A STUDY OF STREPTOCOCCAL GLYCOSYLATION AND ITS IMPACT ON HOST-MICROBE INTERACTIONS

by

SILKE ANDRESEN

(Under the Direction of Christine M. Szymanski)

ABSTRACT

A prominent feature contributing to pathogenicity within the *Streptococcus* genus is the involvement of surface glycans, including capsular polysaccharides (CPS), cell wall associated polysaccharides and O-glycosylated proteins. In general, these carbohydrates are utilized to block immune detection, serve as decoys or aid in host colonization. In this thesis, we employed *S. pneumoniae* (major cause of bacterial pneumonia) and *S. mutans* (primary etiological agent for caries) to better understand the importance of bacterial glycosylation among Streptococci. The human host produces a poorly studied immune protein, human intelectin-1 (hIntL-1), that has been shown to bind the CPS of certain *S. pneumoniae* serotypes. Investigation of *S. pneumoniae* interactions with hIntL-1 showed that hIntL-1 causes bacterial agglutination but does not kill bacteria directly or in complement-dependent manner. Instead, binding of this lectin leads to increased bacterial attachment to host cells, a phenomenon that has previously been described for Streptococcal strains capable of forming long bacterial chains. However, in the presence of fresh human blood cells, hIntL-1 causes enhanced killing of *S. pneumoniae* through phagocytosis specifically by neutrophils. In this study, I demonstrated how a bacterial

surface structure typically used for the evasion of host defenses, could in turn be targeted by a specialized innate immune lectin.

In *S. mutans*, the Pgf glycosylation machinery has recently been described to contribute to the pathogenicity of the organism in various ways. Using a truncated version of the Pgf target protein Cnm, we identified its glycosylation sites that are heavily modified with HexNAc2. We further characterized the key Pgf enzyme involved in HexNAc biosynthesis, PgfE. Computational modeling, NMR and capillary electrophoreses confirmed that PgfE is an NAD^+ dependent UDP-GlcNAc/UDP-GalNAc 4-epimerase. The second 4-epimerase expressed by *S. mutans*, GalE, is required for galactose metabolism, but not protein glycosylation. The loss of GalE leads to defects in cell division and septum formation, presumably due to the loss of glucose which is one of the main building blocks of the *S. mutans* cell wall polysaccharide. These results not only provided new insights into protein O-glycosylation in *S. mutans*, but also demonstrate the importance of a single epimerase for proper cell division.

INDEX WORDS: *Streptococcus pneumoniae*, capsular polysaccharide, CPS, human intelectin-1, hIntL-1, innate immunity, neutrophils, *Streptococcus mutans*, protein O-glycosylation, Cnm, 4-epimerases, PgfE, GalE.

A STUDY OF STREPTOCOCCAL GLYCOSYLATION AND ITS IMPACT ON HOST-MICROBE INTERACTIONS

by

SILKE ANDRESEN

B.Sc., Hochschule Darmstadt, Germany, 2014

M.Sc., Goethe University Frankfurt, Germany, 2017

A Dissertation Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2022

A STUDY OF STREPTOCOCCAL GLYCOSYLATION AND ITS IMPACT ON HOST-MICROBE INTERACTIONS

by

SILKE ANDRESEN

Major Professor:	Christine M. Szymanski
Committee:	Fikri Y. Avci
	J. Michael Pierce
	Vincent J. Starai
	M. Stephen Trent

Electronic Version Approved:
Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2022

DANKSAGUNG

Meine Doktorarbeit widme ich meinen Eltern, den Standsäulen meines Lebens. Wie auch jeder andere Mensch auf dem Planeten, stände ich heute nicht hier, gäbe es meine Eltern nicht. Wie ich jedoch lernen musste, können nur die wenigsten von sich behaupten von ihren Erschaffern und Erziehern nicht verdorben worden zu sein. Ganz im Gegenteil, habe ich zum ersten Mal das Büro eines Therapeuten betreten, weil ich meine Eltern vermisst habe, und diese die Grundwerte meines Seins auf eine solch hohe Ebene gesetzt haben, das es mir, sowie dem Rest der Menschheit schwerfiel diesen gerecht zuwenden. Wo meine Mutter mein Vorbild in Besonnenheit, Einfühlungsvermögen und Fürsorge ist, so strebe ich meinem Vater in seinem Perfektionismus, Realismus und seiner Bedachtsamkeit nachzukommen. Bis heute mache ich den Verlauf meines Erwachsenenlebens an einem einzigen Moment fest, in dem mein Vater darauf bestanden hat, ich solle am Mentoren Programm für Frauen in der Wissenschaft teilnehmen. Dieser Entschluss führte zu meiner Bachelorarbeit, meinem Praktikum in Irland und schließlich meinem Doktor in den USA. Mein Vater hat mir vorgelebt mich nie mit weniger zufrieden zu geben als das zu dem ich fähig bin, meine Leistungen an meinem inneren (und damit seinem) Standard zu messen und für das zu kämpfen das ich mir wünsche. Mit solch hohen und hart zu erstrebenden Zielen wäre ich jedoch zugleich schon vor Jahren gescheitert, hatte mir nicht ebenso meine Mutter zur Seite gestanden und mich gelehrt die Hürden zu meistern die mir in den Weg gelegt wurden.

ACKNOWLEDGEMENTS

My Ph.D. has been 10 years in the making and I have had the fortune to work under as well as with many great scientists. Among them are not only the professors and students who taught me during my Bachelor and Master studies, but also the mentors I learned from during my Bachelor thesis at Heraeus, my internship in Ireland and in the Neidle laboratory. Since these however are the acknowledgements for my Ph.D. thesis, I want to begin with my supervisor Christine Szymanski, who has provided support and never-ending inspiration, while at the same time letting me take the lead on my studies. I feel immense gratitude for the original crew that accepted me into the Szymanski lab, including Justin Duma, Cody Thomas, Clay Crippen, Jolene Gaber, Jessica Sacher, Harald Nothaft, Cory Wenzel and especially Rob Patry who stayed my mentor even years after my rotation. I also want to thank the current Szymanski lab residents Bibi Zhou, Ashley Rogers and Oliver Gregory, who were always ready to hear me complain about science or help me out with experiments. Since moving to the US, I have made many good friends and several great friends whom I lovingly call my emotional support friends, among them Katie Bellissimo, Paige Lamore, Kayla Fantone, Neely Palacio, Ronnie Fulton and Alex Frank. Further, acknowledgements to Janice Stuart, Andrea Barnett and the Micro staff, since I fully believe that nothing in the department would get done without them. Lastly, my husband Chris Moxley, my union with whom was forged in the shared sufferings of graduate school for the past 5 years. I would never have wanted to do it without you and our two emotion support cats.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1. An overview of the human immune system

When defending the host against microbial invasion, the human immune system reacts via two distinct but synergistic mechanisms, the innate and the adaptive immune response. The innate immune response is activated in the first minutes of injury or infection and represents a defense of low specificity to a broad range of pathogens. The adaptive immunity, on the other hand, acquires a memory over time, making its response slow, but highly specific. Both responses rely heavily on the distinction between self and non-self in order to target attacks against invading microbes (1). Innate immunity is achieved through general antimicrobial proteins and enzymes (lectins, lysozyme, phospholipases, lactoferrin, etc.), pattern recognition receptors (PRRs) and the complement pathway (Figure 1.1) (2, 3). These players of the innate response all bind bacteria, while their specificities avoid recognition of host cells and structures. While antimicrobial enzymes exhibit their effects independently (eg degradation of peptidoglycan), PRRs often activate an immune response upon binding of a microbial component (4). Examples include Toll-like receptors, which are membrane-bound and recognize exosomal patterns like lipopolysaccharides and nucleic acids, and NOD and RIG-1 like receptors, which are present in the cytoplasm and bind to peptidoglycan and viral RNA respectively (5, 6). Another mechanism of recognizing and killing microbes is through the three pathways of complement (7, 8) –

classical, lectin, and alternative. In the classical pathway pathogen binding is achieved by antibodies, while in the lectin pathway C-type lectins like the mannose-binding lectin or ficolin-2 recognize surface carbohydrates of a bacterium (9). In the alternative complement pathway, protein C3b, a member of the complement cascade binds directly to certain polysaccharides in yeast and Gram-negative bacteria, leading to their elimination. After the initial recognition of the microbe, the complement cascade is activated, in which over 20 different serum glycoproteins lead to the formation of the membrane attack complex which forms a transmembrane pore in the targeted microbe, leading to its osmotic lysis (7, 8). The binding of complement proteins to the microbial surface additionally serves as a physical barrier, blocking binding sites on the invader and, in the case of antibodies and lectins, can lead to their agglutination due to the presence of multiple binding sites on these proteins (10). Further, both the lectins of the lectin complement pathway and a number of complement-cascade proteins (e.g. C3b) exhibit opsonizing actions that facilitate phagocytosis (11). Together the antimicrobial proteins, receptors and complement comprise the humoral response of the innate immune system, while the cellular response consists of phagocytes, dendritic cells, natural killer cells and mast cells (12). Phagocytes are white blood cells that engulf and kill microbes and exist as two types: neutrophils and macrophages. After the recruitment of neutrophils to the site of an infection, they recognize the microbe or the immune opsonins with various membrane bound receptors or lectins. The initial contact leads to the engulfment of the invader and the formation of the phagosome in which reactive oxygen species and antimicrobial proteins (released from granules) are utilized to kill bacteria (13). In addition to phagocytosis, neutrophils can kill bacteria by releasing neutrophil extracellular traps (NETs) which are composed of a mesh

of chromatin fibers, decorated with antimicrobial enzymes (14, 15). Macrophages engulf microbes and digest them via enzymatic and peroxide degradation as well, but they also act as professional antigen-displaying cells that display antigens from digested microbes on their major histocompatibility complex 2 (MHC2) and present them to helper T-cells (16). Dendritic cells also engulf microbes, but with the sole purpose of processing and presentating microbial antigens to T-cells on their MHC2 (17). Natural killer cells are non-phagocytic and survey the body cells for their MHC1 complexes, which all nucleated cells express, to present small peptides as antigens on their surface (18, 19). The natural killer cells are able to detect infected or cancerous cells by the abnormal antigens presented on their MHC1 and induce apoptosis (19). To promote an increased immune response, mast cells can be activated through a variety of cell surface bound receptors and release inflammatory signaling molecules such as cytokines, growth factors or lipid mediators (12, 20).

If an infection persists beyond the first line of innate immune defenses, the adaptive immune response comes into play. While the innate immune response relies on pattern recognition via certain immune proteins to detect invading organisms, the adaptive response employs antigen presenting cells and antigen recognizing cells which in turn are supervised by control cells. The three antigen presenting cells are macrophages, dendritic cells and B-lymphocytes or B-cells (21). B-cells display specific antibodies, whose binding sites are randomly generated during germline development, leading to a multitude of B-cells with different epitope-binding specificities. This leads to a multitude of B-cells with varying antibodies, which will bind to their target regardless of its origin, engulf the target and display their antigens on their MHC2 (22). Due to this nonspecific binding of the B-

cells, a control mechanism is needed to verify whether the particle bound and displayed by the B-cell belongs to self or non-self. This is achieved by helper T-cells, which – similar to the B-cell – display an endless variety of receptors, but all of which only recognize foreign antigens (23). If a helper T-cell recognizes an antigen displayed by a B-cell it is thereby identified as part of an invader, which triggers the T-cell to multiply into more helper T-cells and memory T-cells and release cytokines that, in turn, both stimulate the B-cell to multiply into effector B-cells and memory B-cells and activate T-cells to survey the body for infected cells via their MHC1 (21, 23). The effector B-cells produce an abundance of the specific antibody, while the memory B-cells are long lived and preserve the genetic information for the antibody for future infections. These specific antibodies secreted by the B-cells affect the invading pathogen in several ways: firstly, the antibodies bind to the surface of the organism, building a physical barrier while at the same time blocking receptors on its surface (24). Secondly, the two epitope binding sites on the antibodies allow them to bind to more than one invading organism at once, thereby agglutinating them (10). Lastly, the antibodies opsonize the bacteria, marking them for uptake by phagocytic immune cells (11, 24).

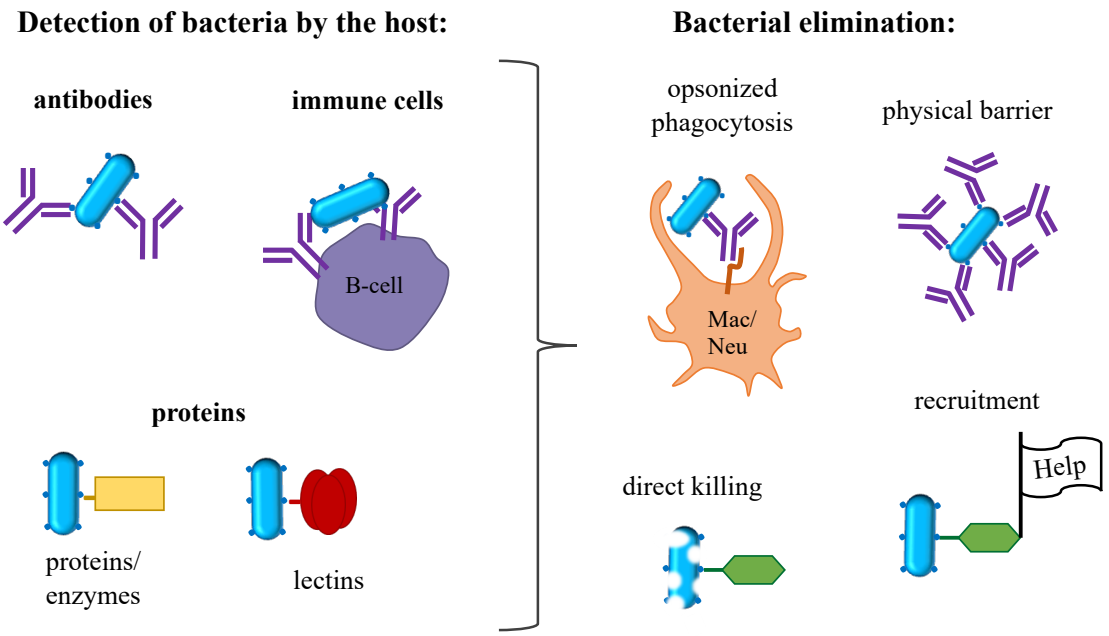


Figure 1.1: Overview of the initial recognition of bacteria through the immune response followed by their elimination. Free, as well as cell surface bound antibodies, antimicrobial enzymes and immune lectins specifically recognize various bacterial antigens. This initial contact leads to bacterial elimination through different mechanisms: Direct bacterial killing can be executed by immune proteins (green) or cells (like B-cells). Indirect killing is achieved by immune proteins (including lectins and antibodies) through the recruitment of proteins or phagocytes (macrophages, Mac and neutrophils, Neu). Lastly immune proteins can form a physical barrier around the bacterium to block contact to the host tissue.

2. Bacterial Glycobiology

The well-accepted dogma of molecular biology describes how the genetic information of all living organisms is encoded in their DNA, which is transcribed into RNA to serve as the building instructions for all the proteins that a given cell can produce. Therefore nucleic acids and proteins, along with lipids constitute three of the four major macromolecules in the cell with glycans being the fourth. The 4 nucleotides present in DNA or RNA can be arranged into 64 different triplets encoding for a total of 22 amino acids, all of which are attached to each other in the same linear fashion, 5' end to 3' end and N-terminus to C-terminus respectively. This, however, is not true for glycans, since

the hundreds of known monosaccharides can be connected linearly or branched in a variety of different ways depending on the connections to each atom (eg. 1→3) and the stereochemistry of the linkage (alpha vs. beta) (25). Additionally, glycans can be found as a free form or as glycoconjugates bound to proteins or lipids. This makes glycans the most statistically diverse group of macromolecules any cell harbors (25). Glycans have been shown to have a wide variety of functions ranging from subtle effects to loss of function in their absence. Their roles can be divided into three primary categories: structure and modulation, recognition by other molecules and host mimicry. All eukaryotic and prokaryotic cells are covered in glycans, which serve as a structural component as well as a physical protective barrier. The addition of glycans to newly synthesized proteins can aid their folding process and stability, which often influences their ability to function properly. Modulation of proteins can occur for example through direct phosphorylation/dephosphorylation or the presence/absence of glycans neighboring an enzyme's target, thereby tuning the activity of the protein. Due to the vast presence of glycans on proteins and cell surfaces, they play a major role in a variety of recognition events. Glycans can be part of cell-cell or cell-microbe interactions, cell-matrix interactions, receptor recognition events and even carbohydrate-carbohydrate interactions. Lastly, glycans can be used by microbes and parasites to mimic the cell surface glycans of their host to evade immune recognition. This short overview of functions executed by glycans illustrates their involvement and importance in all areas of life (26).

Cellular glycosylation can generally be divided into two forms, the attachment to lipids or proteins. While both can be found in eukaryotes as well as prokaryotes, the synthesis mechanisms are different. Glycosylation generally occurs in close proximity to

cellular membranes, which in the case of eukaryotes involves compartmentalization in the endoplasmic reticulum (ER) and Golgi. This review will focus primarily on glycosylation systems in prokaryotes and more specifically in bacteria, with some comparisons to the processes found in higher life forms.

2.1 Protein glycosylation

2.1.1 N-linked glycosylation

It has been proposed that more than two-thirds of all eukaryotic proteins are glycosylated (27), with some glycans comprising the majority of the protein's molecular mass (28), making glycosylation the most important post-translational protein modification. Less is known about the prevalence of glycans on bacterial proteins, due to the vast variety of glycan structures found in bacteria. The two main types of protein glycosylation are N-linked and O-linked glycosylation, which, above all, differ in the glycan-receiving amino acid. During N-linked glycosylation, glycans are attached to the amide group of asparagine residues within a consensus sequence. The bacterial consensus sequence consists of Asp/Glu-X1 -Asn-X2 -Ser/Thr (X1 and X2 represent any amino acid except Pro)(29) and is larger than in eukaryotes (Asn-X-Ser/ Thr). The traditional mechanism for N-glycosylation, that is also in eukaryotes, was found in *Campylobacter jejuni* and consists of block transfer of sugars from a membrane embedded lipid anchor (30). On the cytosolic side of the bacterial cell membrane, sugars are sequentially added onto an undecaprenyl pyrophosphate lipid anchor by glycosyltransferases to build a heptasaccharide, which is then flipped to the periplasmic face of the cell membrane by a flippase, where the oligosaccharide is transferred onto the receiving asparagine of the target

protein (or released as a free oligosaccharide) by an oligosaccharyltransferase (31). While block transfer is the most commonly observed mechanism for protein N-glycosylation in bacteria, a sequential transfer mechanism has been described in *Haemophilus influenzae* (32), wherein galactose or glucose residues are sequentially transferred one-by-one onto the protein in the cytoplasm by a single enzyme onto a sequon not requiring the -2 Asp/Glu. Once glycosylated, the protein is transported through the cell membrane by a Sec transporter, to which it stays bound (32). While different versions of these two N-glycosylation mechanisms were described for a variety of Gram-negative bacteria, N-glycosylation has not yet been shown in any Gram-positive bacteria (33, 34).

2.1.2 O-linked glycosylation

During O-linked glycosylation, glycans are attached to the hydroxyl group of serine or threonine residues and in contrast to N-glycosylation, an absolute consensus sequence for this process has not been defined (35). In a eukaryotic context, O-linked glycans are considered to be simpler in structure and less commonly found than N-linked glycans (27, 35). In bacteria, O-glycosylation has been found in both Gram-positive and Gram-negative organisms to proceed by one of two major mechanisms, which are similar to N-glycosylation. The less common mechanism consists of a block transfer in which the saccharides are assembled onto a lipid anchor on the cytoplasmic face, flipped to the periplasmic face of the cytoplasmic membrane, and transferred onto the protein. In the second mechanism for O-glycosylation, monosaccharides are sequentially transferred directly onto the serine or threonine residue of the receiving peptide, using activated glycan donors. In the following assembly of the polysaccharide, monomers are added to the

growing O-glycan before the fully glycosylated protein is typically transported to the cell surface (31, 36).

Combined, glycosylation can account for up to 90 % of the molecular weight of glycosylated proteins (28), making it evident how crucial the presence of sugars is to the stability and/or function of many proteins.

2.2 Bacterial surface glycans

2.2.1 Lipooligosaccharides and lipopolysaccharides

Protein glycosylation is not the only kind of glycan modification found in and on cells – all eukaryotes and many prokaryotes also exhibit a variety of glycans on their surface. In bacteria, many types of surface glycosylation depend on their Gram-staining type. Gram-negative bacteria are enveloped by an inner phospholipid bilayer membrane, followed by a thin layer of peptidoglycan in the periplasm and an outer membrane (37). The inner leaflet of the outer membrane is made up of phospholipids, while the outer leaflet primarily consists of lipid A, a structure constructed of two glucosamine molecules (usually phosphorylated) and a varying number of fatty acid chains (38). Lipid A, which is unique to Gram-negative bacteria, was shown to have an endotoxic character capable of triggering inflammation and immune responses in the host (38, 39). Lipid A also acts as the membrane anchor region of two of the major surface glycans found in Gram-negative bacteria: lipooligosaccharides (LOS) or lipopolysaccharides **(**LPS). The oligosaccharide in LOS consists of an inner core region, which is rather conserved among species of a given genera, and a variable outer core region. The inner core is attached to lipid A via one or two 2-keto-3-deoxyoctulosonic acids (KDO) residues (40), often followed by heptose residues.

LPS differ from LOS in that they contain an O-antigen attached to their outer core, which consists of a long glycan chain with repeating-unit structure. LOS and LPS are also highly endotoxic, primarily due to their lipid A anchor (39). During LPS synthesis, KDO is attached to its membrane anchor lipid A in the cytosol and monosaccharides are then sequentially added onto KDO to build the core structures. The simultaneous assembly of the O-antigen repeating unit typically proceeds according to one of the two following LPS-biosynthesis pathways. In the ABC-transporter dependent pathway the entire O-antigen is assembled sequentially onto undecaprenylpyrophosphate (UndPP) before being transported to the periplasmic space by an ABC-transporter (41). In the Wzy-dependent LPS synthesis pathway, repeating units are assembled onto UndPP and flipped to the periplasm where the new repeating unit is added to the reducing end of the growing LPS polymer. Regardless of the biosynthesis pathway, the completed O-antigen polysaccharide is attached to the flipped LPS core before the LPS is transported to the outer leaflet (42). A pictorial overview of the synthesis of the different surface glycans can be found in Figure 1.2. A more detailed discussion of these two pathways (as well as the sole example of a synthase-dependent pathway), can be found in the review by Whitfield et al. (43)

2.2.2 Lipoteichoic acids and wall-teichoic acids

In contrast to Gram-negative bacteria, Gram-positive bacteria have a single cell membrane that is enveloped in a thick layer of peptidoglycan that provides the structural integrity for the cell. The surface glycans found on these bacteria are teichoic acids (TA), which contain a linker portion as well as long polymer regions consisting of a) phosphodiester linked polyol repeats of glycerol phosphate or ribitol phosphate, b)

glycosyl-phosphate polymers, or c) glycan-polymer backbones modified with phospho-moieties (eg phosphocholine, phosphoethanolamine, and phosphoglycerol) (44). The teichoic acids can be divided into two groups, wall-teichoic acids (WTA) and lipoteichoic acids (LTA), which differ in their attachment to the bacterium. For wall TAs, the linker and long polymer are assembled onto UndPP in the cytosol, before the complete teichoic acid is flipped to the cell surface and attached to a MurNAc residue within the peptidoglycan (45). For LTAs on the other hand, six distinct types have been described based on their structure as well as different synthesis pathways. The two best studied LTA types are type I and type V, with the latter primarily being found in *Pneumococcal* species (46). While the assembly of type V LTA is similar to many other surface polysaccharides and is executed sequentially and in the cytosol, type I LTA are assembled outside the bacterium, where the linker is added onto the lipid anchor followed by polymerization of the polyol repeats (46, 47). Additional information on WTAs, can be found in the review by Brown et al. (44), and the six LTA types that have been identified to-date have been summarized in a review by Percy and Grundling (46) and Wenzel et al. (48).

2.2.3 Capsular polysaccharide

The outermost structure found on many bacteria is the capsule, or capsular polysaccharide (CPS). This long polysaccharide structure is made of repeating glycan units that are linked to the cell surface. CPS can be highly diverse even within a single strain. These homo- and heteropolymers differ in their glycan composition, their linkages, and decoration with a wide array of non-stoichiometric modifications (eg. acetylation, phosphate and phospho-linked modifications, and O-methylation, etc.), leading to the

occurrence of over 80 distinct capsular serotypes of *E. coli* and over 97 serotypes in *S. pneumoniae* (49, 50) alone. The genes responsible for the synthesis of the CPS are usually organized in discreet clusters within the genome. While very different in composition, the basic steps of capsule synthesis are similar in most bacteria. The first glycan moiety is attached to a lipid carrier on the cytosolic side of the cell membrane, this can be molecules like UndPP (also known as bactoprenyl-pyrophosphate) or diacylglycerol (DAG) which are often found in Gram-negative and Gram-positive bacteria (51, 52). The following attachment of glycans is executed sequentially by different glycosyltransferases. The completed repeating unit is either a) flipped across the membrane, where the repeating units are polymerized in a Wzx/Wzy-dependent manner through addition of repeats at the reducing end of the polysaccharide (as described for LPS), or b) further extended in the cytoplasm at the non-reducing end and translocated to the outer surface by an ABC transporter, or c) synthesized and transported through the use of a synthase complex. The CPS is then transferred to either the outer membrane in Gram-negatives or the peptidoglycan in Gram-positive bacteria (51, 52). Detailed discussion of the aforementioned methods of CPS biosynthesis are discussed in more detail in the review by Whitfield et al. (53).

Bacterial capsules are widely recognized as virulence factors, contributing to their pathogenesis during infection in a variety of ways. Some of the advantages provided by the CPS lie in desiccation protection, host or surface adherence, and resistance to the host immune system (50). Due to the inherent chemical composition of the polysaccharides, capsules attract water molecules and are highly hydrated. This external water reservoir enables bacteria to resist desiccation and survive outside their host for elongated times (54,

55). The location of the CPS as the outermost structure of most bacterial cell walls often makes the capsule the first structure of the bacterium that is detected by its environment. Thus, it plays a role during attachment to both biotic and abiotic surfaces. One way to establish initial contact with host cells is via binding of the capsule to host cells, followed by multifactorial adhesion and potential biofilm formation (56). An area in which the CPS fulfills a wide array of functions as a virulence factor is in evasion of the host immune response. The best-studied example of this is the evasion of the complement pathways, whereby this thick and branched layer of polysaccharide serves as a physical barrier, shielding the bacterial surface from recognition by host proteins and decreasing complement deposition (57, 58). Lastly, any surface-bound complement proteins are shielded from recognition by phagocyte receptors by the CPS, reducing the risk of opsonophagocytosis (59).

