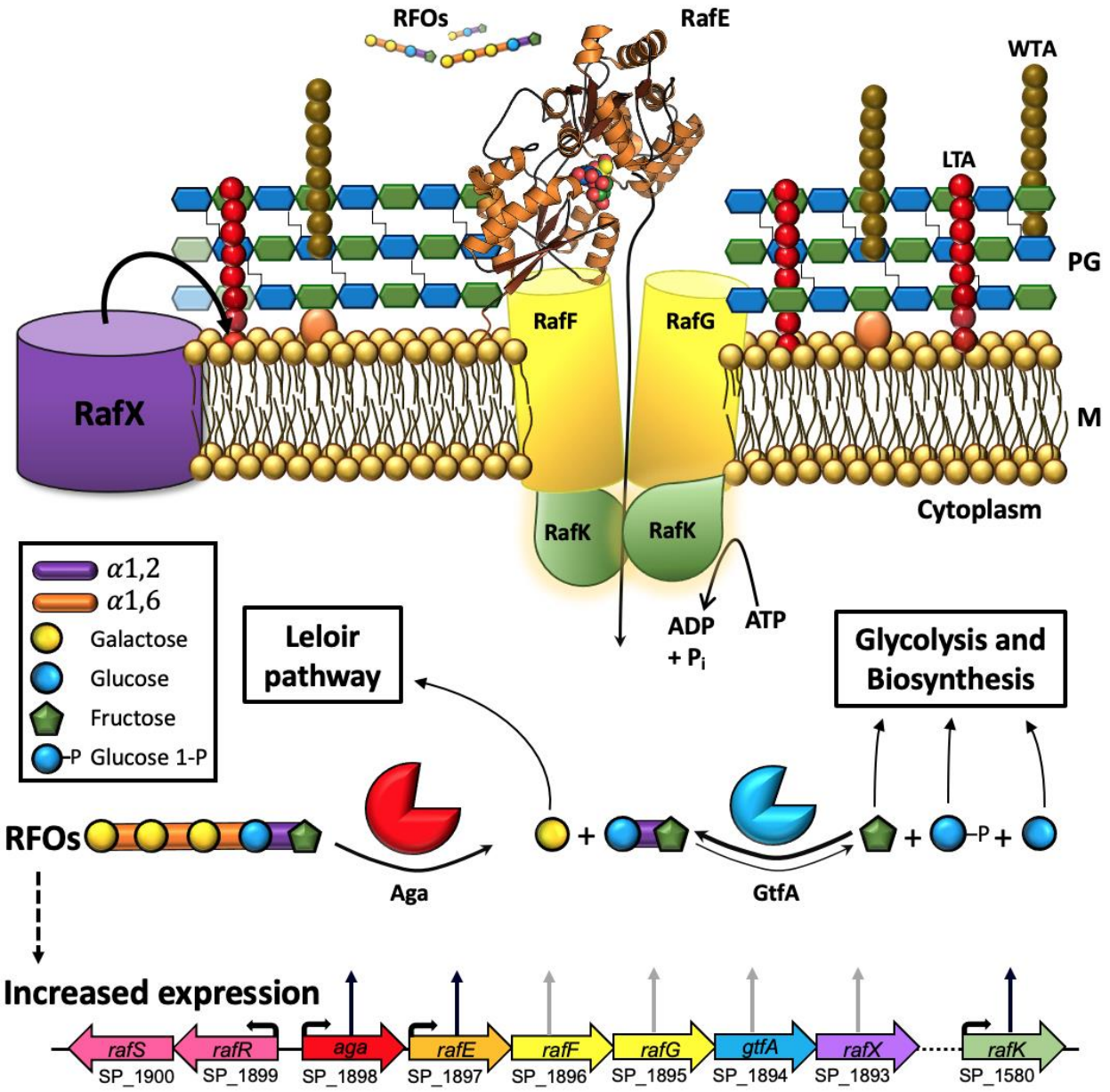


Figure 2-16. An overview of the *S. pneumoniae raf* locus. Extracellular substrates such as RFOs, panose, isomaltotriose and possibly others are recognized by RafE and presented to the membrane permeases RafF and RafG. The substrates are then transported into the cytoplasm by conformation changes driven by RafK ATP hydrolysis. Once in the cytoplasm the substrates are depolymerized by various carbohydrate processing enzymes or they may interact with intracellular binding partners. The presence of RFOs directly increases the expression of the *raf* locus genes, and the shared ATPase RafK$^{\alpha}$. $^{\alpha}$Black vertical arrows indicate genes measured by RT-qPCR or western blots in the presence of RFOs, gray arrows refer to genes which are co-

transcribed with a gene that was directly measured, thus in theory are also expressed. Abbreviation: RFOs, raffinose family oligosaccharides; WTA, wall teichoic acid; LTA, lipoteichoic acid; PG, peptidoglycan; M, plasma membrane.

2.5 CONCLUSIONS

In this study we characterized several properties of the SBP RafE and the glycoside hydrolases, Aga and GtfA. We have described several distinct structural features of RafE bound to the most prominent RFOs; additionally, we revealed RafE is a multi-substrate binding protein that is capable of binding to RFOs, panose and isomaltotriose. We have also demonstrated that Aga has exo-(α1,3-, α1,6-)-galactosidase activity, and that it is capable of depolymerizing the RFOs into galactose and sucrose (Figure 2-4). Lastly, we have found GtfA has trace phosphorylase activity against sucrose. Collectively, the data presented here somewhat substantiates the putative roles of RafE, Aga and GtfA in RFO processing. However, due to the high conservation of the *raf* locus, the significant inefficiencies of the studied proteins towards their putative substrates, and the presence of a lipoteichoic acid ligase that is transcribed with the ABC transporter (Figure 2-1); our results suggest that the *raf* locus within *S. pneumoniae* has evolved an unknown role that may include but is not limited to RFO uptake.

Table 2-4. X-ray crystallography data collection and model statistics for RafE.

	RafE		
	Raffinose	**Stachyose**	**Verbascose**
Data Collection			
Beamline	Home Beam	Home Beam	Home Beam
Wavelength (Å)	1.540	1.540	1.540
Space Group	$P2_1$	$P2_12_12_1$	$P2_1$
Cell Dimensions			
a, b, c (Å)	48.86, 145.82, 118.68	49.28, 118.84, 146.01	48.92, 145.88, 118.37
Resolution (Å)	30.00-2.65 (2.70-2.65)	30.00-2.35 (2.39 - 2.35)	30.00-2.40 (2.44-2.40)
R_{merge}	0.142 (0.565)	0.084 (0.181)	0.152 (0.530)
R_{pim}	0.083 (0.369)	0.036 (0.095)	0.076 (0.344)
CC 1/2	0.987 (0.729)	0.995 (0.966)	0.984 (0.706)
$\langle I/\sigma I \rangle$	11.0 (2.0)	20.2 (6.7)	9.4 (2.2)

Completeness (%)	99.9 (100.0)	99.9 (99.8)	99.6 (94.3)
Redundancy	4.0 (3.6)	5.0 (3.6)	4.0 (3.1)
No. of reflections	191428	183357	255814
No. Unique	47919	36605	64353
Refinement			
Resolution (Å)	2.65	2.35	2.40
R_{work}/R_{free}	0.24/0.29	0.22/0.24	0.23/0.26
No. of atoms			
Protein	2968 (A), 2967 (B), 2972 (C), 2882 (D)	3001 (A), 2986 (B)	2991 (A), 2979 (B), 3000 (C), 2934 (D)
Ligand	34 (raffinose-A), 34 (raffinose-B), 34 (raffinose-C), 34 (raffinose-D)	45 (stachyose-A), 45 (stachyose-B), 26 (CIT), 16 (EDO)	56 (verbascose-A), 56 (verbascose-B), 56 (verbascose-C), 56 (verbascose-D), 4 (EDO)
Water	239	363	304
B-factors			
Protein	36.4 (A), 40.0 (B), 39.3 (C), 49.0 (D)	28.0 (A), 30.0 (B)	32.8 (A), 36.6 (B), 34.9 (C), 47.5 (D)
Ligand	31.3 (raffinose-A), 33.7 (raffinose-B), 37.3 (raffinose-C), 37.5 (raffinose-D)	25.0 (stachyose-A), 29.9 (stachyose-B), 48.7 (CIT), 49.8 (EDO)	33.4 (verbascose-A), 36.6 (verbascose-B), 34.6 (verbascose-C), 41.8 (verbascose-D), 32.1 (EDO)
Water	31.3	32.2	31.4
R.M.S. deviations			
Bond lengths (Å)	0.006	0.006	0.006
Bond angles (°)	1.07	1.11	1.06
Ramachandran (%)			
Favored	96.9	97.8	96.7
Allowed	3.1	2.2	3.3
Outliers	0.0	0.0	0.0

The values for highest resolution shells are shown in parentheses. CIT, citric acid; EDO, ethylene glycol.

2.6 MATERIALS AND METHODS

2.6.1 Gene cloning and plasmids

The genes encoding the *Streptococcus pneumoniae* TIGR4 RafE (lacking the signal sequence and lipid-anchoring motif residues 24-419; Sp_1897) and full length GtfA (Sp_1894) were PCR amplified from the TIGR4 genomic DNA using the primers: RafE_F (5'-GCCGCGCGGCAGCCATTCCAACTATGGTAAATCTGCGGATG-3') and RafE_R (5'-GCTCGAATTCGGATCCTAATCCACATCCGCTTTCATCG-3'); or GtfA_F (5'-GCCGCGCGGCAGCCAAATGCCAATTCAAAATAAAACCATGTTGATTACCT-3') and GtfA_R (5'-GCTCGAATTCGGATCTCAAAATGATACTTCAACTCCATTCTCAATTACC-3'). PCR products were cloned

into a pET28a plasmid between the NdeI and XhoI sites by In-Fusion cloning following supplier protocols (Takara Bio, USA) producing pET28a_RafE and pET28a_GtfA. The desired proteins, fused to a N-terminal 6x-histidine tag, were verified with bi-directional Sanger sequencing (Sequetech, USA). The plasmid containing the full-length Aga gene (Sp_1898) was provided.

2.6.2 Transformations, protein expression and purification

Plasmids were transformed into *E. coli* BL21 (DE3) cells for protein expression. RafE and GtfA expression cultures (3 x 2L) were inoculated with 2 mL of a 10 mL 37 °C overnight LB culture containing 50 µg/mL kanamycin. At an OD_{600} = 0.6 the temperature was lowered to 16 ºC, cultures were induced roughly one hour later to a final concentration of 0.5 mM isopropyl thio-β-galactoside. After 16 hours of shaking at 16 ºC the cells were centrifuged at 8,000 x g for 10 minutes. Cells were chemically lysed by resuspension in 35% (w/v) sucrose, 1% (w/v) deoxycholate, 1% ($^{V}/_{V}$) Triton X-100, 500 mM NaCl, 10 mg lysozyme, and 0.2 µg/mL DNase in 20 mM Tris pH 8.0. Lysate was then centrifuged at 16,000 x g for 30 minutes. The cleared supernatant was then loaded onto a Ni^{2+}-NTA immobilized metal affinity chromatography column. Protein was eluted using binding buffer (150 mM NaCl, 20 mM Tris pH 8.0) at increasing imidazole concentrations (0 to 500 mM). Proteins were concentrated and buffer exchanged into 20 mM Tris pH 8.0, 150 mM NaCl using an Amicon stirred ultrafiltration device. Purified protein was further purified using size exclusion chromatography with either a S100 or S200 HiPrep 16/60 Sephacryl column (GE Healthcare). Protein was again concentrated using an Amicon ultrafiltration device. Aga was expressed using the autoinduction method[190] and all other purification steps are as forementioned. Protein concentrations were determined by UV absorbance at 280 nm using the extinction coefficients of 73,800 M^{-1} cm^{-1} (RafE), 72,540 M^{-1} cm^{-1} (GtfA), and 96,000 M^{-1} cm^{-1} (Aga).

2.6.3 Dynamic light scattering

DLS was performed on RafE following size-exclusion chromatography using a DynaPro plate reader (Wyatt Technology, Santa Barbara, CA, USA) with a wavelength of 833.78 nm at 25˚C. Three replicates of 40 µL were loaded and 10 acquisitions of 5 s were performed for each

sample. Data were collected and analyzed using the accompanying Dynamics V7.1 software under the globular proteins model.

2.6.4 Thin layer chromatography substrate specificity assays

TLC screening of Aga linkage specificity was performed with 1 mM substrate and 100 nM enzyme in Aga assay buffer at 37 °C for 18 hours. Reactions and standards were spotted onto precoated POLYGRAM SIL G/UV254 TLC sheets (Thermo Fisher Scientific), separated in a solvent of 2:1:1 butanol: acetic acid: distilled H2O, and visualized with 0.2% (w/v) napthoresorcinol in acidified ethanol followed by heating at 90 °C. For GtfA-containing reactions, 10 – 30 mM of substrate (pentasaccharide to disaccharide) was digested with 30 µM GtfA and/or 100 nM Aga in 100 mM phosphate pH 6.5 buffer at 37 °C for 18 hours. Reactions and standards were spotted and separated in a solvent of 6:7:1 chloroform: acetic acid: distilled H2O and TLC plates were visualized with acidified ethanol followed by heating at 90 °C.

2.6.5 Aga α-galactosidase kinetics assays

All Aga enzyme assays were performed at 37 °C. Aga was screened for activity against a dozen monosaccharide analogs composed of glycoside-α-pNPs and glycoside-β-pNPs using 200 nM enzyme in 50 mM phosphate buffer (NaH_2PO_4/K_2HPO_4 pH 6.5) with 1 mM substrate. Release of pNP was monitored at 405 nm in a SpectraMaxM5 plate reader (Molecular Devices, San Jose, CA). Kinetic constants for Aga against pNP-α-Gal were determined with 1 nm Aga in phosphate buffer by following the release of pNP directly at 405 nm. For all other substrates, kinetic constants were measured by quantifying the release of α-galactose in a stopped assay format. Reactions contained 40 nM Aga and varying concentrations of substrate in Aga phosphate buffer. Samples were taken every 3 to 5 minutes and stopped by heating to 95 °C for at least 10 min. Once cooled, 25 µL from each time point and concentration were mixed with 75 µL components from the L-arabinose/D-galactose assay kit (Megazyme Inc., Chicago, IL) according to manufacture recommendations. Detection of galactose production through the Megazyme assay kit was achieved through the following: conversion of α-galactose to β-galactose (galactose mutarotase), followed by the reduction of $NADP^+$ to NADPH (β-galactose dehydrogenase), and measuring the increase at 340 nm. Kit-assay reactions were incubated at 25 °C and read at 340

nm every 10 seconds until the absorbance stabilized. Final absorbances were converted to galactose concentrations according to the manufacturer's instructions and accounting for the dilution factor of the original Aga reaction.

2.6.6 RafE isothermal titration calorimetry

ITC was performed using a VP-ITC (MicroCal, Northampton, MA) instrument. RafE was extensively dialyzed with 50 mM phosphate buffer pH 6.6, at 4 °C. Dialysis buffer was stored, and used to dilute RafE and ligand to the appropriate concentrations prior to ITC analysis. Both the protein and ligands were filtered and degassed prior to analyzing their binding interactions. Assays were carried out at 25 °C with 2.5 mM ligand titrated into 100 µM RafE. ITC experiments included a 2 µL pre-injection, which was discarded from analysis, followed by 25 x 10 µL injections into an ITC cell volume of 1400 µL. Data were fit to a one-site binding model due to a relatively low experimental c value (c <5).

2.6.7 RafE crystallization and structure determination

Initial complex structures were obtained by co-crystallization using a sitting drop vapor diffusion method at 18 °C. Optimization screens were then carried out using hanging drop vapor diffusion at 18 °C. Complex structures were obtained by co-crystallizing the protein at 28 mg/mL in the presence of 10 mM sugar. Co-crystals were grown in 0.1 M Tris pH 8.0, 0.2 mM CsCl, 20% (w/v) PEG 3350 for the raffinose and stachyose complexes and the addition of 0.1 M $MnCl_2$ was used with verbascose.

Diffraction data for structure determination was collected on the "in-house" beamline. Diffraction data were collected at a wavelength of 1.540 Å using Pilatus 200K 2D detector coupled to a MicroMax-007HF X-ray generator with a VariMaxTM-HF Arc/Sec Confocal Optical System and an Oxford Cryostream 800. Diffraction data were processed using *HKL-2000*[165], and statistics were compiled in Table 2-4. The structure of RafE in complex with stachyose was solved by molecular replacement using a previously published RafE model (PDB 2I58) as a search model. The structures of RafE in complex with raffinose and verbascose were solved by molecular replacement using the refined RafE-stachyose model (PDB 6PRG) as a search model. All

structures were solved using *PHASER*[168] and iterative cycles of manual building in *COOT*[169] and refinement with *REFMAC*[170]. For all structures, the addition of water molecules was performed in *COOT* with *FINDWATERS* and manually checked after refinement. For all datasets, refinement procedures were assessed for accuracy by flagging 5 % of all observations as "free"[191]. Model validation was performed in *PHENIX*[171] using *MOLPROBITY*[192]. The structures were visualized using *PyMOL*[193].

CHAPTER 3: SpGH92 inhibitor screening assays

Contributions: I performed the protein production, purification, activity assay, differential scanning fluorimetry, HDX-MS set-up and data analysis in this chapter. Brandon Moeller and Kaelin Fleming kindly performed the mass-spectroscopy instrument operation.

3.1 ABSTRACT

The success of *S. pneumoniae* as a deadly opportunistic human pathogen is greatly attributed to its ability to catabolize a wide array of important host glycans. The degradation of High Mannose N-Glycans (HMNGs) that are present on numerous important glycoconjugates, is one process that is vital for hematogenous spread. The enzyme SpGH92 is recognized as a potent virulence factor that initiates the depolymerization of HMNGs and is vital for *S. pneumoniae* pathogenesis in models of pneumonia and sepsis. The function of SpGH92 as an exo-α1,2-mannosidase is to trim the terminal α1,2-mannoside linkages from HMNGs, thereby initiating their full deconstruction. Due to SpGH92s profound impact on pathogenesis, the development of protocols using a variety of biophysical tools to study its interaction with inhibitors in a rapid and unbiased manner is essential for developing future therapeutics. Utilizing HDX-MS, we have developed a system to study the structural dynamics of SpGH92 in solution and in response to inhibitors, as demonstrated using the characterized α-mannosidase inhibitor, mannoimidazole. Initial attempts at examining the binding interface of a novel inhibitor, LIPS343, were attempted; however, it appears that the 2 μM inhibition constant found for this compound is not driven through protein binding. Instead, it is likely working through an indirect effect with the initial screening assay. Overall, this study reveals novel structural insights into the virulence factor controlling the deconstruction of mammalian HMNGs and sets the framework for further inhibitor screening using HDX-MS.

3.2 INTRODUCTION

The influence of host-glycan processing has far reaching implications on the virulence phenotype of *Streptococcus pneumoniae* (pneumococcus)[62]. Indeed, several of the microbe's glycan-processing enzymes have been structurally and functionally characterized, including several which are involved in N- and O-linked glycan degradation[45,47,52,194–196]. The enzymes involved in complex N- and O-linked glycan degradation are important for effective mucus breakdown and colonization, however their significance in hematogenous spread is conflicting[195,196]. On the other hand, the degradation of HMNGs is strongly implicated in the progression of hematogenous spread[47]. Within the gene cluster(s) responsible for HMNG degradation, a fundamental gene encoding the *S. pneumoniae* exo-α1,2-mannosidase, SpGH92 (*S. pneumoniae* glycoside hydrolase family 92), is recognized as the master regulator of HMNG degradation[47]. SpGH92 serves the critical role of trimming terminal α1,2-mannobiose linkages found on HMNGs (Figure 3-1), thereby resulting in the conversion of $Man_9GlcNAc_2$-Asn to $Man_5GlcNAc_2$-Asn. The next enzyme of this pathway is an endo-β-N-acetyl-glucosaminidase (EndoD). Notably, EndoD participates in the final stages of both HMNG and complex N-glycan degradation[47]. In the case of HMNG degradation, EndoD recognizes the product of SpGH92 catalysis and cleaves between the conserved N-glycan chitobiose core (GlcNAc-β1,4-GlcNAc), releasing $Man_5GlcNAc$ (Figure 3-1). Free $Man_5GlcNAc$ is accepted by the solute binding protein NgtS and is subsequently transported into the cytosol by the ABC_{NG} transporter to be further metabolized.