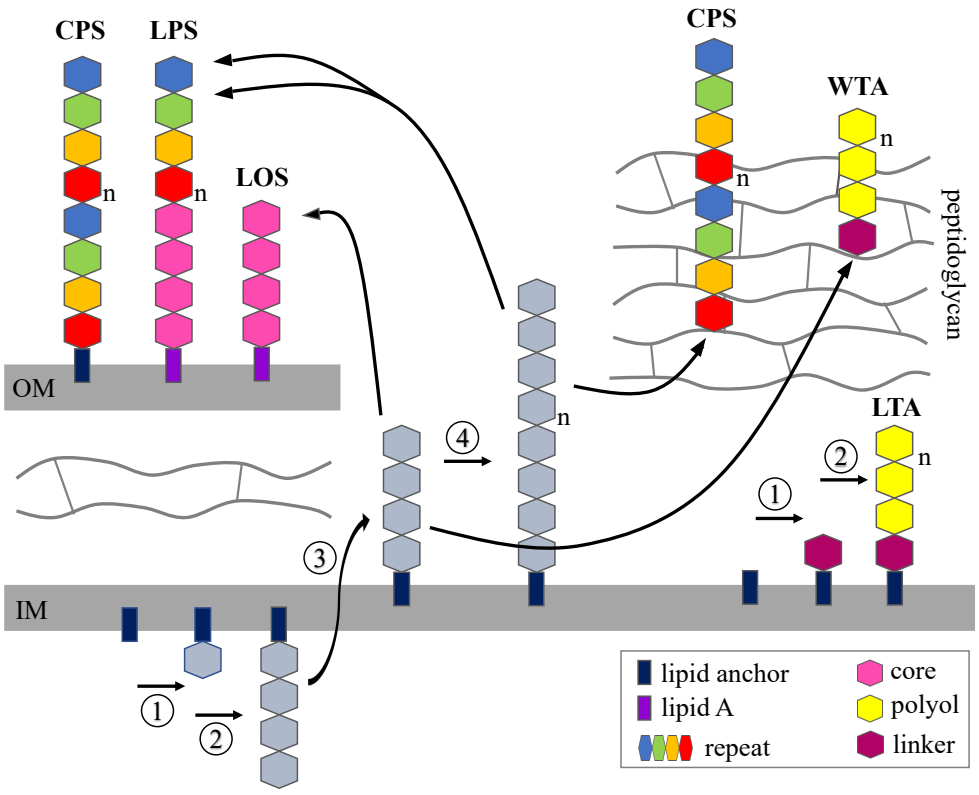


Figure 1.2: Synthesis of bacterial surface glycans. The left side of the image depicts the surface structures found in Gram-negative bacteria and the right-side shows that of Gram-positive bacteria. The starting glycan is transferred onto the lipid carrier (1) followed by the sequential synthesis of the complete glycan repeating unit by glycosyltransferases (2), this process takes place in the cytosol for CPS and LOS. In the case of LPS, the wzy dependent biosynthesis pathway is depicted and for LTA the synthesis of an LTA type I is shown for which the entire glycan assembly occurs outside the cell. The oligosaccharide is flipped across the membrane (3) where potential polymerization occurs (4). Depending on the organism and its polysaccharide the polymerization in step (4) can either be achieved through the addition of new repeating units to the non-reducing end of the growing polysaccharide or through the transfer of the growing polysaccharide chain to the non-reducing end of the newly synthesized repeat unit. Lastly the different glycan structures are transferred to their designated positions on the cell surface. CPS = capsular polysaccharide, LPS = lipopolysaccharide, LOS = lipooligosaccharide, WTA = wall teichoic acid, LTA = lipoteichoic acid.

3. Glycan-binding proteins

Glycan binding proteins can be divided into three different groups, glycan-binding antibodies, lectins, and glycosaminoglycan (GAG) binding proteins (35). Glycan binding

antibodies strictly follow the same rules as all antibodies: they are Y-shaped and recognize glycan moieties as the specific antigen epitopes and are part of an immune response. Both lectins and GAG binding proteins are their own classes of specialized glycan-binding proteins. GAGs are long and highly sulfated polysaccharides that interact with proteins and are major components of the extracellular matrix. The most well-studied GAGs are heparan sufate and hyaluronic acid. The binding of GAGs by GAG-binding proteins is achieved mostly through surface charge interactions between the protein and glycan without the presence of a defined glycan binding pocket. Lectins, on the other hand, are able to bind a wide variety of glycans while exhibiting a shallow, but defined binding pocket (35).

To date, over 15 lectin families are known, which are spread through all kingdoms and found in animals, plants, bacteria, fungi and viruses (60). Lectins were first discovered in plants, like the R-type lectin (known as ricin) in 1888, which can be found in the castor beans of *Ricinus communis* and exhibits a toxic function by inhibiting protein biosynthesis in mammalian cells (61). Ricin, like many of these early discovered lectins, share a function that is common to a wide variety of lectins: agglutination. The ability of lectins to agglutinate proteins, cells and bacteria is possible, in large part, due to the oligomeric/multimeric nature of many lectins, which conveys a high binding avidity in comparison to the low monomer affinity due to the ability to multivalently bind multiple targets, which leads to agglutination. The glycan-binding site in lectins follows the lock-and-key mechanism, with a shallow, but defined binding pocket whose conformation shows only slight changes upon ligand binding. Many lectins additionally require the presence of a divalent cation in their binding site, which directly coordinates hydroxyl groups of the target carbohydrate (60, 62). Due to their ability to recognize glycans

classified as pathogen-associated molecular patterns (PAMPs) or danger-associated molecular patterns (DAMPs), lectins are often referred to pattern recognition receptors (PRRs).

3.2 Immune lectins

Since lectins have been found in all domains of life, a variety of functions have been associated with them, ranging from P-type lectins that direct newly-synthesized lysosomal enzymes to the lysosome to C-type lectins, many of which are involved in mammalian immune processes (60, 62). In order to keep this review concise, it will focus on lectins that play important roles in the human immune response. The following section will therefore showcase how a variety of human lectins execute one of the most vital functions during the immune response: distinguishing between self and non-self through the recognition of microbial surface glycans.

3.2.1 C-type lectins

The largest family of lectins are the C-type lectins, whose name is derived from the fact that many lectins in this family require coordination of calcium ions in their binding sites. These glycan binding proteins feature at least one C-type lectin-like carbohydrate-binding domain for the recognition of mostly mannose-type or galactose-type glycans (63), and can be secreted proteins or transmembrane lectins that serve as receptors (64). One of the most prominent functions found in C-type lectins is the role they play in the mammalian immune response, wherein they are often expressed directly by immune cells or recruit other immune proteins to evoke an inflammatory and antimicrobial cellular response (64).

While many of the well-known immune lectins are soluble, an example of transmembrane lectins are the macrophage inducible C-type lectins (Mincles). Mincle expression is induced upon cytokine release during infection in a variety of immune cells, including macrophages, monocytes, dendritic cells and neutrophils (65–67). The recognition by Mincles has been shown for various bacterial species including binding to the lipoteichoic acid anchor of group A *Streptococcus.* Upon binding through the Mincle extracellular carbohydrate-binding domain, the intracellular tail of the Mincle then undergoes phosphorylation, recruiting a kinase, which in turn activates a signaling cascade that triggers an immune response (68).

The mannose-binding lectin (MBL) is likely the best-known and well-studied C-type lectin member belonging to the collagen-containing C-type lectins, or collectins. This soluble protein is most active when organized as multimers of homotrimers, which recognize glycans containing D-mannose, L-fucose and N-acetylglucosamine in a calcium ion-dependent manner (69). As the initiating protein of the lectin complement pathway, MBL recognizes and binds to glycans on the surface of microbes. This initial binding then recruits the MBL-associated serine proteases (MASPs), which constitute the start of the complement cascade in which over 20 immune proteins lead to the assembly of a transmembrane pore, lysing the recognized bacterium (70).

3.2.2 Ficolins

It had long been believed that the only lectin able to activate the lectin complement pathway was MBL, but in recent years it has been shown that at least two other collectins are able to activate the pathway, as well as a different family of lectins known as the

ficolins. Collectins and ficolins are structurally similar, with the subunits of their trimers possessing a cysteine-rich N-terminus, a collagen-like domain, a coiled-coil neck domain and a globular carbohydrate recognition domain (Figure 1.3) (71). The N-terminus and the collagen-like domain mediate oligomerization and engage in interaction with MASPs, which activate the complement cascade (72). Additionally, both require calcium ions for glycan recognition. What distinguishes ficolins from collectins is their carbohydrate recognition domain, which in the case of ficolins is fibrinogen-like (71).

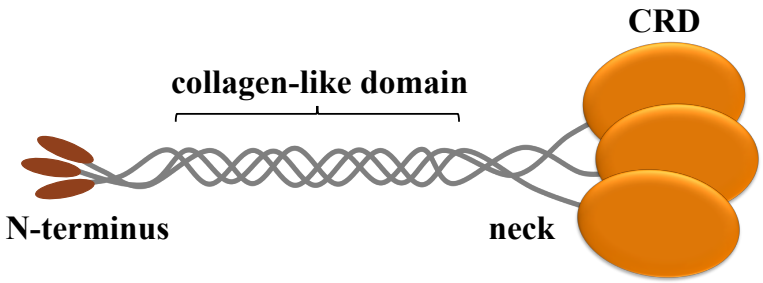


Figure 1.3: Schematic of the structural components of collectin and ficolin. The carbohydrate recognition domain (CRD) in collectins is a C-type lectin like CRD, while ficolins have a fibrinogen-like motif in their CRD.

Since lectins have different glycan binding specificities, the complement-activating collectins and ficolins target different pathogens. An example of this is the human ficolin-2, which recognizes over 10 *Streptococcus pneumoniae* serotypes, including serotype 11A (9). Ficolin-2 binds to the surface glycans on serotype 11A, which leads to both activation of the complement pathway and opsonization that triggers phagocytosis of the bacterium. Serotype 11A shows a comparably low invasive potential in human infection compared to the 17 other serotypes that were assessed, such as serotypes 1, 5, 7, 14 and 18 C. This outcome is partially due to its specialized recognition by the innate immune response through ficolin-2 (9, 73). Apart from *S. pneumoniae* serotype 11A, bacterial recognition by

ficolin-2 has been reported for other bacteria, such as *Pasteurella pneumotropica*, enteropathogenic *Escherichia coli* and *Mycobacterium tuberculosis* (74, 75).

3.2.3 Galectins

While collectins and ficolins eliminate microbes by recruiting a cascade of other immune proteins, a more direct mechanism of bacterial elimination is observed with another family of lectins – the galectins. A total of 15 members of the galectin family have been identified in mammals, eleven of which are present in humans (62). Galectins feature a β-sandwich or jelly-roll configuration in their highly conserved carbohydrate recognition domain. They bind to galactose-containing glycan structures in a calcium-dependent manner and are known for their hemagglutinin activity (62). In humans, galectins 3, 4 and 8 have been shown to recognize blood group antigens A and B, which are comprised of carbohydrates on the membrane of red blood cells, and all three galectins bind to *Escherichia coli* 086, which mimics the blood group B antigen as part of its surface glycans (76). Furthermore, galectins 4 and 8 have been shown to kill *E. coli* 086 directly and independently of complement through interference with motility and disruption of the membrane integrity (76). The mimicry executed by *E. coli* 086 enables it to remain undetected in an individual of the blood group B type who do not possess antibodies against blood group B antigens (77, 78), which makes the specialized recognition by the galectins even more essential to the mechanisms of immune defense. Additionally, the galectins 3, 8 and 9 were shown to activate autophagy in damaged vesicles. Healthy vesicles display their membrane carbohydrates exposed to the lumen, this changes in damaged vesicles, where glycans are flipped and exposed to the cytoplasm. When this change occurs, the

galectins can recognize the exposed glycan structures and trigger autophagy of the vesicle (79, 80).

3.2.4 RegIII lectins

While some lectins, like the ficolins or galectins exhibit a narrow glycan-binding specificity and only affect a small range of bacteria, others have a much broader spectrum of glycan recognition. This is true for the RegIII lectins, which are part of the C-type lectin family. This lectin is one of five Reg proteins known in humans and recognizes Gram-positive bacteria via their peptidoglycan structures (81). In addition to its inherent qualities as a lectin, the mouse lectin RegIIIα, as well as its human counterpart HIP/PAP, has been shown to form pores in the membrane of bacteria, leading to death through osmolysis. The lectin is expressed and stored in its inactive form, in Paneth cells in the small intestine. This inactive form contains a pro-peptide at the N-terminus that inhibits pore formation, and upon release from the Paneth cell the pro-peptide is cleaved from RegIIIα, which binds to bacteria and directly kills them (82). The expression and release of this lectin are triggered by the colonization of the gut epithelium by bacteria. The Paneth cells sense the increased microbial-epithelial contact through their toll-like receptors and release RegIIIα (81, 83). This reveals the role of lectins not only in the defense against pathogens, but against all bacteria in general, as RegIIIα contributes to the sequestration between the intestinal epithelium and microbiota and participates in the maintenance of a symbiotic host-microbe relationship (83).

This review of human immune lectins illustrates the different courses of action lectins have to achieve the common goal of defending the host against invading microbes.

Lectins like ficolin-2 require the assistance of the complement cascade, while galectin-4 kills bacteria directly. Some lectins exhibit a very narrow binding specificity and only bind a limited number of bacteria, while others target entire groups of microbes. As exemplified by RegIIIα, lectins do not distinguish between pathogens and commensals, but rather recognize any microorganism that presents a threat to the host under a specific circumstance.

3.2.5 Siglecs

Antibodies and lectins can conceptually be grouped together as immune molecules that can establish the initial contact between host and invading microbes, build a physical barrier and act as opsonins. This resemblance is especially true for siglecs; these "immunoglobulin-type" (I-type) lectins are a subset of the immunoglobulin superfamily (84, 85). Siglecs are membrane-bound and contain an N-terminal immunoglobulin domain that recognizes sialic acids. A large number of siglecs have been discovered, nine of which are present in humans and are expressed primarily on blood and immune cells (among others such as B cells, basophils, eosinophils, macrophages, monocytes, neutrophils, natural killer cells, oligodendrocytes and trophoblasts) (86). Each siglec is unique in its sialic acid recognition preferences, showing specificity for different linkages and neighboring glycan moieties (87, 88). As is true for most glycan-binding proteins, these lectins demonstrate increased glycan binding avidity when present in clusters or multimers.

Sialic acids are commonly present on various mammalian cell surfaces and, as a result, are involved in a wide range of immune regulatory functions (86), including apoptosis, adhesion, cellular activation and proliferation, and cytokine release (89–91) .

These lectins have also been shown to mediate both signaling mechanisms and cell-cell interactions, including host-host and host-microbe contacts.

4. Human Intelectin-1

4.1 History

There is also a new family of lectins that has been discovered in the last 25 years, known as the X-type lectins or intelectins. The first member of the X-lectin family, XL35 (92), was found in the South African toad *Xenopus laevis*, and its function initially described in 1974, followed by later isolation and characterization (93). This lectin exists as a glycosylated homotrimer, composed of 35 kDa monomers and binds to D-galactopyranosides in a calcium-dependent manner (92, 94, 95). The XL35 intelectin is utilized during fertilization of the toad's eggs to prevent polyspermy. *X. laevis* eggs are covered in the easily-penetrable extracellular vitelline envelope, but upon fertilization, XL35 is released from the cortical granules of the oocytes into the space between the egg and the vitelline envelope (93, 96, 97). There the lectin binds to its ligands on the proteins of the surrounding jelly coat, crosslinking macromolecules and forming the fertilization membrane/barrier which shields the egg from polyspermy (93, 94, 98). Afterward, XL35 expression levels stay high during embryo formation until the tadpole stage (92).

4.2 Structure

Since the discovery of XL35, members of the X-lectin family have been found in the human, mouse, lamprey trout, ascidian worm, sheep and other species (99). Many of these proteins require calcium ions for glycan recognition, but none contain a C-type lectin-

like carbohydrate recognition domain. Instead, they share a conserved fibrinogen-like domain that, in contrast to the ficolins, is not the carbohydrate recognition domain (99). The most studied, yet still poorly understood X-lectin is human intelectin-1 (hIntL-1). It is a secreted glycoprotein of 295 amino acids with a molecular weight of 40 kDa per monomer (100) (35 kDa without glycosylation) and 74 % amino acid similarity to the original X-lectin XL35 (101). The active form of hIntL-1 is a disulfide-linked homotrimer with three calcium ions complexed within each monomer. X-ray crystallography has shown that two of these calcium ions are embedded in the protein structure, while the third is surface exposed and part of the carbohydrate-binding site (102). The glycan-binding site is shallow and surface exposed, containing five conserved amino acid residues that coordinate with the calcium ion and two water molecules to bind the carbohydrate ligand. The conserved features involved in glycan binding start at the very end of the fibrinogen-like domain but lie mostly outside of it (102) (Figure 1.4). The fibrinogen like domain, as well as the N-terminus, are not involved in the carbohydrate recognition and are available for protein-protein interaction (103).

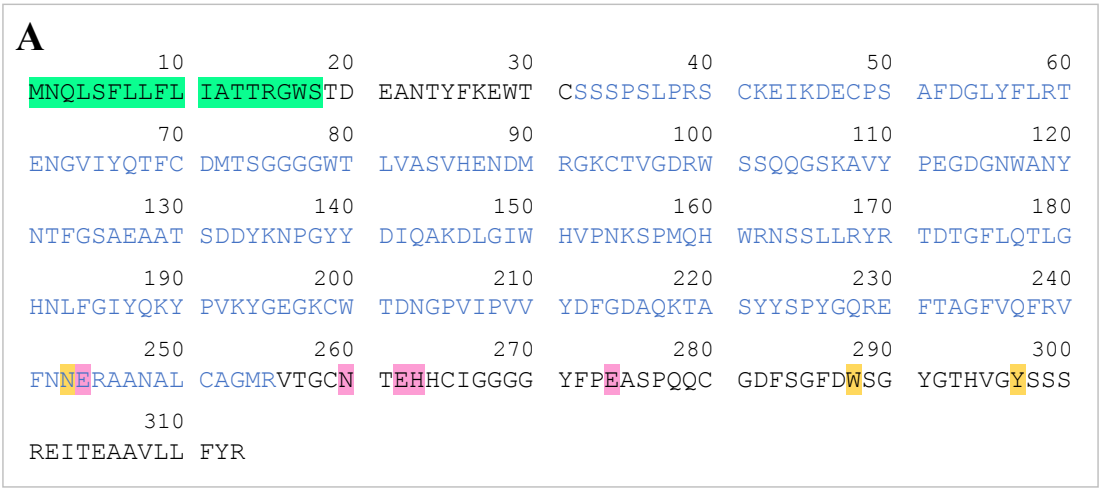


Figure 1.4: Amino acid sequence of hIntL-1. The amino acid sequence of hIntL-1 including the 18 amino acid long signaling peptide (green). The fibrinogen-like domain is

shown in blue letters and the amino acids directly involved in coordinating the calcium ion or the carbohydrate ligand are marked in pink. Additionally, other amino acids that are part of the binding pocked are marked yellow. (UniProt.org)**.**

4.3 Glycan-binding specificity

The glycan-binding specificity of hIntL-1 was first investigated using the Consortium for Functional Glycomics glycan arrays to allow for screening of a wide variety of glycans. While the lectin was shown not to recognize any mammalian glycans, it bound to a variety of microbial glycans (102). The microbial glycan array included over 300 lipopolysaccharides isolated from 12 different pathogens, and in this array hIntL-1 recognized LPS containing β-linked D-galactofuranose (β-Gal*f*), D-phosphoglycerol-modified glycans (Gro-1-P), heptoses, D-*glycero*-D-*talo*-oct-2-ulosonic acid (KO) or 3-deoxy-D-*manno*-oct-2-ulosonic acid (KDO) (102). The common feature among these glycans is the presence of an exocyclic 1,2-diol, which was confirmed as the recognized moiety via X-ray crystallography, in which it was visible that the two hydroxyl groups of the vicinal diol are coordinated by the calcium ion in the binding site (Figure 1.5). McMahon et al. in the same research group later discovered that out of these five ligands for hIntL-1, only β-Gal*f*, Gro-1-P, KO and KDO were recognized with high affinity while D-heptose and L-heptose were bound weakly due to stereoelectronic effects caused by the preferred conformation of the exocyclic 1,2-diol in these two sugars. This study also showed that the binding affinity of hIntL-1 to its favored ligands is increased if the glycan moieties are present in a multivalent form (eg. β-Galf-β-Galf-β-Galf-β-Galf$_n$) (104). Interestingly, sialic acid (α-Neu5Ac), which is widespread in humans (35), was not recognized by hIntL-1 although it displays an exocyclic 1,2-diol. The example of sialic acid demonstrates the limit of our knowledge regarding the binding specificities of hIntL-

1. While the glycan moiety recognized by the binding site of the lectin is an exocyclic 1,2-diol, not every glycan with this moiety is bound. For sialic acid, computational analysis (Figure 1.5) showed anion-anion repulsion of sialic acid by the carboxyl group of glutamic acid 274 in the binding site of hIntL-1 (modeling data generated in cooperation with Dr. Robert Woods, 2019) (102).

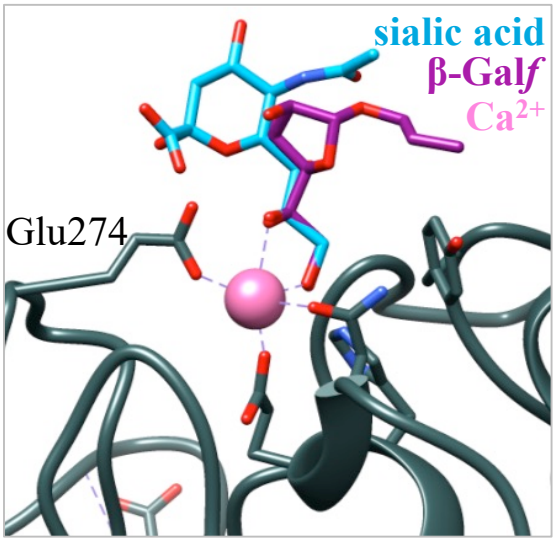


Figure 1.5: Computational model of hIntL-1 in interaction with sialic acid and β-galactofuranose. Sialic acid (Neu5Ac) was superimposed onto a β-galactofuranose (β-Galf) residue within the binding pocket of hIntL-1 (grey). Oxygen atoms are depicted in red, nitrogen atoms in blue. The image was generated using the UCSF Chimera software.

4.4 Bacteria bound by hIntL-1

The first bacterium shown to be bound by hIntL-1 was *S. pneumoniae* (102). This bacterium is a Gram-positive coccus that colonizes the nasal cavity of up to 45 % of adults and 90 % of children, is the leading cause of pneumonia in children, and also causes meningitis, sepsis, bacteremia and otitis media (105, 106). The bacterium exhibits over 97 different serotypes that are distinguished by the composition of their capsular polysaccharides (49). Using fluorescent microscopy, binding of labeled hIntL-1 was

detected to *S. pneumoniae* serotypes 20, 43, and 70. Serotype 20 and 70 display β-Gal*f*, while the CPS of serotype 43 contains Gro-1-P as the proposed exocyclic 1,2-diol containing recognition sites for hIntL-1. Serotype 8, which does not display an exocyclic 1,2-diol in its CPS, was used as a negative control and was not recognized by the lectin (102).

Since the first discovery of the interaction between hIntL-1 and *S. pneumoniae*, several other human pathogens have also been shown to be bound by hIntL-1. *Vibrio cholerae*, the causative agent of the potentially lethal diarrheal disease cholera, was shown to be recognized by the lectin via a simple co-incubation, followed by washes and detection through western blotting. The same group also demonstrated the recognition of the Gram-negative bacteria: *E. coli*, *Vibrio parahaemolyticus* and *Salmonella enterica* as well as the Gram-positive bacteria: *Listeria monocytogenes* and *Staphylococcus aureus* (107). Another bacterium, *Helicobacter pylori*, was found to be recognized in 2019, along with the first discovery of an effect on the microbe carried out by hIntL-1 (108). *H. pylori*, a risk factor for gastric cancer, was found to co-localize with the intelectin in the gastric glands of the intestine and to be recognized by it. While no killing effect of the bacteria could be determined, microscopy showed concentration- and calcium-dependent agglutination of the bacteria in the presence of hItnL-1, and this agglutination lead to increased phagocytosis of *H. pylori* (108).

4.5 Expression

hIntL-1 is expressed by several cell types in many different tissues in the healthy human body, as well as in the serum, where it is present at a concentration of 0.37 µg/ml

(109). Washimi et al. (2012) did an extensive analysis of hIntL-1 expression in human tissues (110), and showed that the highest expression level is found in goblet cells in the digestive system; more specifically the stomach, duodenum, small intestine, and colon, where the intelectin can be localized along the enterocyte brush border (110, 111). Mesothelial cells in the mesothelium and greater cementum, as well as umbrella cells in the bladder, produce hIntL-1 at a moderate level, while low-level expression was found in collecting tubule cells in the kidney, in cortex cells of the adrenal glands, and bronchial mucus glands in the bronchus (110, 112, 113). Other research groups also found expression of intelectin in the heart, thymus, ovary, and testis (101), as well as in basal stem cells (108), and in omental adipose tissue, where it is expressed by stromal vascular cells (114).

Several research groups have examined expression levels of hIntL-1, to verify whether the lectin is constitutively expressed or is reliant on induction through unknown factors. While very little research focused on this intelectin in interaction with human tissues, this work showed that the presence of IL-13 increases hIntL-1 expression by goblet cells in bronchial epithelial cells (115). This dependency on IL-13, as well as Stat6, a pathway activated by IL-13, has since been confirmed in human goblet-like colon adenocarcinoma cells as well (J. Viola, M. Pierce, unpublished). Additionally, hIntL-1 was shown to be upregulated during overexpression of Rspo3 in human basal stem cells (Rspo3 mediates the Wnt pathway) (108). Other work has focused on intelectins native to other mammals, such as sheep and mice, and while these results can be used as a guideline for hIntL-1, it is unclear how well they correspond to the processes in humans. It was shown that intelectin expression in the mouse lung is induced by Stat6 (116) and by IL-4 in the sheep trachea (117). Additionally, in both sheep (118) and mice (119), upregulation of

intelectin has been observed during parasite infections of the lung and intestine. Collectively, these studies have revealed IL-13, Stat6 and Rspo3 as regulators for hIntL-1 expression and, furthermore, they showed that while the expression of the lectin can be induced, it is also present at low levels in different tissues in the absence of these inducers. This suggests a low constitutive expression of hIntL-1 as well as need-controlled induction, although the possibility of additional as-of-yet unidentified inducers cannot be ruled out.

4.7 hIntL-1 as the lactoferrin receptor

Although hIntL-1 has been found to be a factor in several diseases as well as an adipokine involved in many pathophysiological processes, not many proteins are known to directly interact hIntL-1 (AMPK, P13K, NOX and the Akt pathway during inflammatory disease or diabetes (120–123)). This makes it all the more intriguing that hIntL-1 has been shown to bind the multifunctional and antimicrobial protein lactoferrin. Lactoferrin is part of the transferrin family and was first discovered in bovine milk in 1939 (124, 125). This 80 kDa glycosylated protein binds reversibly to iron with a high affinity and can be found in most human secretions, including saliva, tears, mucus and urine (126–128). Lactoferrin is involved in regulating iron absorption, but also has major antimicrobial activities and is considered part of the human innate immune response (129, 130). The antimicrobial function of this protein can mostly be attributed to two mechanisms. The first involves sequestration of iron by lactoferrin at sites of infection, thereby depriving microbes of the essential nutrient (131, 132). Secondly, the protein can interact directly with microbes via charged amino acids, causing a multitude of effects like the release of lipopolysaccharides from the outer membrane or interference with biofilm formation by the bacteria (133, 134).

The binding between hIntL-1 and lactoferrin has been shown via a number of techniques by different groups and earned the lectin the name lactoferrin-receptor (135, 136). The interaction between these two proteins is most likely to be charge-based and was not inhibited by the addition of hIntL-1 ligands, suggesting that its glycan binding site is not involved in the binding of lactoferrin (137). Some groups have also tried to show killing effects of hIntL-1 on bacteria in the presence of lactoferrin, but have failed (102). Nevertheless, lactoferrin is one of the few human immune proteins that have been found to bind hIntL-1 for unknown reasons.

4.8 hIntL-2, the ortholog of hIntL-1

While the studies regarding hIntL-1 and its function have only just begun, even less is known about its human orthologue, hIntL-2. The two intelectins share 91 % amino acid identity (138) and both are encoded on the same locus on chromosome 1q23 (101). hIntL-2 is 325 amino acids in length and is expressed and secreted exclusively by Paneth cells in the crypts of the small intestine where the stem cells reside (101). The only work done on the binding specificity of this lectin is yet unpublished and was conducted by the research group of Dr. Michael Pierce supported by computational analysis from Dr. Robert Woods. Their data showed that hIntL-2 exhibits a very narrow binding specificity that overlaps with that of hIntL-1, it recognizes *S. pneumoniae* serotype 43 and *Proteus mirabilis* O54a, 54b, both of which display a glycerol-1-phosphate moiety in their CPS. While Gro-1-P is a glycan moiety hIntL-1 recognizes as well, hIntL-2 was not able to bind to other exocyclic 1,2-diol containing ligands of its homolog, like β-Gal*f* (J. Viola and M. Pierce, manuscript in preparation). In accordance with this, three amino acid changes in the binding sites of

the two lectins were found. The glutamic acid 274 in hIntL-1 is replaced by glutamine 274 in hIntL-1 which both form an equivalent hydrogen bond with the ligand. Secondly, the binding pocket shows a change from tyrosine in hIntL-1 to serine in hIntL-2 for position 297, but neither of these amino acids directly interact with the carbohydrate or the calcium ion. Lastly, hIntL-2 contains glutamine in position 279 forming an additional hydrogen bond with the ligand, while the corresponding amino acid in hIntL-1 is not part of the glycan-binding site (Figure 1.6) (modeling by Dr. Amika Sood and Dr. Robert Woods). The limited information accessible for hIntL-2 so far demonstrates an overlap of its binding specificity and site of expression with hIntL-1. The current data suggests a specialization of hIntL-2 for the regulation of specific microbes present in the crypts, thereby protecting the stem cell niche from invasion.

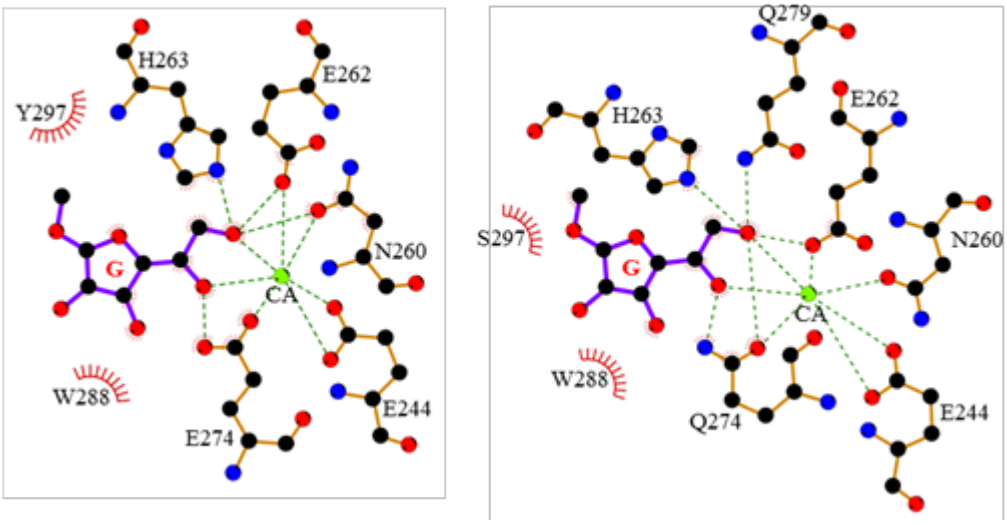


Figure 1.6: Ligplot contact analysis of hIntL-1 (left) and hIntL-2 (right) bound to β-Gal*f* (G). Shown in black are carbon atoms, red are oxygen atoms and blue are nitrogen atoms. The central calcium ion is shown in green and direct interaction between atoms is indicated by dashed lines. (By Amika Sood in the research group of Dr. Robert Woods)

5. The *Streptococcus* genus

Streptococci are Gram-positive cocci or ovoids that are nonmotile, do not form spores and are catalase negative. Contradictory to their name, the bacteria do not exclusively display chain arrangements, but also commonly occur as diplococci. As many *Streptococci* exploit human or animal hosts, they thrive best at 37 °C and are generally facultative anaerobes (139). *Streptococci* are heavily linked with pathogenicity due to some of the genus' most prominent members, such as *S. pnuemoniae* and *S. pyogenes*, also called Group A *Streptococcus*. Despite this infamous reputation, *Streptococci* comprise a major component of our commensal microbiota, inhabiting mucosal membranes of various tissues, as well as the skin and even breast milk (140). Just a few of these species include *S. oralis*, *S. parasinguinis* and *S. mitis*, which inhabit the oral cavity, as well as *S. salivarius* and *S. thermophilus*, which can be found on gut mucosal surfaces (141, 142). The oral cavity and its adjacent tissues are niches commonly colonized by *Streptococci*, and many of their inhabitants have the potential for opportunistic pathogenicity (140). Diseases caused by classified pathogenic strains, however, are accompanied by more intensified symptoms due to their multitude of virulence factors that are specialized for evasion of human immune responses. These infections tend to originate in a specific tissue, but have the potential to spread and become systemic (143).

5.1 Classification of *Streptococci* through surface glycosylation

As is typical for Gram-positive bacteria, *Streptococci* are enveloped in a thick layer of peptidoglycan that is considered the primary component of the bacterial cell wall. In addition to peptidoglycan these bacteria usually display at least one other type of surface

glycan structures such as CPS, WTA, LTA and rhamnose-glucose polysaccharides (RGP) (144). Only a few species throughout the genus are known to produce a CPS, such as the hyaluronic acid capsule of *S. pyogenes* or the many serotype-specific CPS of *S. pneumoniae* (140). Surface glycans that are bound to the peptidoglycan however are very common, but highly diverse within the genus and even among certain species. Due to the physical characteristics of the different polysaccharides that have been found attached on the peptidoglycan, these structures cannot all be grouped under the same name. *S. pneumoniae*, for example produces typical WTA, while *S. pyogenes* is decorated with "Group A carbohydrate", and *S. mutans* displays RGP (144, 145). Due to this diversity, the classification of *Streptococci* is not straightforward. Nevertheless, a common theme within the genus are polysaccharides that are attached to the cell wall and contain a rhamnose backbone, such as in *S. pyogenes* and *S. mutans*. This feature has served as the main criterium for classification of *Streptococci* prior to genome sequencing, alongside of biochemical tests for catalase activity, hemolytic activity and the like. The Lancefield classification of *Streptococci* separates Streptococci into serological groups dependent on their surface glycans and was introduced by Rebecca Lancefield in 1933 (146, 147). During the classical experiment to determine the Lancefield grouping of a *Streptococcus* sample, the bacteria are mixed with specific antisera from animals immunized with formalin-fixed cultures and monitored for agglutination of the bacteria via microscopy. With this method, *Streptococci* are classified into 20 Lancefield groups, designated A-H and K-V (148). This is also the origin of the name Group A S*treptococci*, used to refer to *S. pyogenes*, which reacts with the Lancefield Group A antiserum. The best-studied Lancefield Groups include A (*S. pyogenes*), B (*S. agalactiae*), C (*Streptococcus equisimilis* and others) and G (*S.*

anginosus). Some *Streptococcus* species however are not as easily grouped by Lancefield and are therefore considered the viridans Group. *S. pneumoniae,* for example, does not react with the Lancefield antisera (148). Other species, such as *S. mutans, S. suis* and *S. salivarius*, are simply too diverse due to the display of a wide variety of different polysaccharide antigens depending on the strain. Some of these bacteria can be classified by Lancefield antisera strain by strain; for example, the serotype c *S. mutans* strain XC corresponds to Lancefield Group C (145).

5.2 Protein O-glycosylation in *Streptococci*

While both protein N-glycosylation and O-glycosylation have been described in other bacteria, only O-glycosylation has been observed within the genus *Streptococcus*. The attachment of glycans to serine or threonine residues in *Streptococci* has been shown for various proteins and the glycosylation machineries of several species have been described in detail. This overview will largely be focusing on the best-studied O-glycosylation systems of *S. pneumoniae* and *S. parasanguinis*. In *S. pneumoniae*, the serine rich repeat protein PsrP is glycosylated with GlcNAc from a nucleotide-activated donor by the glycosyltransferases GtfA/B in a cooperative mechanism (36). This is followed by the addition of glucose residues by Gtf3, creating a glucosylated protein that is recognized by a variety of glycosyltransferases. Therefore, in the following glycosylation steps, an additional glucose can be attached by $GlyD_{DUF1792}$ or GlyG, or the addition of galactose can be executed by GlyE or $GlyD_{DUF1792}$ (the DUF1792 domain of GlyD can execute transfer of both glucose and galactose). Similarly, further glucosylation (by $GlyD_{DUF1792}$ or GlyG) or galactosylation (by $GlyD_{DUF1792}$, GlyE, GlyA or $GlyD_{GT8}$) can occur creating four

types of tetrasaccharide chains on PsrP. Following this sequential glycosylation of the protein using exclusively UDP-activated sugar donors, PsrP is transported out through its own dedicated secretion system (Figure 1.7A) (36).

In *S. parasangionis*, the O-glycosylation of the serine rich repeat protein Fap1 is initiated in the same manner as shown for *S. pneumoniae*. The enzymes GtfA/B attach GlcNAc to Fap1, thereby utilizing GtfA as a glycosyltransferase and GtfB to interact with the acceptor protein (149, 150). In the next step glucose is added by GtfC, followed by the generation of a branching point in the glycan structure through the addition of Glc and GlcNAc by dGT1. Lastly the glucose branch is fucosylated by GalT2 and the GlcNAc branch is elongated by the addition of glucose through Gly. The protein is then secreted to the cell surface where it is anchored to the cell wall (Figure 1.7B) (149, 150).

Notably, both *Streptococcal* O-glycosylation systems employ sequential glycosylation (rather than *en block* attachment), use UDP-activated donor glycans and target serine rich proteins that are secreted to the cell surface.

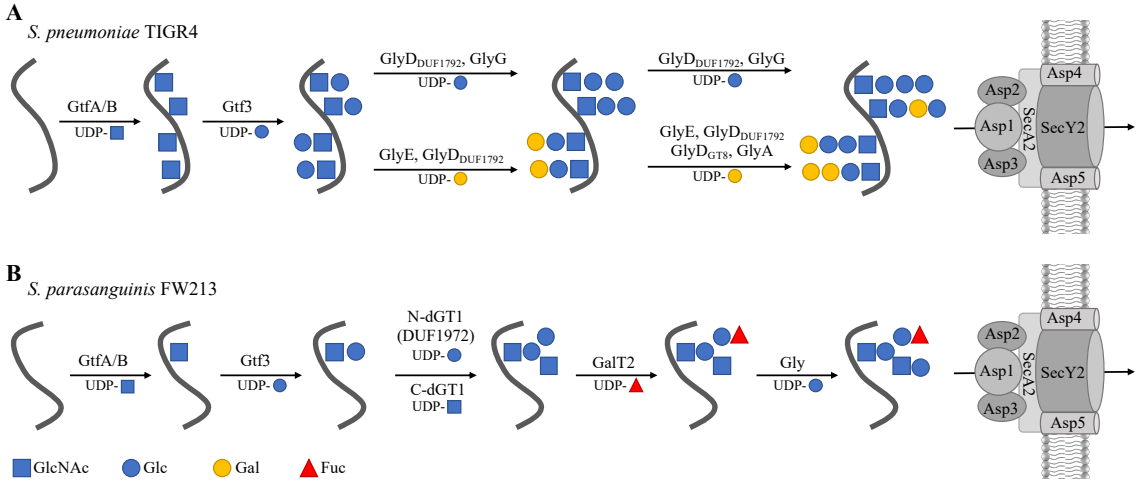


Figure 1.7: Protein O-glycosylation in *S. pneumoniae* and *S. parasanguinis*. The serine rich repeat proteins PsrP and Fap1 are glycosylated in *S. pneumoniae* TIGR4 **(A)** and *S. parasanguinis* FW213 **(B)** respectively, followed by transport to the cell surface through a dedicated SecA2/Y2 secretion system (36, 150).

5.3 *Streptococcus pneumoniae*

5.3.1 Epidemiology

S. pneumoniae is a Gram-positive coccoid bacterium that is enveloped by a capsular polysaccharide (CPS). The species *S. pneumoniae* is subdivided into serotypes, which are differentiated via differences in CPS structure. To date, over 97 different serotypes have been identified (49). The serotypes differ in their geographical prevalence, the disease caused, and the age group affected, which constitutes one of the major obstacles in the fight against this pathogen (49). *S. pneumoniae* colonizes the nasal tract of 1-10 % of adults and up to 95 % in children (106). This pathogen can cause meningitis, sepsis, bacteremia, otitis media and pneumonia, which are commonly referred to as invasive pneumococcal diseases (IPD) (105). According to the Centers for Disease Control (CDC), *S. pneumoniae* constitutes the leading cause of pneumonia, bloodstream infections, meningitis and middle ear infections in young children (https://www.cdc.gov/pneumococcal/about/infection-types.html). In 2011, around 4 million people were infected with invasive pneumococcal diseases in the United States, among them mainly children, elderly and immunosuppressed patients (151).

5.3.2 Infection

During an infection, the human host has defenses designed to clear pathogens from its tissues, and *S. pneumoniae* employs a variety of specialized mechanisms to evade these clearance approaches. To help protect against attack by the host innate immune system, the first layer of defense employed by *S. pneumoniae* is its negatively charged CPS which

serves as a major factor in evading complement (see: 5.3.4 Evasion of the complement pathway). Additionally, as part of the adaptive immune response, IgA antibodies recognize pathogens and induce endocytosis of the opsonized particle by binding to the membrane polymeric immunoglobulin receptor (PIGR) (152), but these antibodies are degraded by pneumococcal proteases secreted by the bacterium (Figure 1.8) (153). The thick mucus layer of host airways also represents an obstacle that immobilizes bacteria and contains many antimicrobial proteins to attack the invader and, once trapped in the mucus, the bacteria are transported out of the body through the mucociliary system (153). To combat this, *S. pneumoniae* expresses different enzymes that deglycosylate the mucins in the mucus layer and inhibit ciliary beating, thereby exposing the host epithelial surface. To enter the host cells, the pathogen displays phosphorylcholine (ChoP) and the choline-binding protein A (CbpA) on its surface, which can bind to the host cell via the platelet-activating factor receptor (PAFR) and PIGR respectively (152). Both receptors enable the bacterium to be endocytosed into the host epithelial cell, followed by intracellular migration across the cell and invasion into other tissues. In addition to receptor-mediated endocytosis, *S. pneumoniae* also produces pneumolysin and hydrogen peroxide, which damage the epithelium and allow for paracellular invasion through the space between the host cells (153).

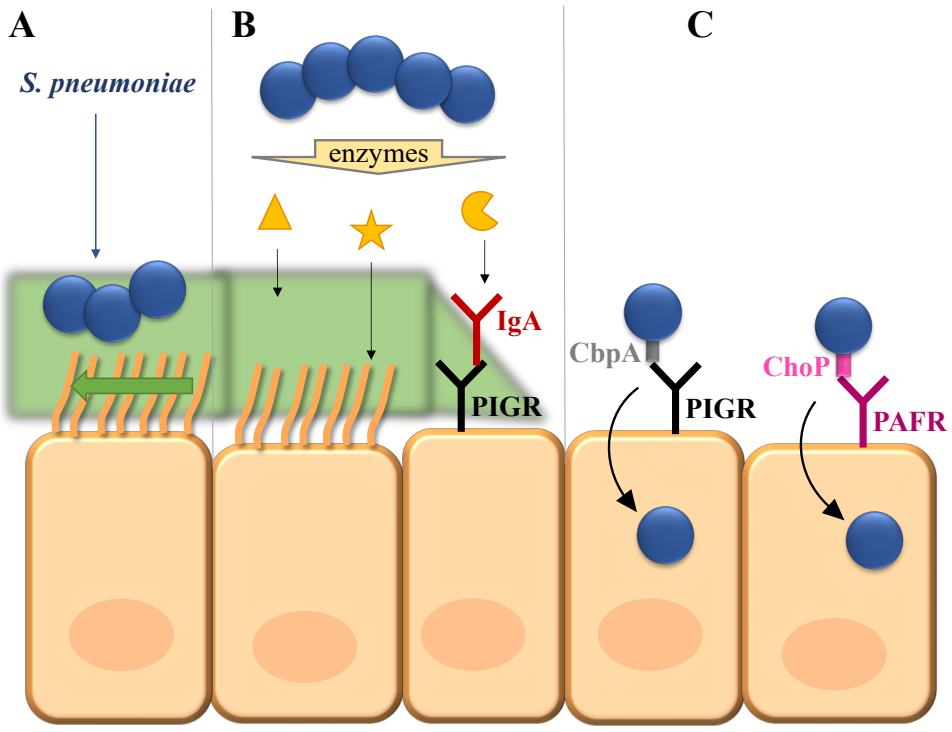


Figure 1.8: Model of human airway infection with *S. pneumoniae*. A) When *S. pneumoniae* enters the airway, it is trapped in the mucus layer (green) lining the epithelial cells and is then transported out of the body through the mucociliary transport system of cilia and microvilli. **B)** The host's immune defenses are evaded by *S. pneumoniae* through a variety of expressed enzymes that deglycosylate (triangle) the mucins, inhibit the movement (star) of the cilia and degrade (Pac-Man) host antibodies. **C)** To invade the host cells, *S. pneumoniae* binds to two different membrane receptors, which mediate endocytosis of the pathogen.

5.3.3 Capsule

5.3.3.1 History and background

One of the major virulence factors of *S. pneumoniae*, which enables it to invade human tissue and evade phagocytosis, is its capsular polysaccharide (154). The CPS is also the distinguishing factor used by scientists to differentiate the pneumococcal strains into serotypes. In the US, this work began during WWII, in which the newly identified serotypes were simply numbered based on their date of discovery (155). Simultaneously similar work was executed in Denmark, where researchers named serotypes according to

their serogroups, which are composed of all serotypes sharing the same antibody cross-reactivity (49). This led to the two *S. pneumoniae* serotype nomenclatures used today, the US and the Danish version, the latter of which is typically preferred. At the time of their discovery, serotyping was carried out using the quelling reaction. Bacterial strains were mixed with rabbit antisera, causing visible agglutination of the pneumococci due to the binding of the multivalent antibodies to the CPS (156). Today, the categorization of *S. pneumoniae* is realized through a variety of methods. MALDI-TOF and especially NMR are the most commonly used biochemical approaches (157), and genome sequencing-based methods are also commonly used. The latter includes PCR with specialized primers (158, 159), microchip assays with DNA probes and whole-genome sequencing (49). Combined this work has led to the discovery of over 97 serotypes of *S. pneumoniae* subdivided into 48 serogroups (49). A comprehensive list of all known serotypes in both nomenclatures can be found in Table 1.1.

5.3.3.2 Synthesis

The CPS of all *S. pneumoniae* serotypes is built from repeating units of 2-7 glycan residues, and is in most cases negatively charged (160). Among all the serotypes, only two biosynthetic pathways have been described for the CPS structures to date. With the exception of serotypes 3 and 37 (161), all other serotypes follow the Wzy-dependent mechanism of CPS synthesis, in which only the genes differ that are responsible for order, type and linkage of the assembled carbohydrates, leading to the wide variety of CPS structures. The genes for this CPS synthesis mechanism are arranged in a single operon located between the genes *dexB* and *aliA* (162). Using serotype 11A as a representative of

the Wzy-dependent pathway, CPS synthesis begins on the cytoplasmic side of the bacterial cell membrane (49). As is common during polysaccharide synthesis, all sugar moieties are transferred from their nucleotide-activated forms. In this manner, glucose-1-phosphate is transferred from UDP-glucose onto an undecaprenyl-phosphate acceptor that is part of the inner leaflet of the cell membrane. This glucose-PP-Und serves as the first moiety of the repeating unit, also known as the initiating sugar, onto which the following glycan structures are built by sequential addition of sugar residues by a variety of specific glycosyltransferases. Once one repeating unit is complete, it is modified with acetyl groups by acetyltransferases before it is flipped to the outer leaflet of the cell membrane by the flippase, Wzx. There, the polymerase, Wzy, transfers the previously made polymer chain from its undecaprenyl-pyrophosphate carrier to the non-reducing end of the newly synthesized repeating unit (demonstrated in *S. aureus* and *S. enterica*) (52, 163). Lastly, the newest repeating unit is further acetylated through the transmembrane O-acetyltransferase WcjE (Figure 1.9) (49).

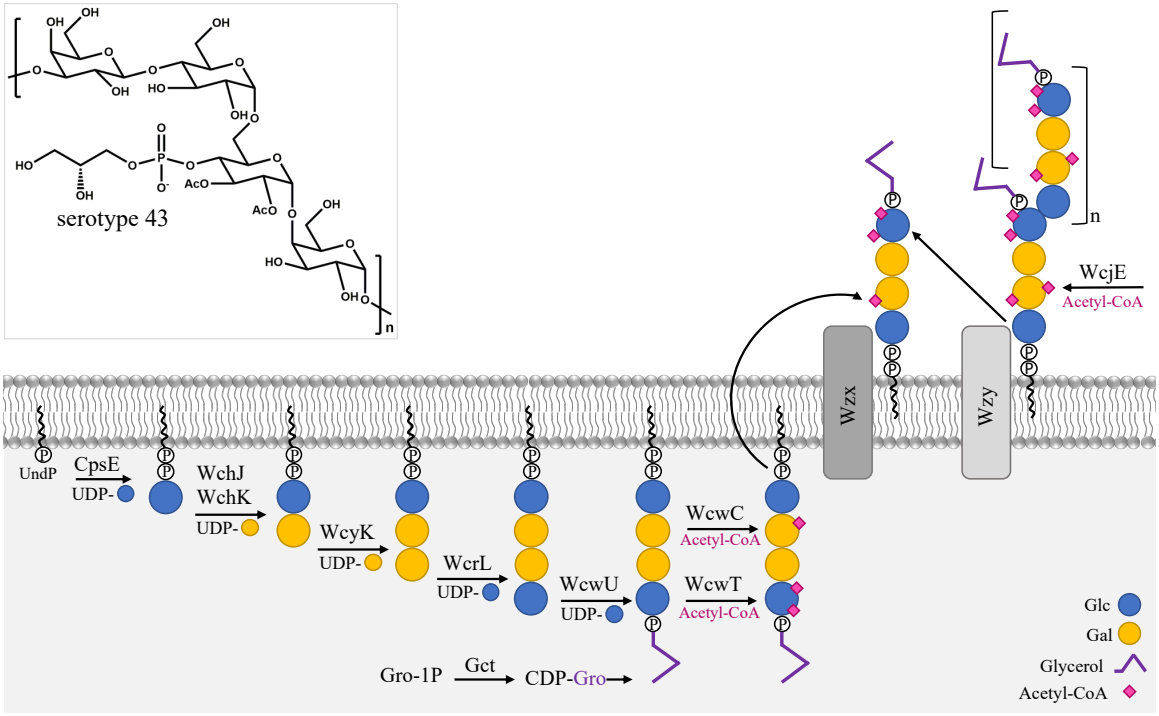


Figure 1.9: CPS structure and synthesis of *S. pneumoniae* serotype 43. In *S. pneumoniae* serotype 43, the CPS is sequentially assembled. onto an undecaprenyl-phosphate lipid carrier. After the glycan transfer from UDP-activated donors, the repeating unit is decorated with glycerol-phosphate and acetyl groups before being flipped to the cell surface and polymerized with further CPS repeats (49).

In addition to being the defining feature of the multitude of different *S. pneumoniae* serotypes, the CPS is also responsible for another phenotype. Dependent on phase variable regulation of the capsule genes, the pneumococcal colonies can appear in either an opaque or transparent form (164). The transparent form was shown to exhibit a thinner CPS layer when associated with the colonialization of the nasopharynx (165, 166). The opaque form, on the other hand, results from a thicker CPS layer that is typically observed in the planktonic state and is generally more virulent (165–167).