Strikingly, knockout studies revealed SpGH92 is the primary cause of the profound virulence phenotype associated with HMNG degradation. While EndoD, which is responsible for removing a large portion of the HMNG also contributes to pathogenesis, its impact is significantly less[47]. Intriguingly, this implies that the conversion of $Man_9GlcNAc_2$-Asn to $Man_5GlcNAc_2$-Asn is primarily responsible for the observed pathogenicity, and not the complete removal of HMNGs. Although it's too soon to rule out an indirect binding interaction between SpGH92 and a component of the host, this observed virulence phenotype is likely due to a disruption of important HMNG-containing glycoconjugates *in vivo*. HMNGs are found on many important

human glycoproteins including a main component of the complement amplification loop, complement component 3 (C3), which contains two HMNG attachment sites in its structure[197]. Several other pneumococcal glycoside hydrolases are also implicated in immune evasion[132]. Thus, it is possible that the degradation of key HMNGs, like those found on immune proteins may be altering the host's immune response, leading to a profound reduction in pneumococcal clearance. Due to its profound impact on the *S. pneumoniae* virulence phenotype, SpGH92 is a prime candidate for future drug and vaccine developments.

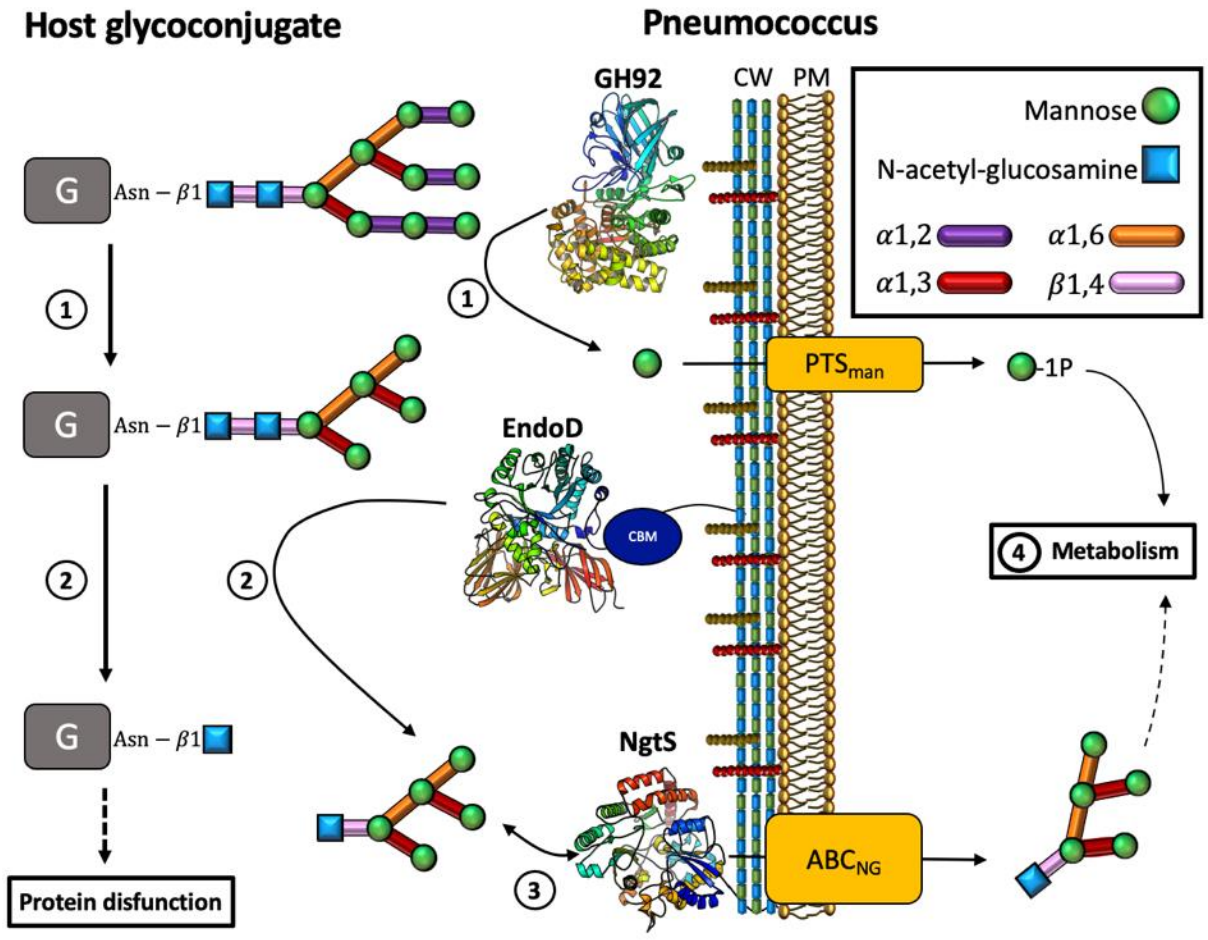


Figure 3-1. High-Mannose N-Glycan degradation by *S. pneumoniae*. In the initiating step of HMNG degradation **(1)** a $Man_9GlcNAc_2$-Asn-glycoconjugate is degraded by SpGH92 producing $Man_5GlcNAc_2$-Asn-G. Subsequently, EndoD recognizes the product of SpGH92 and cleaves the $Man_5GlcNAc_2$-Asn-G from its corresponding glycoconjugate **(2)** releasing it into the extracellular space. The substrate binding protein NgtS can then bind the released $Man_5GlcNAc$ **(3)** and

present it to the ABC_{NG} transporter where it is subsequently taken into the cytoplasm and further metabolized **(4)**. Abbreviations, G, glycoconjugate; CW, cell wall; PM, plasma membrane; CBM, carbohydrate binding module.

The current selection of α-mannosidase inhibitors have relatively weak binding affinities. Zhu et al., found that the selection of common α-mannosidase inhibitors have inhibition constants ranging from 5 µM to 12 mM using Bt3990, a close homolog of SpGH92 found in *Bacteroides thetaiotamicron* VPI-5482[40]. In addition, these inhibitors are not very specific, thereby targeting a broad range of α-mannosidases including eukaryotic lysosomal, endoplasmic reticulum, and Golgi α-mannosidases that results in immunomodulatory effects[198–201]. Recently, a novel SpGH92 inhibitor (LIPS343) was identified in a high-throughput screening assay and showed a low initial inhibition constant towards SpGH92 (K_i = 2 µM; Vocadlo et al., unpublished). In contrast to the K_i-values found with Bt3990, the 2 µM K_i of LIPS343 is approximately 2.5-fold lower than the expected K_i of swainsonine (5 µM), 48-fold lower than mannoimidazole (96 µM), 70-fold lower than kifunensine (140 µM), and 6000-fold lower than deoxymannojirimycin (12000 µM)[40]. While the apparent K_i of LIPS343 is promising especially considering it has not been optimized, one drawback is the lack of available information regarding the structure and binding interface of LIPS343. To overcome this challenge numerous attempts to form crystal complexes of SpGH92 bound with LIPS343 were attempted, but after exhausting multiple approaches the inhibitor did not incorporate into SpGH92 crystals. Therefore, a new approach was required to study the binding interface of SpGH92 with the novel inhibitor LIPS343. Utilizing Hydrogen Deuterium eXchange Mass-Spectrometry (HDX-MS), we have developed a new approach for characterizing the binding interfaces between SpGH92 and candidate inhibitors. Due to the abundance of structural information available regarding the binding interface of Bt3990 bound to mannoimidazole (Figure 3-2), we integrate mannoimidazole as positive control into the development of a new system for studying the binding interactions between SpGH92 and inhibitors.

HDX-MS is used as it a powerful tool for measuring the conformational dynamics of proteins in solution and the structural changes occurring in response to environmental changes[202]. HDX-

MS provides an adaptable platform that can be used to study a variety of interactions and conformational changes including those related to protein-protein complex formation[203], protein-ligand interactions[204,205], membrane interactions[206], viral capsid stability[207], and protein glycosylation[208]. This technique takes advantage of the protein's polypeptide backbone, which readily exchanges its amide protons (N-H) with the aqueous environment when they are not secured by secondary structure hydrogen bonding. Therefore, when a protein of interest is incubated in a buffer containing heavy water (D_2O), mass-spectrometry can accurately quantify the degree of deuterium incorporation across the peptide backbone over multiple time points. The rates of deuterium exchange vary among regions of a protein, with regions containing stable secondary structure incorporating less deuterium than regions containing disordered loops or flexible secondary structure. Thus, HDX-MS provides valuable insight into the structural dynamics of proteins in solution. Because the conformational changes upon small molecule binding can be accurately quantified and illustrated, HDX-MS is a valuable tool for identifying the binding interfaces of inhibitors to SpGH92.

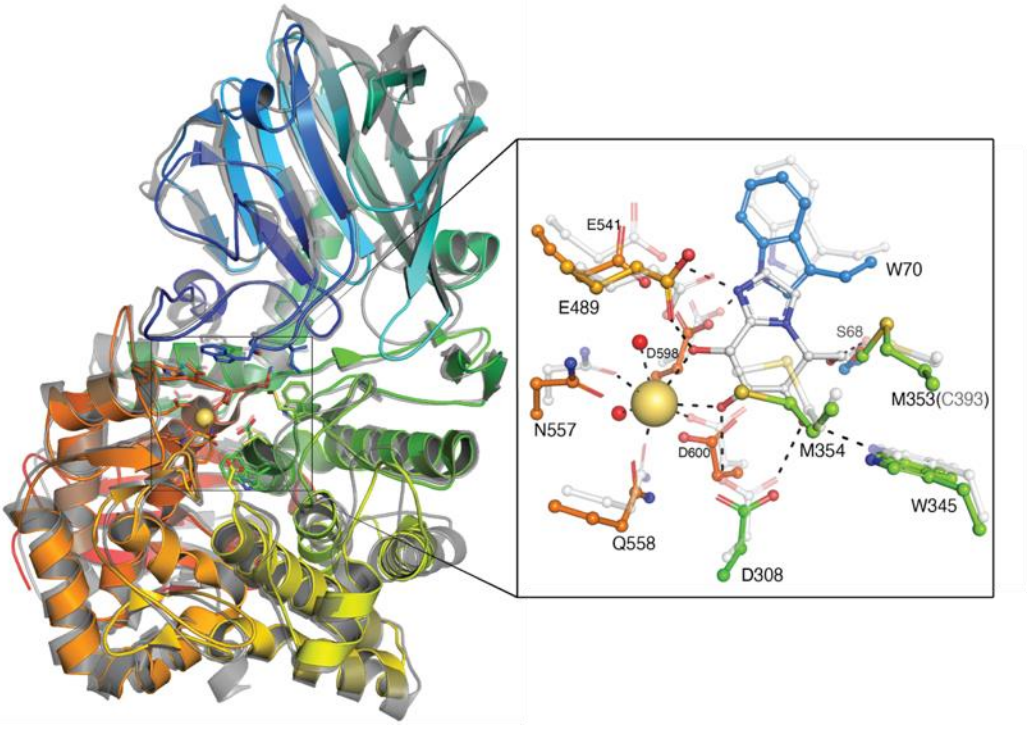


Figure 3-2. Structural comparison of SpGH92 and Bt3990. An overlay of SpGH92 (PDB 5SWI; rainbow) and Bt3990 (PDB 2WZS; transparent gray) bound to mannoimidazole (white) shows the

high structural similarity between the two enzymes. GH92s share a characteristic two-domain structure comprising an N-terminal β-sandwich domain and a larger C-terminal $(\alpha/\alpha)_6$ barrel domain. The two domains frame a roughly 10 Å deep active site containing a vital calcium ion (yellow sphere) near the -1 subsite that is required for catalysis and aids in coordinating the O2 and O3 hydroxy groups of the bound substrate[40]. Within the -1 subsite of SpGH92 the substrate coordinating residues are S68, D308, W345, M353, M354, E489 and D600; while in the +1 subsite the key residues are W70, H540 and E541. Notably, a substitution of C393 to M353 in SpGH92 introduces a longer hydrophobic residue into the -1 subsite that likely alters the inhibitor binding interaction slightly by increasing the hydrophobic interface along the C5-C6 plane of the bound mannoside. The catalytic mechanism of GH92s follow a one-step, inverting displacement whereby the glycosidic bond attaching the non-reducing end mannoside is cleaved by a nucleophilic water molecule through a single displacement, subsequently releasing the new non-reducing end sugar and a β-mannose monosaccharide[202]. Overall rmsd = 1.58, SS_{ID} = 87%, and Seq_{ID} = 30%. The black dashes represent hydrogen bonds (2.2-3.2 Å) that correspond to the interactions identified in the Bt3990-mannoimidazole structure. While the amino acid labels correspond to SpGH92 with the single Bt3990 variant (C393) shown in parenthesis.

Here we present a system for studying the structural dynamics of SpGH92 in solution and the conformational changes occurring in response to inhibitor binding. Using the methodology outlined here we have uncovered the structural dynamics of apo-SpGH92 in solution, as well as the changes occurring in response to the characterized α-mannosidase inhibitor mannoimidazole and the novel inhibitor LIPS343. In the presence of mannoimidazole we observed a large increase in stability of SpGH92, particularly around the active site region. However, in the presence of LIPS343 we did not observe any significant changes in the structural dynamics of SpGH92. This study, in addition to highlighting the relative stability of SpGH92 in solution and the key areas undergoing change in response to inhibitor binding, describes the use of HDX-MS in studying the binding interactions of small molecule inhibitors to the deadly virulence factor SpGH92.

3.3 RESULTS AND DISCUSSION

3.3.1 SpGH92 protein expression and purification

To investigate the structural dynamics of SpGH92, the full-length protein was expressed and purified from *E. coli* (BL21) with an additional N-terminal 6x-histidine tag. SpGH92 was first purified by immobilized-Ni^{2+} affinity chromatography followed by size exclusion chromatography before flash freezing in liquid nitrogen until further use (Figure 3-3). The protein identity was established by tandem mass spectrometry and enzyme activity was confirmed towards α1,2-mannobiose using the D-mannose detection kit (Megazyme).

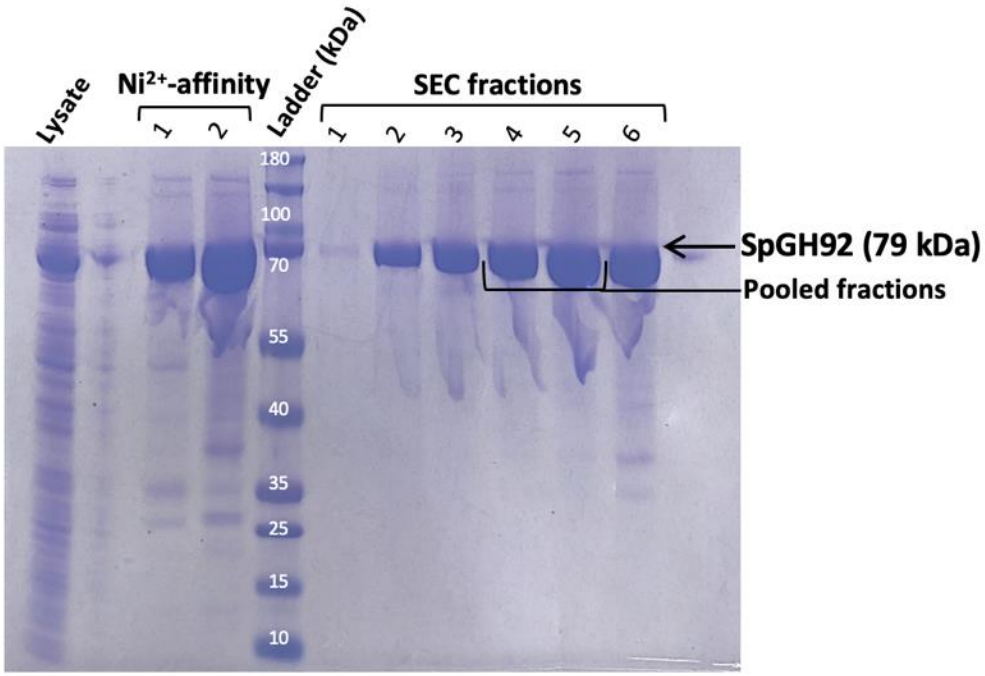


Figure 3-3. SDS-PAGE analysis of SpGH92 purification. SpGH92 was purified by immobilized Ni^{2+}-affinity chromatography and size exclusion chromatography. The protein ladder corresponds to a 180 kDa protein ladder (Page Ruler™ Thermofisher).

3.3.2 Designing an HDX-MS methodology

To elucidate the structural dynamics of SpGH92, we began our HDX-MS analysis with a timecourse assay. Using a buffer containing 20 mM HEPES pH 7.5, 100 mM NaCl and 75.5% D_2O the secondary structural dynamics of apo-SpGH92, SpGH92 with 200 µM mannoimidazole and SpGH92 with 200µM LIPS343 were examined using five time points of exchange (0.3, 3, 30, 300, and 3000 s) at 18°C. Following D_2O incubation, each reaction was stopped using an ice-cold acidic quench buffer (0.6M guanidine-HCl, 0.9% formic acid and pH ~2.5 final concentration) before being flash frozen in liquid nitrogen until mass-analysis. A key step in HDX-MS analysis is the generation of a peptide map that is used to determine the locations of deuterium incorporated across the protein. To produce peptide maps for SpGH92, the protein samples were injected onto an ultra-performance liquid chromatography system containing an in-line pepsin column, C-18 trap column, C-18 analytical column, and a QTOF mass spectrometer. The levels of deuterium incorporation throughout the protein samples were calculated using *HDExaminer* (Sierra Analytics) and an initial "heat map" of the SpGH92 corresponding to the levels of deuterium incorporation was produced. All peptides were manually inspected for the correct charge state and peak selection. Following data processing, the refined heat maps for all conditions showed suitable peptide overlap, high coverage and redundancy, and a low variation in deuterium incorporation for each individual time point (Figure 3-4). Surprisingly, we saw a trend of low deuterium incorporation throughout the majority of peptides with these initial buffer conditions. Even at the 3000 s post D_2O incubation time point, our SpGH92 sample contained <10% deuterium (%D) in ~20% of the amide backbone indicating the protein is inherently quite stable.

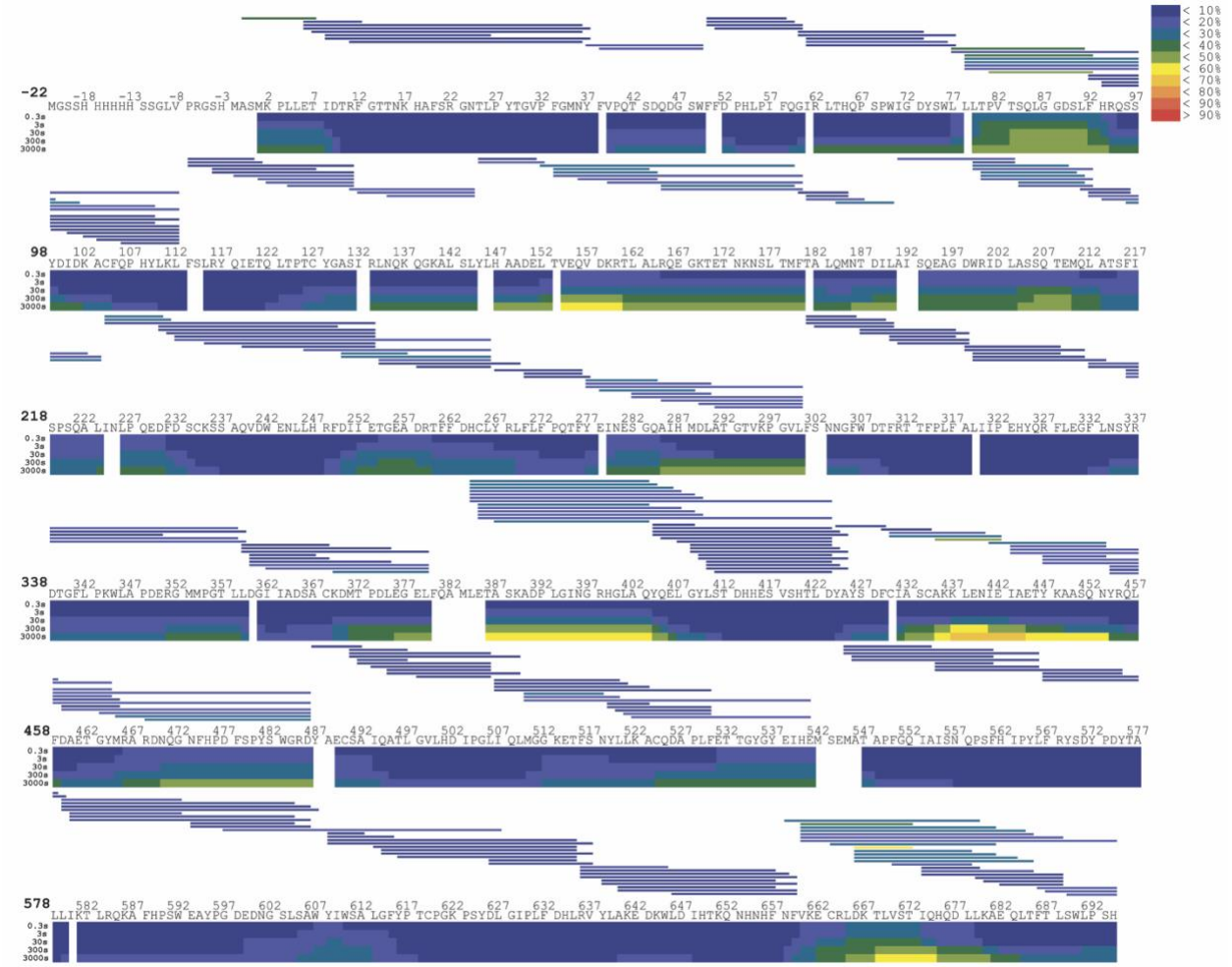


Figure 3-4. SpGH92 heat map at pH 7.5 reveals low deuterium incorporation. A pepsin digest and sequence coverage map displaying the level of deuterium incorporation (legend, top right) at each time point for each amino acid. The thin bars on top correspond to peptides produced from the pepsin digest, and the thicker rows below represent the %D incorporation for each individual amino acid at each time points. Pepsin peptides and amino acid residues are colored according to the average %D incorporation. HDX-MS statistics for apo-SpGH92: Coverage = 96.1%, average peptide length = 16 $\pm$ 7.6 residues, redundancy = 5.9, replicates = 3, repeatability = 0.4 %D.