Table 1.1: Overview of all *S. pneumoniae* serotypes including the US and Danish nomenclatures and information regarding CPS structure.

USA	Denmark	USA	Denmark	USA	Denmark	USA	Denmark	USA	Denmark
1	1	21	21	41	**34** Gal*f* 0.5, Rib	61	**35C** Gal*f*	81	44
2	2	22	**22F** Gal*f*	42	**33B** Gal*f**, Rib	62	**35A** Gal*f**	82	48
3	3	23	23F	43	**11A** Gro1P	63	22A	83	12A
4	4	24	24F	44	**18A** D-Gro1P	64	23B	84	**47A** Rib
5	5	25	25F	45	40	65	24A	6C	6C
6	**6A** Rib	26	**6B** Rib	46	23A	66	**35B** Gal*f**, Rib	6D	6D
7	7A	27	27	47	35A	67	32A	ND	6E
8	8	28	28F	48	7B	68	9V	6F	6F
9	9N	29	**29** Gal*f**, Rib	49	9L	69	**39** Gal*f* 0.4, Rib	6G	6G
10	**10F** Gal*f*	30	15A	50	7C	70	**33F** Gal*f*	6H	6H
11	**11F** Rib	31	**31** Gal*f*	51	7F	71	38	ND	**10B** Gal*f*, Rib
12	12F	32	32F	52	**47F** Rib	72	**45** Gro1P	ND	**10C** Rib
13	**13** Gal*f*, Rib	33	9A	53	**11C** Gro1P	73	46	ND	**11D** Gro1P
14	14	34	**10A** Gal*f*, Rib	54	15B	74	**41A** Gal*f*	ND	**11E** Gro1P
15	15F	35	**35F** Gal*f*, Rib	55	**18B** D-Gro1P	75	43	ND	12B
16	16F	36	36	56	**18C** Gro1P	76	**11B** Rib	ND	16A
17	17F	37	37	57	19A	77	15C	ND	**20B** Gal*f*
18	**18F** Gro1P	38	**41F** Gal*f*	58	19B	78	**17A** Gal*f*	ND	25A
19	19F	39	**33C** Rib	59	19C	79	28A	ND	**33D** Rib
20	**20A** Gal*f*	40	33A	60	24B	80	42		

The *S. pneumoniae* CPS were listed based on the publications from Mavroidi et al., 2007 and Geno et al., 2015 (49, 168). Indicated with red background in the US nomenclature are the serotypes that have been tested negative for hIntL-1 binding by the group of Dr. Michael Pierce (unpublished) or Wesener et al., 2015 (102), Green background indicates positive binding of hIntL-1 by Dr. Michael Pierce (unpublished) or Wesener et al., 2015 (102) (all green or red serotypes were assessed by Dr. Michael Pierce, Wesener et al.,2015 confirmed serotypes 8, 20, 43, 70). Yellow background in the Danish nomenclature indicates the serotypes of the PCV13 vaccine. Serotypes with an exocyclic 1,2-diol in their CPS structure are shown in bold with the indication if they contain the diol in a glycerol-1-phosphate (Gro1P), a galactofuranose (Gal*f*), or a ribitol (Rib). The denoted 1,2-diol can have the following variations: asterisk* indicating that the 1,2-diol of the galactofuranose is only free on the last (outermost) repeating unit, the number following Gal*f* specifies the percentage at which the Gal*f* residue and its 1,2-diol are free and Gro1P stands for D-glycerol-1-phosphate and is indicated when the specific stereo structure of the glycerol is known.

5.3.4 Evasion of the complement pathway

As different as the CPS of the many pneumococcal serotypes are, they all provide essential protection from the human immune system. During *S. pneumoniae* colonization

and infection, the major mechanism of bacterial clearance by the host immune system is phagocytosis by neutrophils following detection and opsonization by complement (8, 169). As a result of this, *S. pneumoniae* is highly specialized to evade detection and killing through the complement cascade. Due to the thick peptidoglycan layer inherent to the Gram-positive pneumococci, the bacteria are largely protected against the effects of the membrane attack complex as the final step of the complement (170). Therefore, the opsonizing effects of many cascade proteins, and subsequent phagocytosis are responsible for the clearing of the pneumococcal pathogens (171, 172). In the evasion of these effects, the CPS plays a major factor in inhibiting the classical and alternative pathway, since their initiating proteins target the cell surface. The thick layer of CPS makes it challenging to penetrate the surface glycans of the bacterium for the antibodies as well as C3b (154). Even if those pathogen recognition molecules bind to their targets on the cell surface, they are occluded by the CPS and cannot be detected by host cell receptors. It has additionally been shown that the capsule inhibits the degradation of C3b, preventing the progression of the complement cascade (154).

However, the capsule of *S. pneumoniae* is not the only means by which the bacterium promotes evasion of the complement pathways; the bacteria also express ~15 pneumococcal proteins that target different cascade and immune proteins (173), the most prominent and well-studied include the pneumococcal surface proteins A and C (PspA and PspC respectively). PspA, which is present in all *S. pneumoniae* serotypes, limits complement deposition on the cell surface and has additionally been shown to limit complement deposition by binding to the phosphocholine moieties of the bacterial cell wall, thereby blocking them from recognition by host immune proteins (174–176).

Similary, PspC interacts with a variety of immune proteins, such as factor H and the complement protein C3 and it has further been shown to inhibit the assembly of the membrane attack complex (177, 178). Combined, the CPS and the pneumococcal defense proteins provide a multitude of protection mechanisms to the bacterium against the human immune response and complement by specifically targeting the focal points between the three pathways (173).

In addition to the advantages conveyed by the capsule, *S. pneumoniae* also benefits from its unique cell morphology. While most Streptococci occur as long chains of coccoid bacteria, clinical *S. pneumoniae* isolates are typically found as diplococci or small chains (169). It was shown that the pathogens benefit from this untypical arrangement since increased chain length enhances the susceptibility of the bacteria to complement deposition, opsonization and neutrophil uptake. The presentation as short pneumococcal chains conveys a competitive advantage during infections *in vivo* (169).

5.3.5 Vaccines

Due to its high degree of immune evasion, *S. pneumoniae* has been one of the major pathogens of our time, being responsible for an estimated 14.5 million cases of IDP and 825,000 deaths in 2000 in the US (179). Accounting for the bulk of these numbers are infections in children under the age of 5, due to the child's immature immune response and high colonization rate. Furthermore, individuals over 65 and immunosuppressed patients are highly affected by the pathogen. Today, infection rates have decreased significantly – from 14.5 million cases to approximately 4 million annual cases already in 2011 (151) – due to the introduction of pneumococcal conjugate vaccines. One of the earliest

pneumococcal vaccines, brought onto the market in 1983, was the pneumococcal polysaccharide vaccine PPSV23, which is a non-conjugate vaccine containing the purified CPS of 23 of the most virulent *S. pneumoniae* serotypes (180). While the vaccine was efficient in adults, it did not show protection in children under the age of two, due to the inability of non-conjugated carbohydrate structures to elicit the T-cell dependent immune response needed for protection in children (181–183). As children are one of the major risk groups for IPD, the first pneumococcal glycoconjugate vaccine, PCV7, was introduced in 2000. This vaccine contained the CPS of the 7 most common IDP causing serotypes conjugated to a carrier protein (182). Presented in this manner, the carbohydrate portion of the vaccine elicits a T-cell dependent immune response needed for immunization in children (181). After vaccination, B-cells recognize the carbohydrate portion of the vaccine with their transmembrane receptors and internalize the glycoconjugate. Inside the B-cell endosome, both the carrier protein and the polysaccharide are processed into small fragments of either glycans, peptides, or glycopeptides. Only the glycopeptide fragments can elicit the desired immune response of the vaccine. The MHCII of the B-cell binds to the peptide portion and presents the conjugate to a T-cell, which recognizes the carbohydrate. This interaction then triggers the maturation of the B-cell into memory B-cells and effector B-cells, the latter of which produce antibodies specific against the carbohydrate portion of the vaccine (181). In 2013, 10 years after the introduction of PCV7, PCV13 was released and included CPS of the 13 most virulent serotypes. By 2015 the use of the PCV vaccines had yielded a 93 % reduction in pneumococcal diseases in children caused by the serotypes contained in the vaccines and a 64 % decrease in IPD in children overall (184). Similar success rates could be registered in adults and immunosuppressed

people since the introduction of the PPSV23 vaccine. Therefore the Advisory Committee on Immunization Practices (ACIP) recommends that children between 2 and 59 months, as well as immunocompromised patients of 6 years or older receive PCV13 as a routine vaccination (185). PCV13 and PPV23 are recommended for adults 65 years and older (186). But even after the PPV23 and conjugate vaccines have been on the market for 39 and 22 years respectively, IPD still presents a major threat to the general public health and, while the number of IPD caused by vaccine serotypes decreased significantly, the IPD cases caused by non-vaccine serotypes increased by 22 % over the last years (183). These problems can partially be attributed to limitations inherent in the construction of the PCVs, because the number of serotypes included in the vaccine is limited by its conjugation chemistry. Additionally, the obtained immune response is decreased by the carrier protein, which is commonly used for immunizations and elicits a low immune response in pre-exposed patients (187). However, the major complication in the fight of IPDs is presented by the pathogen itself. The many serotypes show varying virulence in different tissues, age groups, and global geographics, making it difficult to design a universal vaccine. Additionally, non-vaccine serotypes expand or change their CPS to be more virulent, which is respectively referred to as serotype replacement and serotype switching. Lastly, non-encapsulated *S. pneumoniae* strains, which are unaffected by the vaccines have been shown to cause infections in humans as well (188).

5.4 *Streptococcus mutans*

5.4.1 Epidemiology and Infection

Streptococcus mutans is a facultative anaerobe, Gram-positive, coccoid bacterium that inhabits the oral cavity of humans and causes dental caries. The pathogen is particularly well equipped to thrive in the oral cavity due to its ability to adhere to and colonize the hard tooth surface, metabolize a wide range of carbohydrates and thrive under environmental stress (189). Among the oral microbiota and its many streptococcal inhabitants, *S. mutans* is considered the major etiological agent for dental caries within the group of the tooth colonizers. The infectious disease of tooth decay, or caries, affects 60 – 90 % of the population worldwide (190). While oral infections caused by this pathogen are highly common, the organism can enter the blood stream during dental procedures and travel to the heart where it has been correlated with infective endocarditis (191, 192). *S. mutans* strains are categorized into four serotypes (c, e, f and k) based on the structure of their rhamnose-glucose polysaccharide (RGP) which is attached to the peptidoglycan, with 70 % of strains isolated from dental plaque belonging to serotype c, 20 % to serotype e, and the remaining 10 % evenly divided between the serotypes f and k (193). Due to this variability in surface glycosylation, *S. mutans* is considered part of the viridans group of Streptococci and as a species, does not fall into any Lancefield group.

5.4.2 Rhamnose-Glucose-Polysaccharide (RGP)

Gram-positive bacteria typically display WTA attached to their peptidoglycan; these sugar structures are considered acids due to the presence of negatively charged phosphate groups. Based to this definition, the carbohydrate structure produced by *S.*

mutans and attached to the peptidoglycan wall is not considered a WTA (194). *S. mutans* instead is decorated with neutral RGP consisting of a rhamnose backbone that contains glucose moieties with varying linkages that define the serotype of the strain. RGP have been shown to affect *S. mutans* tolerance to acid and oxidative stress, as well as biofilm formation and virulence (195). Responsible for the synthesis of RGP are the *rgp* genes, consisting of four glycan transferases and an ABC transporter, as well as the gene for rhamnose biosynthesis (dTDP-4-keto-L-rhamnose-reductase, *rmlD*). During RGP biosynthesis, the rhamnose backbone is assembled onto the lipid carrier bactoprenol, which is linked to GlcNAc-1-P by RgpG (Figure 10) (196, 197). Glucose residues are then attached to the rhamnose polysaccharide at different linkages depending on the glycosyltransferase present on the genome. While serotype k lacks a glucosyltransferase and is not modified with glucose, the proteins RgpE, RgpH and RgpI determine the linkage of glucose to rhamnose in the other serotypes. Serotype c displays an α-1,2 linkage, serotype e is modified with a β-1,2 linkage and in serotype f the glucose and rhamnose residues are connected by a α-1,3 linkage (193). The RgpC and RgpD proteins together form an ABC transporter that transports the RGP to the cell surface where its lipid carrier remains in the outer leaflet of the phospholipid bilayer and subsequently flipped back to the cytoplasmic face. In the final step, the completed RGP is covalently bound to the peptidoglycan by LCP proteins (LytR-CpsA-Psr family) (196, 197).

The impact of RGP as a surface glycan of *S. Mutans* becomes clear when RGP synthesis is disrupted. RGP mutant strains fail to localize cell division complexes properly and display defects in septum formation (198). Furthermore, the mutants display the formation of long bacterial chains, indicating an alteration in efficient cell

division (197). The latter can also be observed in *S. mutans* strains with deficient rhamnose biosynthesis (*ΔrmlD*) (199).

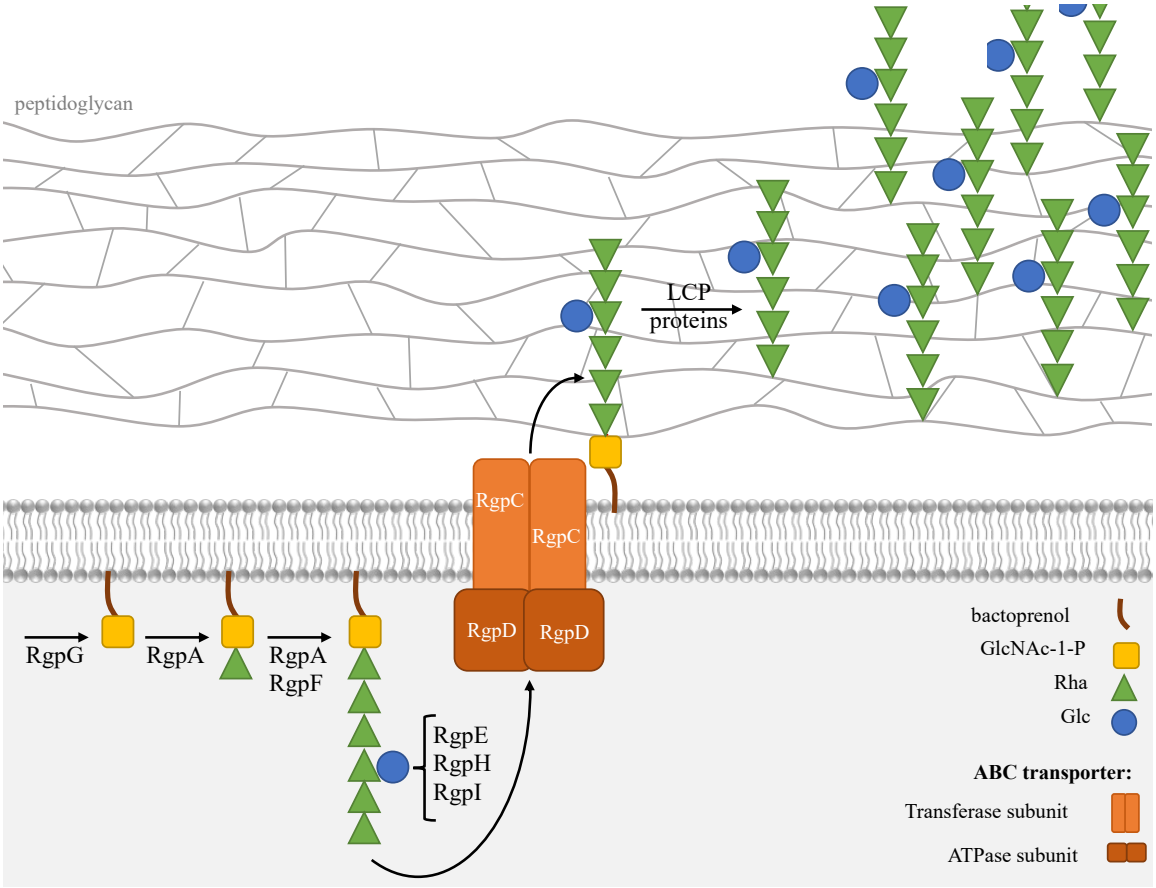


Figure 1.10: Schematic overview of the RGP structure and synthesis in *S. mutans*. The RGP in *S. mutans* is assembled by sequential addition of rhamnose residues onto bactoprenol and GlcNAc-1-P, some of whom are then decorated with glucose. Following glucose modification of the rhamnose backbone at different linkages, the RGP is flipped to the cell surface and then attached to peptidoglycan (197).

5.4.3 Virulence factor Cnm

S. mutans, as many other Streptococci and pathogens in general, possesses a wide variety of virulence factors. These enable the organism to thrive in its natural habitat, outcompete other microbes and evade the human immune response. The extracellular matrix of the oral cavity is comprised of proteoglycans, as well as the fibrous proteins collagen, laminin, fibronectin and elastin (200), and *S. mutans* has evolved a variety of

specialized proteins for the binding of one or several of these extracellular matrix components including the collagen binding proteins (CBP). This group of virulence factors includes SpaP, WapA, Cnm and Cbm, all of which are localized to the bacterial cell surface, where they are anchored via a C-terminal LPXTG motif (201). SpaP (also termed P1, PA, or antigen I/II), a factor in sucrose-independent biofilm formation, is a surface adhesin that specifically binds to collagen, laminin and fibronectin (201, 202). The wall associated protein A, WapA, undergoes proteolytic cleavage to generate antigen A (Aga), which has also been shown to bind collagen (203, 204). Both SpaP and WapA are part of the core genome of *S. mutans*, while either Cnm or Cbm are only encoded in 10-20 % of isolates (205, 206). Cnm is present in the vast majority of that subset, and while Cnm and Cbm demonstrate a 78 % amino acid sequence identity and both bind to collagen (207), Cnm has additionally been shown to bind laminin (208). Apart from the pervasive LPXTG cell wall-anchoring motif at the C-terminus and the collagen binding domain at the N-terminus, Cnm contains a signal sequence and large threonine rich repeating unit close to the C-terminus, containing 21 threonine rich repeats of the sequence TTTEAP, totaling 64 threonine residues (201).

The importance of Cnm in the initial phases of oral colonialization by *S. mutans* has been shown in a wide variety of assays. Impaired adhesion and invasion of oral and heart cells upon loss of Cnm has been shown in tissue culture experiments with human coronary artery epithelial cells, human gingival fibroblasts (HGF-1) and human oral keratinocytes as cell line models, as well as a number of *ex vivo* and *in vivo* studies (191, 209). Additionally, experiments using *ex vivo* tooth sections have shown reduced bacterial binding to dentinal and root tissue in the absence of Cnm, and a live rat model demonstrated

measurable decreases in *S. mutans* oral colonization and cariogenic potential (191, 209). Combined these data illustrate the importance of Cnm has for both initial binding of collagenous oral surfaces and subsequent implantation of dental biofilms by *S. mutans*.

When inspected for its neighboring genes in *S. mutans* OMZ175, *cnm* was found to be co-transcribed with the four *pgf* genes of the core genome: *pgfS*, *pgfM1*, *pgfE* and *pgfM2*. Deletion of the *pgf* genes *pgfS*, *pgfM1*, or *pgfE* results in a large drop in molecular weight of Cnm from 120 kDa to 90 kDa (with a predicted MW of 58 kDa), while deletion of *pgfM2* leads to a more moderate reduction in protein mass to 110 kDa. Cnm also becomes highly susceptible to degradation by proteinase K (ProK) in the absence of the *pgf* operon, but retains its collagen binding activity (204, 210).

In addition to these effects on Cnm, deletion of the *pgf* genes also has a drastic impact on the pathogenicity of *S. mutans*, was evidenced by the *Δpgf* strain displaying lowered binding to collagen, decreased invasion of human coronary artery epithelial cells (HCAEC) and an attenuated ability to kill in a *Galleria mellonella* model (204, 210).

5.4.4 The Pgf machinery

The role of the highly impactful *pgf* operon was quickly identified as that of a protein glycosylation machinery. As mentioned, studies showed a drop in the molecular weight of Cnm and a lowered resistance to ProK in the absence of the *pgf* genes, both of which are indicators for heavy glycosylation. Furthermore, the use of wheat germ agglutinin (WGA) detected the presence of GlcNAcylated proteins only in the WT, not in the *pgf* deletion strains (204, 210).

Apart from Cnm, which is only present in a subset of *S. mutans* strains, the core genome-encoded WapA was also identified as a target for Pgf glycosylation (204), and similarly showed both a reduction in molecular weight and heightened susceptibility towards ProK in the absence of the *pgf* operon. NetOGlyc 4.0 glycoprotein prediction identified the following surface associated *S. mutans* proteins as additional potential targets for Pgf glycosylation: GtfC, SpaP, PrsA, GbpC, GbpB and WapE , most of which are known for their involvement in early oral colonization, as discussed earlier (204).

The Pgf machinery likely catalyzes O-glycosylation of its target proteins. O-glycosylation has been described for bacteria in general and Gram-positive bacteria specifically where N-glycosylation has not yet been described. For the *Streptococcus* genus, protein glycosylation machineries have been found in several species, such as *S. pneumoniae*, *Streptococcus gordonii* and *S. parasanguinis* (36, 149, 211). In all three Streptococcal glycosylation machineries, the protein targets contain serine or threonine rich repeating units that are subject for O-glycosylation. Additionally, in Streptococci, only sequential transfer of glycans to serine or threonine residues has been described (ie. not transferred *en bloc*, as observed in *C. jejuni* for N-glycosylation). Both Cnm and WapA contain threonine rich regions as potential targets for the Pgf machinery, and sequential glycan transfer is suggested through the partially glycosylated Cnm and WapA intermediates found in the absence of the *pgfM2* gene.

The *pgf* operon consists of four genes arranged in the order *pgfS* (~1000bp), *pgfM1* (~3000 bp), *pgfE* (~1000 bp) and *pgfM2* (~3000 bp) (204). PgfS is proposed to be a membrane bound GT-A type glycosyltransferase with its catalytic N-terminal domain located intracellularly. Both PgfM1 and PgfM2 are integral membrane proteins of

unknown function, although PgfM2 has been implicated in the final steps of glycosylation based on incomplete glycosylation of Cnm when *pgfM2* is deleted. PgfE is predicted to be an intracellular 4-epimerase that functions as a homodimer and in NAD^+ dependent manner (204).

Research Aims

The genus *Streptococcus* is well known for its many pathogenic members, such as *S. pneumoniae*, which constitutes the world leading cause of bacterial pneumonia, and *S. pyogenes*, which is well known for causing the infectious disease "Strep throat" (https://www.cdc.gov/pneumococcal/about/infection-types.html). What makes many *Streptococcal* species so pathogenic is their ability to evade the human immune response. The bacteria have evolved a large array of mechanisms to outlast, block or inhibit the cellular and molecular strategies employed by the host to counteract them (170, 172). A prominent role during this molecular warfare is played by the glycans present on the bacterial surface. In this dissertation, the glycosylation present on the bacterial surface is studied from multiple angles in two prominent *Streptococcal* species. I have focused on (I) the interactions between the *S. pneumoniae* capsular polysaccharide and the human immune system, (II) the novel *S. mutans* Pgf protein O-glycosylation machinery and (III) the role of cell wall glycosylation during cell division in *S. mutans*. (I) In *S. pneumoniae*, the capsular polysaccharide (CPS) is well known to be a major virulence factor, protecting the bacterium from immune detection and aiding in host colonization (212). The human innate immune system, however, has evolved in parallel with its microbial intruders and found mechanisms to exploit the diversity of the pneumococcal CPS. Here we investigate

the interaction of hIntL-1, a poorly studied innate immune lectin, with different *S. pneumoniae* serotypes and different human immune components. (II) Long chain polysaccharide structures like CPS or teichoic acids are not the only glycans present on the bacterial surface, and glycosylated proteins on the surface of bacteria also play an important role during infection and colonization (204, 210). The Pgf protein O-glycosylation machinery in *S. mutans* has only recently been discovered, and neither the core features of its biosynthesis nor the individual roles of its four proteins have been characterized to-date (204, 210). I therefore aimed to not only identify the glycan moieties attached through the Pgf system, but also work towards understanding the biosynthetic pathway for this glycan, starting with characterization of the putative 4-epimerase PgfE. While working with PgfE, we further expanded our studies to the second 4-epimerase in *S. mutans,* GalE. (III) GalE, as a protein of the Leloir pathway for galactose metabolism, is essential for the production of glucose from galactose under glucose-deficient conditions (213, 214). Since glucose is a major building block of the peptidoglycan attached rhamnose-glucose polysaccharide that decorates the surface of *S. mutans*, we aimed to investigate the effects of GalE loss on cell division and cell wall biosynthesis of the pathogen.

I) Determine the interactions between the *S. pneumoniae* capsule and human immune components. It has previously been shown that some serotypes of *S. pneumoniae* are bound by hIntL-1, a lectin of the innate immune response (102). This project aimed to understand the interplay between binding of hIntL-1 to *S. pneumoniae* and the response of the human immune system to this pathogen. This will not only shed light on the function of a poorly characterized human protein, but further our understanding of the effect the

CPS of bacteria can have during infection and immune evasion. In particular, the aims were to:

a) Confirm the binding of hIntL-1 to *S. pneumoniae* serotype 43, and not serotype 8, using fluorescent microscopy and a round bottom well agglutination assay.
b) Evaluate whether hIntL-1 binding has any bactericidal or bacteriostatic effects on *S. pneumoniae*, either directly or in a complement-dependent manner.
c) Determine if hIntL-1 binding has any effect on the ability of *S. pneumoniae* to form biofilms and attach to A549 cells.
d) Assess the phagocytic ability of PBMCs towards *S. pneumoniae*, both in the presence and absence of hIntL-1 and serum.
e) Determine whether hIntL-1 binding of *S. pneumoniae* affects human neutrophil NET, ROS, and elastase production/release and/or their bacterial killing efficiency.