For simplicity of analysis, the heat map created for apo-SpGH92 was transposed onto the protein model (PDB 5SWI). Using this recapitulated view, the differences in deuterium incorporation across the protein were easily viewed in a three dimensional format (Figure 3-5).

We saw clear regions in both the N-terminal β-sandwich domain (residues 130∼230) and the C-terminal $(\alpha/\alpha)_6$ barrel domain (residues 370∼440, 430∼480, 512∼540, 607∼612, 660∼694) that contained regions of relatively fast exchange rates, with low to moderate secondary structure flexibility colored light blue to yellow (Figure 3-5). Notably, many of the regions around the active site where we would expect mannoimidazole to bind incorporated almost no deuterium, even at our latest time point. Consequently, this lack in deuterium incorporation severly limited our ability to detect any changes in stability occuring in response to inhibitor binding. Nevertheless, we proceeded to analyze the differences between the apo- and inhibitor bound states of SpGH92.

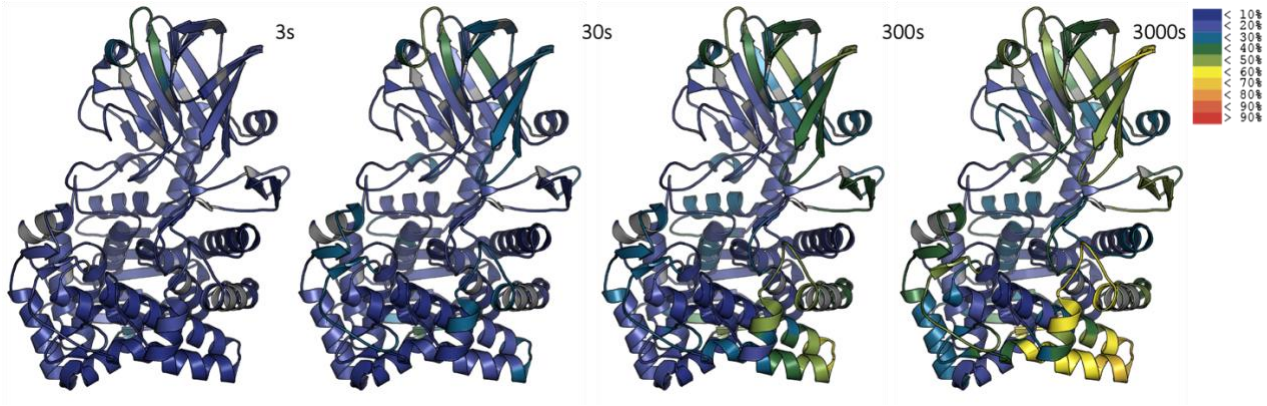


Figure 3-5. Heat map transposition reveals low deuterium exchange around active site. A tertiary view of apo-SpGH92 representing the deuterium incorporation as a heat scale. Areas of low coverage are shown in grey. Regions of relatively higher deuterium uptake are seen at the apexes (green-yellow). Labelling conditions correspond to 2 µM SpGH92, 75.5 %D, 20 mM HEPES pH 7.5, and 100 mM NaCl.

3.3.2.1 Mapping inhibitor binding interfaces

After we compared the deuterium incorporation profiles between apo-SpGH92 and the two inhibitor-bound states, our data revealed a subtle decrease in deuterium incorporation within the presence of mannoimidazole. However, there were no significant changes seen upon the addition of LIPS343 using our significance cut-off criteria of 3% deuterium (>3 %D), 0.3 deuterium atom (>0.3 DA), and the two data sets must pass a two-tailed Student's *t*-test with a p-value less than 0.01.

As anticipated, mannoimidazole binding induced protection in multiple peptides close to the active site including the peptides spanning residues 61-73, 336-359 and 384-406 (Figure 3-6). These results were consistent with the known Bt3990-mannoimidazole binding interactions, as the majority of peptides contained one or more active site residues (Figure 3-6). Notably, the peptide that spanned residues 384-406 does not contain any known active site residues, however the protection seen in this region was likely caused by the more prominent conformational change occurring in the neighboring peptide 336-359 (Figure 3-6).

Surprisingly, the few changes that were observed using our initial assay conditions were small (3 - 5 %D) which is extremely close to our significance cut-off values (Figure 3-6C). In addition, many regions of the amide backbone never incorporated more than 10% deuterium, even at our longest time point. Therefore, many regions of the protein that may have experienced conformational changes were invisible to our assay under these initial conditions as we cannot detect changes in conformational dynamics of those very stable amides undergoing little D_2O incorporation. These initial results indicated that a new methodology was needed to gain better insight into the conformational changes occurring in response to SpGH92-inhibitor binding.

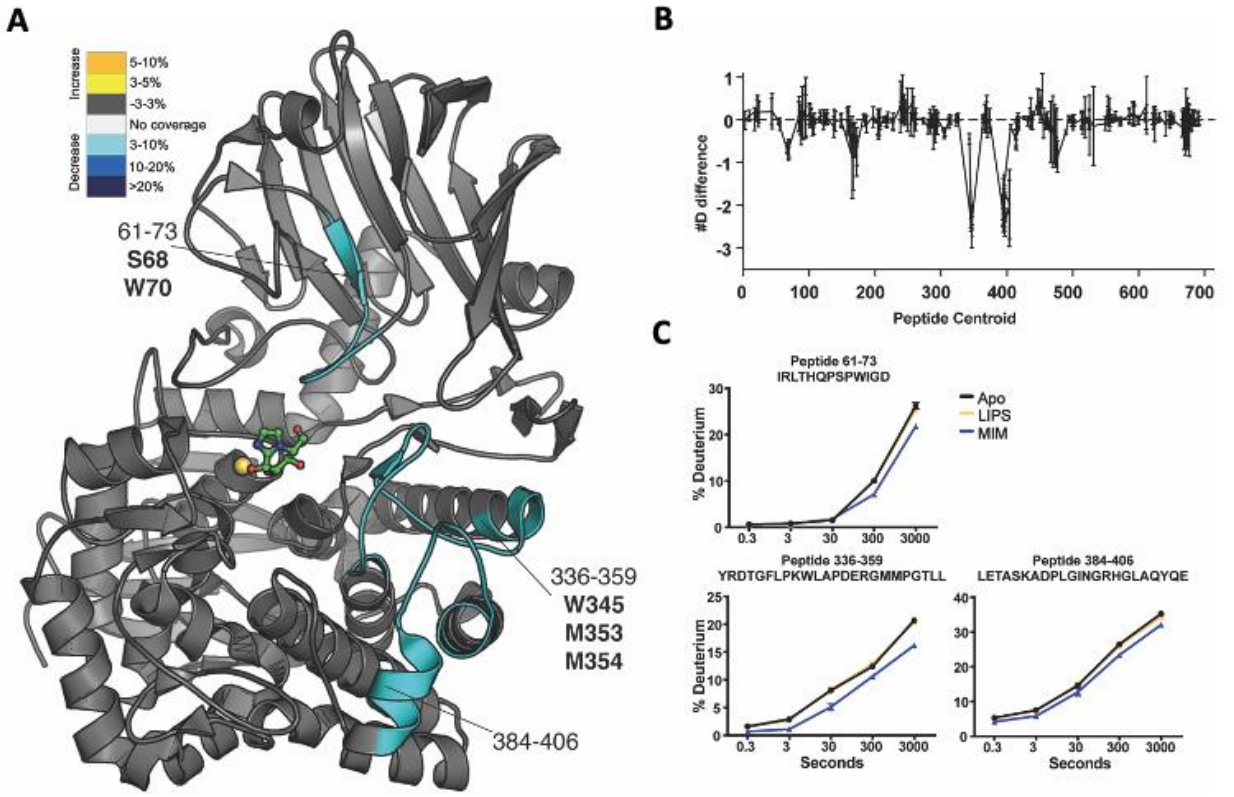


Figure 3-6. Differences in HDX reveal an increased protection factor near the SpGH92 active site in the mannoimidazole group. (A) Regions of SpGH92 that show significant changes of exchange (defined as >3 %D, >0.3 DA., and two-tailed Student's *t*-test p <0.01) in the presence of mannoimidazole. Peptides around the active site show the most protection from exchange. **(B)** The sum of the deuterium atoms (#D) protected from exchange in the presence of mannoimidazole across all time points. Each point represents a single peptide where the centroid residue of that peptide is plotted on the x-axis. Error bars represent standard deviation (n = 3). **(C)** The %D incorporation of selected peptides at different time points (0.3, 3, 30, 300 and 3000 s). Error bars represent the standard deviation (n = 3). The active site calcium ion is shown as a yellow sphere, with mannoimidazole in green.

3.3.3 HDX-MS methodology and buffer optimization

Many of the regions that were expected to bind to mannoimidazole contained very stable secondary structure binding networks that resulted in very little deuterium incorporation at pH 7.5. Therefore, to increase the low deuterium exchange rate we reworked the buffering

conditions, timescale and inhibitor concentrations. To address the first aim of increasing the exchange rate, we raised the alkalinity of the buffer by one pH unit using 100 mM Tris pH 8.5, 150 mM NaCl, and 1 mM $CaCl_2$. At pH 8.5, the deuterium exchange rate was expected to increase approximately one order of magnitude in relation to the roughly ten-fold increase in hydroxide ions present thereby increasing the rate of OH-catalyzed exchange[210,211]. Additionally, 1 mM $CaCl_2$ was added to the buffer to ensure the active site of SpGH92 was saturated with a Ca^{2+} ion, thereby eliminating the possibility that calcium deficiency could be affecting the inhibitor binding in our HDX-MS assays. Furthermore, the concentration of mannoimidazole was increased to from 200µM to 400 µM to compensate for any potential decrease in the binding affinity of SpGH92 relative to the inhibition constant of 96 µM reported for the Bt3990-mannoimidzaole interaction. Lastly, the maximum time point was increased to 10,000 s to encourage a high degree of deuterium saturation, while the previously ineffective 0.3 s and 3 s time points were removed.

3.3.3.1 Thermal shift assay analysis

As a precautionary measure to ensure the stability of SpGH92 was not negatively impacted upon the pH shift, differential scanning fluorimetry (DSF) was utilized. DSF is a biophysical technique that is commonly used in buffer optimization and drug discovery[212]. The reason DSF is a great tool for optimizing buffer conditions and screening ligands is because it is relatively rapid to set up and it can be used to simultaneously measure numerous different conditions at once[213]. The technique takes advantage of a fluorescent dye that increases in fluorescence when bound to hydrophobic surfaces (of an unfolding protein), yet it is quenched in aqueous environments. Thus, while measuring the fluorescence as the temperature increases, when the energy input begins to reach and surpass the amount of energy holding the protein in a folded state, the melting temperature (T_m) can be measured. The melting temperature, or the temperature at which 50% of the protein is folded and 50% in unfolded, is a good indicator of protein stability; therefore, by comparing the shifts in melting temperature between varying buffer conditions or inhibitors, the corresponding values can be used to determine optimal buffer conditions and inhibitor binding[214].

Using a StepOnePlus™ Real-Time PCR (Applied Biosciences) to monitor the changes in fluorescence intensity corresponding to the fluorescent dye SYPRO Orange, we determined the melting temperatures of SpGH92 in the pH 7.5 and pH 8.5 buffer systems. The analysis was performed using a 1.7 °C/min ramp rate that revealed a 3.74 °C thermal shift (ΔT_m) between the apo-SpGH92 at pH 8.5 (T_m = 56.49 °C) and apo-SpGH92 at pH 7.5 (T_m = 52.75 °C) samples. Thus, our results revealed SpGH92 was stabilized at pH 8.5 relative to pH 7.5 (Figure 3-7).

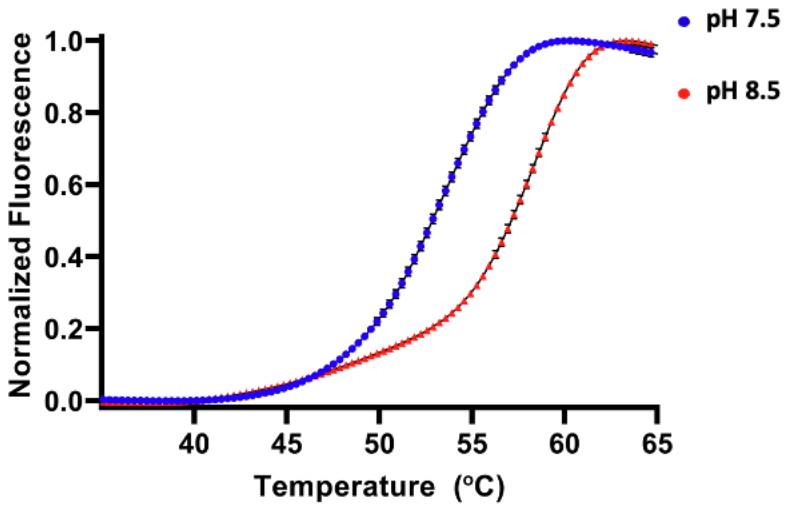


Figure 3-7. The T_m of SpGH92 increases at pH 8.5. Thermal unfolding of SpGH92 monitored using SYPRO Orange (max fluorescence = 586 nm) indicates SpGH92 has a higher stability at pH 8.5 (T_m = 56.49 °C) relative to pH 7.5 (T_m = 52.75 °C). The melting temperatures were calculated using the StepOnePlus™ Real-Time PCR system software. Data are representative of triplicates.

3.3.4 HDX-MS analysis of SpGH92 at pH 8.5

To determine the structural dynamics of apo-SpGH92 and the changes incurred upon mannoimidazole and LIPS343 exposure we examined the deuterium incorporation at pH 8.5 using four time points (10, 100, 1000 and 10,000 s) and a final D_2O concentration of 70.7%. Following the same general methodology as previously (Chapter 3.3.2), the refined heat maps of apo-SpGH92 and SpGH92 bound to each inhibitor revealed a high sequence coverage and redundancy, with a low variation in deuterium incorporation between replicates from all three groups at each time point (Figure 3-8). Importantly, by using the new buffering conditions we

saw a significant increase in deuterium incorporation across the entire protein, therefor we anticipated an increase in detection throughout the protein.

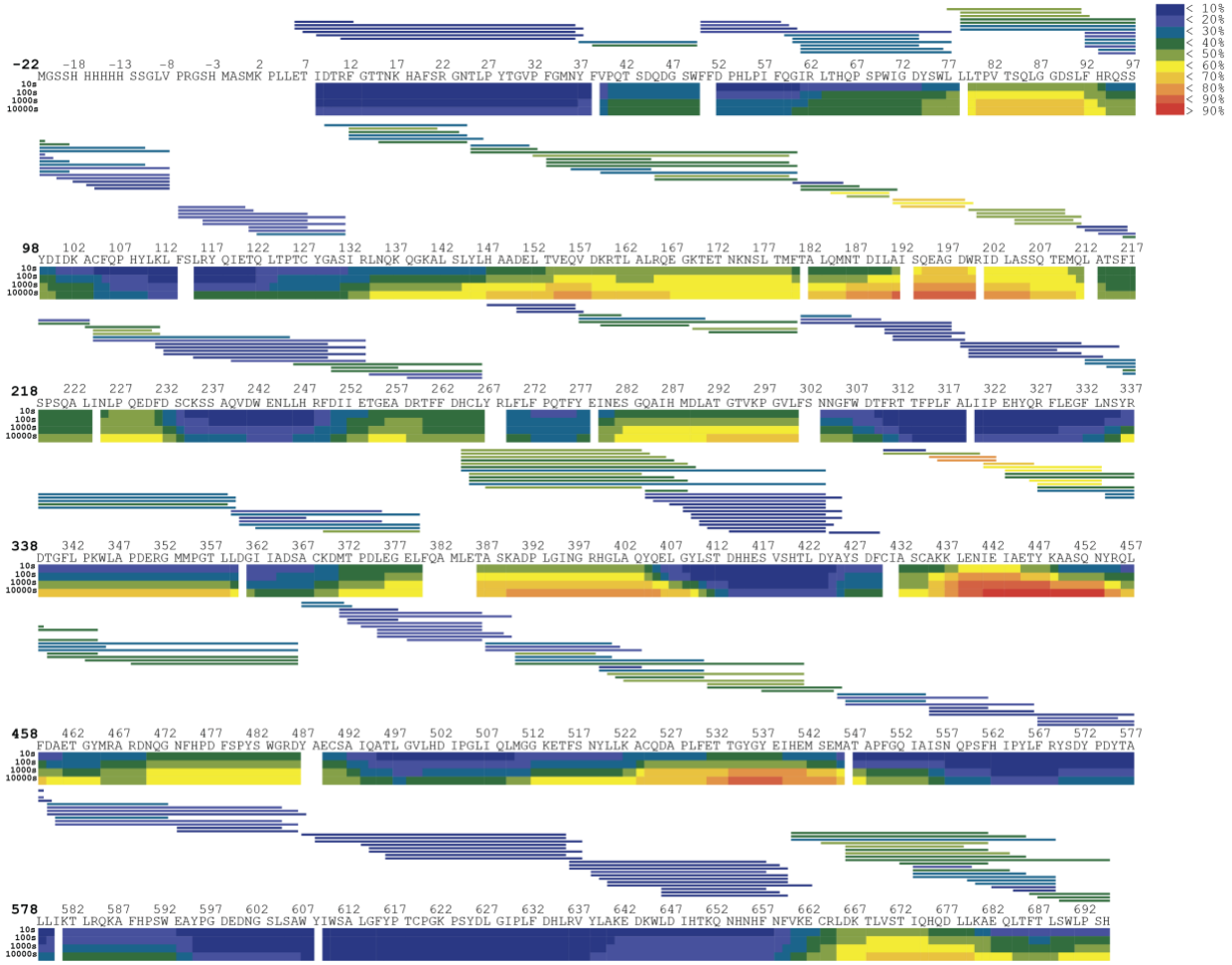


Figure 3-8. SpGH92 heat map at pH 8.5 shows increased levels of deuterium exchange. The peptide and sequence coverage map shows the level of deuterium incorporation (legend, top right) for each time point at every residue. The thin bars on top correspond to the peptides produced from pepsin digest and the thicker rows beneath describe the %D for the individual amino acid residues at each time point. Peptides and individual residues are rendered in color showing the calculated deuterium incorporation. Statistics for apo-SpGH92; coverage = 95.5%, average peptide length = 15.7 $\pm$ 7.7 residues, redundancy = 5.7, replicates = 3, repeatability = 0.6 %D.

Comparing the deuterium incorporation profiles of the two buffering conditions revealed that the 3000 s time point of the pH 7.5 timecourse assay looked the most similar to the 100 s time point at pH 8.5. This suggests that the exchange rate increased nearly 30-fold relative to the pH 7.5 values. Furthermore, the regions that contained the greatest flexibility within the N-terminal β-sandwich and C-terminal $(\alpha/\alpha)_6$ barrel domains at pH 7.5 were increasingly evident at pH 8.5. The greatest exchange was seen in the apexes of SpGH92, as seen on the apo-heat map model (Figure 3-9). However, despite the dramatic increase in deuterium incorporation throughout the protein, several regions of low incorporation persisted across the N-terminus (residues 8~37) and C-terminus (313~332, 413~423, and 607~637) at the final time point of 10,000 s (Figure 3-9). Notably, the regions of highly stable secondary structure occur around the core of the C-terminal $(\alpha/\alpha)_6$ barrel domain close to where the vital calcium ion is bound.