II) Identify the glycans added to target proteins by the Pgf O-glycosylation machinery of *S. mutans*. The Pgf system has been shown to have an impact on the virulence of *S. mutans* (204, 210). Cnm, the first known target of this putative glycosylation system, is the obvious candidate for characterizing the modification(s) added by the Pgf machinery. However, due to the presence of a large threonine rich repeat region in Cnm that has the potential to limit the information that can be obtained by MS analysis, the goal was to determine the identity of the modification(s) through a multi-faceted approach combining results from MS, lectin-blotting, and composition analyses of both natively-expressed full-length Cnm and an engineered Cnm with a shortened threonine-rich-repeat region,

III) Determine the substrate specificity and biological roles of the two *S. mutans* 4-epimerases. The function of the four proteins of the Pgf glycosylation machinery have not yet been characterized. PgfE is predicted to be a 4-epimerase, constituting the second protein of this function encoded in *S. mutans*, besides GalE (204). These experiments therefore aimed to characterize the first enzyme of the Pgf system, as well as compare its function and role in the organism to another *S. mutans* 4-epimerase, with the following specific aims:

a) Use computational modeling to compare the two 4-epimerases and predict their respective substrate specificities.
b) Determine the substrate specificities of recombinantly expressed and purified PgfE and GalE via enzymatic assays with UDP-glycans (monitored by NMR and capillary electrophoresis).
c) Verify the involvement of PgfE in protein glycosylation and test for the involvement of GalE by monitoring for molecular weight differences between anti-Cnm antibody-reactive bands from whole-cell lysates of wildtype vs 4-epimerase deletion mutants in immunoblots.
d) Confirm the role of GalE in galactose metabolism and assess the impact of PgfE by comparing growth of both wildtype and 4-epimerase mutants in chemically defined media containing only glucose or galactose as the predominant carbon source.
e) Assess the importance of the two 4-epimerases on cell wall biosynthesis and cell division, by analyzing both the cellular and cell-surface morphologies of *pgfE* and *galE* deletion mutants grown with galactose as the main carbon source (via both brightfield microscopy and transmission electron microscopy).

CHAPTER 2

HUMAN INTELECTIN-1 PROMOTES CELLULAR ATTACHMENT AND NEUTROPHIL KILLING OF STREPTOCOCCUS PNEUMONIAE IN A SEROTYPE-DEPENDENT MANNER[1]

[1]Silke Andresen, Kayla Fantone, Digantkumar Chapla, Balázs Rada, Kelley W. Moremen, Michael Pierce, Christine M. Szymanski. 2022. *Infection and Immunology* 90:5. Reprinted here with permission of publisher. Author contributions: SA and CMS designed the study with input from MP and BR. KF conducted the isolation of neutrophils and the operated the flow cytometer. DC and KWM executed the transformation of HEK-293F cells with pTracer for hIntL-1 expression and harvested the cells for protein purification. MP provided the pTracer plasmid for hIntL-1 expression and the β-galactofuranose agarose column for protein purification.

Abstract

Human intelectin-1 (hIntL-1) is a secreted glycoprotein capable of binding exocyclic 1,2-diols within surface glycans of human pathogens such as *Streptococcus pneumoniae*, *Vibrio cholerae*, and *Helicobacter pylori*. For the latter, lectin binding was shown to cause bacterial agglutination and increased phagocytosis, suggesting a role for hIntL-1 in pathogen surveillance. In this study, we investigated the interactions between hIntL-1 and *S. pneumoniae*, the leading cause of bacterial pneumonia. We show that hIntL-1 also agglutinates *S. pneumoniae* serotype 43, which displays an exocyclic 1,2-diol moiety in its capsular polysaccharide but is unable to kill in a complement dependent manner or to promote bacterial killing by peripheral blood mononuclear cells. In contrast, hIntL-1 not only significantly increases serotype-specific *S. pneumoniae* killing by neutrophils but also enhances the attachment of these bacteria to A549 lung epithelial cells. Taken together, our results suggest that hIntL-1 participates in host surveillance through microbe sequestration and enhanced targeting to neutrophils.

Introduction

Human immune lectins are known for their unique abilities to distinguish between self and non-self and play major roles during the innate immune response. During this response, lectins are used to initiate first contact with the invading microbes through recognition of specific carbohydrate structures. Subsequently, there are a plethora of effects the lectins can exert that lead to the elimination of the unwanted bacteria. For example, the binding of microbial surface mannans through the mannose-binding lectin (MBL), one of the best-studied lectins, initiates a cascade of binding events involving approximately 20

serum proteins prior to bacterial destruction by complement-mediated pore formation (69). The MBL and the later-discovered ficolin-2 are very similar in structure yet diverse in their target glycans (215). Nevertheless, the recognition of pathogens by either lectin leads to the recruitment of MASPs, which constitutes the activating step of the complement pathway (9, 69). Galectins are a distinct lectin family exhibiting a more direct mode of killing pathogens. These lectins are known for their agglutinating and membrane-disrupting activity through recognition of human blood group antigen-like structures. Galectins 4 and 8 specifically have been shown to bind the *Escherichia coli* strain 086, which mimics blood group B antigens on its surface (76). Lectin recognition causes interference with bacterial motility and disruption of membrane integrity (76). Apart from their killing mechanisms, lectins differ in their glycan binding specificity, playing different roles during the immune response. Both ficolin-2 and the galectins have very narrow carbohydrate specificities recognizing the O-acetyl group of carbon-6 of beta-galactose in *S. pneumoniae* 43 (also called 11A) and the blood group B antigen epitope displayed by *E. coli* 086, respectively (9, 76, 216). These ligands contribute to lectin specificity during the innate immune response, leading to recognition of their pathogen of choice down to certain serotypes within a species. A lectin with a much broader target range is RegIII or its human equivalent, HIP/PAP. RegIII recognizes peptidoglycan and forms pores in the membranes of Gram-positive bacteria (81, 82). This broad killing mechanism is associated with a role in host-microbe sequestration and general maintenance of homeostasis rather than specific pathogen surveillance and is likely the reason for the highly regulated manner of RegIII release and activation (81, 83). The mechanisms used by MBL/ficolin-2, the galectins, and RegIII to achieve bacterial clearance illustrate the

diversity among secreted human innate immune lectins. Apart from the different processes employed by lectins for bacterial recognition and killing, many lectins show commonalities based on their structural characteristics. For example, the multimeric assembly of active forms of most lectins leads to increased avidity for their targets, since individual carbohydrate binding affinities commonly involve weak interactions. These lectins can also act in the same fashion as antibodies, in which their binding leads to the agglutination of microbes that alone can act as a physical barrier around the bacteria, blocking their access to the host epithelium and facilitating the removal of bacterial aggregates in the mucosal layer through the action of cilia and microvilli (7, 12). Alternatively, lectin binding, similar to antibodies, can also opsonize bacteria for phagocytic uptake by neutrophils and macrophages as part of the inflammatory processes that are activated during immune responses (9).

One lectin that has been proposed to be an important member of the innate immune response is human intelectin-1 (hIntL-1). hIntL-1 is part of the X-lectin, or intelectin, family, whose first member was identified in 1974 in the African toad *Xenopus laevis* (92, 217). hIntL-1 is a glycoprotein that is expressed by several cell types in many different tissues as well as in serum, where it is present in micrograms-per-milliliter concentrations (121). The highest hIntL-1 expression levels originate from goblet cells in the digestive system (stomach, duodenum, small intestine, and colon), where it is secreted into the mucus (110). It can further be found in the mesothelium, bladder, bronchial mucus glands, kidney, heart, testis, and basal stem cells, making it widely distributed throughout the human body (101, 108, 110, 112, 218). In the airways, in particular, the presence of hIntL-1 has been studied in connection with asthma and smoking, where its expression can be stimulated by

interleukin-13 (IL-13) (115, 219). The active form of hIntL-1 consists of a disulfide-linked homotrimer complexing three calcium ions in each monomer, one of which is part of the glycan binding site and essential for ligand binding (102). The glycan binding specificity of hIntL-1 was originally analyzed by glycan arrays and showed that only microbial surface glycans, no mammalian glycans, are recognized (103). More specifically, β-linked D-galactofuranose (β-Gal*f*), D-phosphoglycerol-modified glycans (G-1-P), D-*glycero*-D-*talo*-oct-2-ulosonic acid (KO), and 3-deoxy-D-*manno*-oct-2-ulosonic acid (KDO) are bound by the lectin, all of which have an exocyclic 1,2-diol moiety in common, constituting the ligand for hIntL-1 (102, 104). Despite this specificity, it was found that sialic acid, a sugar common to mammals, is not recognized by hIntL-1 due to anion-anion repulsion of sialic acid by one of the carboxyl groups in the binding site (102). Similarly, heptoses possess exocyclic 1,2-diols but are only weakly bound by hIntL-1 due to stereoelectronic effects (104). This demonstrates the degree of specificity in glycan recognition that allowed intelectins to differentiate self and non-self in phyla from placozoa to chordata (99, 220).

The first bacteria shown to be bound by hIntL-1 were *S. pneumoniae* serotypes 20, 43, and 70, which contain either β-galactofuranoses or glycerol-1-phosphates in their capsular polysaccharide (CPS) repeating units (102). This discovery was closely followed by a publication that noted recognition of a variety of Gram-negative (*Vibrio cholerae*, *E. coli*, *Vibrio parahaemolyticus*, and *Salmonella enterica*) and Gram-positive (*Listeria monocytogenes* and *Staphylococcus aureus*) human pathogens by hIntL-1 (107). The only bacterium for which lectin effects on host biology have been described is *Helicobacter pylori*. The binding by hIntL-1 led to the agglutination of the bacteria in a calcium- and concentration-dependent manner as well as to an increase in phagocytic uptake by

macrophages (108). Notably, all of the bacteria bound by hIntL-1 are considered human pathogens, leading to the proposed role of hIntL-1 in pathogen surveillance.

Beyond the effects of hIntL-1 observed for *H. pylori*, it is unknown how hIntL-1 interacts with other microbes it recognizes or how the lectin protects from infection. We hypothesized that *S. pneumoniae* serotypes expressing exocyclic 1,2-diols would also be agglutinated by hIntL-1 and that this would be coupled to microbial clearance by host immune cells. Our results demonstrate that hIntL-1 specifically targets *S. pneumoniae* serotype 43 to polymorphonuclear leukocytes (PMNs) but not peripheral blood mononuclear cells (PBMCs) and also contributes to lung epithelial cell binding.

Results

hIntL-1 binds and agglutinates *S. pneumoniae* serotype 43.

hIntL-1 was purified from the supernatant (conditioned medium) of transfected HEK-293F cells using a β-galactofuranose agarose column. Subsequently, the elution fractions shown in Figure 2.1A were combined, concentrated, and used in the experiments described in this study. Previous studies have shown that hIntL-1 binds *S. pneumoniae* serotypes specifically (serotypes 20, 43, and 70), depending on their CPS structure (102). Therefore, we first verified this finding for *S. pneumoniae* serotypes 8 and 43 (Fig. 2.1B). The CPS repeating units of both serotypes are shown in Figure 2.1C. For serotype 8, the CPS consists of glucose and galactose repeating units that do not include any exocyclic 1,2-diols. In contrast, *S. pneumoniae* serotype 43 displays a glycerol-1-phosphate moiety in its CPS that is recognized by hIntL-1. Fluorescent microscopy with 488 Alexa-Fluor-labeled hIntL-1 showed lectin binding only to *S. pneumoniae* serotype 43

(Fig. 2.1B). As previously observed for *H. pylori*, microscopy further indicated that serotype 43 is agglutinated by hIntL-1 (108). Agglutination was also verified in a round-bottom well agglutination assay in which an agglutinating agent cross-links bacterial cells, acting as a spacer between them and preventing the formation of a small pellet that is characteristic of the lack of agglutination in a well. As shown in Figure 2.1D, the presence of hIntL-1 leads to bacterial agglutination (lack of pellet formation) in a concentration-dependent, serotype-specific, and calcium-dependent manner.

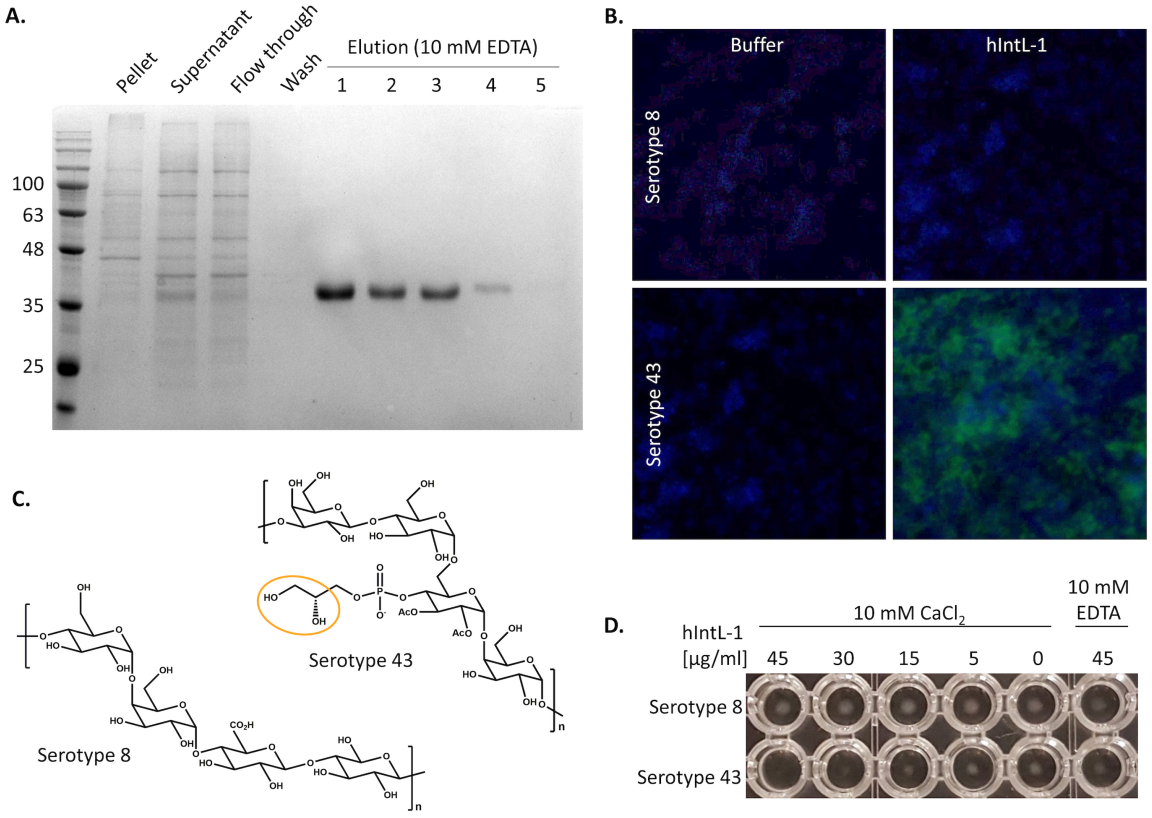


Figure 2.1: hIntL-1 binds and agglutinates *S. pneumoniae* serotypes specifically. (A) Coomassie-stained SDS-PAGE of the hIntL-1 purification. hIntL-1 was purified from the supernatant of transfected HEK-293F cells using a β-galactofuranose agarose column and eluted with EDTA. Monomeric hIntL-1 has a molecular mass of 35 kDa. Molecular masses, in kDa, are shown. (B) *S. pneumoniae* was labeled with 4',6-diamidino-2-phenylindole (DAPI, blue) and incubated with or without Alexa Fluor 488-labeled human intelectin-1 (hIntL-1, green) prior to viewing by fluorescence microscopy at ×40 magnification. (C) The capsular polysaccharide of *S. pneumoniae* serotype 8 consists of GlcA-1,4Glc-1,4Glc-1,4Gal-1,4 repeating units and does not contain an exocyclic 1,2-diol.

Serotype 43 displays the hIntL-1 ligand, glycerol-1-phosphate (yellow circle), that is part of the GlcNAc3Ac[-4Gro-1-P]-1,4Gal-1,3Gal-1,4Glc-1,6 CPS repeating unit. (D) *S. pneumoniae* agglutination assay in calcium-containing buffer using indicated concentrations of hIntL-1. EDTA was used to inhibit hIntL-1 as a control. The assay was performed in biological triplicate, and representative results are shown.

hIntL-1 does not kill bacteria directly or through complement activation.

After the initial validation of the *S. pneumoniae* model, the direct biological effect of hIntL-1 on the bacteria was tested. The addition of hIntL-1 onto an *S. pneumoniae* soft-agar overlay did not affect bacterial growth (Fig. 2.2A). While some lectins, such as galectins 4 and 8, can kill bacteria directly through contact, others, like MBL and ficolin-2, require the complement cascade for the same effect (9, 70, 76). We therefore examined the ability of *S. pneumoniae* to survive in pooled normal human serum in the presence and absence of hIntL-1 and complement. To confirm the presence of specific antibodies for both *S. pneumoniae* serotypes in the pooled serum, *S. pneumoniae* whole-cell lysates were probed with the pooled human serum (used for the serum survival assay), followed by detection with an anti-human IgG4 antibody. As shown in Figure 2.2B, several proteins were detected in the pooled serum. During the serum survival assay, both serotypes showed robust survival in the presence and absence of hIntL-1 (Fig. 2.2C), demonstrating that hIntL-1 does not promote killing by the complement cascade.

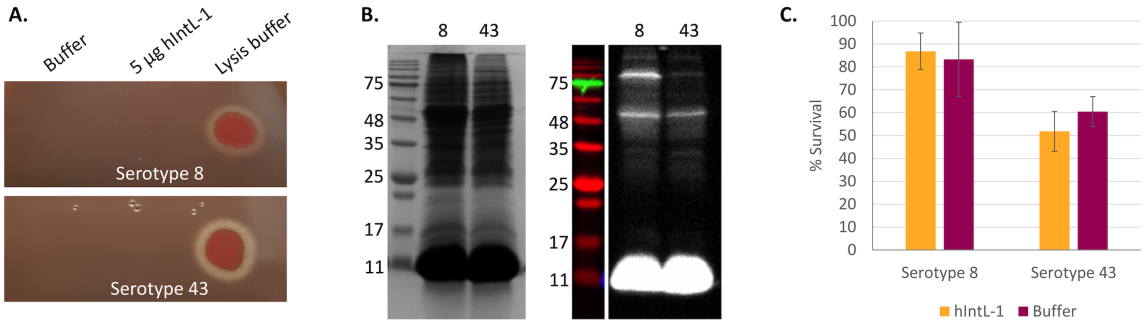


Figure 2.2: hIntL-1 does not kill *S. pneumoniae* directly or in a complement-dependent manner. (A) Spot assay showing indicated *S. pneumoniae* serotypes in soft agar overlay with either hIntL-1, buffer alone (control), or lysis buffer (control) spotted

onto the agar prior to overnight growth. Blood agar beneath agar overlay is visible in lysis buffer control. (B) Verification of the presence of *S. pneumoniae*-specific antibodies in pooled human complement-containing serum. Equal concentrations (OD_{600} of 1) of bacterial whole-cell lysates of serotypes (indicated as 8 or 43) were Coomassie stained (left) or probed with pooled human complement serum, followed by detection by mouse-anti-human IgG4 Fc HRP antibodies in a Western blot (right). Molecular masses, in kDa, are shown. (C) In the serum bactericidal assay, *S. pneumoniae* was incubated with undiluted pooled human serum containing complement in the presence or absence of hIntL-1. Numbers of CFU per milliliter were determined after 2 h of incubation at 37°C, and percent survival was calculated relative to CFU numbers obtained with heat-inactivated serum. The error bars represent the standard errors of the means from quadruplicate independent experiments.

hIntL-1 increases bacterial attachment to host cells.

While hIntL-1 did not show direct killing of *S. pneumoniae*, the agglutinating phenotype caused by the lectin is likely to interfere with a variety of processes during bacterial infection. A major aspect of bacterial pathogenesis is the ability of the microbe to attach to host tissues and form biofilms. We first tested the effect of hIntL-1 on *S. pneumoniae* biofilm formation in a 96-well biofilm assay with spectrophotometric quantification. This assay showed no significant difference in biofilm formation in the presence of hIntL-1 (Fig. 2.3A). We further tested the human lung epithelial A549 cell line that is commonly used in *S. pneumoniae* host cell studies. Here, serotype 43 showed a significantly higher ability to attach to the host cells. This finding was verified by microscopy and Hema3 staining (Fig. 2.3C). The attachment of *S. pneumoniae* serotype 43, but not serotype 8, showed an unexpected increase in A549 cell attachment in the presence of hIntL-1 (Fig. 2.3B).

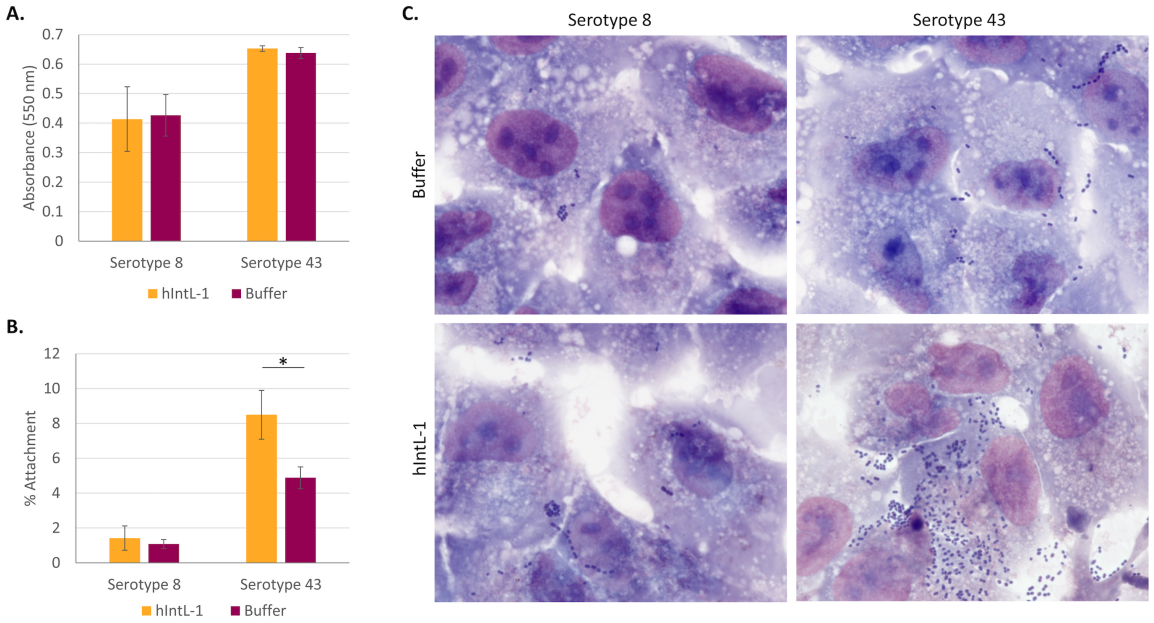


Figure 2.3: hIntL-1 does not alter *S. pneumoniae* biofilm formation but increases A549 cell attachment. (A) *S. pneumoniae* with or without hIntL-1 was incubated in 96-well plastic plates for 24 h. After removal of planktonic bacteria, biofilms were stained with crystal violet and absorbance readings were measured as shown. Error bars represent the standard errors of the means from five replicates. (B) *S. pneumoniae* attachment to confluent A549 lung cells in the presence and absence of hIntL-1. The assay was done at an MOI of 100 for 2 h. Percent attachment was determined after washing unbound bacteria relative to wells without A549 cells. The error bars represent the standard errors of the means from four replicates, and an asterisk indicates significant difference ($P < 0.05$, Student's *t* test). (C) For microscopy, the attachment assay was modified by incubating *S. pneumoniae* with or without hIntL-1 with A549 cells grown to 75% confluence on glass slides. Hema3 staining was used to visualize the bacteria and A549 cells. Images are shown at ×100 magnification in each panel.

hIntL-1 does not cause increased bacterial killing by PBMCs.

We next sought to determine whether hIntL-1 was capable of opsonizing and subsequently targeting bacteria toward phagocyte elimination. In this phase of host defense, a variety of immune cells are recruited for pathogen elimination. One group of immune cells is the PBMCs containing the phagocytic cells, monocytes. To test this, we performed an opsonophagocytosis assay (OPA) by infecting freshly isolated PBMCs with *S. pneumoniae* in the presence or absence of hIntL-1. We also included pneumococcal

pool B serotype 8-specific antisera and pool D serotype 43-specific antisera as assay controls (commonly used for *S. pneumoniae* serotyping) and verified their binding and agglutinating abilities through fluorescence microscopy with 4′,6-diamidino-2-phenylindole (DAPI)-stained bacteria (Fig. 2.4A). A low multiplicity of infection (MOI) of 0.05 (bacteria to PBMCs) was selected due to the extensive ability *S. pneumoniae* possesses to evade clearance strategies employed by the immune system, allowing the microbe to survive even when outnumbered by human immune cells (170, 172, 221). *S. pneumoniae* showed robust survival and even a trend toward increased proliferation when incubated with PBMCs. Thus, the addition of hIntL-1 did not show statistically significant changes in bacterial numbers, although incubation with serotype-specific antibodies showed a significant decrease in bacterial survival (Fig. 2.4B). The experiment was then repeated in the presence of 20% heat-inactivated human serum to supply autologous antibodies to accompany hIntL-1 during bacterial opsonization. The addition of serum did not lead to changes in bacterial killing in the presence of hIntL-1 (Fig. 2.4C).

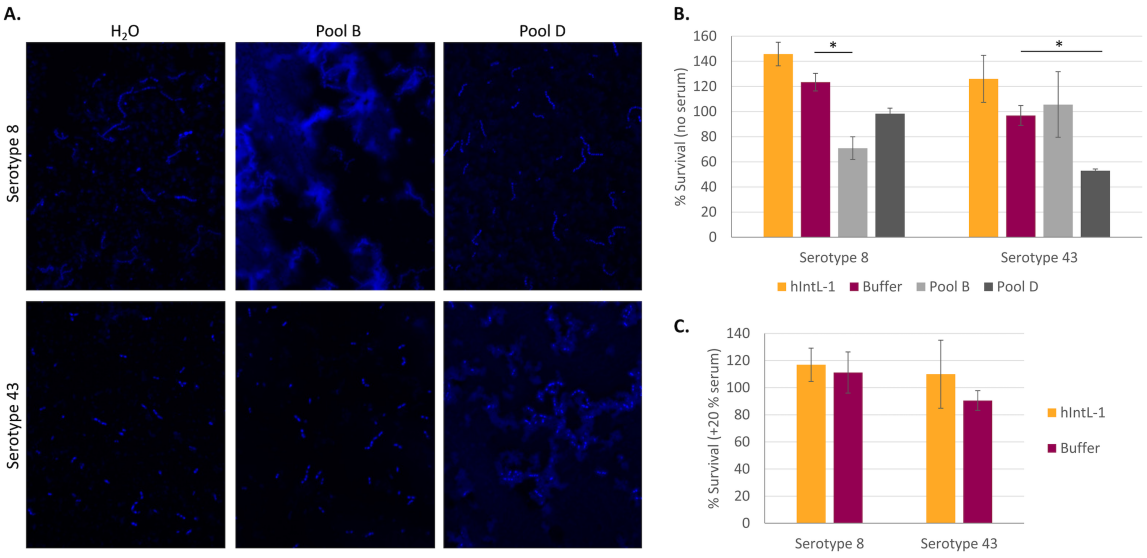


Figure 2.4: hIntL-1 does not promote *S. pneumoniae* phagocytosis by human PBMCs. (A) Labeled *S. pneumoniae* (blue) was incubated with pneumococcal antiserum pools B and D, or water, before imaging via fluorescence microscopy at ×64 magnification. (B) *S. pneumoniae* opsonophagocytosis assay (OPA) with or without hIntL-1. Human peripheral blood mononuclear cells (PBMCs) were freshly isolated from different donors for each of the 7 replicates shown. *S. pneumoniae* serotype-specific antisera were added as bacterial agglutination controls. Antiserum pool B and pool D bind to *S. pneumoniae* serotypes 8 and 43, respectively. The OPA was done at an MOI of 0.05 for 2 h. Percent survival was quantified by determining colony counts relative to samples with heat-inactivated PBMCs. Error bars represent the standard errors of the means, and the asterisks indicate significant differences (*, $P < 0.05$, Student's *t* test). (C) OPA as described in panel B with the addition of 20% fresh autologous human serum in each sample. Error bars represent the standard errors of the means for four replicates.

hIntL-1 causes a drastic increase in bacterial killing by neutrophils.

PMNs are another prominent class of immune cells during bacterial infection and are particularly abundant during *S. pneumoniae* infections. PMNs are equipped with a variety of mechanisms to kill bacteria, including phagocytosis, intracellular killing in the phagosome, and release of reactive oxygen species (ROS), as well as extracellularly through employment of neutrophil extracellular traps (NETs) and serine proteases (222, 223). To test the impact of hIntL-1 during PMN-mediated killing, *S. pneumoniae* and PMNs were incubated at a multiplicity of infection (MOI) of 0.01 (bacteria to PMNs) with 5% autologous serum and in the presence or absence of hIntL-1. As shown in Figure 2.5A, while serotype 8 was unaffected by the presence of hIntL-1, the presence of the lectin led to a significant drop in serotype 43 survival, from 83% to 31% (Fig. 2.5A). The use of pneumococcal antiserum pools B and D as controls for bacterial clumping showed drastic and serotype-specific decreases in bacterial survival during the PMN killing assay. To assess the ability of hIntL-1 to serve as an opsonin during the PMN killing process, the assay was repeated with heat-inactivated serum and in the absence of serum. Through incubation of serum at 56°C, complement components were degraded, resulting in serum

that contains only antibodies as the opsonizing agents. PMNs were unable to kill either serotype with hIntL-1 in the absence of complement. Furthermore, no decrease in bacterial survival could be noted in the presence of hIntL-1 in samples lacking serum altogether. These data suggest that hIntL-1 cannot serve as an opsonin itself. Rather, the increase in killing of *S. pneumoniae* serotype 43 in the presence of hIntL-1 is due to bacterial agglutination in conjunction with complement opsonization.

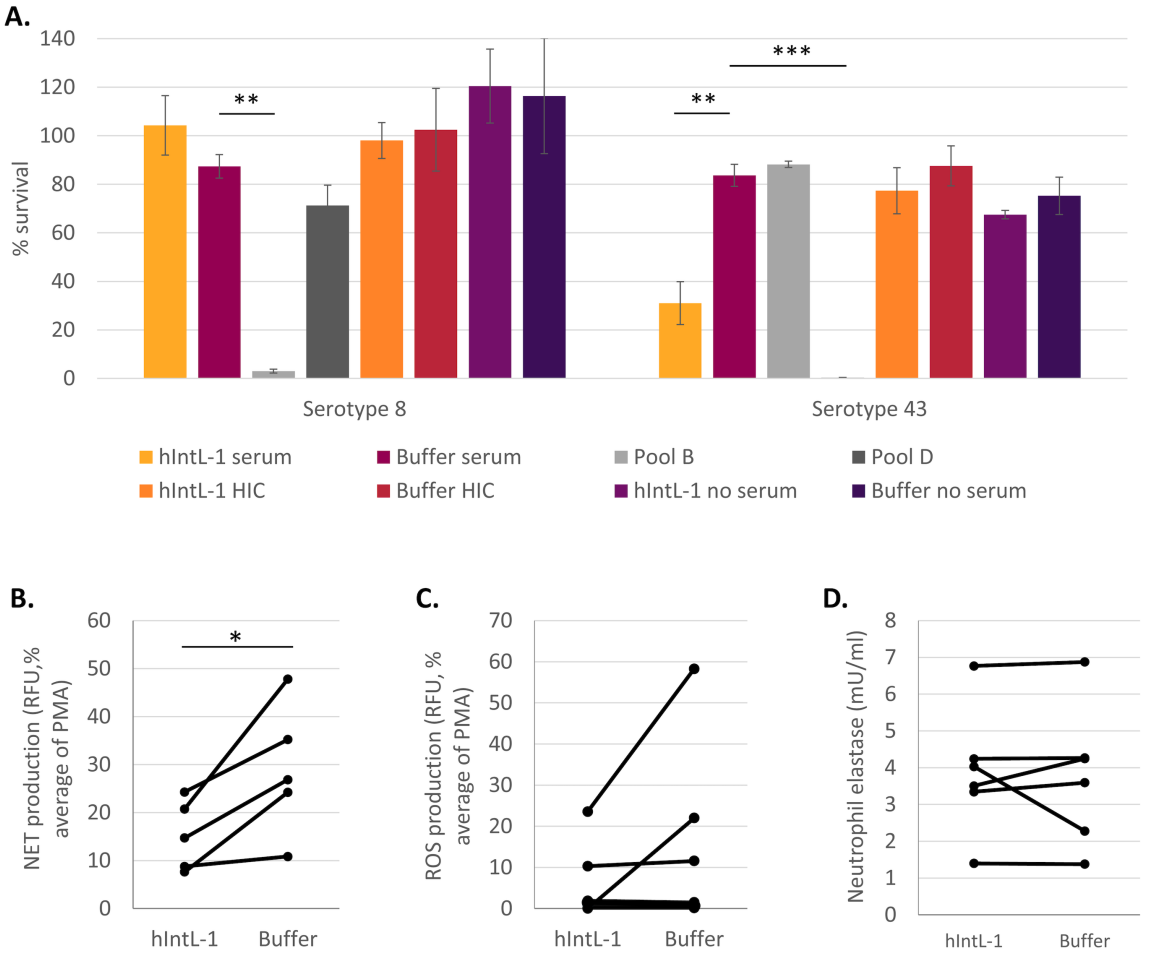


Figure 2.5: hIntL-1 leads to increased *S. pneumoniae* serotype 43 killing by human PMNs. (A) PMNs and autologous serum were freshly isolated from the blood of healthy human donors ($n \geq 3$). *S. pneumoniae* serotype-specific antisera were added as bacterial agglutination controls. Pool B and pool D antisera bind to *S. pneumoniae* serotypes 8 and 43, respectively. During the neutrophil killing assay, cells and bacteria were incubated at an MOI of 0.01 for 2 h in the presence of 5% autologous serum, 5% heat-inactivated autologous serum (56°C for 1 h for heat-inactivated complement, HIC), or no serum.

Percent survival was determined by comparing colony counts relative to samples without PMNs as controls or to samples with inactivated PMNs for samples without serum. The error bars represent the standard errors of the means, and the asterisks indicate significant differences (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ by Student's *t* test). (B) Neutrophil extracellular trap (NET) production using the Sytox Orange nucleic acid stain. PMNs isolated from five different donors were incubated with and without hIntL-1. Percent fluorescence was calculated relative to measurements with phorbol 12-myristate 13-acetate (PMA)-activated PMNs. The error bars represent the standard errors of the means, and an asterisk indicates significant differences (*, $P < 0.05$, Student's *t* test). (C) Measurement of reactive oxygen species production using the Diogenes-based chemiluminescence kit. PMNs from different donors for each of 7 replicates were incubated with and without hIntL-1. Percent luminescence was calculated relative to measurements with phorbol 12-myristate 13-acetate (PMA)-activated PMNs. (D) The enzymatic activity of neutrophil elastase released from PMNs in the presence of hIntL-1 was measured using the neutrophil elastase activity assay kit. PMNs isolated from six different donors were incubated with hIntL-1. Neutrophil elastase activity, in milliunits per milliliter, was calculated using a neutrophil elastase standard curve.

We aimed to further investigate the mechanism of killing leading to hIntL-1-mediated bacterial elimination. Direct incubation of PMNs with hIntL-1 led to a decrease of DNA release by the cells, indicating that hIntL-1 does not affect PMN viability or increased NET formation (Fig. 2.5B). Similarly, the production of ROS (H_2O_2) and the release of enzymatically active neutrophil elastase from PMNs were not stimulated by hIntL-1 (Fig. 2.5C and D).

To establish whether bacterial phagocytosis by PMNs is the primary killing mechanism, two approaches were utilized. The percentage of phagocytosing PMNs was assessed using flow cytometry. Bacteria were stained using the pH-sensitive dye pHrodo, which fluoresces green only in the acidic environment of the phagosome. A representative image of the flow gating strategy is shown in Figure 2.6A. The collected data from four replicates showed the same amount of PMN phagocytosis for both serotypes in the presence and absence of hIntL-1 (Fig. 2.6B). Since the data only represent the number of PMNs that have phagocytosed bacteria but not the amount of bacteria phagocytosed per

PMN, we further employed fluorescence microscopy. To quantify the bacteria phagocytosed by PMNs, we used pHrodo-labeled *S. pneumoniae* to visualize phagocytosis. Using ImageJ software, 150 PMNs were analyzed for their green fluorescence intensity, resulting in a significant increase in bacterial uptake of *S. pneumoniae* serotype 43 in the presence of hIntL-1 (Fig. 2.6C). The same trend can be seen in the representative fluorescence microscopy image in Figure 2.6D showing hIntL-1-mediated serotype-specific phagocytosis.

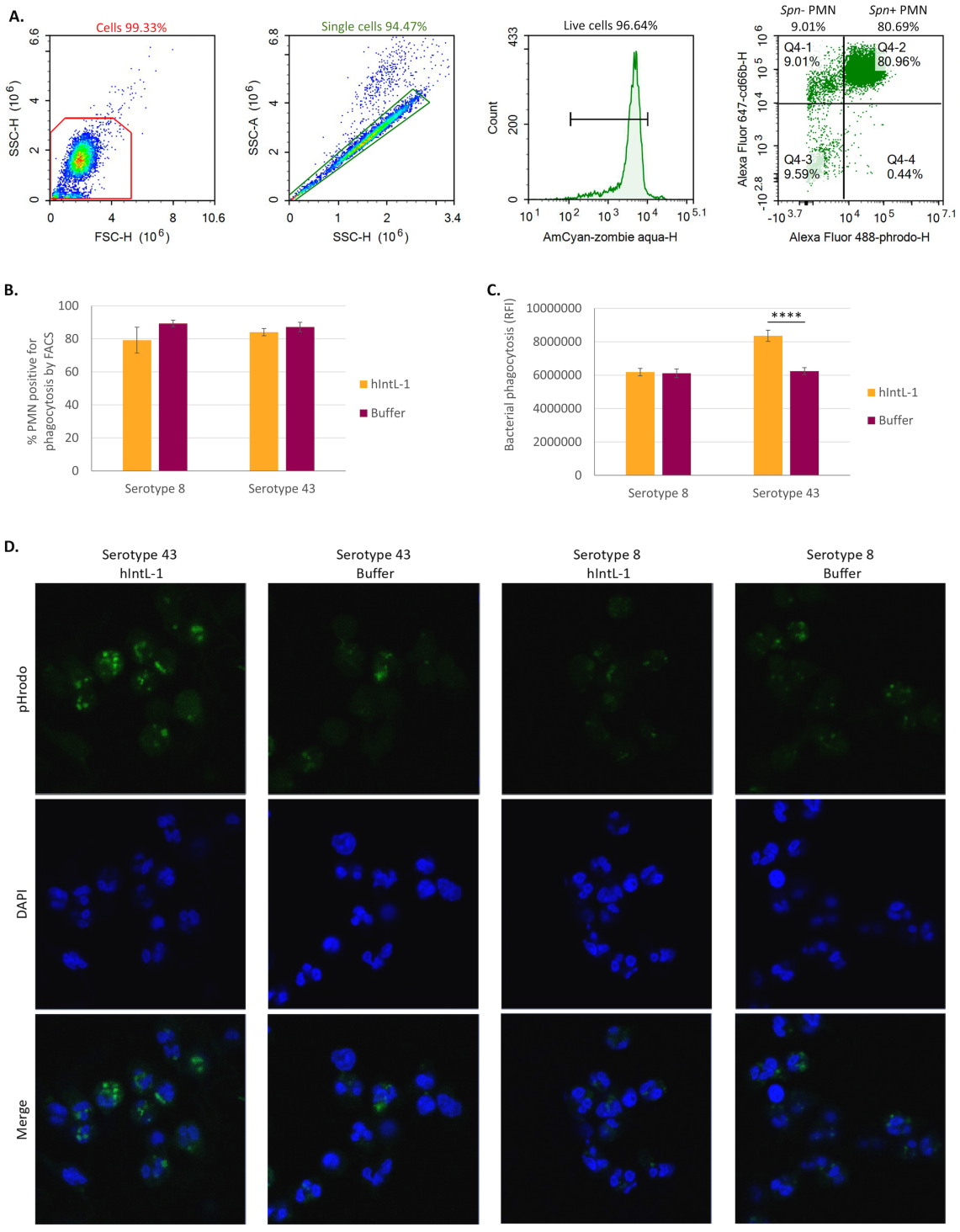


Figure 2.6: PMNs display increased bacterial uptake of *S. pneumoniae* serotype 43 in the presence of hIntL-1. (A) Flow cytometry was used to quantify the percentage of PMNs that have phagocytosed *S. pneumoniae* (*Spn*) in the presence and absence of hIntL-1. The gating strategy for the percentage of PMNs positive for phagocytosis (*Spn*+ PMN) is shown for one representative sample (*S. pneumoniae* serotype 43 in the presence of

hIntL-1). (B) Flow cytometry analyses for phagocytic PMNs were executed as four replicates with isolated PMNs from four different donors. Means and standard errors are shown. FACS, fluorescence-activated cell sorting. (C) pHrodo-labeled *S. pneumoniae* (green) was incubated with PMNs at an MOI of 5 for 2 h in the presence or absence of hIntL-1. Samples were fixed with paraformaldehyde, applied to microscopy slides, and stained with DAPI (blue). Using the ImageJ software, the green fluorescence intensity per PMN was quantified for a total of 150 cells from three independent experiments. Means and standard errors are shown, and the asterisks represent significant difference (****, $P < 0.0001$, Student's *t* test). (D) Representative images of fluorescence microscopy showing the phagocytosis of pHrodo-labeled *S. pneumoniae* by PMNs in the presence and absence of hIntL-1. Images were obtained using fluorescence microscopy at ×100 magnification.

Discussion

Our research aimed to better understand the proposed role of hIntL-1 in pathogen surveillance as a part of the human innate immune response. As shown previously for *S. pneumoniae* and *H. pylori*, hIntL-1 is unable to kill bacteria directly, and our data further demonstrated that the presence of the lectin does not lead to bacterial killing through the complement-mediated membrane attack complex (102, 108). hIntL-1 is composed of a homotrimer containing a glycan binding site in the fibrinogen-like domain with three bound calcium ions in each monomer, but it lacks the collagen-binding domain present in MBL and ficolins. Thus, it is not surprising that hIntL-1 is unable to recruit MASPs to engage the complement cascade (102). This activity has only been shown for lectins such as MBL, ficolins, and other C-type lectins (collectins), all of which contain the appropriate fibrinogen-like carbohydrate binding domains in their N-termini and collagen-like domains for association with MASPs (71, 72, 215). Nevertheless, it was important to exclude the possibility of hIntL-1 association with the complement cascade, particularly since MBL was long believed to be the only lectin to activate the lectin complement pathway and now several other lectins have been demonstrated to possess this ability (71, 72).

Previous research accompanied by the findings presented here shed further light on the role hIntL-1 plays within the large group of human innate immune lectins. While lectins like RegIII and the galectins demonstrate direct killing effects, others engage various immune proteins and host cells to assist in bacterial elimination (76, 81, 224). Especially in the context of Gram-positive bacteria with thick peptidoglycan layers, for which the membrane attack complex of the complement pathway is only minimally effective, the most potent method of microbial killing involves phagocytosis (170). Lectins have evolved to utilize phagocytic cells in different ways. The presence of soluble lectins like MBL and ficolin-2 lead to increased phagocytosis upon bacterial binding through opsonization with recruited proteins of the complement cascade (9). A direct method for lectins to facilitate phagocytosis is observed with membrane-bound lectins, such as the mannose receptor present on macrophage surfaces, where pathogen binding to the lectin leads to direct phagocytic uptake (225, 226). In the case of hIntL-1, our data show no killing through complement alone but enhanced bacterial killing by PMNs in the presence of serum. This suggests bacterial phagocytosis is initiated through opsonization by complement but increased through the agglutinating effect of the lectin binding (Fig. 2.7). Agglutination is a common feature for lectins involved in innate immunity due to their multimeric structures, which increase binding avidity and allow for increased bacterial clearance. Bacterial aggregates are typically hampered in their ability to attach to and invade the host epithelium and are more easily removed by the mucociliary system (3, 7, 153). Combined, we suggest the primary function for hIntL-1 is in bacterial agglutination and facilitating killing by specific immune cells.

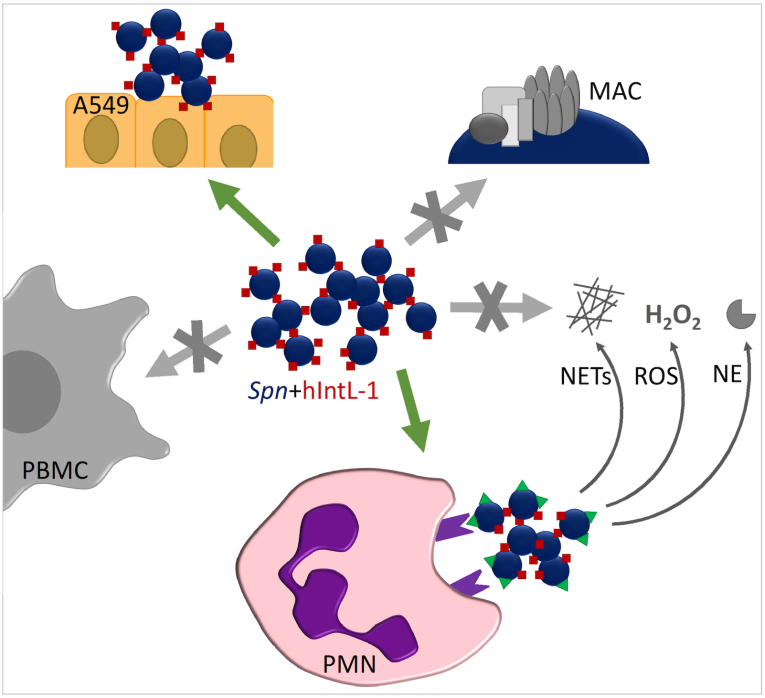


Figure 2.7: Proposed role of hIntL-1 during *S. pneumoniae* serotype 43 infection and immune-mediated clearance. hIntL-1 (red square) recognizes *S. pneumoniae* serotype 43 (*Spn*43, dark blue circle) through exocyclic 1,2-diols displayed on its CPS, leading to bacterial agglutination and increased attachment to lung A549 host cells. The presence of hIntL-1 does not result in an increase in bacterial killing by PBMCs, and complement (green triangle) alone does not kill *Spn*43 in the presence or absence of hIntL-1 through a membrane attack complex (MAC) mechanism. Rather, complement opsonization together with hIntL-1-mediated *Spn*43 agglutination leads to PMN uptake. The presence of hIntL-1 does not increase production of neutrophil extracellular traps (NETs), reactive oxygen species (ROS), H_2O_2, or neutrophil elastase (NE) by PMNs; instead, the increased killing is caused by phagocytosis.

The exocyclic 1,2-diol glycan ligand is widely distributed among bacteria and is present as surface polysaccharides, including glycerol-1-phosphates, β-galactofuranose, KO, and KDO (102). This gives hIntL-1 the potential to bind Gram-positive and Gram-negative bacteria, commensals, and pathogens alike. While MBL and RegIII have broad glycan binding specificities, the galectins and ficolin-2 display narrow ligand ranges (9, 69, 81). Ficolin-2 has been shown to recognize *S. pneumoniae* serotype 43 (also called 11A), leading to the activation of the complement cascade and increased phagocytosis. The

authors of this study note the comparatively low invasive potential of serotype 43 among other *S. pneumoniae* serotypes and speculate that the recognition through ficolin-2 is the cause of this lowered pathogenicity (9). This phenotype can be further explained by the specific targeting of this bacterium through not only ficolin-2 but also hIntL-1. This demonstrates the intricacies employed by our immune system to ensure effective and redundant defenses.

hIntL-1 is secreted into intestinal and airway mucus, where it can exercise its agglutinating effects on a variety of bacteria that can penetrate the mucosal layer and present a risk to the host (110, 115, 219). It is anticipated that the resulting bacterial aggregates will be more easily removed through the mucociliary system or phagocytosed by immune cells. Although our data indicate an unexpected increase in *S. pneumoniae* serotype 43 attachment to A549 cells, this does not necessarily negate the role for hIntL-1 in host protection *in vivo*. One possible explanation for the increase in host cell attachment may be a direct result from bacterial agglutination while still leaving some of the bacterial surfaces exposed for host contact. This would lead to attachment of large bacterial aggregates rather than short bacterial chains or diplococci. This phenomenon has been observed before with *S. pneumoniae* mutants that are capable of forming elongated bacterial chains showing increased attachment to A549 cells compared to the wild-type (WT) strain, with predominantly diplococcal cells (227). This relationship also held true in a mouse model where longer-chain-forming mutants exhibited higher degrees of colonization (227). Additionally, *S. pneumoniae* is well known for its ability to adapt to and evade human defense mechanisms (170, 172). Thus, it is also possible that bacteria like *S. pneumoniae* have evolved methods to use immune strategies of their hosts to their

own benefit. Nevertheless, although the A549 cell line is a commonly used pulmonary epithelial model, it does not possess a thick mucus layer or the complete architecture of the more complex lung environment, so the increased attachment of agglutinated *S. pneumoniae* serotype 43 to host cells should be explored further. Future work could test this hypothesis using *in vitro* cultures of human bronchial epithelial cells, lung organoids, or mouse models.

Our data demonstrate that hIntL-1 enhanced killing of *S. pneumoniae* serotype 43 by PMNs in the presence of complement, suggesting the main mechanism of microbial elimination is hIntL-1-mediated bacterial agglutination followed by elimination through PMN phagocytosis. Previous studies have shown increased phagocytosis of *H. pylori* by bone marrow-derived macrophages in the presence of hIntL-1 (108). For *S. pneumoniae*, we found that hIntL-1 did not increase killing by freshly isolated human PBMCs. This discrepancy could be due to a variety of factors, including the differences in the bacterium, experimental conditions, and immune cells. Pneumococcal infections are controlled by host neutrophils, which kill this pathogen via opsonophagocytosis in collaboration with the complement system (221). Given the use of *S. pneumoniae* as our model pathogen, it is not surprising that we observed PMNs as the primary mechanism for bacterial killing together with hIntL-1 (rather than PBMCs). During the immune response to invading *S. pneumoniae*, all three pathways of complement are activated, resulting in the deposition of C3b on the bacterial surface (173, 228). While in Gram-negative bacteria the formation of the membrane attack complex would play a major role in bacterial elimination, this effect excludes Gram-positive bacteria due to the thick layer of impenetrable peptidoglycan (170). Instead, the deposited C3b is cleaved to iC3b, which is recognized by complement

receptor 3 on neutrophils, stimulating phagocytosis (229–231). Therefore, killing of *S. pneumoniae* is largely facilitated through neutrophil uptake. We have observed similar bacterial killing mechanisms in our studies in the presence of hIntL-1. We observed no increase in NET, hydrogen peroxide, or neutrophil elastase release. Instead, we noted hIntL-1-mediated increases in killing of *S. pneumoniae* serotype 43 only in the presence of complement and an enhanced uptake of bacteria, while the number of phagocytic PMNs remained unchanged. The presence of hIntL-1 did not increase the number of phagocytic events; rather, the bacterial count for each phagocytosed unit was increased through bacterial clumping. Our data further suggest that hIntL-1 acts as an agglutinating agent, but not an opsonin, leading to an increase of bacterial killing through bacterial clumping followed by complement opsonization and phagocytic uptake. This finding is supported by studies demonstrating enhanced phagocytosis of *S. pneumoniae* mutants with increased chain lengths. To investigate this, the authors used an *S. pneumoniae* mutant with long-chain morphology and showed increased complement deposition as well as uptake by human neutrophils compared to the wild type. The fitness disadvantage observed for long bacterial chains of *S. pneumoniae* can be phenotypically compared to the bacterial aggregation observed in the presence of hIntL-1 (221). It can be speculated that hIntL-1, in interaction with neutrophils, has a similar killing effect on other *S. pneumoniae* serotypes containing the exocyclic 1,2-diol ligand. This could include the serotypes 20 and 70 that have been shown to be bound by hIntL-1 as well as the 21 other serotypes (10, 13, 18, 22, 31, 34, 35, 41, 44, 53, 55, 56, 61, 69, 72, 74, 78, 10B, 11D, 11E, and 20B) that contain β-galactofuranoses or glycerol-1-phosphates in their CPS repeating unit (49, 102, 168).

The full potential for intelectin function as part of an immune system is best exemplified by looking at one of arguably the simplest free-living animals on earth. As a placozoan, Trichoplax measures 1 to 2 mm in diameter and consists of two epithelial layers that sandwich multinucleated fiber cells (232, 233). The organism only contains a small number of cell types, none of which appear to be immune cells, suggesting an adaptive immune response is not possible (234, 235). Instead, Trichoplax employs a wide variety of Toll-like and NOD-like extracellular scavenger receptors as well as lectins, composing its innate immune system. Computational analyses identified 31 intelectins in the genome of Trichoplax, which, in combination with the other pattern recognition molecules, ensure differentiation between self and non-self (220). Humans, in comparison, possess only two intelectins. While the field has been working on elucidating the role of hIntL-1, much less is known about hIntL-2, which shows a more restricted glycan binding specificity and is only expressed by Paneth cells in the small intestine (113). Nevertheless, our data serve to exemplify the evolution of intelectins from a potentially key player in the Trichoplax innate immune response to a microbial pattern recognition receptor involved in pathogen surveillance.