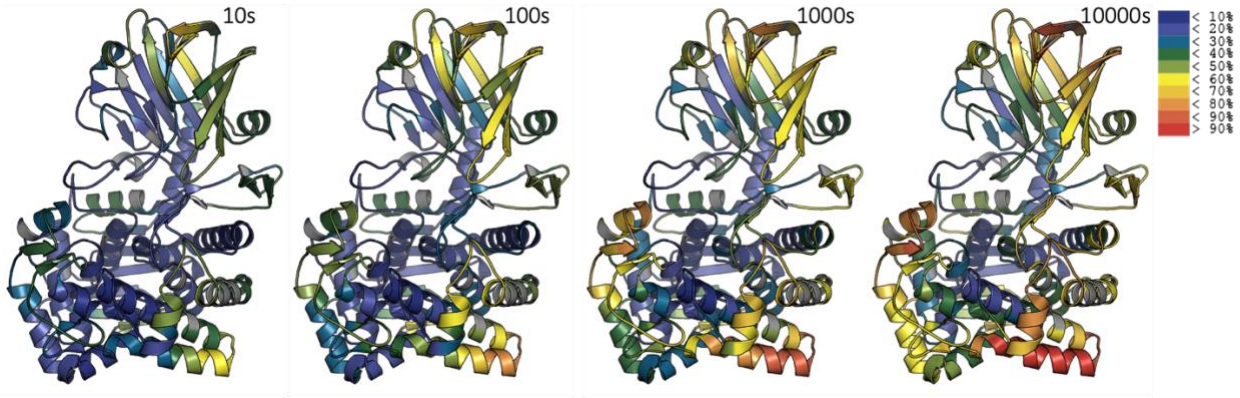


Figure 3-9. At pH 8.5 a dramatic increase in deuterium incorporation is seen across SpGH92. A tertiary view of apo-SpGH92 showing the deuterium incorporation as a heat scale. Areas of low coverage are shown in grey. The areas of highest %D uptake are at the apexes. Labelling conditions correspond to 2 µM SpGH92, 70.7% deuterium, 100 mM Tris pH 8.5, 150 mM NaCl and 1 mM $CaCl_2$.

3.3.4.1 Mannoimidazole induces a significant conformational change in SpGH92 at pH 8.5

To identify any additional conformational changes that were made detectable using the pH 8.5 buffer, we compared both the apo-SpGH92 and inhibitor bound states. Our results revealed conformational changes occurring in the presence of both inhibitors using our restructured buffer and assay format. In addition, the difference in deuterium between the apo and inhibitor bound states were considerably larger than that seen in our first timecourse assay.

In the presence of mannoimidazole we saw a huge increase in the number of peptides that underwent a significant change around the active site. In addition, these results were consistent with the initial timecourse assay, as the highest levels of protection occurred within the peptides containing active site residues that were known to interact with the inhibitor (Figure 3-10). The regions that showed the highest protection factors were seen in peptides 61-73 (S68 and W70; 12 %D), 301-309 (D308; 17.25 %D), 336-359 (W345, M553, and M554; 15.1 %D), 487-592 (E489; 14.2 %D), 537-544 (H540 and E541; 26.1 %D), and 545-561 (N557 and Q558; 9.6 %D). Though, many additional regions of protection were seen across various other regions of the enzyme, especially within the C-terminal domain between residues 300-550. Interestingly, mannoimidazole binding also resulted in destabilization of the alpha helix comprising amino acids 443-453 (Figure 3-10).

Taken together, the results of this timecourse assay revealed the conformational changes incurred throughout SpGH92 upon mannoimidazole binding, whereby the active site tightens around the bound inhibitor and consequently invokes a large stabilizing conformational change around the active site region that also results in exposure of the alpha helix comprising residues 443-453 (Figure 3-10).

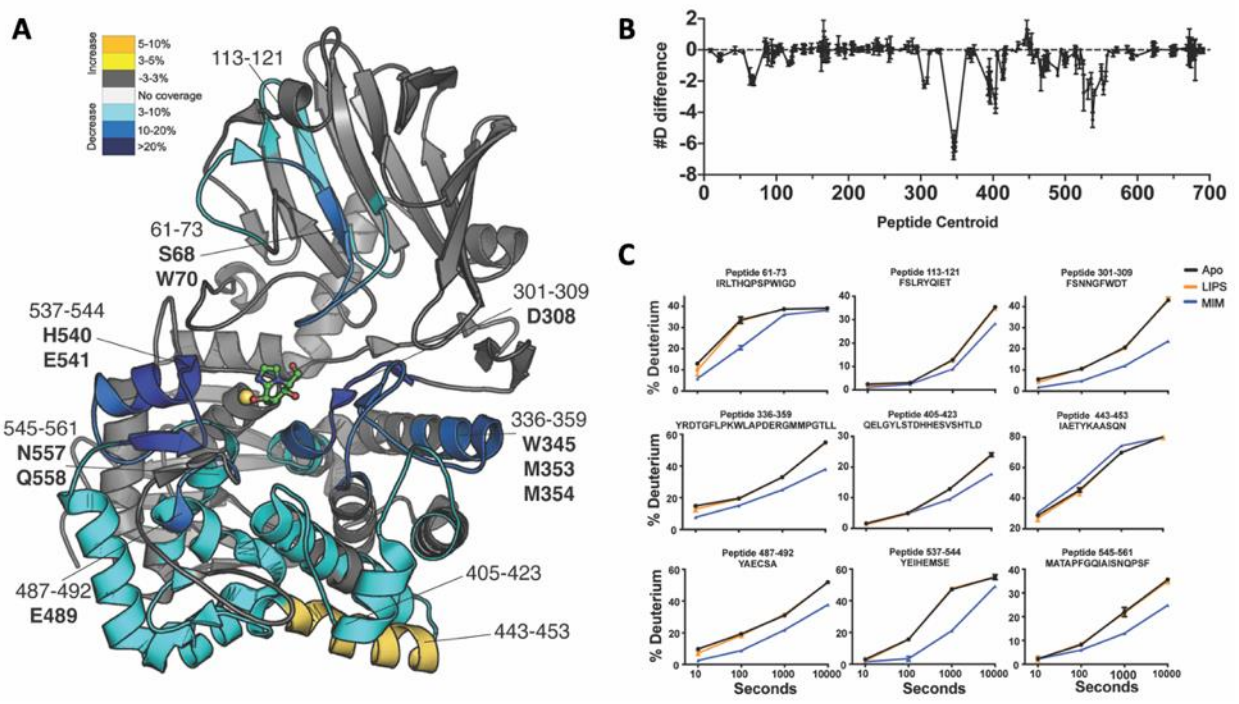


Figure 3-10. Mannoimidazole binding substantially decreases deuterium uptake around the SpGH92 active site. (A) Regions of SpGH92 that show significant changes in exchange (defined as >3 %D, >0.3 DA, and Student's *t*-test p-value<0.01) in the presence of mannoimidazole. Peptides in the C-terminus demonstrate the most protection from exchange. **(B)** The sum of the deuterium atoms protected from exchange in the presence of mannoimidazole across all time points is shown. Each point represents a single peptide where the centroid residue of that individual peptide is plotted on the x-axis. **(C)** The %D incorporation of selected peptides at different time points (10, 100, 1000 and 10,000 s). Error bars represent standard deviation (n = 3). The active site calcium ion is shown as a yellow sphere, with mannoimidazole in green.

3.3.4.2 Subtle differences occur in the presence of LIPS343 at pH 8.5

The presence of LIPS343 resulted in two small, yet significant regions of protection in the C-terminal domain using the new timecourse assay format. The significant changes were found in the peptides comprising residues 430-440 and 686-694. Furthermore, the changes were only seen at the 10 s time point, and both peptides contained roughly 5% less deuterium than the apo state. Notably, these values were very close to our significance cut-off values (

), and they did not appear to be connected through any central binding region. Thus, leading to doubts in the authenticity of the observed binding interactions.

The reasoning behind the small changes could be due to a limited solubility of the compound, as we were restricted to using 200 µM LIPS343 (in 1% DMSO). However, the concentration of LIPS343 that was used significantly exceeded the reported 2 µM K_i (Vocadlo et. al., unpublished). Therefore, we would not expect the observed changes to have been caused by a low occupancy of the inhibitor bound to SpGH92. As the results from this assay did not provide any clear evidence of a binding interface, like that seen upon mannoimidazole binding, we were unable to make any strong conclusions regarding the changes seen here.

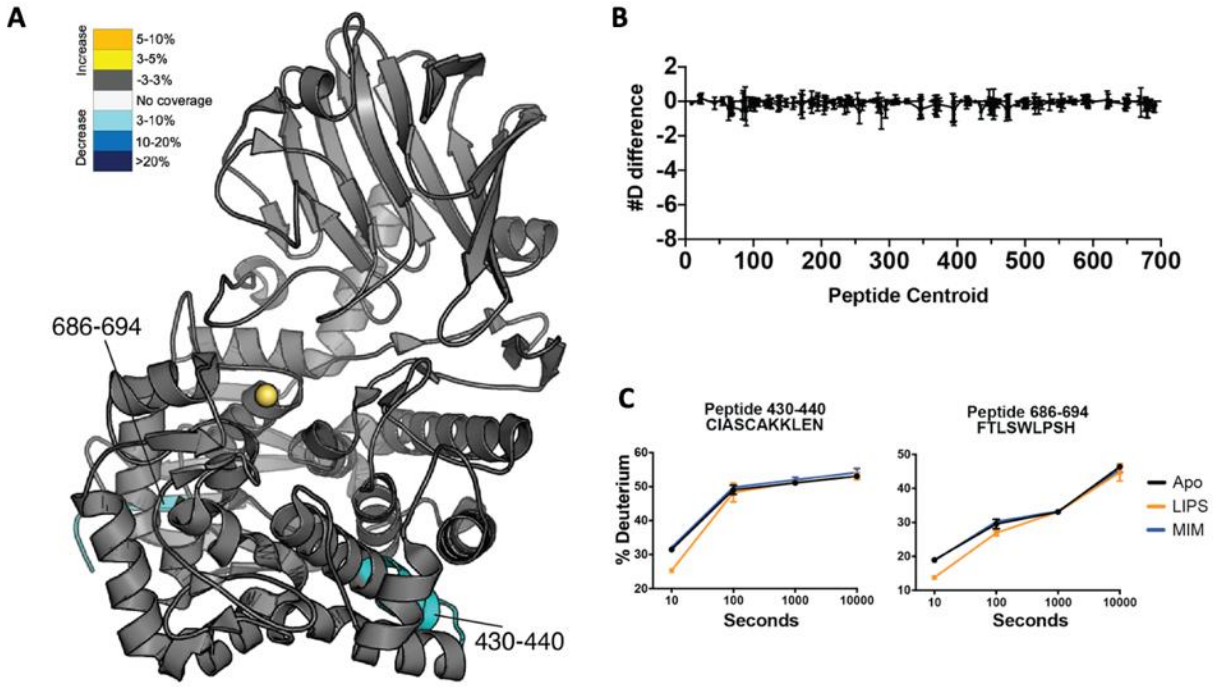


Figure 3-11. LIPS343 produces subtle changes in deuterium uptake at pH 8.5. (A) Regions of SpGH92 that show significant changes in exchange in the presence of LIPS343 (defined as >3 %D, >0.3 DA, and two-tailed Student's *t*-test p <0.01). Peptides in the C-terminus demonstrate the most protection from exchange. **(B)** The sum of the deuterium atoms protected from exchange in the presence of LIPS343 across all time points is shown. Each point represents a single peptide where the centroid residue of that individual peptide is plotted on the x-axis. **(C)** The %D

incorporation of selected peptides at different time points (10, 100, 1000 and 10,000 s). Error bars represent standard deviation (n = 3). The active site calcium ion is shown as a yellow sphere.

3.3.5 Mapping dose-dependent changes induced upon inhibitor binding

To verify that the changes seen in the LIPS343 pH 8.5 timecourse assay were not an artifact, we designed a dose-response experiment. Using this layout, we measured the conformational changes occurring in SpGH92 in response to increasing concentrations of inhibitor at a single time point. The time points which provided the most information from the previous timecourse assays were selected to be 10 s for LIPS343 and 480 s for mannoimidazole. For LIPS343, the optimal time point was clear, as all significant changes dissipated after the 10 s time point (Figure 3-11). However, for mannoimidazole, the optimal time point had to reflect both the average deuterium exchange rate (i.e., how fast the residues comprising a particular peptide in apo-SpGH92 became saturated with deuterium), as well as the protection factors (i.e., the difference in exchange rates between the apo and inhibitor bound states). As the peptides present in Figure 3-10C had the greatest protection factors seen out of all 254 peptides created during the pepsin digestion, it came down to selecting the time point that reflected, on average, the greatest difference in deuterium incorporation between the apo and mannoimidazole groups. Thus, with this criteria in mind, the 480 s time point was selected as it allowed for the most information to be gathered from the areas containing fast exchange rates like the region spanning residues 61-73 that reached >90% deuterium saturation at around 100 s; as well as areas with slower exchange rates like the region spanning residues 545-561 within the C-terminal domain that only began displaying significant differences between the apo and mannoimidazole groups around 200 s (Figure 3-10C).

An additional precaution was taken to increase the chances of detecting any changes induced by LIPS343. For this, we increased the inhibitor concentration from 200 µM to 400 µM which resulted in a 10% decrease in the final D_2O concentration (62% D_2O). The reactions were performed as mentioned previously, however now instead of a constant inhibitor concentration, the time point was fixed, and the inhibitor concentrations used were 0, 0.64, 3.2, 16, 80, and 400 µM.

3.3.5.1 Mannoimidazole induces a dose-dependent conformational change

Following data processing, a clear dose-response was seen in the SpGH92-mannoimidazole group with several key areas undergoing significant change (Figure 3-12). Protection became visible between 3.2 µM and 16 µM mannoimidazole and continued to grow exponentially up until 400 µM (Figure 3-12B). As expected, the greatest protection occurred in the shorter peptides containing active site residues. Within the C-terminal domain, the penta-peptide that spanned residues 537-541 and contained active site residues H540 and E541 underwent the largest protection seen with a 28.5% decrease in deuterium observed in the 400 µM mannoimidazole sample. On the other hand, in the N-terminal domain the peptide that covered residues 61-73 and contained the active site residues S68 and W70 underwent the largest protection with a 9.5% decrease in deuterium seen in the 400 µM mannoimidazole sample. Taken together these peptides along with several others demonstrated a clear dose-dependent conformational change occurring in SpGH92 upon mannoimidazole binding (Figure 3-12).

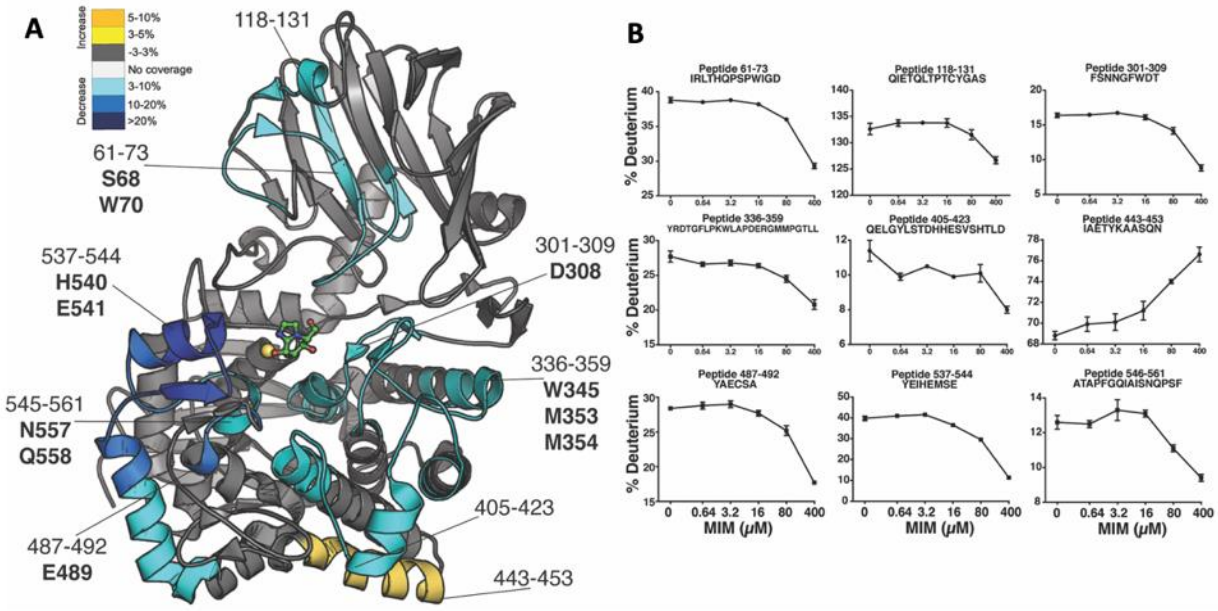


Figure 3-12. Mannoimidazole induces dose-dependent conformational changes in SpGH92. (A) Regions of SpGH92 showing significant changes in exchange at pH 8.5, after 480 s of D_2O incubation (defined as >3 %D, >0.3 DA, and two-tailed Student's *t*-test p <0.01) in the presence of 400 µM mannoimidazole. **(B)** The %D incorporation of representative peptides at different

mannoimidazole concentrations. Error bars represent standard deviation (n = 3). Experimental statistics are coverage = 99.3%, average peptide length = 15.1 $\pm$ 7.5 residues, redundancy = 7.9, replicates = 3, repeatability = 0.5–0.6 %D.

Another useful feature of HDX-MS is the ability to approximate the binding dissociation constant of a particular interaction. In this work the data from the mannoimidazole dose response assay was processed into % deuterium graphs that were used to approximate the dissociation constant (K_d) of inhibitor binding. When analyzed using the total-binding non-linear regression function in *Prism,* the data from several peptides fit nicely to a one-site binding curve. However, the 0.64 µM and 3.2 µM groups did not provide meaningful data points, thus the experimental K_d values could not be properly calculated. To obtain a fair approximation of the SpGH92-mannoimidazole K_d using HDX-MS in future experiments, higher concentrations of inhibitor (>15 µM) will be required. Although, techniques that directly measure the thermodynamic properties of protein-inhibitor interactions such as differential scanning calorimetry and isothermal titration calorimetry are better suited for determining binding constants.

3.3.5.2 LIPS343 does not produce a dose-dependent conformational change

The characterized inhibitor, mannoimidazole, demonstrated a clear downward trend in deuterium exchange with increasing concentrations of inhibitor. However, when we searched for trends in deuterium exchange corresponding to an increasing concentration of the novel inhibitor LIPS343 we could not detect any differences (Figure 3-13).

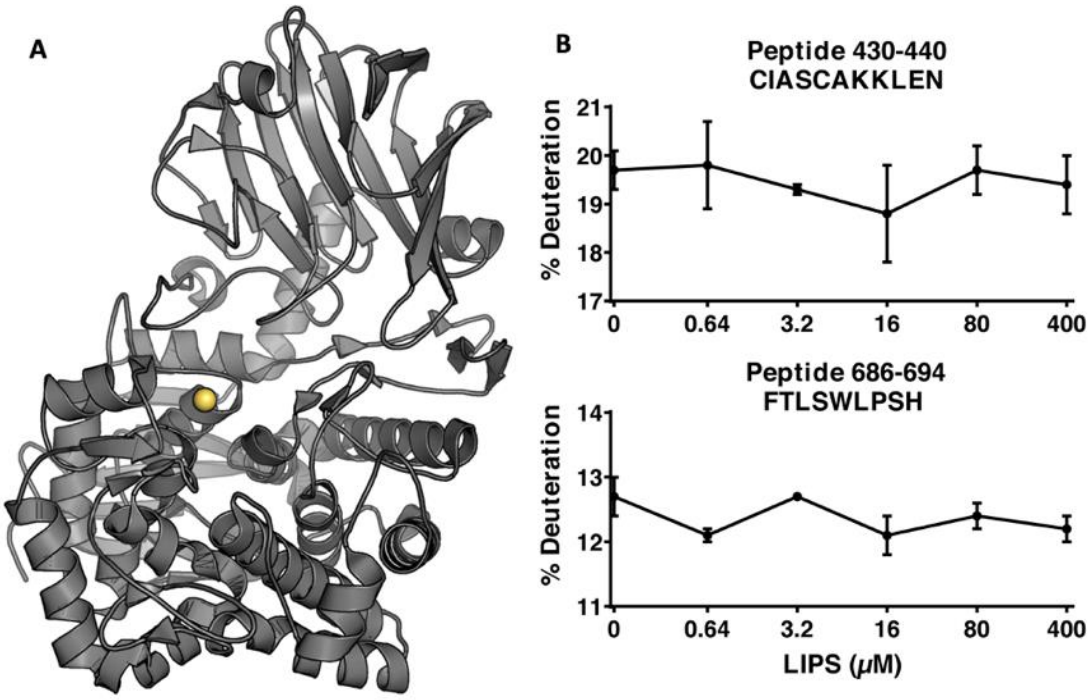


Figure 3-13. LIPS343 does not induce a dose-dependent change in SpGH92. (A) Regions of SpGH92 showing significant changes in exchange at pH 8.5, after 10s of D_2O incubation (defined as >3 %D, >0.3 DA, and Student's *t* test p <0.01) in the presence of 400 μM LIPS343. No significant changes were observed upon an increasing LIPS343 concentration **(B)** The %D incorporation of representative peptides at different LIPS343 concentrations. Error bars represent standard deviation (n=3). Experimental statistics are coverage= 96.5%, average peptide length= 15.3 $\pm$ 7.5 residues, redundancy= 5.9, replicates= 3, repeatability= 0.4 – 0.6 %D.