Collectively, we suggest the main mechanism of bacterial killing by hIntL-1 lies in its ability to agglutinate followed by targeting microbes for elimination through neutrophil phagocytosis. Considering the prevalent nature of the ligand for hIntL-1 on various microbes, along with our data showing enhanced neutrophil phagocytosis, hIntL-1 is most likely involved in general host defense, preventing the breach of host barriers.

Material and Methods

Bacterial strains and growth conditions

The *S. pneumoniae* serotypes 8 (ATCC 6308) and 43 (ATCC 10343) were grown on tryptic-soy agar (Hardy Diagnostics) plates enriched with 10% defibrilized sheep blood (TSA plates) or grown statically in liquid Bacto Todd Hewitt broth (Becton, Dickinson, and Company) with 0.5% yeast extract (THY) at 37°C under aerobic conditions.

Expression and purification of native human intelectin-1

Standard protocols were used for the generation of the hIntL-1 expression plasmid restriction enzyme digests, DNA ligations, PCR, and other recombinant DNA procedures. cDNA encoding the open reading frame of hITLN-1 was cloned by reverse transcription-PCR. Primers were designed on the basis of a reported hITLN-1 sequence, 5′-CCACTAGTATTACAATGAACCAACTCAGCTTCC-3′ and 5′-CCTCTAGACTCTCAACGATAGAATAGAAGCACA-3′ (113). cDNA was synthesized using previously isolated hITLN-1, cloned from human small intestinal total RNA (113). A single band was amplified by PCR using these primers, cDNA, and Platinum high-fidelity *Taq* DNA polymerase (Life Technologies) (35 cycles of 94°C for 30 s, 60°C for 30 s, and 72°C for 60 s). The PCR product was digested with the enzymes SpeI and XbaI and then cloned into the pTracer vector (Life Technologies). The DNA sequence was confirmed by the University of Georgia Molecular Genetics Instrumentation Facility.

Large amounts of the pTracer-hIntL-1 expression vector were generated in *Escherichia coli* and purified using the PureLink HiPure Expi Plasmid Megaprep kit (Invitrogen). The pTracer_hIntL-1 vector was then introduced into HEK-293F suspension cultures

(FreeStyle 293-Fcells; Thermo Fisher Scientific) maintained at 2.5×10^6 to 3.0×10^6 cells/mL in a humidified CO_2 platform shaker incubator at 125 rpm at 37°C and 50% humidity. Transient transfection was performed using HEK293-F cells in the expression medium comprised of a 9:1 ratio of Freestyle 293 expression medium (Thermo Fisher Scientific) and EX-Cell expression medium including GlutaMAX (Sigma-Aldrich). Transfection was initiated by the addition of plasmid DNA and polyethyleneimine as the transfection reagent (linear 25-kDa polyethyleneimine; Polysciences, Inc.) at a ratio of 4:9. Twenty-four hours posttransfection, the cell cultures were diluted with an equal volume of fresh medium supplemented with valproic acid (2.2 mM final concentration), and protein production was continued for an additional 5 days at 37°C. The cell cultures were harvested, clarified by sequential centrifugation at 1,200 rpm for 10 min and 3,500 rpm for 15 min at 4°C, and passed through a 0.8-μm filter (Millipore).

For the purification of hIntL-1, the culture supernatant was treated with protease inhibitor tablets (cOmplete, Mini, EDTA-free; Roche Diagnostics GmBH) and DNase/RNase (Turbonuclease from *Serratia marcescens*; Sigma-Aldrich) and adjusted to 10 mM $CaCl_2$. The purification was performed at 4°C using a β-galactofuranose agarose column equilibrated with 20 mM Tris, pH 7.4, 150 mM NaCl, 0.1% Tween 20, and 10 mM $CaCl_2$. hIntL-1 in the culture supernatant was allowed to bind to the column resin during a 30-min batch incubation before being slowly passed through the column. The column was washed using 20 mM Tris, pH 7.4, 150 mM NaCl, and 10 mM $CaCl_2$ and hIntL-1 was eluted in 20 mM Tris, pH 7.4, 150 mM NaCl, and 10 mM EDTA. The resulting eluent was concentrated using a 30-kDa molecular-weight-cutoff Ultra-15 centrifugal filter unit (Amicon) and subsequently dialyzed into 20 mM HEPES, pH 7.4, 150 mM NaCl,

500 mM D-glucose (control buffer for samples without hIntL-1). The purity of hIntL-1 was assessed by SDS-PAGE and Coomassie staining, and the concentration was measured using absorbance at 280 nm with a hIntL-1-specific extinction coefficient of ε = 237 400 cm^{-1} M^{-1} and molecular mass of 101,400 Da (102).

Agglutination assay

Bacterial strains were grown on their respective media and harvested into binding buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 10 mM $CaCl_2$) or EDTA buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 10 mM EDTA). Next, 100 µL of the bacterial suspension adjusted to an optical density at 600 nm (OD_{600}) of 2 was added per well of a round-bottom 96-well plate (Cooke Microtiter). hIntL-1 at various concentrations, or equivalent volumes of control buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 500 mM D-glucose), were added. The plate was incubated for 2 h at 4°C under agitation before allowing the bacteria to settle overnight at 4°C.

Fluorescence microscopy with labeled hIntL-1

The fluorescent labeling of the purified hIntL-1 was done using the Alexa Fluor 488 microscale protein labeling kit (Thermofisher Scientific), resulting in a final concentration of labeled hIntL-1 of 0.2 µg/µL at a calculated average of two molecules of Alexa Fluor per trimer of hIntL-1.

For the preparation of microscopy samples, bacteria were grown on their respective media. The harvested bacteria were washed once in binding buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 10 mM $CaCl_2$) before adjusting their OD_{600} to 0.2. Next, 100 µL of bacterial

suspension was incubated with 1.5 µg hIntL-1 or control buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 500 mM D-glucose) for 2 h at 4°C in the dark and under agitation. Subsequently, the bacteria were centrifuged at 8,000 rpm for 4 min, washed three times with binding buffer, and resuspended to 15 µL. The resulting bacterial suspension was applied to a coverslip, dried at 37°C, and heat fixed. The slides were stained with DAPI (2.5 µg/mL), rinsed in H_2O, and mounted onto a microscope slide with 5 µL Vectashield antifade mounting medium (Vector Laboratories). The margins of the coverslip were sealed with nail polish before storage at 4°C in the dark. Confocal microscopy was executed on a Zeiss LSM 710 at ×40 magnification.

Biofilm formation assay

The conditions for the biofilm formation assay with *S. pneumoniae* were adapted from a previously described method (236). *S. pneumoniae* serotypes 8 and 43 were grown in THY to mid-exponential phase (OD_{600} of approximately 0.6). The cultures were diluted at 1:20 with THY. In a 96-well plate, 200 µL of bacteria suspension per well was incubated with 45 µg/mL of hIntL-1 or its control buffer for 24 h at 37°C. The bacterial suspensions were removed from the wells by vigorous inversion of the plate, followed by a wash in H_2O and drying of the plate. All washes in the assay were executed by submerging the 96-well plate in H_2O entirely, followed by the emptying of all wells simultaneously via vigorous inversion of the plate. The biofilms were stained with 250 µL/well of a 1% crystal violet solution for 15 min at room temperature (RT). Subsequently, the plate was washed twice and left to dry. To dissolve the crystal violet stain, 200 µL of 30% acetic acid was added per well and incubated for 15 min. Next, 150 µL of the stained acetic acid solution

was transferred into a fresh 96-well plate for photometric analysis using the Synergy H1 microplate reader at an absorbance of 550 nm (BioTek Instruments, Inc.). The assay was executed as five biological replicates; means and standard errors of the means are displayed in the graph.

Attachment to A549 lung cells

The A549 lung epithelial cell line was cultured in Dulbecco's modified Eagle medium (DMEM) containing 4.5 g/liter glucose and L-glutamine, without pyruvate (Corning) replaced by 10% fetal bovine serum (FBS) (VWR), 100 U/mL of penicillin, and 100 μg/mL of streptomycin (Gibco).

The attachment assay was performed as described previously (237). Briefly, A549 cells between passage numbers 6 and 20 were used. The cells were seeded into 24-well plates with approximately 1×10^5 cells/well at 95 to 100% confluence. Two hours before the assay, the cells were washed twice with phosphate-buffered saline (PBS) before 500 μL DMEM with 5 mM $CaCl_2$ was added. *S. pneumoniae* was grown to mid-exponential phase and adjusted to 1×10^9 CFU/mL in DMEM with 5 mM $CaCl_2$. To start the attachment assay, an appropriate amount of medium was removed from each well to yield a final reaction volume of 200 μL/well. Next, 45 μg/mL hIntL-1 and 1×10^7 CFU of bacteria were added for an MOI of 100:1 (bacteria to A549 cells). The plate was incubated for 2 h at 37°C with 5% CO_2. Each well was washed 3× with PBS, followed by the addition of 100 μL trypsin-EDTA solution (Sigma) and incubation at 37°C for 5 min. To detach the cells, the suspension was pipetted up and down vigorously before serial diluting and plating for determination of number of CFU per milliliter. Samples without A549 cells served as

100% attachment controls. These samples were diluted and plated directly after the 2-h incubation period. Shown are means and standard errors of the means for 4 biological replicates.

Microscopy of A549 cells with Hema3 stain

The attachment assay for microscopy was executed as described above, with the following changes. A549 cells were seeded into 24-well plates with sterile glass slides at the bottom. The attachment assay was done at 75% confluence for better visualization of the cellular margins. After the attachment period of 2 h at 37°C, paraformaldehyde was added to a final concentration of 5% and incubated at RT for 30 min. The wells were washed 3× with PBS before the glass slides were removed, rinsed with H_2O, and dried. The Hema3 Staining System (Fisher Scientific) was used by following the manufacturer's instructions. To stain the samples, three dips in each staining reagent were performed. The slides were imaged using a Zeiss AxioScope A1 at ×100 magnification.

Serum survival assay

A serum survival assay was performed to assess the ability of *S. pneumoniae* to resist complement-mediated killing in the presence of hIntL-1 (238, 239). Pooled human complement serum was purchased from Innovative Research. Prior to the serum survival assay, the presence of human antibodies specific to *S. pneumoniae* serotypes 8 and 43 was verified by Western blotting. Briefly, bacterial whole-cell lysates were separated by SDS-PAGE and transferred onto a nitrocellulose membrane. The membrane was blocked overnight at 4°C in 5% bovine serum albumin (BSA). The pooled human serum was used

as the primary antibody (1:1,000) and mouse-anti-human IgG4 Fc, horseradish peroxidase (HRP) (1:2,000; Invitrogen), as the secondary antibody. The serum survival assay was largely executed as described by Lees-Miller et al. (239), with some modifications. Briefly, *S. pneumoniae* was grown to mid-exponential phase in THY and was brought to an OD_{600} of 1. In a noncoated 96-well plate, 15 µL bacterial suspension (for a final OD_{600} of 0.1) was combined with hIntL-1 or buffer (for a final concentration of 45 µg/mL hIntL-1) and brought to a final volume of 150 µL with normal human serum (NHS) or heat-inactivated human serum (HIS). HIS was prepared by incubating NHS for 1 h at 56°C. Tenfold serial dilutions of all samples were prepared in 150 µL NHS or HIS. The 96-well plate was incubated at 37°C for 2 h before plating on TSA plates. The survival of *S. pneumoniae* in NHS in the presence and absence of hIntL-1 was calculated as a percentage of its survival in HIS. Four biological replicates of the experiment were performed, and the resulting mean and its standard error are displayed. The serum survival assay was verified using the *Acinetobacter baumannii* 5075 wild-type and *A. baumannii* 5075 *ΔpglC* strains for which survival and complete killing, respectively, were published previously (238). Under the exposure to normal human complement serum for 2 h as described for the assay above, 37.2% survival was observed for *A. baumannii* WT and complete killing for the *ΔpglC* mutant.

Isolation of fresh human PBMCs

Whole fresh human blood was drawn from adult volunteers at the Health Center or the Clinical Translational Research Unit of the University of Georgia under informed consent according to procedures approved by the Institutional Review Boards at the

University of Georgia (UGA number 2012-10769-06). For cell isolation, blood was drawn into EDTA-coated tubes. Peripheral blood mononuclear cells (PBMCs) were isolated using a gradient of Histopaque-1077 Hybri-Max (H8889-100ML; Sigma Life Science). Briefly, equal amounts of blood were layered onto Histopaque 1077 and centrifuged at 350 × *g* for 30 min with the centrifuge deceleration set to 1. The resulting PBMC layer was collected and washed twice with PBS before resuspending in DMEM. The cells were stained with methylene blue for quantification via hemocytometer.

Isolation of autologous serum

Whole fresh human blood was drawn from adult volunteers at the Health Center or the Clinical Translational Research Unit of the University of Georgia under informed consent according to procedures approved by the Institutional Review Boards at the University of Georgia (UGA number 2012-10769-06). Next, 10 mL of blood was drawn into a silicone-coated tube without anticoagulant, and the blood was allowed to clot at RT for 30 min. The tube was centrifuged twice at 2,000 × *g* for 5 min, and the resulting serum was stored on ice until use.

Fluorescent microscopy of *S. pneumoniae* with pneumococcus antiserum pools

S. pneumoniae was grown fresh to mid-exponential phase in THY. The bacteria were harvested and brought to an OD_{600} of 0.35. Next, 40 µL of bacterial suspension was incubated with 2 µL pneumococcal antiserum pool B or D or H_2O for 20 min at RT under agitation (Cedarlane Laboratories). Next, 20 µL was applied to a microscopy coverslip, air dried, and heat fixed. The bacteria were stained with DAPI (2.5 µg/mL final concentration)

in 5 µL Vectashield antifade mounting medium (Vector Laboratories) and sealed to a microscopy slide using nail polish. Confocal microscopy was executed on a Zeiss LSM 710 at ×100 magnification.

Opsonophagocytosis assay

S. pneumoniae was grown to mid-exponential phase in THY prior to the assay. Bacterial stock solutions were prepared in DMEM containing 5 mM $CaCl_2$. The OPA was executed at an MOI of 0.05 in an uncoated 96-well plate. A total of 3.3×10^4 CFU, 45 µg/mL hIntL-1, or its buffer and 6.6×10^5 PBMCs were combined and filled to a reaction volume of 100 µL with DMEM containing 5 mM $CaCl_2$. Where indicated, 1.3 µL *Pneumococcus* antiserum pool B and pool D (Cedarlane Laboratories) were added as a control for *S. pneumoniae* agglutination. During the assay, samples were incubated at 37°C for 2 h under 5% CO_2 before serial dilution in DMEM containing 5 mM $CaCl_2$ and plating for CFU quantification. The percentage of survival was calculated relative to samples to which heat-killed PBMCs were added (15 min at 95°C). For the OPA in the presence of 20% serum, 20% autologous serum was added to all samples and controls. All experiments were executed at least in triplicates; means and standard errors of the means are depicted.

Neutrophil killing assay

Whole fresh human blood was drawn from adult volunteers at the Health Center or the Clinical Translational Research Unit of the University of Georgia under informed consent according to procedures approved by the Institutional Review Boards at the

University of Georgia (UGA number 2012-10769-06). For cell isolation, blood was drawn into EDTA-coated tubes. For the neutrophil killing assay, human PMNs were isolated from the peripheral blood of healthy, consenting volunteers using the EasySep Direct Human PMN isolation kit (Stem Cell Technologies). Isolated PMNs were >98% viable as assessed by Trypan blue staining and >95% pure as determined by CD66b flow cytometry. The assay was adapted from previously described methods (240, 241).

S. pneumoniae serotypes 8 and 43 were grown to mid-exponential phase in THY and adjusted to 1×10^7 CFU/mL in 1× Hanks' balanced salt solution (HBSS) containing 5% autologous serum, 5 mM glucose, 10 mM HEPES, and 5 mM $CaCl_2$. Purified human PMNs were washed and resuspended in assay medium (1× HBSS containing 5% [vol/vol] autologous serum of the PMN donor, 5 mM glucose, 10 mM HEPES, and 5 mM $CaCl_2$). For samples with differing serum conditions, the PMNs were washed 3 times in serum-free assay medium and resuspended in assay medium containing either no serum or 5% heat-inactivated serum (1 h at 56°C for complement inactivation). The assay was executed as a batch-killing assay with no prior opsonization of the bacteria. Eppendorf tubes (1.5 mL) were blocked with 500 µL 5% BSA in PBS followed by a wash with PBS. A total of 4×10^6 PMNs were combined with 45 µg/mL hIntL-1 or an equivalent volume of hIntL-1 buffer (20 mM HEPES, 150 mM NaCl, 500 mM glucose, pH 7.4). For the agglutination controls, 1.3 µL *Pneumococcus* antiserum pool B and pool D (Cedarlane Laboratories) was used. Assay medium was used to reach a volume of 95 µL before the addition of 5 µL bacterial suspension containing 4×10^4 CFU to start the assay at an MOI of 0.01. The samples were incubated for 2 h at 37°C, during which the tubes were mixed every 5 to 10 min to prevent the settling of the PMNs. To lyse the PMNs, 20 µL of each sample was

incubated with 180 µL ice-cold saponin (1 mg/mL in HBSS) for 5 min, followed by a dilution series in HBSS on ice. Samples were plated onto TSA plates containing 5% sheep blood and incubated at 37°C overnight before the quantification of survival via CFU count. The percentage of survival was calculated relative to control samples that did not contain PMNs. For samples containing no serum, the percentage of survival was calculated relative to control samples that contained heat-killed PMNs (20 min at 65°C). All experiments were executed at least in triplicate; shown are means and standard errors.

Measurement of NET release

NET production by PMNs was determined using 0.2% Sytox Orange nucleic acid stain (Thermo Fischer Scientific). The assay was adapted from a method described previously (240, 242). In a 96-well black transparent-bottom plate, 250,000 PMNs were combined with 45 µg/mL hIntL-1 and assay medium (1× HBSS containing 5% autologous serum, 5 mM glucose, 10 mM HEPES, and 5 mM $CaCl_2$). Fluorescence measurements were taken for 8 h at 37°C using a Varioskan Flash microplate luminometer (Thermo Scientific, Waltham, MA, USA) (excitation, 530 nm; emission, 590 nm). The relative fluorescence units (RFU) were accumulated over the entire 8 h, normalized to PMNs activated with PMA (100 nM, positive control), and are shown as percent. The experiment was performed in five biological replicates.

Measurement of ROS release

Production of ROS by PMNs was analyzed using the Diogenes-based chemiluminescence kit (National Diagnostics, Atlanta, GA). The assay was adapted from

a method described previously (240, 243, 244). A total of 250,000 PMNs and 45 μg/mL hIntL-1 were combined in assay medium (1× HBSS containing 5% autologous serum, 5 mM glucose, 10 mM HEPES, and 5 mM $CaCl_2$) in a solid white 96-well plate. Luminescence measurements were taken for 2 h at 37°C using a Varioskan Flash microplate luminometer (Thermo Scientific, Waltham, MA, USA). The relative luminescence units (RLU) were accumulated over the entire 2 h and normalized to PMNs activated with PMA (100 nM) and are shown as percent. The experiment was performed in seven biological replicates.

Measurement of neutrophil elastase release

The release of enzymatically active neutrophil elastase by PMNs was measured using the neutrophil elastase activity assay kit (Cayman Chemical) by following the manufacturer's instructions. First, 250,000 PMNs and 45 μg/mL hIntL-1 were combined in assay medium (1× HBSS containing 5% autologous serum, 5 mM glucose, 10 mM HEPES, and 5 mM $CaCl_2$) in a black clear-bottom 96-well plate. Fluorescence measurements were then taken for 2 h at 37°C using a Varioskan Flash microplate luminometer (Thermo Scientific, Waltham, MA, USA). Relative fluorescence units were accumulated over the entire 2 h, and activity of neutrophil elastase (in microunits per milliliter) were calculated using a standard curve. The experiment was performed in six biological replicates.

Flow cytometry

S. pneumoniae was stained with 5 mM pHrodo iFL green STP ester (Thermofisher) in assay medium (1× HBSS containing 5% autologous serum, 5 mM glucose, 10 mM HEPES, and 5 mM $CaCl_2$) with 45 μg/mL hIntL-1 or buffer for 1 h at 37°C. Bacteria were added to 1×10^6 PMNs in assay medium at an MOI of 0.1 (bacteria to PMN) in the presence or absence of 45 μg/mL hIntL-1. The infection mixture was incubated for 1 h at 37°C with regular perturbation to avoid settling. Samples were washed and resuspended in PBS before staining with the Zombie Aqua Fixability kit (BioLegend) at a final dilution of 1:5,000 for 15 min at RT. The samples then were washed and resuspended in 1% BSA in PBS before incubation with Fc blocking reagent (Miltenyi Biotec) at a final dilution of 1:100 for 10 min at RT, followed by the addition of Alexa Fluor 647 mouse anti-human CD66b antibody (BD Biosciences) to a final concentration of 1 μg/mL and further incubation at RT for 25 min. In the final step, all samples were washed and resuspended in 300 μL 1× BD stabilizing fixative. During the entire preparation, samples were kept in the dark. For flow cytometry, the same number of PMNs was analyzed for all conditions. The following lasers and filters were used during analysis: Zombie Aqua at 405 nm with the 530/30 filter, CD66B Alexa Fluor at 647 nm with the 660/20 filter, and pHrodo green at 488 nm with the 530/30 filter. The flow cytometry experiments were performed at the University of Georgia College of Veterinary Medicine Cytometry Core Facility using a NovoCyte Quanteon 4025 with NovoSamplerQ utilizing NovoExpress software v.1.4.1. The gating strategy is shown for one representative sample. For each sample, an identical control was included that was left on ice during infection to inhibit phagocytosis, and all controls showed less than 20%

PMN positive for bacterial uptake. The assay was performed in four biological replicates; means and standard errors are shown.

Microscopy and quantification of bacterial phagocytosis

To visualize and quantify bacterial phagocytosis by PMNs, the neutrophil killing assay was executed as described above, with the following changes. *S. pneumoniae* was stained with the pHrodo green AM intracellular pH indicator (Thermo Fisher Scientific) for 30 min at 37°C by following the manufacturer's instructions. Next, 4×10^7 bacteria and 4×10^6 PMNs were incubated at an MOI of 5 (bacteria to PMN) for better visualization of the bacteria. After the 2-h incubation period at 37°C, paraformaldehyde was added to the samples to a final concentration of 5% for 15 min at RT. Samples were applied to a microscopy slide, dried at RT, rinsed in H_2O, and stained with DAPI. During the assay and microscopy preparation, samples were kept in the dark. Under equal exposure conditions, neither pure PMNs nor pHrodo-labeled *S. pneumoniae* showed any green fluorescence. Confocal microscopy was performed on a Zeiss LSM 710 at ×100 magnification. For quantification of bacterial phagocytosis, the microscopy images were analyzed for green fluorescence intensity with ImageJ software. A total of 150 PMNs for each condition from three independent replicates were analyzed; shown are means and standard errors.

Acknowledgements

We thank Fikri Avci for providing access to the cell culture room for A549 lung cell attachment assays. We also thank the healthy volunteers for their blood donations and the staff of the UGA Health Center and the UGA Clinical and Translational Research Unit

for drawing blood used in studies, supported by the National Institutes of Health (5R01HL136707 and R21AI147097 to B.R.). S.A. was supported by NIGMS training grant award number T32GM107004 and NIDCR grant number R01DE022559 to C.M.S.

We declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

CHAPTER 3

INVOLVEMENT OF THE *STREPTOCOCCUS MUTANS* PGFE AND GALE 4-EPIMERASES IN PROTEIN GLYCOSYLATION, CARBON METABOLISM AND CELL WALL BIOSYNTHESIS[2]

[2] Silke Andresen, Nicholas de Mojana di Cologna, Stephanie Archer-Hartmann, Ashley M. Rogers, Sandip Samaddar, Tridib Ganguly, Ian M. Black, John Glushka, Kenneth S. Ng, Parastoo Azadi, José A. Lemos, Jacqueline Abranches and Christine M. Szymanski, 2022. *Submitted to Glycobiology*
Author contribution: SA and CMS designed the study in collaboration with JAL and JA. SS constructed the *S. mutans* OMZ175 strain for tCnm expression, SS and TG expressed and purified Cnm and tCnm from *S. mutans*. SA-H and PA conducted MS analysis and IMB executed compositional analysis of Cnm glycosylation. KSN conducted computational analysis and modeling on PgfE, GalE and Gne. AMR constructed the *E. coli* expression vectors for PgfE, GalE and Gne. JG operated the NMR spectrometers and NMC performed the growth experiments of *S. mutans* strains in different media.

Abstract

Streptococcus mutans is a key pathogen associated with dental caries and is often implicated in infective endocarditis. This organism forms robust biofilms on tooth surfaces and can use collagen binding proteins (CBPs) to efficiently colonize collagenous substrates including dentin and heart valves. One of the best characterized CBPs of *S. mutans* is Cnm that contributes to adhesion and invasion of oral epithelial and heart endothelial cells. These virulence properties were subsequently linked to post-translational modification (PTM) of the Cnm threonine rich repeat region by the Pgf glycosylation machinery, which consists of four enzymes: PgfS, PgfM1, PgfE and PgfM2. Mutation of the *S. mutans pgf* genes leads to decreased collagen binding, reduced invasion of human coronary artery endothelial cells and attenuated virulence in the *Galleria mellonella* invertebrate model. The present study aimed to better understand Cnm glycosylation and to characterize the predicted 4-epimerase, PgfE. Using a truncated Cnm variant containing only two threonine rich repeats, mass spectrometric analysis revealed extensive glycosylation with HexNAc2. Compositional analysis, complemented with lectin blotting, identified the HexNAc2 moieties as GlcNAc and GalNAc. Comparison of PgfE with the other *S. mutans* 4-epimerase GalE through structural modeling, NMR and capillary electrophoresis, demonstrated that GalE is a UDP-Glc-4-epimerase while PgfE is a GlcNAc-4-epimerase. While PgfE exclusively participates in protein O-glycosylation, we found that GalE affects galactose metabolism and cell wall biosynthesis. This study further emphasizes the importance of O-linked protein glycosylation as well as surface glycosylation in *S. mutans* and identifies the PTM modifications of the key CBP, Cnm.