The contradicting results seen between the timecourse assay and dose response curve, raise concerns about the efficacy of LIPS343. Though, considering the changes seen in the pH 8.5 timecourse assay were nearing our significance cut-off values (Figure 3-11), the results from our previous assay may have simply been an artifact. Therefore, we attempted one last approach to detect any changes that could be occurring in the SpGH92 in response to LIPS343.

3.3.6 Differential scanning fluorimetry inhibitor binding assay

Utilizing DSF we analyzed the thermostability of SpGH92 in the presence of LIPS343 and mannoimidazole. Using methods similar to our initial DSF assay, we measured the thermal

stability of SpGH92 in response to inhibitor binding (Figure 3-7). Our analysis revealed the largest thermal shift occurred within the pH 7.5 mannoimidazole group relative to apo-SpGH92 (ΔT_m = 5.05 °C). While the second largest thermal shift occurred when increasing the pH of apo-SpGH92 from pH 7.5 to pH 8.5 (ΔT_m = 3.74 °C; Figure 3-7). All of the remaining DSF assays revealed markedly low changes in thermostability ($\Delta T_m \leq 1.60$ °C; Table 3-1). Notably, the LIPS343 groups displayed a small increase in protein stability. However, the changes seen in the pH 7.5 LIPS343 group were significantly lower than those observed with mannoimidazole. Interestingly, the pH 8.5 inhibitor groups and the pH 7.5 mannoimidazole group all appeared to reach a similar melting temperature around 58 °C suggesting an unfolding event is occurring near this temperature that may be unaffected by inhibitor binding.

Table 3-1. Differential scanning fluorimetry analysis of SpGH92 inhibitor binding

SpGH92 condition	Thermal shift relative to apo	T_m (°C)	$\pm$ SD (°C)
pH 7.5; apo		52.75	0.027
pH 7.5; 400 µM mannoimidazole	ΔT_m = 5.05 °C	57.53	0.019
pH 7.5; 200 µM LIPS343	ΔT_m = 1.60 °C	54.35	0.053
pH 8.5; apo	ΔT_m = 3.74 °C (relative to apo pH 7.5)	56.49	0.098
pH 8.5; 400 µM mannoimidazole	ΔT_m = 1.43 °C	57.92	0.029
pH 8.5; 200 µM LIPS343	ΔT_m = 1.25 °C	57.73	0.055

Values are representative of triplicates.

3.4 CONCLUSIONS

HDX-MS is a well-established technique for effectively identifying conformational changes that occur upon protein-small molecule interactions. In this study, we utilized HDX-MS to map conformational changes occurring in SpGH92 that are associated with the binding of a characterized and unknown inhibitor. The protocol developed in this study has demonstrated the efficacy of using HDX-MS to study inhibitor binding interfaces with SpGH92. One of the major

drawbacks of our initial experiments using pH 7.5 was the low global deuterium incorporation. However, at pH 8.5 we found the rate of deuterium exchange increased approximately 30-fold, thus allowing us to gather abundant information regarding the conformational dynamics of SpGH92 following inhibitor binding. Notably, even in our optimized buffer conditions some regions of stable secondary structure persisted within the core of the C-terminal $(\alpha/\alpha)_6$ barrel domain that extended into the C-terminal 5-stranded antiparallel β-sheet and into the two adjacent alpha-helices (Figure 3-14). Although, in many of the regions where detection was feasible mannoimidazole induced a clear dose-dependent conformational change that originated from the active site.

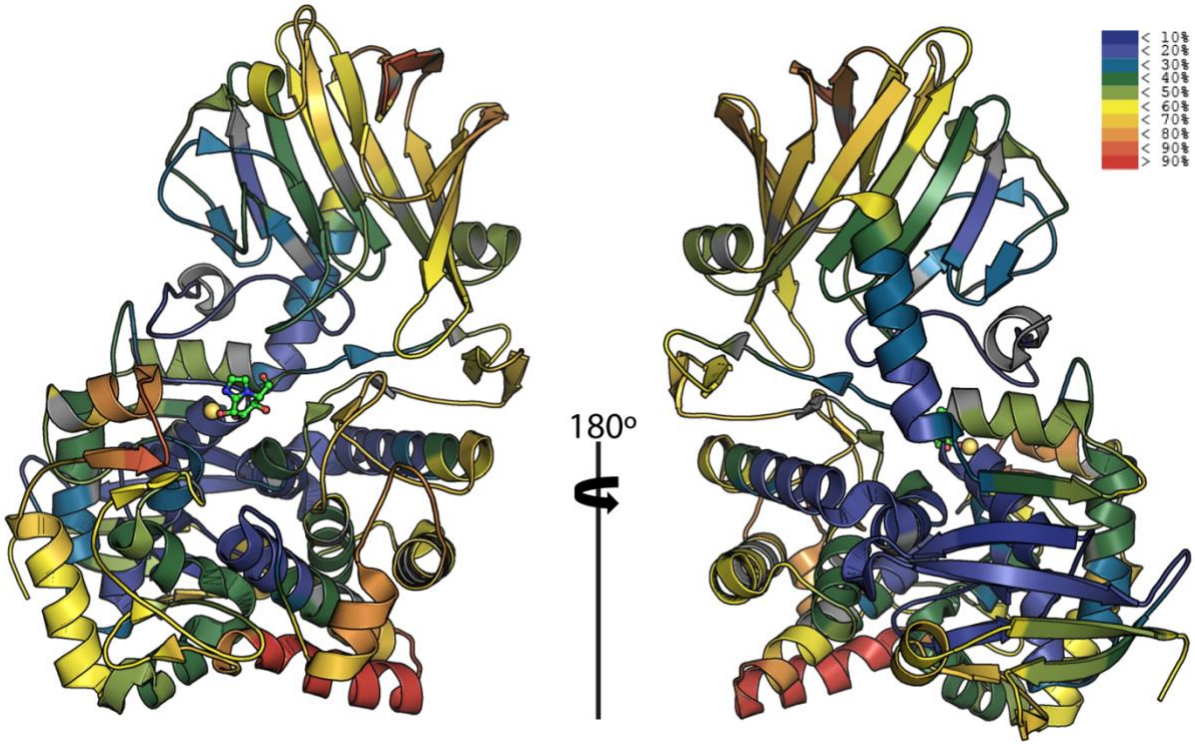


Figure 3-14. A significant region of low deuterium incorporation remains unexamined. The pH 8.5 apo-SpGH92 heat map at 10,000s reveals a considerable area of low deuterium incorporation in the core of the C-terminal $(\alpha/\alpha)_6$ barrel and the 5-stranded antiparallel β-sheet. Areas of low or no coverage are shown colored in grey.

In contrast to the results seen with mannoimidazole we were unable to obtain any consistent results regarding LIPS343 binding. After multiple assays and buffer conditions were tested, we could only detect two small regions near the cutoff of our significance threshold. Further attempts to recapitulate the results using a dose response assay were unsuccessful. Thus, the HDX-MS and DSF data collected here, along with the inability to form crystal complexes, suggests that LIPS343 does not interact with SpGH92 in a significant manner. Therefore, before continuing structural studies, additional inhibition assays using different assay formats should be performed to confirm the previously reported K_i value. A possible alternative assay is the Megazyme monosaccharide detection kit that was used in Chapter 2 to determine the kinetics of Aga (Chapter 2; Figure 2-6). Notably, the Megazyme kits detects the monosaccharides released from *exo*-glycoside hydrolase activity and thus does not require a fluorescent-sugar analog or a secondary β-mannosidase like the previous assay (Vocadlo et al., unpublished). We would however need to perform counter screens against the Megazyme kit's dehydrogenase and mutarotase. If no inhibition is found using alternative assay(s), further screening will be required to identify suitable SpGH92 inhibitors.

3.5 MATERIALS AND METHODS

3.5.1 SpGH92 expression and purification

Plasmid (pET28a) containing the full sequence of SpGH92 with an additional N-terminal six-histidine tag were transformed into the expression strain Escherichia coli BL21 Star (DE3). Recombinant SpGH92 was produced by the autoinduction method by incubating inoculated 2 L cultures at 37 °C for 96 hours. Pellets were lysed with 20 mM Tris pH 8.0, 100 mM NaCl, 20 mM Imidazole, 5% Glycerol, 10 mM bME, 0.1% v/v Triton X-100 buffer and underwent ultrasonication for 5 min with cycles consisting of 10 seconds on, 15 seconds off. Lysed cells were centrifuged at 25,000 x rcf in an ultracentrifuge for 45 min at 6 °C in a J45 rotor. Lysate was continuously circulated through a 5 mL NiNTA column equilibrated with 20 mM Tris pH 8.0, 100 mM NaCl, 20 mM imidazole, 5% glycerol. Bound SpGH92 was wash with 20 mL of high salt buffer

(20 mM Tris pH 8.0, 1 M NaCl, 5% glycerol) and plumbed into an AKTA Pure FPLC system to be eluted on an imidazole gradient. Fractions containing SpGH92 were located via SDS-PAGE and pooled before being concentrated. Concentrated protein was further purified by size exclusion chromatography using a Superose 6 Increase 10/300 column equilibrated with binding buffer (20 mM HEPES pH 7.5, 100 mM NaCl and 5 mM TCEP). Purified SpGH92 was aliquoted into 50 µL fractions, flash frozen, and stored at -80^{o}C.

3.5.2 SpGH92 activity assay

SpGH92 activity was tested against α1,2-mannobiose using a buffer containing 20 mM Tris pH 8.0 and 150 mM NaCl. First 10 µM SpGH92 and 1 mM α1,2-mannobiose were mixed together and left at 37 °C for 1 hour. The production of mannose monosaccharide was detected follow the manufacturer's protocol for the D-mannose/D-Fructose/D-glucose assay kit (Megazyme Inc., Chicago, IL), which contains a hexokinase, phosphomannose isomerase, phosphoglucose isomerase, and a $NADP^+$ dependent glucose-6-phosphate dehydrogenase. Mannose detection reactions contained 77 µL of SpGH92-mannobiose solution and 3 µL of kit enzymes, 10 µL kit buffer, 10 µL $NADP^+$. The kit reactions were incubated at 25 °C and the production of NADPH was monitored at 340 nm.

3.5.3 Differential scanning fluorimetry assay

SpGH92 was used at a final concentration of 2 mg/mL (2 µM) for this assay. SYPRO Orange was used at a final concentration of 2.5X. Experiments were carried out in either 20 mM Hepes pH 7.5, 150 mM NaCl and 1% DMSO; or 100 mM Tris pH 8.5, 150 mM NaCl, 1 mM $CaCl_2$ and 1% DMSO. Inhibitor concentrations of 400 µM mannoimidazole and 200 µM LIPS343 were examined where indicated. All DSF experiments were carried out with a StepOnePlus Real-Time PCR system (Applied Biosystems). Each sample was divided into three 15 µL replicates. Samples were dispensed into a 96-well optical reaction plate and the plate was sealed with an optical PCR plate sheet. Instrument settings were set to measure the change in fluorescence using SYPRO Orange. Temperature was continuously increased at 1.7 °C/min. Melting curves were exported directly from the instrument, and then were analyzed with Prism 6 (GraphPad Software Inc.).

3.5.4 HDX-MS analysis of SpGH92 at pH 7.5

Purified protein was diluted to 20 µM SpGH92, 20 mM HEPES pH 7.5, 100 mM NaCl ± 200 µM inhibitor and equilibrated at room temperature (16 °C) for 30 minutes. Hydrogen–deuterium exchange reactions were conducted in 10 µL reactions with a final concentration of 2 µM SpGH92 (20 pmol) per reaction. Reactions were initiated by the addition of 8 µL deuterium buffer (20 mM HEPES pH 7.5, 100 mM NaCl, 94% deuterium oxide) to 2 µL of protein solution, giving a final deuterium concentration of 75.5%. Exchange was carried out over 6 time points (0.3, 3, 30, 300 and 3000s at 16 °C). Exchange was terminated by the addition of acidic quench buffer giving a final concentration 0.6 M guanidine HCl and 0.8% formic acid. All experiments were carried out in triplicate. Samples were immediately frozen in liquid nitrogen and stored at -80°C until mass analysis.

3.5.5 HDX-MS analysis of SpGH92 at pH 8.5

Purified protein was diluted to 20 µM SpGH92, 100 mM TRIS pH 8.5, 150 mM NaCl ± 200 µM LIPS or 400 µM mannoimidazole and equilibrated at room temperature (16 °C) for 30 minutes. Hydrogen–deuterium exchange reactions were conducted in 10 µL reactions with a final concentration of 2 µM SpGH92 (20 pmol) per reaction. Reactions were initiated by the addition of 8 µL deuterium buffer (100 mM TRIS pH 8.5, 150 mM NaCl, 88% D_2O) to 2 µL of protein solution, giving a final deuterium concentration of 71%. Exchange was carried out over 4 time points (10, 100, 1000 and 10,000 seconds at 16 °C). Exchange was terminated by the addition of acidic quench buffer giving a final concentration 0.6 M guanidine HCl and 0.8% formic acid. All experiments were carried out in triplicate. Samples were immediately frozen in liquid nitrogen and stored at -80 °C until mass analysis.

3.5.6 SpGH92-mannoimidazole dose response assay

Purified protein was diluted to 10 µM SpGH92, 100 mM Tris pH 8.5, 150 mM NaCl ± or 400 µM mannoimidazole and equilibrated at room temperature (16 °C) for 30 minutes. Hydrogen–deuterium exchange reactions were conducted in 10 µL reactions with a final concentration of 0.76 µM SpGH92 (20 pmol) per reaction. Reactions were initiated by the addition of 24 µL

deuterium buffer (100 mM Tris pH 8.5, 150 mM NaCl, 88% D_2O) to 6 µL of protein solution, giving a final deuterium concentration of 74%. Exchange was carried out over 4 time points (10, 100, 1000 and 10,000 s at 16 °C). Exchange was terminated by the addition of acidic quench buffer giving a final concentration 0.6 M guanidine HCl and 0.8% formic acid. All experiments were carried out in triplicate. Samples were immediately frozen in liquid nitrogen and stored at -80 °C until mass analysis.

3.5.7 SpGH92 LIPS343 dose response assay

Purified protein was diluted to 10 µM SpGH92, 100 mM Tris pH 8.5, 150 mM NaCl $\pm$ or 400 µM LIPS343 and equilibrated at room temperature (16 °C) for 30 minutes. Hydrogen–deuterium exchange reactions were conducted in 10 µL reactions with a final concentration of 2 µM SpGH92 (20 pmol) per reaction. Reactions were initiated by the addition of 7 µL deuterium buffer (100 mM Tris pH 8.5, 150 mM NaCl, 88% D_2O) to 3 µL of protein solution, giving a final deuterium concentration of 62%. Exchange was carried out over 4 time points (10, 100, 1000 and 10,000 at 16 °C). Exchange was terminated by the addition of acidic quench buffer giving a final concentration 0.6 M guanidine HCl and 0.8% formic acid. All experiments were carried out in triplicate. Samples were immediately frozen in liquid nitrogen and stored at -80 °C until mass analysis.

3.5.8 HDX-MS data analysis

Protein samples were rapidly thawed and injected onto an ultra-performance liquid chromatography (UPLC) system kept in a cold box at 2 °C. Protein samples were run over two immobilized pepsin columns (Applied Biosystems; Porosyme 2-3131-00) and the peptides were collected onto a C18 VanGuard Pre-column 1 mm x 300 µm internal diameter (Waters). The C18 (Trap) column was eluted in line with an ACQUITY 1.7 µm particle, 100 x 1 mm^2 C18 UPLC column (Waters), using a gradient of 5 - 36% buffer B (Buffer A 0.1% formic acid, Buffer B 100% acetonitrile) over 16 minutes. Mass spectrometry experiments were performed on an Impact ESI-QTOF (Bruker), and peptide identification was done by running tandem mass spectrometry (MS/MS) experiments run in data-dependent acquisition mode. The resulting MS/MS datasets

were analyzed using PEAKS7 (PEAKS), Peptides were identified using a target decoy database search in the software program PEAKS 7. The database was composed of common affinity purification protein contaminants, other proteins purified in the lab, and pepsin. The search parameters were set with a precursor tolerance of 20 ppm, fragment mass error 0.02 Da, charge states from 1 to 8, a false discovery rate of 1%, leading to a selection criterion of peptides that had a -10logP score of 20.2. HDExaminer Software (Sierra Analytics) was used to automatically calculate the level of deuterium incorporation into each peptide. All peptides were manually inspected for correct charge state and presence of overlapping peptides. Deuteration levels were calculated using the centroid of the experimental isotope clusters. Exchange differences in a peptide were considered significant if they met all three of the following criteria: >3 %D change in exchange, >0.3 DA difference in exchange, and a p-value <0.01 using a two tailed student *t*-test. The raw HDX data are shown in two different formats. To allow for visualization of deuterium incorporation progression across all peptides, we utilized the number of deuteron atom difference (#D) plots. These plots show the total difference in deuterium incorporation over the entire H/D exchange timecourse, with each point indicating a single peptide. Samples were only compared within a single experiment and were never compared to experiments completed at a different time with a different final D_2O level.

CHAPTER 4: Conclusions and Future Directions

4.1 SUMMARY

The work completed in this thesis contributes to our current understanding of carbohydrate processing in *S. pneumoniae*. Carbohydrate processing is an essential component of pneumococcal pathogenicity and thus is an important area of research. Many of the enzymes involved in deconstructing both dietary saccharides and human glycans that influence virulence have been discussed. Regarding enzymes that deconstruct important host glycans, like those found on important immune glycoproteins, the connection between glycan degradation and pathogenicity is logical. However, the relationship between pathogenicity and dietary saccharides processing, such as plant RFOs, is far more ambiguous. Ultimately, this work illuminates multiple characteristics of the proteins involved in RFO utilization and the deadly virulence factor SpGH92 that initiates the deconstruction of HMNGs.

4.1.1 The peculiar case of RFO processing

The first research objective of this thesis was aimed at uncovering the biochemical and structural features of several proteins involved in RFO processing. In total, the *S. pneumoniae raf* locus is composed of nine genes, an activator and repressor (*rafR* and *rafS*), an α-galactosidase (*aga*), a sucrose phosphorylase (*gtfA*), a lipoteichoic acid ligase (*rafX*), and an ABC transport system. The ABC transporter consists of the substrate binding protein (*rafE*), two membrane permeases (*rafF*/*rafG*), and the shared ATPase (*rafK*). Numerous signature-tagged mutagenesis and knockout studies have linked *rafK*[54,60,147,188], *rafF*[54], *rafR*[60], *aga*[54,148], and *rafX*[159,160] to pneumococcal pathogenesis.

It has now been over two decades since the enigmatic relationship between RFO utilization and virulence was first described, yet the relationship is still poorly understood. Impeding our understanding is a lack of biochemical and structural data regarding the gene products. Thus, to broaden our understanding of RFO processing in *S. pneumoniae* we investigated the biochemical

properties of RafE, Aga and GtfA towards multiple carbohydrates. We identified several important features regarding the substrate specificity of RafE, Aga and GtfA. Also, we determined the kinetic parameters of Aga towards several α-galactoside substrates and the binding affinities of RafE towards several oligosaccharides. Using X-ray crystallography, we produced three structure models of RafE bound to raffinose, stachyose and verbascose; all of which provided information regarding the substrate binding interface of RafE.

The binding affinity of RafE towards all of the examined RFOs averaged a K_d between 199 μM to 479 μM (Table 2-2). Notably, the values obtained from our ITC analyses were much higher than the average substrate binding proteins' K_d of 1 μM. Interestingly, we found that RafE possessed the lowest K_d (80 μM) towards panose, which is a trisaccharide derivative of glycogen containing an α1,4-linked reducing end glucose. We also determined that Aga and GtfA are active towards their putative substrates. However, Aga was relatively inefficient at cleaving α1,6-galactoside linkages like those in RFOs. Instead, it was more efficient towards α1,3-galactosides, like those in the linear type II B blood group antigen. With regards to GtfA, we were able to detect activity towards sucrose. However, only trace amounts of cleavage products could be seen using our TLC analysis which suggests that the enzyme is not well adapted for sucrose degradation. Collectively, our results did not support the notion that RFO utilization is contributing to invasive disease as all the proteins examined showed significant inefficiencies. This compounds with the expectation that *in vivo* concentrations of RFOs are likely quite low. Thus, our results hint of an unorthodox role for the *raf* locus in *S. pneumoniae* pathogenesis which will need to be investigated in future studies.

4.1.2 SpGH92: The structurally sound virulence factor

The second research objective was aimed at establishing a method using HDX-MS to characterize the binding interface of inhibitors to the deadly virulence factor SpGH92. A model of sepsis and pneumonia has highlighted the role of SpGH92 in initiating the degradation of mammalian HMNGs[47]. Due to its enormous impact on pneumococcal pathogenesis, we wanted to establish a methodology for identifying the binding interface of SpGH92 with inhibitors. The known α-mannosidase inhibitor, mannoimidazole, was used as a comparative control for

establishing an HDX-MS protocol. Mannoimidazole binding demonstrated a clear time and dose-dependent protection from deuterium exchange in several key areas of the enzyme. Conversely, when probing for the binding interface of the novel inhibitor LIPS343 that was found to have a K_i of 2 μM by our collaborators (Vocadlo et al., unpublished), we did not detect any binding in a concentration ranging from 0.27 to 167-fold the reported K_i value. Thus, our results suggest that LIPS343 is not an SpGH92 inhibitor.

Several possibilities exist for why we obtained these results. The novel inhibitor does not have a known or suspected binding interface, therefore if it were to interact with the enzyme through a region undetectable by HDX-MS such as a region of low deuterium-exchange, we would not be able to see any interactions. It could also be interacting in a fashion that does not produce a significant change in secondary structure. Another possibility is that the inhibitor may be dependent upon a SpGH92 substrate-bound intermediate state being present, which we did not test. Though the most likely possibility is that the initial inhibitor screening produced a false-positive result due to an unexpected interaction between the inhibitor and one or more components of the assay. Overall, this portion of the thesis uncovered the stable nature of SpGH92 in solution and also established a highly-reproducible and unbiased method for detecting inhibitor binding interfaces with SpGH92 for future drug screening and characterization.

4.2 FUTURE DIRECTIONS

This thesis work has unveiled several previously unknown characteristics of the *raf* locus proteins. In addition, it has provided information regarding the stability of SpGH92 in solution, the feasibility of using HDX-MS to identify binding interfaces between inhibitors, as well as contradictory evidence regarding the binding of interaction of a novel inhibitor LIPS343. With what we now know, several follow-up experiments are recommended.

4.2.1 A mysterious link between the *raf* locus and lipoteichoic acid metabolism

We have elucidated the biochemical and structural properties of several genes involved in RFO

processing and our results do not appear to support the notion that RFOs utilization *in vivo* is important for pathogenesis. Due to the low affinities and catalytic efficiencies of the proteins in this study our findings raise many questions. For example, if not for RFO utilization why is the *raf* locus within the pneumococcal core genome present in over 98% of pneumococcal strains? Similarly, if RFO utilization is important for virulence, why do all the proteins appear inefficient towards their putative substrates?

One possible answer may hiding in a study by Heß et al[160]. The authors found that the gene *rafX* is critical in lipoteichoic acid formation and that it is likely a lipoteichoic acid ligase. Their finding are supported by the fact that RafX contains a WzyC domain which is commonly associated with O-antigen synthesis in Gram negative bacteria[161,215]. Lipoteichoic acids are key structures used to adhere to host cells, dock important choline binding proteins to the pneumococcal cell surface, and for moderating inflammation[159]. Thus, this information provides possible explanation to the *raf* locus's impact on virulence as well as its presence within the core genome. However, despite the importance of lipoteichoic acid synthesis in many aspects of virulence, no studies to date have mentioned *rafX* in the context of the *raf* locus. In addition, many of the enzymes involved in lipoteichoic acid synthesis have not been elucidated; thus, it's possible that other components of the *raf* locus are also involved in lipoteichoic acid metabolism.

Pneumococcus is notorious for repurposing its proteins which often serve functions similar to their putative functions; however, are found to have alternative roles *in vivo*[53,216,217]. It would be interesting for future studies to look for alternative roles of the *raf* genes towards the synthesis, modification, and degradation of lipoteichoic acid constituents. Recalling from Chapter 1.7.2, lipoteichoic acids are known to contain four to eight repeating units of -(PCho-O6-)GalNAc-α1,3-(PCho-6O-)GalNAc-β1,1-ribitol-5-phosphate-6-Glc-β1,3-AATGalp-α1,4-[64].

There are a variety of biophysical tools that we could utilize to probe for high affinity interactions between lipoteichoic acid derivates and RafE including ITC, NMR, and HDX-MS . Finding an interaction between RafE and a component of lipoteichoic acid would be a significant discovery and we would absolutely want to pursue structural studies to further elucidate the novel binding interactions. Experiments aimed at uncovering the role of the interaction in bacterial fitness would also need to be completed. For example, it would be important to see if

the RafEFG proteins comprising the ABC transporter are involved in lipoteichoic acid metabolism or "recycling" that would in theory result from pneumococcal autolysis. If mutations in RafE or other components of the *raf* locus interfered with lipoteichoic acid synthesis similar to that seen with *rafX* knockouts, the corresponding proteins may make promising therapeutic targets.

Furthermore, the α-galactosidase Aga demonstrated a preference towards α1,3-galactoside linkages over α1,6-galactosides which is interesting considering the terminal α1,3-linkages of pneumococcal teichoic acids. However, lipoteichoic acids contain GalNAc-α1,3-, not Gal-α1,3-, thus Aga would appear to be inactive towards lipoteichoic acids according to our pNP-assays. Nevertheless, it's possible that Aga could be involved in some component of lipoteichoic acid regulation. Additionally, GH36 subgroup I enzymes have been shown to possess a variety of functions including hydrolase, transferase and kinase activities[132,153,180,218]. Therefore, we could probe for an alternative Aga activity, with an initial focus on the synthesis and depolymerization of lipoteichoic acids.

Lastly, the sucrose phosphorylase GtfA only demonstrated trace activity towards sucrose. This could mean that GtfA has a preference towards the formation of sucrose. However, it could mean GtfA has an alternative function which we did not test for. Indeed, several studies investigating GH13 enzymes have revealed a large diversity in the functions of these enzymes[63,219]. Structural comparisons of GtfA using the HMMER search engine revealed homology to several protein families, including the sucrose phosphorylases, sucrose-6'-phosphoylases, glucosyl-glycerate phosphorylases and α1,4 glucan/maltose-1-phosphate transferases[220–224]. Thus, it could be beneficial to search for possible alternative functions. To test this, we could probe for the synthesis and catalysis of the common substrates of the aforementioned protein families. In addition, we could probe for the synthesis and degradation of structures related to lipoteichoic acids such as Glc-α-diacylglycerol and ribitol-5P-6-Glc in a similar manner as other groups[220–224].

4.2.2 A potential binding site in the RafE binding pocket

Upon examining the binding pocket of RafE, it is hard to miss a conspicuous cavity contiguous with the subsite 1 binding site. To see if an additional sugar could fit within this cavity, a sialic acid residue was manually built-in (Figure 4-1). Intriguingly, the cavity appeared to form well with

the functional groups of a sialic acid. The amino acids forming the cavity containing the glycerol functional group are mostly charged and polar groups which would make good hydrogen bonding partners with the sialic acid functional group (shown as surface electrostatics; Figure 4-1). The amino acids forming the pocket housing the N-acetyl functional group form a hydrophobic base with plenty of potential hydrogen bonding partners present for the C=O and NH functional groups. As we did not examine the binding of RafE to sialic acid containing glycans in this study, future binding assays could incorporate a wider range of substrate targets including α2,3- and α2,6-sialic acid containing glycans like those found on several O- and N-linked glycans[48,225].

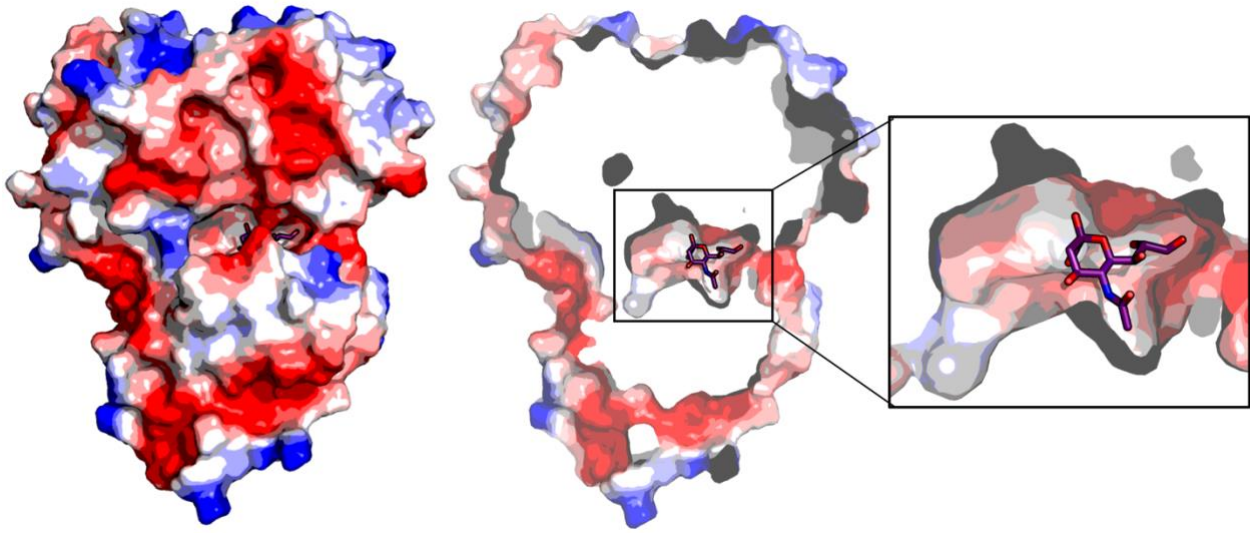


Figure 4-1. Electrostatic surface map of RafE containing an α-sialic acid. The binding pocket of RafE appears to have sufficient room to house a bulky sialic-acid sugar residue next to the subsite 1 binding site.

4.2.3 A surface exposed RGD binding motif

Sequence analysis using the MyHits search engine (myhits.sib.swiss) revealed several putative motifs within the RafE sequence which should be experimentally tested. The motifs identified include a C-terminus EF-Hand-like calcium binding motif in the C-terminus (residues 163-167); and an RGD motif (residues 263-265; Figure 4-2). RGD motifs are found on many human extracellular matrix proteins where they function as natural targets of integrin binding and

contribute immensely to the strength and structure of the cell surface[226–228]. Several pathogens including group B *Streptococcus* express proteins containing RGD-motifs which facilitate their virulence[226–228]. Intriguingly, many integrins contain multiple calcium ions and glycans near their RGD binding pockets that are exploited by pathogens to create a strong binding interface[229]. Thus, it is possible that pneumococcus is using its SBP RafE to adhere to host integrins, thereby facilitating adhesion and during infection.

To investigate this possible role, we could perform ITC experiments with RafE and a collection of integrins found within the host ECM. If any high affinity binding interactions are detected, we could look into the conformational changes within RafE and the integrin(s) using HDX-MS and X-ray crystallography. In addition, it would be interesting if we observed conformational changes in the calcium binding motif (α-helix 6) on RafE as comes into close proximity with the integrins proteins bound calcium ions. HDX-MS is the perfect tool for probing the conformational changes that may occur between RafE and integrins, as the structures of RafE and many integrins are already known and the areas of change could be accurately mapped. As a follow up, any significant conformational changes seen could be validated by mutating amino acids within the binding interfaces to pinpoint which amino acids play a key role in binding.

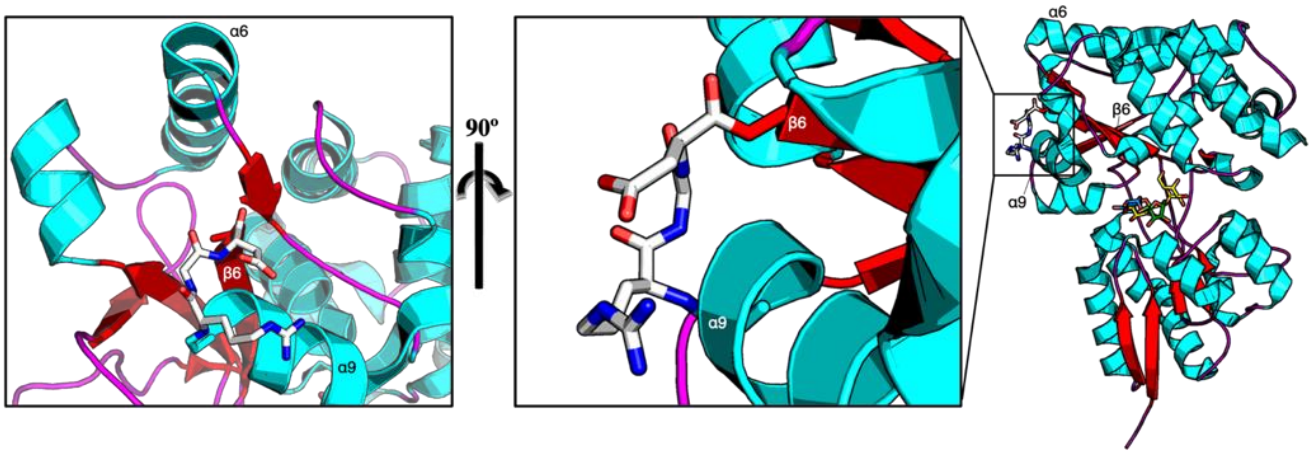


Figure 4-2. The integrin-binding RGD motif of RafE. The surface exposed RGD binding motif in RafE (middle) is located in the C-terminal domain within a loop region linking α-helix 9 (α9) and β-strand 7 (β7). The EF-Hand-like calcium binding motif is predicted to be along α6 located above the RGD sequence. The location of the active site is shown bound to stachyose.

4.2.4 Mapping the binding interface of potential SpGH92 inhibitors

Overall, this area of the thesis has led to a deeper understanding of the structural dynamics of SpGH92 in solution. We have shown here that SpGH92 contains regions of very stable secondary structure. Using the protocols developed in Chapter 3 we were able to clearly see conformational changes upon MIM binding. Unfortunately, we were unable to detect any binding interfaces between SpGH92 and LIP343. As we have attempted both X-ray crystallography and HDX-MS without success and others have recently attempted ITC without detecting binding, we should consider other options.

Going forward with LIPS343 it is crucial to first re-examine the results of the fluorophore-labeled substrate assay using another approach. The initial fluorophore coupled assay used to determine the K_i utilized a mannobiose (disaccharide analog) containing a fluorophore attached to the reducing end mannose. In their assay SpGH92 cleaved the terminal mannose off of the mannobioside-fluorophore analog creating a mannoside-fluorophore (monosaccharide analog) which would subsequently be cleaved by a secondary enzyme releasing the fluorophore. Attempting to analyze the activity of SpGH92 without a secondary glycoside hydrolase enzyme may be helpful. To do this, the Megazyme detection kit that was used for determining Aga kinetics in Chapter 2 could prove useful. The Megazyme kits couple monosaccharide production with NAD-reduction which perhaps would prove beneficial as the secondary enzyme would be a dehydrogenase. Additionally, examining the cleavage products using mass-spectrometry we could quantify monosaccharide production in the presence of inhibitors, and this would completely avoid the use of a secondary enzyme, thus avoiding any complications arising from the use of multiple enzymes.

Regardless of the method we should perform future binding assays in the presence and absence of a substrate or a non-hydrolysable substrate-mimic such as thiomannobioside, as we have not inspected if LIPS343-binding is dependent on a conformational intermediate induced upon substrate binding. In addition, future studies could examine mannosidase inhibitors from drug-repositioning screens, like those used in cancer research, to inhibit ER mannosidase (Man1B1) and Golgi mannosidase II (Man2A1) in humans[230]. Using the protocol developed in this thesis, we could test a multitude of inhibitors with SpGH92 to identify their binding interfaces as

well as their approximate binding affinities. Identification of any moderate-high affinity ligands may then be used as templates for rounds of iterative optimization, as well as *in vitro* assays and *in vivo* assays.

REFERENCES

1. Watson, D. A., Musher, D. M., Jacobson, J. W. & Verhoef, J. A Brief History of the Pneumococcus in Biomedical Research: A Panoply of Scientific Discovery. *Clin. Infect. Dis.* **17**, 913–924 (1993).
2. Paster L. Sur une maladie nouvelle, provoquee par la salive d'enfant mort de la rage. *Compt. Rend. Acad. d sci.* (1881).
3. Neufeld, F. Ueber die Agglutination der Pneumokokken und über die Theorieen der Agglutination. *Zeitschrift für Hyg. und Infekt.* **40**, 54–72 (1902).
4. Griffith, F. The Significance of Pneumococcal Types. *J. Hyg. (Lond).* **27**, 113–159 (1928).
5. Avery, O. T., MacLeod, C. M. & McCarty, M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *J. Exp. Med.* **79**, 137–158 (1944).
6. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).
7. Weiser, J. N., Ferreira, D. M. & Paton, J. C. Streptococcus pneumoniae: Transmission, colonization and invasion. *Nat. Rev. Microbiol.* **16**, 355–367 (2018).
8. De Ste Croix, M. *et al.* Phase variation in pneumococcal populations during carriage in the human nasopharynx. *Sci. Rep.* **10**, 1803 (2020).
9. Public Health Agency of Canada. Invasive Pneumococcal Disease - Canada.ca. *2016* https://www.canada.ca/en/public-health/services/immunization/vaccine-preventable-diseases/invasive-pneumococcal-disease/health-professionals.html.
10. Zafar, M. A., Wang, Y., Hamaguchi, S. & Weiser, J. N. Host-to-Host Transmission of Streptococcus pneumoniae Is Driven by Its Inflammatory Toxin, Pneumolysin. *Cell Host Microbe* **21**, 73–83 (2017).
11. Hu, T., Weiss, T., Owusu-Edusei, K. & Petigara, T. Health and economic burden associated with 15-valent pneumococcal conjugate vaccine serotypes in children in the United States. *J. Med. Econ.* 1–8 (2020) doi:10.1080/13696998.2020.1840216.
12. Thornton, R. B. *et al.* Australian Aboriginal Children with Otitis Media Have Reduced Antibody Titers to Specific Nontypeable Haemophilus influenzae Vaccine Antigens. *Clin.*

Vaccine Immunol. **24**, 1–10 (2017).

13. Kaplan, A. *et al.* Vaccine strategies for prevention of community-acquired pneumonia in Canada. *Candadian Fam. Physician* **65**, 625–633 (2019).
14. Randle, E., Ninis, N. & Inwald, D. Invasive pneumococcal disease. *Arch. Dis. Child. Educ. Pract. Ed.* (2011) doi:10.1136/adc.2010.191718.
15. OMS. World Health Report Life in the 21st century A vision for all Report of the Director-General. *51st World Heal. Assem.* (1998).
16. Matanock, A. *et al.* Use of 13-Valent Pneumococcal Conjugate Vaccine and 23-Valent Pneumococcal Polysaccharide Vaccine for Adults With Immunocompromising Conditions: Recommendations of the Advisory Committee on Immunization Practices (ACIP). *Am. J. Transplant.* **13**, 232–235 (2013).
17. Scott, J. A. G. The preventable burden of pneumococcal disease in the developing world. *Vaccine* **25**, 2398–2405 (2007).
18. Mukerji, R. & Briles, D. E. New Strategy Is Needed to Prevent Pneumococcal Meningitis. *Pediatr. Infect. Dis. J.* **39**, 298–304 (2020).
19. Pneumococcal Disease | Clinical | Features | CDC. https://www.cdc.gov/pneumococcal/clinicians/clinical-features.html.
20. McNeil, S. A. *et al.* A Retrospective Study of the Clinical Burden of Hospitalized All-Cause and Pneumococcal Pneumonia in Canada. *Can. Respir. J.* **2016**, 1–10 (2016).
21. Wahl, B. *et al.* Burden of Streptococcus pneumoniae and Haemophilus influenzae type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000–15. *Lancet Glob. Heal.* **6**, e744–e757 (2018).
22. Dobay, O. The complexity of serotype replacement of pneumococci. *Human Vaccines and Immunotherapeutics* vol. 15 2725–2728 (2019).
23. Troeger, C. *et al.* Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Infect. Dis.* **18**, 1191–1210 (2018).
24. Pneumococcal Disease | Global Pneumococcal Disease and Vaccine | CDC.

https://www.cdc.gov/pneumococcal/global.html.

25. Ayoola, M. B. *et al.* Polyamine Synthesis Effects Capsule Expression by Reduction of Precursors in Streptococcus pneumoniae. *Front. Microbiol.* **10**, 1–14 (2019).
26. Geno, K. A. *et al.* Pneumococcal Capsules and Their Types: Past, Present, and Future. *Clin. Microbiol. Rev.* **28**, 871–899 (2015).
27. Smillie, W. G., Warnock, G. H. & White, H. J. A Study of a Type I Pneumococcus Epidemic at the State Hospital at Worcester, Mass. *Am. J. Public Heal. Nations Heal.* (1938) doi:10.2105/ajph.28.3.293.
28. Austrian, R., Douglas, R. M. & Schiffman, G. Prevention of pneumococcal pneumonia by vaccination. *Trans. Assoc. Am. Physicians* (1976) doi:10.9783/9781512800135-009.
29. KAUFMAN, P. PNEUMONIA IN OLD AGE. *Arch. Intern. Med.* **79**, 518 (1947).
30. Brooks, L. R. K. & Mias, G. I. Streptococcus pneumoniae's virulence and host immunity: Aging, diagnostics, and prevention. *Front. Immunol.* **9**, (2018).
31. Bettinger, J. A. *et al.* The effect of routine vaccination on invasive pneumococcal infections in Canadian children, Immunization Monitoring Program, Active 2000–2007. *Vaccine* **28**, 2130–2136 (2010).
32. McIntosh, E. D. G. & Reinert, R. R. Global prevailing and emerging pediatric pneumococcal serotypes. *Expert Rev. Vaccines* **10**, 109–129 (2011).
33. Sempere, J., de Miguel, S., González-Camacho, F., Yuste, J. & Domenech, M. Clinical Relevance and Molecular Pathogenesis of the Emerging Serotypes 22F and 33F of Streptococcus pneumoniae in Spain. *Front. Microbiol.* **11**, 1–16 (2020).
34. Masomian, M., Ahmad, Z., Gew, L. T. & Poh, C. L. Development of next generation streptococcus pneumoniae vaccines conferring broad protection. *Vaccines* **8**, 1–23 (2020).
35. WHO publishes list of bacteria for which new antibiotics are urgently needed. https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed.
36. Canadian Antimicrobial Resistance Surveillance System Report 2016 - Canada.ca. https://www.canada.ca/en/public-health/services/publications/drugs-health-

products/canadian-antimicrobial-resistance-surveillance-system-report-2016.html.

37. Wang, C. *et al.* Antibiotic resistance profiles and multidrug resistance patterns of Streptococcus pneumoniae in pediatrics. *Medicine (Baltimore).* **98**, e15942 (2019).
38. Lo, S. W. *et al.* Pneumococcal lineages associated with serotype replacement and antibiotic resistance in childhood invasive pneumococcal disease in the post-PCV13 era: an international whole-genome sequencing study. *Lancet Infect. Dis.* **19**, 759–769 (2019).
39. Varki, Ajit; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. & E. *Essentials of Glycobiology, 3rd edition*. *Cold Spring Harbor (NY)* (2015).
40. Zhu, Y. *et al.* Mechanistic insights into a Ca2+-dependent family of α-mannosidases in a human gut symbiont. *Nat. Chem. Biol.* **6**, 125–132 (2010).
41. Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W. & Brumer, H. Polysaccharide utilization loci: Fueling microbial communities. *J. Bacteriol.* **199**, 1–15 (2017).
42. Varki, A. Biological roles of glycans. *Glycobiology* **27**, 3–49 (2017).
43. Santos, A. L. & Lindner, A. B. Protein Posttranslational Modifications: Roles in Aging and Age-Related Disease. *Oxid. Med. Cell. Longev.* **2017**, (2017).
44. Marion, C. *et al.* Identification of a Pneumococcal Glycosidase That Modifies O-Linked Glycans. *Infect. Immun.* **77**, 1389–1396 (2009).
45. Pluvinage, B. *et al.* Inhibition of the Pneumococcal Virulence Factor StrH and Molecular Insights into N-Glycan Recognition and Hydrolysis. *Structure* **19**, 1603–1614 (2011).
46. King, S. J. Pneumococcal modification of host sugars: A major contributor to colonization of the human airway? *Mol. Oral Microbiol.* **25**, 15–24 (2010).
47. Robb, M. *et al.* Molecular Characterization of N-glycan Degradation and Transport in Streptococcus pneumoniae and Its Contribution to Virulence. *PLOS Pathog.* **13**, e1006090 (2017).
48. Jia, N. *et al.* The Human Lung Glycome Reveals Novel Glycan Ligands for Influenza A Virus. *Sci. Rep.* **10**, 5320 (2020).
49. Darula, Z. & Medzihradszky, K. F. Analysis of mammalian O-glycopeptides - We have made a good start, but there is a long way to go. *Mol. Cell. Proteomics* **17**, 2–17 (2018).
50. Van Den Steen, P., Rudd, P. M., Dwek, R. A. & Opdenakker, G. Concepts and principles of

O-linked glycosylation. *Crit. Rev. Biochem. Mol. Biol.* **33**, 151–208 (1998).

51. Yesilkaya, H., Manco, S., Kadioglu, A., Terra, V. S. & Andrew, P. W. The ability to utilize mucin affects the regulation of virulence gene expression in Streptococcus pneumoniae. *FEMS Microbiol. Lett.* **278**, 231–235 (2008).
52. Janesch, P. *et al.* Assessing the function of pneumococcal neuraminidases NanA, NanB and NanC in in vitro and in vivo lung infection models using monoclonal antibodies. *Virulence* **9**, 1521–1538 (2018).
53. Andre, G. O. *et al.* Role of Streptococcus pneumoniae Proteins in Evasion of Complement-Mediated Immunity. *Front. Microbiol.* **8**, 1–20 (2017).
54. Hava, D. L. & Camilli, A. Large-scale identification of serotype 4 Streptococcus pneumoniae virulence factors. *Mol. Microbiol.* **45**, 1389–1406 (2002).
55. Chen, H. *et al.* Genetic Requirement for Pneumococcal Ear Infection. *PLoS One* **3**, e2950 (2008).
56. Iyer, R. & Camilli, A. Sucrose metabolism contributes to in vivo fitness of Streptococcus pneumoniae. *Mol. Microbiol.* **66**, 1–13 (2007).
57. Kuo, T. M., VanMiddlesworth, J. F. & Wolf, W. J. Content of raffinose oligosaccharides and sucrose in various plant seeds. *J. Agric. Food Chem.* **36**, 32–36 (1988).
58. Vinson, C. C. *et al.* Characterization of raffinose metabolism genes uncovers a wild Arachis galactinol synthase conferring tolerance to abiotic stresses. *Sci. Rep.* **10**, 15258 (2020).
59. Smith, A. W., Roche, H., Trombe, M. C., Briles, D. E. & Håkansson, A. Characterization of the dihydrolipoamide dehydrogenase from Streptococcus pneumoniae and its role in pneumococcal infection. *Mol. Microbiol.* **44**, 431–448 (2002).
60. Minhas, V. *et al.* Capacity To Utilize Raffinose Dictates Pneumococcal Disease Phenotype. *MBio* **10**, 8–10 (2019).
61. Buckwalter, C. M. & King, S. J. Pneumococcal carbohydrate transport: Food for thought. *Trends Microbiol.* **20**, 517–522 (2012).
62. Hobbs, J. K., Pluvinage, B. & Boraston, A. B. Glycan-metabolizing enzymes in microbe–host interactions: the Streptococcus pneumoniae paradigm. *FEBS Lett.* **592**, 3865–3897

(2018).

63. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
64. Denapaite, D., Brückner, R., Hakenbeck, R. & Vollmer, W. Biosynthesis of Teichoic Acids in Streptococcus pneumoniae and Closely Related Species: Lessons from Genomes. *Microb. Drug Resist.* **18**, 344–358 (2012).
65. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyl transferases: Structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77**, 521–555 (2008).
66. Cuthbertson, L., Kos, V. & Whitfield, C. ABC Transporters Involved in Export of Cell Surface Glycoconjugates. *Microbiol. Mol. Biol. Rev.* **74**, 341–362 (2010).
67. CAZy - GT. http://www.cazy.org/GlycosylTransferases.html.
68. Lombard, V. *et al.* A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochem. J.* **432**, 437–444 (2010).
69. Jongkees, S. A. K. & Withers, S. G. Unusual enzymatic glycoside cleavage mechanisms. *Acc. Chem. Res.* **47**, 226–235 (2014).
70. Li, S., Kelly, S. J., Lamani, E., Ferraroni, M. & Jedrzejas, M. J. Structural basis of hyaluronan degradation by Streptococcus pneumoniae hyaluronate lyase. *EMBO J.* **19**, 1228–1240 (2000).
71. Yadav, G. *et al.* Evidence for Inhibitory Interaction of Hyaluronan-binding Protein 1 (HABP1/p32/gC1qR) with Streptococcus pneumoniae Hyaluronidase. *J. Biol. Chem.* **284**, 3897–3905 (2009).
72. Uechi, K., Kamachi, S., Akita, H., Mine, S. & Watanabe, M. Crystal structure of an acetyl esterase complexed with acetate ion provides insights into the catalytic mechanism. *Biochem. Biophys. Res. Commun.* **477**, 383–387 (2016).
73. Vollmer, W. & Tomasz, A. Identification of the teichoic acid phosphorylcholine esterase in Streptococcus pneumoniae. *Mol. Microbiol.* **39**, 1610–1622 (2001).
74. Vollmer, W. & Tomasz, A. Peptidoglycan N-acetylglucosamine deacetylase, a putative virulence factor in Streptococcus pneumoniae. *Infect. Immun.* **70**, 7176–7178 (2002).

75. Vuong, T. V. & Wilson, D. B. Glycoside hydrolases: Catalytic base/nucleophile diversity. *Biotechnol. Bioeng.* **107**, 195–205 (2010).

76. Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochem. J.* **382**, 769–781 (2004).

77. Koshland, D. E. Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev.* **28**, 416–436 (1953).

78. Luley-Goedl, C. & Nidetzky, B. Carbohydrate synthesis by disaccharide phosphorylases: Reactions, catalytic mechanisms and application in the glycosciences. *Biotechnol. J.* **5**, 1324–1338 (2010).

79. Bidossi, A. *et al.* A Functional Genomics Approach to Establish the Complement of Carbohydrate Transporters in Streptococcus pneumoniae. *PLoS One* **7**, e33320 (2012).

80. Tettelin, H. Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae. *Science (80-.).* **293**, 498–506 (2001).

81. Deutscher, J. *et al.* The Bacterial Phosphoenolpyruvate:Carbohydrate Phosphotransferase System: Regulation by Protein Phosphorylation and Phosphorylation-Dependent Protein-Protein Interactions. *Microbiol. Mol. Biol. Rev.* **78**, 231–256 (2014).

82. McCoy, J. G., Levin, E. J. & Zhou, M. Structural insight into the PTS sugar transporter EIIC. *Biochim. Biophys. Acta - Gen. Subj.* **1850**, 577–585 (2015).

83. Deutscher, J., Francke, C. & Postma, P. W. How Phosphotransferase System-Related Protein Phosphorylation Regulates Carbohydrate Metabolism in Bacteria. *Microbiol. Mol. Biol. Rev.* **70**, 939–1031 (2006).

84. Dawson, R. J. P., Hollenstein, K. & Locher, K. P. Uptake or extrusion: Crystal structures of full ABC transporters suggest a common mechanism. *Mol. Microbiol.* **65**, 250–257 (2007).

85. Paixão, L. *et al.* Host Glycan Sugar-Specific Pathways in Streptococcus pneumonia: Galactose as a Key Sugar in Colonisation and Infection. *PLoS One* **10**, e0121042 (2015).

86. Ferrando, M. L. *et al.* Carbohydrate Availability Regulates Virulence Gene Expression in Streptococcus suis. *PLoS One* **9**, e89334 (2014).

87. Troxler, L. J. *et al.* Carbon source regulates polysaccharide capsule biosynthesis in

Streptococcus pneumoniae. *J. Biol. Chem.* **294**, 17224–17238 (2019).

88. Trappetti, C. *et al.* Autoinducer 2 Signaling via the Phosphotransferase FruA Drives Galactose Utilization by Streptococcus pneumoniae , Resulting in Hypervirulence. *MBio* **8**, 1–18 (2017).
89. Al-Bayati, F. A. Y. *et al.* Pneumococcal galactose catabolism is controlled by multiple regulators acting on pyruvate formate lyase. *Sci. Rep.* **7**, 43587 (2017).
90. Willenborg, J. & Goethe, R. Metabolic traits of pathogenic streptococci. *FEBS Lett.* **590**, 3905–3919 (2016).
91. Chen, L. L. *et al.* Characterization and mutational analysis of two UDP-galactose 4-epimerases in Streptococcus pneumoniae TIGR4. *Biochem.* **83**, 37–44 (2018).
92. Richardson †, A. R., Somerville †, G. A. & Sonenshein †, A. L. Regulating the Intersection of Metabolism and Pathogenesis in Gram-positive Bacteria. in *Metabolism and Bacterial Pathogenesis* 129–165 (American Society of Microbiology, 2015). doi:10.1128/microbiolspec.MBP-0004-2014.
93. Carvalho, S. M., Kloosterman, T. G., Kuipers, O. P. & Neves, A. R. CcpA ensures optimal metabolic fitness of streptococcus pneumoniae. *PLoS One* **6**, (2011).
94. Kay, E. J., Yates, L. E., Terra, V. S., Cuccui, J. & Wren, B. W. Recombinant expression of Streptococcus pneumoniae capsular polysaccharides in Escherichia coli. *Open Biol.* **6**, (2016).
95. Yesilkaya, H. *et al.* Pyruvate Formate Lyase Is Required for Pneumococcal Fermentative Metabolism and Virulence. *Infect. Immun.* **77**, 5418–5427 (2009).
96. Echlin, H. *et al.* Pyruvate Oxidase as a Critical Link between Metabolism and Capsule Biosynthesis in Streptococcus pneumoniae. *PLOS Pathog.* **12**, e1005951 (2016).
97. Echlin, H., Frank, M., Rock, C. & Rosch, J. W. Role of the pyruvate metabolic network on carbohydrate metabolism and virulence in Streptococcus pneumoniae. *Mol. Microbiol.* **114**, 536–552 (2020).
98. Jones, C., Currie, F. & Forster, M. J. N.m.r. and conformational analysis of the capsular polysaccharide from Streptococcus pneumoniae type 4. *Carbohydr. Res.* **221**, 95–121 (1991).

99. Blanchette, K. A. *et al.* Neuraminidase A-Exposed Galactose Promotes Streptococcus pneumoniae Biofilm Formation during Colonization. *Infect. Immun.* **84**, 2922–2932 (2016).

100. Caveney, N. A., Li, F. K. & Strynadka, N. C. Enzyme structures of the bacterial peptidoglycan and wall teichoic acid biogenesis pathways. *Curr. Opin. Struct. Biol.* **53**, 45–58 (2018).

101. Vollmer, W., Massidda, O. & Tomasz, A. The Cell Wall of Streptococcus pneumoniae. *Microbiol. Spectr.* **7**, 3–18 (2019).

102. Filipe, S. R., Pinho, M. G. & Tomasz, A. Characterization of the murMN operon involved in the synthesis of branched peptidoglycan peptides in Streptococcus pneumoniae. *J. Biol. Chem.* **275**, 27768–27774 (2000).

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

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

105. Kim, J. O., Shchepetov, M. & Weiser, J. N. Association of intrastrain phase variation in quantity of capsular polysaccharide and teichoic acid with the virulence of Streptococcus pneumonioe. *Clin. Infect. Dis.* **25**, 427 (1997).

106. Roche, A. M., King, S. J. & Weiser, J. N. Live attenuated Streptococcus pneumoniae strains induce serotype-independent mucosal and systemic protection in mice. *Infect. Immun.* **75**, 2469–2475 (2007).

107. Croucher, N. J., Løchen, A. & Bentley, S. D. Pneumococcal Vaccines: Host Interactions, Population Dynamics, and Design Principles. *Annu. Rev. Microbiol.* **72**, 521–549 (2018).

108. Ganaie, F. *et al.* A New Pneumococcal Capsule Type, 10D, is the 100th Serotype and Has a Large cps Fragment from an Oral Streptococcus. *MBio* **11**, 1–15 (2020).

109. Mavroidi, A. *et al.* Genetic Relatedness of the Streptococcus pneumoniae Capsular Biosynthetic Loci. *J. Bacteriol.* **189**, 7841–7855 (2007).

110. Bentley, S. D. *et al.* Genetic Analysis of the Capsular Biosynthetic Locus from All 90

Pneumococcal Serotypes. *PLoS Genet.* **2**, e31 (2006).

111. Yother, J. Capsules of Streptococcus pneumoniae and Other Bacteria: Paradigms for Polysaccharide Biosynthesis and Regulation. *Annu. Rev. Microbiol.* **65**, 563–581 (2011).
112. Yother, J. Membrane linkage of a Streptococcus pneumoniae Wzy capsular polysaccharide occurs through an acylglycerol. 1–29 (2020) doi:10.1101/2020.09.16.299636.
113. Larson, T. R. & Yother, J. Streptococcus pneumoniae capsular polysaccharide is linked to peptidoglycan via a direct glycosidic bond to β-D-N-acetylglucosamine. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 5695–5700 (2017).
114. Wartha, F. *et al.* Capsule and d-alanylated lipoteichoic acids protect Streptococcus pneumoniae against neutrophil extracellular traps. *Cell. Microbiol.* **9**, 1162–1171 (2007).
115. Park, I. H. *et al.* Genetic, Biochemical, and Serological Characterization of a New Pneumococcal Serotype, 6H, and Generation of a Pneumococcal Strain Producing Three Different Capsular Repeat Units. *Clin. Vaccine Immunol.* **22**, 313–318 (2015).
116. Calix, J. J. *et al.* Biochemical, Genetic, and Serological Characterization of Two Capsule Subtypes among Streptococcus pneumoniae Serotype 20 Strains. *J. Biol. Chem.* **287**, 27885–27894 (2012).
117. Sorensen, U. B. S., Blom, J., Birch-Andersen, A. & Henrichsen, J. Ultrastructural localization of capsules, cell wall polysaccharide, cell wall proteins, and F antigen in pneumococci. *Infect. Immun.* **56**, 1890–1896 (1988).
118. Valentino, M. D. *et al.* Unencapsulated Streptococcus pneumoniae from conjunctivitis encode variant traits and belong to a distinct phylogenetic cluster. *Nat. Commun.* **5**, 5411 (2014).
119. Nelson, A. L. *et al.* Capsule Enhances Pneumococcal Colonization by Limiting Mucus-Mediated Clearance. *Infect. Immun.* **75**, 83–90 (2007).
120. Hyams, C., Camberlein, E., Cohen, J. M., Bax, K. & Brown, J. S. The Streptococcus pneumoniae capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. *Infect. Immun.* **78**, 704–715 (2010).
121. Pérez-Dorado, I., Galan-Bartual, S. & Hermoso, J. A. Pneumococcal surface proteins:

When the whole is greater than the sum of its parts. *Mol. Oral Microbiol.* **27**, 221–245 (2012).

122. Saleh, M. *et al.* Molecular architecture of Streptococcus pneumoniae surface thioredoxin-fold lipoproteins crucial for extracellular oxidative stress resistance and maintenance of virulence. *EMBO Mol. Med.* **5**, 1852–1870 (2013).
123. Abdullah, M. R. *et al.* Structure of the pneumococcal <scp>l</scp> , <scp>d</scp> -carboxypeptidase DacB and pathophysiological effects of disabled cell wall hydrolases DacA and DacB. *Mol. Microbiol.* **93**, n/a-n/a (2014).
124. Meziane-Cherif, D., Stogios, P. J., Evdokimova, E., Savchenko, A. & Courvalin, P. Structural basis for the evolution of vancomycin resistance D,D-peptidases. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 5872–5877 (2014).
125. Kohler, S., Voß, F., Gómez Mejia, A., Brown, J. S. & Hammerschmidt, S. Pneumococcal lipoproteins involved in bacterial fitness, virulence, and immune evasion. *FEBS Lett.* **590**, 3820–3839 (2016).
126. King, S. J., Hippe, K. R. & Weiser, J. N. Deglycosylation of human glycoconjugates by the sequential activities of exoglycosidases expressed by Streptococcus pneumoniae. *Mol. Microbiol.* **59**, 961–974 (2006).
127. Gómez-Mejia, A., Gámez, G. & Hammerschmidt, S. Streptococcus pneumoniae two-component regulatory systems: The interplay of the pneumococcus with its environment. *Int. J. Med. Microbiol.* **308**, 722–737 (2018).
128. Iovino, F. *et al.* pIgR and PECAM-1 bind to pneumococcal adhesins RrgA and PspC mediating bacterial brain invasion. *J. Exp. Med.* **214**, 1619–1630 (2017).
129. Radin, J. N. *et al.* β-Arrestin 1 Participates in Platelet-Activating Factor Receptor-Mediated Endocytosis of Streptococcus pneumoniae. *Infect. Immun.* **73**, 7827–7835 (2005).
130. Bergmann, S., Rohde, M., Chhatwal, G. S. & Hammerschmidt, S. alpha-Enolase of Streptococcus pneumoniae is a plasmin(ogen)-binding protein displayed on the bacterial cell surface. *Mol. Microbiol.* **40**, 1273–1287 (2001).
131. Löfling, J., Vimberg, V., Battig, P. & Henriques-Normark, B. Cellular interactions by LPxTG-

anchored pneumococcal adhesins and their streptococcal homologues. *Cell. Microbiol.* **13**, 186–197 (2011).

132. Dalia, A. B., Standish, A. J. & Weiser, J. N. Three surface exoglycosidases from Streptococcus pneumoniae, NanA, BgaA, and StrH, promote resistance to opsonophagocytic killing by human neutrophils. *Infect. Immun.* **78**, 2108–2116 (2010).
133. Hermoso, J. A. *et al.* Insights into pneumococcal pathogenesis from the crystal structure of the modular teichoic acid phosphorylcholine esterase Pce. *Nat. Struct. Mol. Biol.* **12**, 533–538 (2005).
134. Hakenbeck, R., Madhour, A., Denapaite, D. & Brückner, R. Versatility of choline metabolism and choline-binding proteins in Streptococcus pneumoniae and commensal streptococci. *FEMS Microbiol. Rev.* **33**, 572–586 (2009).
135. Hammerschmidt, S., Bethe, G., Remane, P. H. & Chhatwal, G. S. Identification of pneumococcal surface protein A as a lactoferrin- binding protein of Streptococcus pneumoniae. *Infect. Immun.* (1999) doi:10.1128/.67.4.1683-1687.1999.
136. Yamaguchi, M. *et al.* Identification of evolutionarily conserved virulence factor by selective pressure analysis of Streptococcus pneumoniae. *Commun. Biol.* **2**, 96 (2019).
137. Attali, C. *et al.* Streptococcus pneumoniae Choline-Binding Protein E Interaction with Plasminogen/Plasmin Stimulates Migration across the Extracellular Matrix. *Infect. Immun.* **76**, 466–476 (2008).
138. Johnsborg, O., Eldholm, V., Bjørnstad, M. L. & Håvarstein, L. S. A predatory mechanism dramatically increases the efficiency of lateral gene transfer in Streptococcus pneumoniae and related commensal species. *Mol. Microbiol.* (2008) doi:10.1111/j.1365-2958.2008.06288.x.
139. Eldholm, V., Johnsborg, O., Haugen, K., Ohnstad, H. S. & Havastein, L. S. Fratricide in Streptococcus pneumoniae: Contributions and role of the cell wall hydrolases CbpD, LytA and LytC. *Microbiology* **155**, 2223–2234 (2009).
140. Jacques, L. C. *et al.* Increased pathogenicity of pneumococcal serotype 1 is driven by rapid autolysis and release of pneumolysin. *Nat. Commun.* **11**, 1892 (2020).
141. Olaya-Abril, A. *et al.* Characterization of protective extracellular membrane-derived

vesicles produced by Streptococcus pneumoniae. *J. Proteomics* **106**, 46–60 (2014).

142. Bandara, M. *et al.* The accessory Sec system (SecY2A2) in Streptococcus pneumoniae is involved in export of pneumolysin toxin, adhesion and biofilm formation. *Microbes Infect.* **19**, 402–412 (2017).
143. Bergmann, S., Rohde, M. & Hammerschmidt, S. Glyceraldehyde-3-Phosphate Dehydrogenase of Streptococcus pneumoniae Is a Surface-Displayed Plasminogen-Binding Protein. *Infect. Immun.* **72**, 2416–2419 (2004).
144. Bergmann, S. *et al.* Identification of a novel plasmin(ogen)-binding motif in surface displayed α-enolase of Streptococcus pneumoniae. *Mol. Microbiol.* **49**, 411–423 (2003).
145. Chen, C. *et al.* Effect and cost-effectiveness of pneumococcal conjugate vaccination: a global modelling analysis. *Lancet Glob. Heal.* **7**, e58–e67 (2019).
146. Kilian, M. & Tettelin, H. Identification of Virulence-Associated Properties by Comparative Genome Analysis of Streptococcus pneumoniae , S. pseudopneumoniae , S. mitis , Three S. oralis Subspecies, and S. infantis. *MBio* **10**, 1–32 (2019).
147. Polissi, A. *et al.* Large-Scale Identification of Virulence Genes fromStreptococcus pneumoniae. *Infect. Immun.* **66**, 5620–5629 (1998).
148. Lau, G. W. *et al.* A functional genomic analysis of type 3 Streptococcus pneumoniae virulence. *Mol. Microbiol.* **40**, 555–571 (2001).
149. Culurgioni, S., Harris, G., Singh, A. K., King, S. J. & Walsh, M. A. Structural Basis for Regulation and Specificity of Fructooligosaccharide Import in Streptococcus pneumoniae. *Structure* **25**, 79–93 (2017).
150. Shelburne, S. A., Davenport, M. T., Keith, D. B. & Musser, J. M. The role of complex carbohydrate catabolism in the pathogenesis of invasive streptococci. *Trends Microbiol.* **16**, 318–325 (2008).
151. Shafeeq, S., Kloosterman, T. G. & Kuipers, O. P. CelR-mediated activation of the cellobioseutilization gene cluster in Streptococcus pneumoniae. *Microbiology* **157**, 2854–2861 (2011).
152. Obert, C. *et al.* Identification of a Candidate Streptococcus pneumoniae Core Genome and Regions of Diversity Correlated with Invasive Pneumococcal Disease. *Infect. Immun.*

74, 4766–4777 (2006).

153. Rosenow, C., Maniar, M. & Trias, J. Regulation of the α-galactosidase activity in Streptococcus pneumoniae: Characterization of the raffinose utilization system. *Genome Res.* **9**, 1189–1197 (1999).

154. Tan, M.-F. *et al.* MsmK, an ATPase, Contributes to Utilization of Multiple Carbohydrates and Host Colonization of Streptococcus suis. *PLoS One* **10**, e0130792 (2015).

155. Marion, C., Aten, A. E., Woodiga, S. A. & King, S. J. Identification of an ATPase, MsmK, Which Energizes Multiple Carbohydrate ABC Transporters in Streptococcus pneumoniae. *Infect. Immun.* **79**, 4193–4200 (2011).

156. Aprianto, R., Slager, J., Holsappel, S. & Veening, J. W. Time-resolved dual RNA-seq reveals extensive rewiring of lung epithelial and pneumococcal transcriptomes during early infection. *Genome Biol.* **17**, 1–16 (2016).

157. Song, X. M., Connor, W., Hokamp, K., Babiuk, L. A. & Potter, A. A. Streptococcus pneumoniae early response genes to human lung epithelial cells. *BMC Res. Notes* **1**, 1–10 (2008).

158. Song, X. M., Connor, W., Hokamp, K., Babiuk, L. A. & Potter, A. A. Transcriptome studies on Streptococcus pneumoniae, illustration of early response genes to THP-1 human macrophages. *Genomics* **93**, 72–82 (2009).

159. Wu, K. *et al.* A Novel Protein, RafX, Is Important for Common Cell Wall Polysaccharide Biosynthesis in Streptococcus pneumoniae: Implications for Bacterial Virulence. *J. Bacteriol.* **196**, 3324–3334 (2014).

160. Heß, N. *et al.* Lipoteichoic acid deficiency permits normal growth but impairs virulence of Streptococcus pneumoniae. *Nat. Commun.* **8**, 2093 (2017).

161. Islam, S. T., Taylor, V. L., Qi, M. & Lam, J. S. Membrane topology mapping of the O-antigen flippase (Wzx), polymerase (Wzy), and ligase (WaaL) from Pseudomonas aeruginosa PAO1 reveals novel domain architectures. *MBio* (2010) doi:10.1128/mBio.00189-10.

162. Russell, R. R. B., Mukasa, H., Shimamura, A. & Ferretti, J. J. Streptococcus mutans gtfA gene specifies sucrose phosphorylase. *Infect. Immun.* **56**, 2763–2765 (1988).

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

164. Varki, A. *Essentials of Glycobiology [Internet]*. *Cold Spring Harbor (NY)* (2015).

165. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* (1997) doi:10.1016/S0076-6879(97)76066-X.

166. Wlodawer, A., Minor, W., Dauter, Z. & Jaskolski, M. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS J.* **280**, 5705–5736 (2013).

167. Paterson, N. G., Riboldi-Tunnicliffe, A., Mitchell, T. J. & Isaacs, N. W. Purification, crystallization and preliminary X-ray diffraction analysis of RafE, a sugar-binding lipoprotein from Streptococcus pneumoniae. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* (2006) doi:10.1107/S1744309106021695.

168. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

169. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2004) doi:10.1107/S0907444904019158.

170. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography* (1997) doi:10.1107/S0907444996012255.

171. Adams, P. D. *et al.* PHENIX : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).

172. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* (2018) doi:10.1002/pro.3330.

173. Scheepers, G. H., Lycklama a Nijeholt, J. A. & Poolman, B. An updated structural classification of substrate-binding proteins. *FEBS Lett.* **590**, 4393–4401 (2016).

174. Fukamizo, T., Kitaoku, Y. & Suginta, W. Periplasmic solute-binding proteins: Structure classification and chitooligosaccharide recognition. *Int. J. Biol. Macromol.* **128**, 985–993 (2019).

175. Holm, L. & Laakso, L. M. Dali server update. *Nucleic Acids Res.* **44**, W351–W355 (2016).

176. Ejby, M. *et al.* An ATP Binding Cassette Transporter Mediates the Uptake of α-(1,6)-Linked Dietary Oligosaccharides in Bifidobacterium and Correlates with Competitive Growth on These Substrates. *J. Biol. Chem.* **291**, 20220–20231 (2016).

177. Webb, A. J., Homer, K. A. & Hosie, A. H. F. Two closely related ABC transporters in Streptococcus mutans are involved in disaccharide and/or oligosaccharide uptake. *J. Bacteriol.* **190**, 168–178 (2008).

178. Tan, M.-F. *et al.* The involvement of MsmK in pathogenesis of the Streptococcus suis serotype 2. *Microbiologyopen* **6**, e00433 (2017).

179. CAZy - GH36. http://www.cazy.org/GH36.html.

180. Bruel, L. *et al.* α-Galactosidase/Sucrose Kinase (AgaSK), a Novel Bifunctional Enzyme from the Human Microbiome Coupling Galactosidase and Kinase Activities*. *J. Biol. Chem.* **286**, 40814–40823 (2011).

181. Merceron, R. *et al.* The Molecular Mechanism of Thermostable α-Galactosidases AgaA and AgaB Explained by X-ray Crystallography and Mutational Studies*. *J. Biol. Chem.* **287**, 39642–39652 (2012).

182. Panwar, D., Shubhashini, A., Chaudhari, S. R., Prashanth, K. V. H. & Kapoor, M. GH36 α-galactosidase from Lactobacillus plantarum WCFS1 synthesize Gal-α-1,6 linked prebiotic α-galactooligosaccharide by transglycosylation. *Int. J. Biol. Macromol.* **144**, 334–342 (2020).

183. Fridjonsson, O., Watzlawick, H. & Mattes, R. Thermoadaptation of α-galactosidase AgaB1 in Thermus thermophilus. *J. Bacteriol.* **184**, 3385–3391 (2002).

184. Hobbs, J. K., Pluvinage, B., Robb, M., Smith, S. P. & Boraston, A. B. Two complementary α-fucosidases from Streptococcus pneumoniae promote complete degradation of host-derived carbohydrate antigens. *J. Biol. Chem.* **294**, 12670–12682 (2019).

185. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).

186. Cuyvers, S., Dornez, E., Delcour, J. A. & Courtin, C. M. Occurrence and functional significance of secondary carbohydrate binding sites in glycoside hydrolases. *Crit. Rev. Biotechnol.* **32**, 93–107 (2012).

187. Franceus, J. & Desmet, T. Sucrose Phosphorylase and Related Enzymes in Glycoside Hydrolase Family 13: Discovery, Application and Engineering. *Int. J. Mol. Sci.* **21**, 2526 (2020).

188. Tyx, R. E., Roche-Hakansson, H. & Hakansson, A. P. Role of dihydrolipoamide dehydrogenase in regulation of raffinose transport in Streptococcus pneumoniae. *J. Bacteriol.* **193**, 3512–3524 (2011).

189. Bissaro, B., Monsan, P., Fauré, R. & O'Donohue, M. J. Glycosynthesis in a waterworld: New insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. *Biochem. J.* **467**, 17–35 (2015).

190. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* (2005) doi:10.1016/j.pep.2005.01.016.

191. AT, B. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–5 (1992).

192. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).

193. DeLano, W. L. Pymol: An open-source molecular graphics tool. *{CCP4} Newsl. Protein Crystallogr.* (2002).

194. Van Bueren, A. L., Higgins, M., Wang, D., Burke, R. D. & Boraston, A. B. Identification and structural basis of binding to host lung glycogen by streptococcal virulence factors. *Nat. Struct. Mol. Biol.* **14**, 76–84 (2007).

195. Janapatla, R. P., Chen, C. L., Hsu, M. H., Liao, W. T. & Chiu, C. H. Immunization with pneumococcal neuraminidases NanA, NanB and NanC to generate neutralizing antibodies and to increase survival in mice. *J. Med. Microbiol.* **67**, 709–723 (2018).

196. Kahya, H. F., Andrew, P. W. & Yesilkaya, H. Deacetylation of sialic acid by esterases potentiates pneumococcal neuraminidase activity for mucin utilization, colonization and virulence. *PLoS Pathog.* **13**, 1–21 (2017).

197. Ritchie, G. E. *et al.* Glycosylation and the complement system. *Chem. Rev.* **102**, 305–319 (2002).

198. Butters, T. D., Alonzi, D. S., Kukushkin, N. V., Ren, Y. & Blériot, Y. Novel mannosidase

inhibitors probe glycoprotein degradation pathways in cells. *Glycoconj. J.* **26**, 1109–1116 (2009).

199. Legler, K. *et al.* Reduced mannosidase MAN1A1 expression leads to aberrant N-glycosylation and impaired survival in breast cancer. *Br. J. Cancer* **118**, 847–856 (2018).
200. Van Den Elsen, J. M. H., Kuntz, D. A. & Rose, D. R. Structure of Golgi α-mannosidase II: A target for inhibition of growth and metastasis of cancer cells. *EMBO J.* **20**, 3008–3017 (2001).
201. Vallée, F., Karaveg, K., Herscovics, A., Moremen, K. W. & Howell, P. L. Structural basis for catalysis and inhibition of N-glycan processing class I α1,2-mannosidases. *J. Biol. Chem.* **275**, 41287–41298 (2000).
202. Masson, G. R. *et al.* Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat. Methods* **16**, 595–602 (2019).
203. McPhail, J. A. *et al.* Characterization of the c10orf76-PI4KB complex and its necessity for Golgi PI4P levels and enterovirus replication. *EMBO Rep.* **21**, (2020).
204. Gallagher, E. S. & Hudgens, J. W. *Mapping Protein-Ligand Interactions with Proteolytic Fragmentation, Hydrogen/Deuterium Exchange-Mass Spectrometry*. *Methods in Enzymology* vol. 566 (Elsevier Inc., 2016).
205. Masson, G. R., Jenkins, M. L. & Burke, J. E. An overview of hydrogen deuterium exchange mass spectrometry (HDX-MS) in drug discovery. *Expert Opin. Drug Discov.* **12**, 981–994 (2017).
206. Khayyo, V. I. *et al.* Crystal structure of a lipin/Pah phosphatidic acid phosphatase. *Nat. Commun.* **11**, 1309 (2020).
207. Lim, X.-X. *et al.* Conformational changes in intact dengue virus reveal serotype-specific expansion. *Nat. Commun.* **8**, 14339 (2017).
208. Huang, R. Y. C. & Hudgens, J. W. Effects of desialylation on human α1-acid glycoprotein-ligand interactions. *Biochemistry* **52**, 7127–7136 (2013).
209. Davies, G. & Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853–859 (1995).

210. Konermann, L., Pan, J. & Liu, Y.-H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **40**, 1224–1234 (2011).

211. Oganesyan, I., Lento, C. & Wilson, D. J. Contemporary hydrogen deuterium exchange mass spectrometry. *Methods* **144**, 27–42 (2018).

212. Gao, K., Oerlemans, R. & Groves, M. R. Theory and applications of differential scanning fluorimetry in early-stage drug discovery. *Biophys. Rev.* **12**, 85–104 (2020).

213. Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N. & Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **357**, 289–298 (2006).

214. Bai, N., Roder, H., Dickson, A. & Karanicolas, J. Isothermal Analysis of ThermoFluor Data can readily provide Quantitative Binding Affinities. *Sci. Rep.* **9**, 1–15 (2019).

215. Bengoechea, J. A. *et al.* Functional Characterization of Gne (UDP-N-Acetylglucosamine- 4-Epimerase), Wzz (Chain Length Determinant), and Wzy (O-Antigen Polymerase) of Yersinia enterocolitica Serotype O:8. *J. Bacteriol.* **184**, 4277–4287 (2002).

216. Henderson, B. & Martin, A. Bacterial virulence in the moonlight: Multitasking bacterial moonlighting proteins are virulence determinants in infectious disease. *Infect. Immun.* **79**, 3476–3491 (2011).

217. Zomer, A., Hermans, P. W. M. & Bootsma, H. J. Non-Adhesive Surface Proteins of Streptococcus pneumoniae. in *Streptococcus Pneumoniae* 231–244 (Elsevier, 2015). doi:10.1016/B978-0-12-410530-0.00012-0.

218. Nakai, H. *et al.* Aspergillus nidulansα-galactosidase of glycoside hydrolase family 36 catalyses the formation of α-galacto-oligosaccharides by transglycosylation. *FEBS J.* **277**, 3538–3551 (2010).

219. Stam, M. R., Danchin, E. G. J., Rancurel, C., Coutinho, P. M. & Henrissat, B. Dividing the large glycoside hydrolase family 13 into subfamilies: Towards improved functional annotations of α-amylase-related proteins. *Protein Eng. Des. Sel.* **19**, 555–562 (2006).

220. Franceus, J., Decuyper, L., D'hooghe, M. & Desmet, T. Exploring the sequence diversity in glycoside hydrolase family 13_18 reveals a novel glucosylglycerol phosphorylase. *Appl. Microbiol. Biotechnol.* **102**, 3183–3191 (2018).

221. Franceus, J., Pinel, D. & Desmet, T. Glucosylglycerate phosphorylase, an enzyme with novel specificity involved in compatible solute metabolism. *Appl. Environ. Microbiol.* **83**, 1–13 (2017).

222. Franceus, J., Capra, N., Desmet, T. & Thunnissen, A.-M. W. H. Structural Comparison of a Promiscuous and a Highly Specific Sucrose 6F-Phosphate Phosphorylase. *Int. J. Mol. Sci.* **20**, 3906 (2019).

223. Jeong, J. W. *et al.* Biosynthesis of glucosyl glycerol, a compatible solute, using intermolecular transglycosylation activity of amylosucrase from Methylobacillus flagellatus KT. *Appl. Biochem. Biotechnol.* **173**, 904–917 (2014).

224. Shi, W. W. *et al.* Structure of a novel O-linked N-acetyl-D-glucosamine (O-GlcNAc) transferase, GtfA, reveals insights into the glycosylation of pneumococcal serine-rich repeat adhesins. *J. Biol. Chem.* **289**, 20898–20907 (2014).

225. Byrd-Leotis, L. *et al.* Influenza binds phosphorylated glycans from human lung. *Sci. Adv.* **5**, eaav2554 (2019).

226. Kotecha, A. *et al.* Rules of engagement between αvβ6 integrin and foot-and-mouth disease virus. *Nat. Commun.* **8**, (2017).

227. Brown, C. K. *et al.* Structure of the streptococcal cell wall C5a peptidase. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 18391–18396 (2005).

228. Mu, R. *et al.* Identification of a group B streptococcal fibronectin binding protein, SfbA, that contributes to invasion of brain endothelium and development of meningitis. *Infect. Immun.* **82**, 2276–2286 (2014).

229. Johnson, M. D. L. *et al.* Pseudomonas aeruginosa PilY1 binds integrin in an RGD- and calcium-dependent manner. *PLoS One* **6**, 1–8 (2011).

230. Koyama, R., Hakamata, W., Hirano, T. & Nishio, T. Identification of small-molecule inhibitors of human golgi mannosidase via a drug repositioning screen. *Chem. Pharm. Bull.* **66**, 678–681 (2018).