Introduction

Streptococcus mutans is a natural inhabitant of the human oral cavity and is one of the major etiological agents associated with dental caries (190, 245). Apart from the attachment of the pathogen to hard tooth enamel, *S. mutans* can gain access to the bloodstream and adhere to heart valves causing infective endocarditis (189). To form a biofilm, *S. mutans* employs sucrose-dependent and sucrose-independent mechanisms of colonization. In the extensively studied sucrose-dependent mechanism, biofilm establishment is driven by the activity of glucosyltransferases (GtfB, C and D) that utilize sucrose to synthesize copious amounts of adhesive glucan polymers that become the major component of the biofilm matrix. It follows that glucans promote attachment to solid surfaces, recruit other microorganisms to the growing biofilm, and serve as a scaffold for biofilm development and architecture (246). The sucrose-independent mechanism relies on surface adhesins that recognize specific ligands such as the salivary agglutinin GP340 and components of extracellular matrix such as collagen and laminin (189). For example, several collagen binding proteins (CBPs) have been identified in *S. mutans*, two of them are part of the core genome, SpaP (also known as P1 or AgI/II) and WapA (247–250). The other two CBPs, known as Cnm and Cbm, are present in approximately 20% of clinical isolates, with Cnm being more prevalent than Cbm (>90 %) (205, 251). Notably, the presence of Cnm^+ *S. mutans* has been associated with severe caries, infective endocarditis, IgA nephropathy and cerebral microbleeds (184, 191, 251–258). Moreover, past studies have shown that Cnm is a virulence factor that mediates invasion of endothelial and epithelial cells and colonization of heart valves (191, 250, 259).

Cnm is a surface-associated protein containing an N-terminal secretion signal, a conserved collagen binding domain (A-domain), a threonine rich domain (B-domain), and a C-terminal LPxTG cell-wall anchoring motif (252, 260). A few years ago, we discovered that the *cnm* gene is co-transcribed with a previously uncharacterized glycosylation locus, consisting of genes encoding a glycosyltransferase (PgfS), a UDP-4-epimerase (PgfE), and two transmembrane proteins (PgfM1 and PgfM2) (204, 210). In the absence of the *pgf* operon, Cnm showed a faster migration pattern on SDS-PAGE (lower molecular weight), which might be associated with the loss of protein glycosylation. Indeed, the detection of proteins with wheat germ agglutinin (WGA) in the *S. mutans* wildtype strain lysate, but not in Δ*pgf* lysates, strongly suggested the attachment of *N*-acetylglucosamine (GlcNAc) moieties to target proteins, including Cnm, through the Pgf machinery (204, 210). The deletion of individual *pgf* genes or of the entire *pgf* operon resulted in reduced proteolytic resistance, decreased collagen binding and human coronary artery endothelial cell invasion, and attenuated virulence in the *G. mellonella* invertebrate model (204).

Threonine and serine residues are known targets of protein O-glycosylation in eukaryotes as well as prokaryotes (261). One of the best characterized examples of O-linked glycoproteins are mucins, which contain excessive numbers of serine and threonine rich repeating units leading to a bottle brush-like glycosylation pattern coating the protein surface (262, 263). The threonine rich repeat B-domain of Cnm from *S. mutans* OMZ175 constitutes roughly one-third of the protein and contains a total of 64 threonine residues within 21 repeating units, which is likely the target of Pgf modification (210). Of interest, the four *pgf* genes are part of the *S. mutans* core genome, with subsequent studies identifying the cell-wall associated WapA as a target of Pgf modification (204). The four

proteins of the Pgf pathway have not yet been characterized, but computational analysis predicts that both PgfM1 and PgfM2 are membrane embedded enzymes while PgfS is a membrane associated GT-A type glycosyltransferase. In addition, the active form of PgfE is predicted to be an intracellular homodimer with NAD^+ dependent UDP-4-epimerase activity (204). Four-epimerases are isomerases that catalyze the reversible inversion of configuration of the 4' hydroxyl group in nucleotide activated UDP-glycans. Among others, this reaction leads to the interconversion of UDP-Glc and UDP-Gal, as well as of UDP-GlcNAc and UDP-GalNAc and is typically executed in an NAD^+ dependent manner. As demonstrated by Beerens et al, UDP-4-epimerases can be grouped together depending on their substrate specificity (264). Epimerases involved in the interconversion of UDP-Glc and UDP-Gal are considered group 1 enzymes and will also be referred to as using UDP-Hex substrates for the purpose of this study. Group 3 epimerases catalyze the interconversion of UDP-GlcNAc and UDP-GalNAc, both UDP-N-acetylhexosamines (UDP-HexNAc), while epimerases in group 2 catalyze both reactions. The substrate specificity of the UDP-4-epimerases is determined by the amino acid composition in the sugar-binding pockets. Of special interest within the binding pocket is the "gatekeeper" amino acid, which can be used as a predictor of enzyme catalytic activity. UDP-4-epimerases with bulky amino acids in the gatekeeper position are likely to be specific for smaller UDP-Hex substrates, while small gatekeeping residues allow for the acceptance of UDP-HexNAc substrates (264).

In *S. mutans*, PgfE is not the only UDP-4-epimerase. GalE shares ~39 % protein identity over 98 % of the PgfE amino acid sequence (NCBI protein BLAST) and is encoded by *galE*, a gene that is distant from the *pgf* operon. GalE is a known UDP-Gal/Glc-4-

epimerase that carries out the final step of the Leloir pathway of galactose metabolism (214). The Leloir pathway is catalyzed by three enzymes; GalK (galactokinase), GalT (galactose-1-phosphate uridylyltransferase) and GalE (UDP-galactose 4-epimerase). GalK phosphorylates intracellular galactose to Gal-1-P, followed by a combined reaction in which Gal-1-P is combined with UDP-Glc and converted to UDP-Gal plus Glc-1-P by GalT. Subsequently, GalE epimerizes UDP-Gal back to UDP-Glc (213, 214). Combined, this leaves *S. mutans* with two predicted 4-epimerases, including one of them (PgfE) with unknown substrate specificity. In this study, we aimed to better understand Cnm glycosylation by the Pgf machinery by first determining the enzymatic activity of PgfE, and then elucidating the specific roles of both 4-epimerases (PgfE and GalE) in *S. mutans*.

Results

Cnm is heavily glycosylated with HexNAc2

Previous studies suggested PgfE was involved in Cnm glycosylation in *S. mutans*. Deletion of *pgfE* led to a shift in the molecular weight of Cnm and to the loss of WGA recognition (204). To begin understanding the function of PgfE, we first analyzed Cnm glycosylation by mass spectrometry (MS) to gain a basic understanding of the function of the Pgf pathway as a whole. Glycomics analysis of full length Cnm purified from the *S. mutans* wildtype strain OMZ175 provide a global understanding of the glycosylation patterns for native full-length Cnm. Interestingly we found that the primary glycoform of Cnm was HexNAc2 which provided a starting point for better mapping out the protein. Further mass spectrometric analysis of illustrates the difficulties arising from the properties of heavily glycosylated glycoproteins. The full length Cnm, with its 21 threonine-rich

repeating units as potential targets for O-glycosylation, proved difficult to properly digest and analyze peptides using conventional methods such as trypsin, chymotrypsin, GluC and other enzymatic treatments (data not shown). Due to the threonine-rich repeating region's size and resistance to enzymatic degradation, the majority of the region was left undigested. This results in both heterogeneity of the resulting peptide fragments as well as pre-existing heterogeneity resulting for multiple and variable locations of the glycosylation. Therefore, the Cnm digest only yielded degradation and partial cleavage peptides of a small enough size to be picked up with softwares like Byonic, followed by manual interpretation. Additionally, the utility of commercial software are limited for this protein due to the heavy total amount of glycan modification regarding not only the number of sites, but also their location immediately next to each other. Nevertheless, our HCD data from a fragment peptide of full-length Cnm showed the presence of HexNAc2 attached to at least four different threonine residues (Fig. 3.1A). The glycopeptide analysis for the full length Cnm also indicated the portions of proteins outside of the threonine-rich region were not glycosylated (Fig. 3.7). The glycomics results combined with, albeit limited, mass spectrometry data indicate that full-legth Cnm is heavily modified with sites of glycosylation in the threonine rich region, and that these glycans are comprised of HexNAc2.

For a better understanding of the glycosylation pattern of Cnm, a truncated version of Cnm (tCnm) was generated and purified from its native host (Fig. 3.10). tCnm contains only two true threonine rich repeating units (and one imperfect repeat) with a total of 45 threonine and 28 serine residues in the along the entire truncated protein as potential targets for O-glycosylation (Fig. 3.1B). The removal of multiple threonine-rich repeats greatly

improved the MS analysis (Fig. 3.1D and Fig. 3.8). tCnm digestion resulted in a T-rich peptide small enough for LC-MS/MS analysis with assistance by a commercial software package (Byonic) and a high-quality match with the peptide TTTTTEKPTTTTEAPTTTEAPPNKPSGK containing seven sites of HexNAc2 as shown in Figure 1D. This result prompted us to manually search for additional peptides with variable amounts of HexNAc2 occupancy by extracting ion abundances from the full mass data. Abundances of peptides containing full occupancies (assuming HexNAc2 for each site) of all 12 sites of glycosylation were observed while the majority of analyzed peptides showed glycosylation on 9-11 threonine residues (Fig. 3.1C). The y8 ion observed in Figure 1D corresponding to the peptide not covered by the T-repeated region and was observed in nearly all versions of the peptide in the truncated Cnm. This indicates a potential preference for the threonine-repeating region over the exposed serine residues. A manual deconvolution and inspection of the data showed small amounts of "odd-numbered" additions of HexNAc indicating the addition of either a single HexNAc or (HexNAc)x (where x = 3, 5, etc). Byonic results, previous glycomics data, and manual inspections for oxonium ions infer that the most likely situation is the addition of a single HexNAc in those instances, or are potential losses in the electrospray used for MS analysis. Combined, the data collected for tCnm suggest abundant glycosylation of Cnm with HexNAc2 moieties, particularly in the threonine rich repeat region.

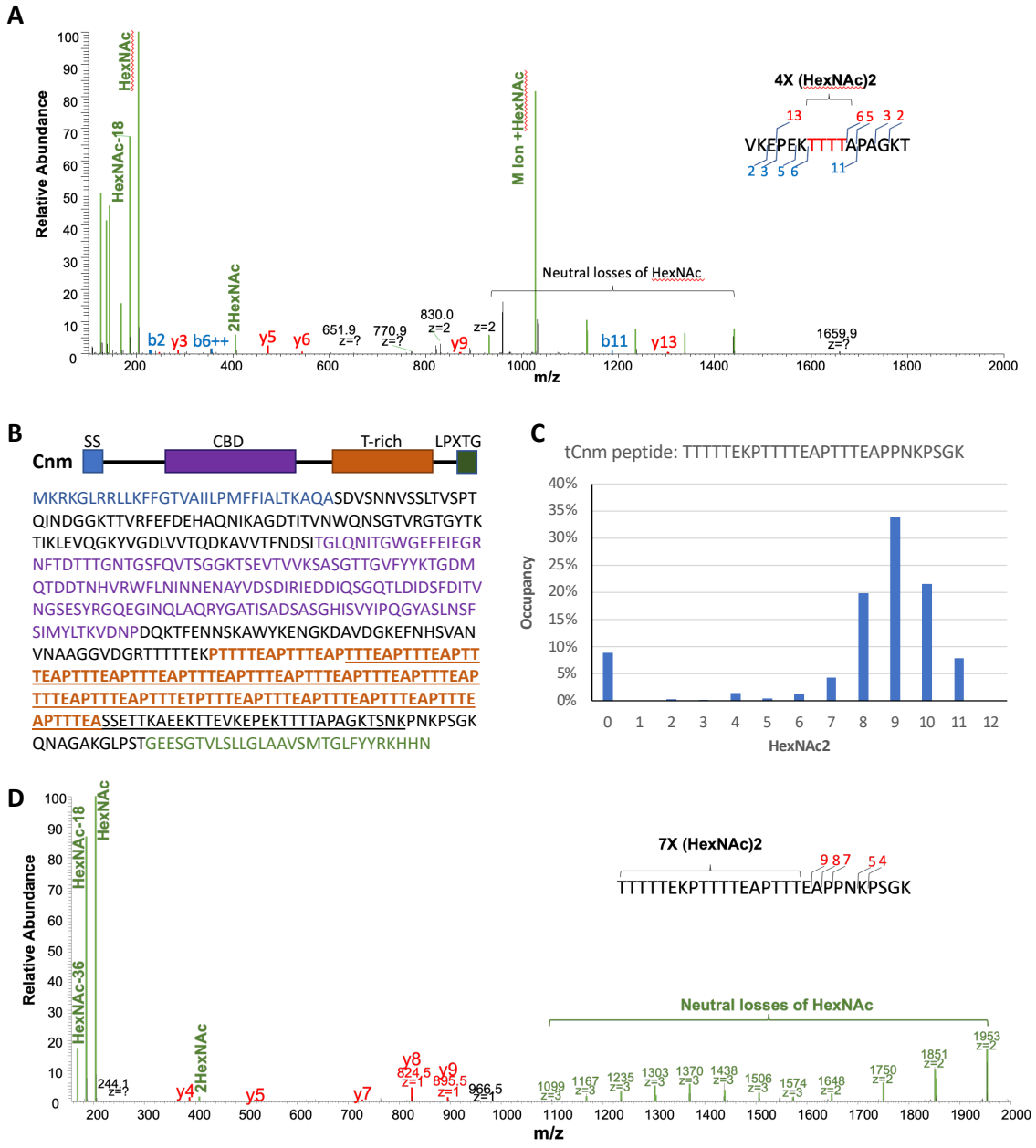


Figure 3.1: Mass spectroscopic analysis of Cnm. (A) Native Cnm was purified from *S. mutans* OMZ175 and analyzed via MS LC-MS/MS following trypsin digestion. Shown is an exemplary MS2 spectrum (HCD fragmentation) for the indicated peptide. **(B)** A schematic of Cnm and the full amino acid sequence of the protein are shown. Cnm contains a signal sequence (SS), collagen binding domain and LPXTG cell wall anchor. The threonine-rich region contains 21 threonine rich repeat regions, totaling 64 threonine residues. The truncated version of Cnm was designed (tCnm) to lack all but two threonine rich repeat regions (deleted region is underlined). tCnm contains only 2 out of 21 repeats and was purified directly from *S. mutans* wildtype. **(C)** tCnm protein was enzymatically digested by trypsin and analyzed via LC-MS/MS. The glycan occupancy of 12 threonine residues in the T-rich region of tCnm is shown as percent occupancy with HexNAc2, which

was manually determined by extracting ion abundances of each glycopeptide. **(D)** An exemplary MS2 spectrum (HCD fragmentation) for the indicated peptide of tCnm is shown.

Cnm is glycosylated with GalNAc and GlcNAc residues

Apart from GlcNAc, GalNAc is another common HexNAc utilized by several bacterial species (particularly those with 4-epimerases). We therefore tested soybean agglutinin (SBA) in a lectin blot to specifically probe for the presence of GalNAc residues on Cnm (Fig. 3.2). SBA shows strong binding to GalNAc, as well as recognition of galactose to a lesser extent. As shown in Fig. 3.2A-B, SBA recognized at least four proteins within whole cell lysates of *S. mutans* OMZ175 and the *ΔpgfM2* mutant, while no GalNAc binding was observed for the *Δpgf* (deletion of entire locus), *ΔpgfS*, *ΔpgfM1* and *ΔpgfE* strains. While the lysate for the Δ*pgf* strain shows the detection of the highest molecular weight band as found in the wildtype, this is most likely caused by overloading of the well as the band was not present in repeats of the experiment with decreased loading volume. The recognition of glycosylated proteins in the whole cell lysate of the *ΔpgfM2* strain suggests that PgfM2 is not involved in GalNAc addition and likely serves another role. PgfS, PgfM1 and PgfE, however, are crucial for the addition of GalNAc residues to Pgf target proteins. Native Cnm, affinity purified from OMZ175 was further examined to confirm the identity of its glycan moieties by compositional analysis. Despite the presence of large quantities of *S. mutans* surface glycans such as rhamnose (likely derived from the rhamnose glucose polysaccharide, RGP), the presence of both GlcNAc and GalNAc in the Cnm sample was verified (Fig. 3.9). And SBA recognition of more than one protein band in the wildtype lysate further suggests that Pgf glycosylation is not only targeting Cnm, but at least three additional proteins by the O-glycosylation pathway. We therefore analyzed

the protein bands excised from the SDS-PAGE via MS (Fig. 3.2A, excised bands are indicated by arrows) and compared the identified proteins to the predicted target proteins in *S. mutans* OMZ175 in a previous study (204). Besides Cnm and WapA, the NetOGlyc 4.0 glycoprotein prediction server predicted the surface associated proteins SpaP and GtfC as targets for O-glycosylation. Both SpaP and GtfC, as well as Cnm and WapA were identified in the excised gel bands, suggesting them as likely targets for the Pgf system (Table 3.2). These proteins were present in multiple gel pieces, likely due to protein degradation during lysate preparation.

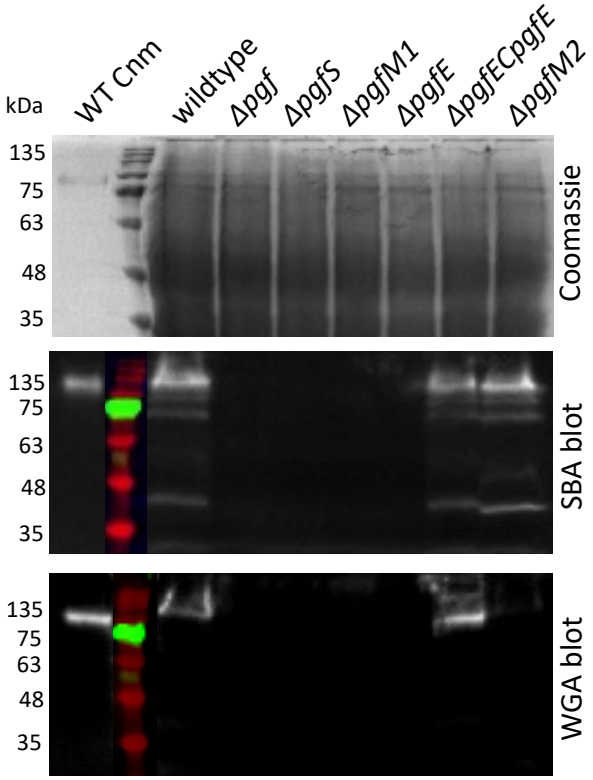


Figure 3.2: Detection of GalNAc and GlcNAc residues in *S. mutans* whole cell lysates using lectins. Whole cell lysates of *S. mutans* strains were prepared and separated by SDS-PAGE. Proteins were transferred onto a nitrocellulose membrane and probed with HRP-conjugated soybean agglutinin. Shown are the Coomassie stained SDS-PAGE, a SBA lectin blot and a WGA lectin blot. In the lectin blot, the BLUEstain protein ladder (goldbio) and sample proteins were visualized using different fluorescent filters, resulting in their representation in color or black and white, but the ladder and the wildtype sample were run in adjacent wells on the same gel, as depicted.

PgfE and GalE belong to different groups of UDP-4-epimerases

The utilization of N-acetylated hexosamines by the Pgf machinery led us to investigate the function of PgfE, the proposed UDP-4-epimerase of the Pgf system alongside GalE, the second UDP-4-epimerase present in *S. mutans.* GalE, which is part of the Leloir pathway, shares a 39 % identity over 98 % amino acid coverage with PgfE (NCBI protein BLAST). In addition to the *S. mutans* epimerases PgfE and GalE, we included the epimerase Gne from *Campylobacter jejuni* as a control. Gne has been shown to catalyze the interconversion of UDP-Glc and UDP-Gal as well as UDP-GlcNAc and UDP-GalNAc and an unpublished crystal structure has been deposited in the PDB (7K3P) (265). Molecular models of the two epimerases from *S. mutans* were constructed using Alphafold-2 and compared to the crystal structure of the well-characterized WbpP epimerase from *P. aeruginosa* (PDB 1SB8), which has a strict specificity for UDP-GlcNAc and UDP-GalNAc (266) (Fig. 3.3A). These comparisons highlight the importance of the side chains of two key residues, including the "gatekeeper" residue lying nearest to the 2-position of UDP-hexose substrates (as reviewed by Beerens et al. 2015a) (Fig. 3.3A-B). The model of PgfE suggests a structure that is similar to that seen in the crystal structure of *P. aeruginosa* WbpP bound to UDP-GalNAc, which is a group 3 epimerase with a small serine residue in gatekeeper position. PgfE shows the formation of a binding pocket with steric and electrostatic complementarity to the N-acetyl modification of the 2-position of the UDP-hexose substrate (Fig. 3.3A-D). The model of *S. mutans* GalE shows how the larger and non-polar side chain of Leu at the equivalent position may close off this pocket. This bulky residue can be found in group 1 epimerases and may help to explain the enzyme's preference of UDP-hexose substrates lacking the N-acetyl modification (Fig.

3.3E). Although *C. jejuni* Gne also has Leu at the same position, the crystal structure of the apo-enzyme indicates that the substrate-binding pocket is expanded, allowing sufficient space to accommodate the N-acetyl modification at the 2-position of UDP-hexose substrates (Fig. 3.3F).

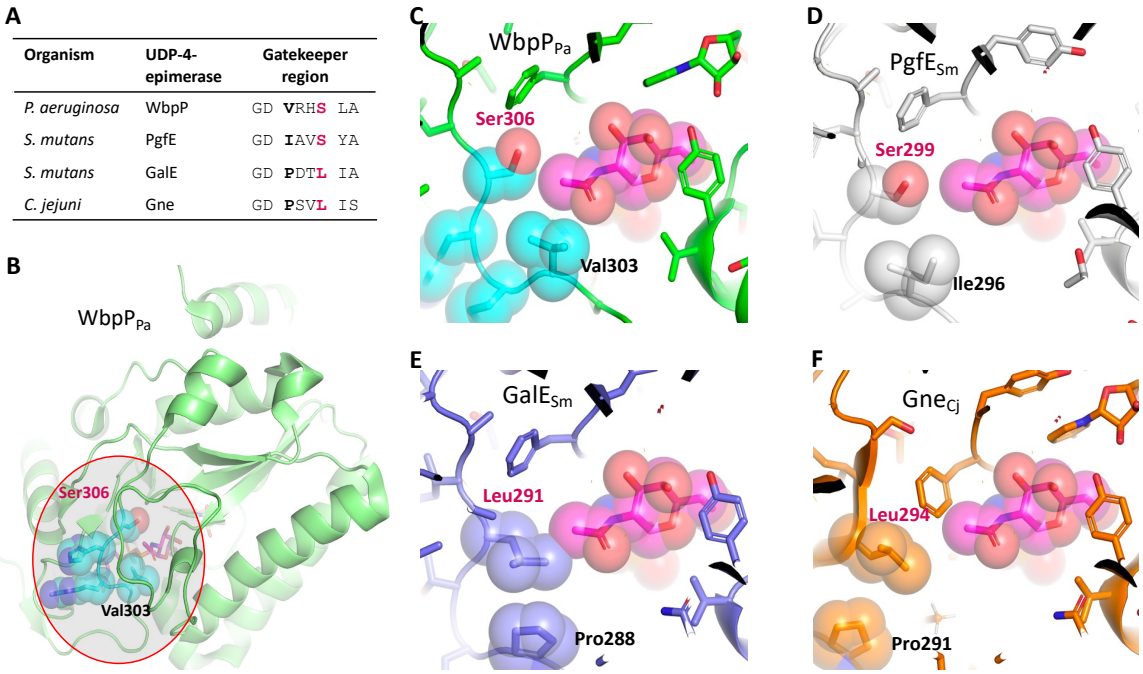


Figure 3.3: Sequences and structures of the "gatekeeper" region in UDP-hexose epimerases. (A) Amino acid sequences corresponding to the "gatekeeper" region; the residues with side chains closest to the 2-N-acetyl group of UDP-GalNAc are highlighted by bold font and the gatekeeping residue in red following the work of Beerens et al. (264). **(B,C)** Crystal structure of *P. aeruginosa* WbpP bound to UDP-GalNAc (PDB 1SB8); the red circle highlights the binding pocket for the hexose moiety and the gatekeeper region. Molecular models of **(D)** PgfE and **(E)** GalE from *S. mutans* were constructed by AlphaFold CoLab v2.1.0 (267). **(F)** Crystal structure of *C. jejuni* Gne (PDB 7K3P). An approximate model for the location of UDP-GalNAc (magenta) was determined by superimposing various structures onto the crystal structure of *P. aeruginosa* WbpP bound to UDP-GalNAc to evaluate potential binding interactions and steric clashes.

PgfE uses UDP-HexNAc while GalE uses UDP-Hex sugars

To assess the substrate specificity of PgfE in comparison to GalE and Gne, the three epimerases were expressed in *E. coli* and purified for *in vitro* analysis. The three

epimerases showed a similar size of approximately 35 kDa as well as the possible formation of dimers, even in reducing SDS-PAGE (Fig. 3.4A).

For NMR analysis, the purified enzymes were incubated with NAD^+ and different UDP-sugars. As expected, the control epimerase Gne from *C. jejuni* showed the conversion of UDP-Glc to UDP-Gal and UDP-GlcNAc to UDP-GalNAc, while PgfE catalyzed the epimerization of UDP-GlcNAc to UDP-GalNAc and GalE converted UDP-Glc to UDP-Gal (Fig. 3.4B).

To confirm the substrate specificity observed with NMR, we further employed capillary electrophoreses (CE) as a more sensitive detection method for nucleotide sugars. As shown in Fig. 3.4C, PgfE interconverts UDP-HexNAc sugars, GalE epimerizes UDP-Hex sugars and Gne catalyzes both reactions. The use of both glucose (UDP-Glc/GlcNAc) and galactose (UDP-Gal/GalNAc) based substrates allowed the visualization of the glycan equilibrium for each reaction. In all mentioned samples, the amount of the glucose sugar at the reaction endpoint was higher than the galactose sugar, regardless of the added substrate. This suggests an equilibrium of the reactions that favors the more stable glucose configuration of the glycan. While CE confirmed all enzymatic specificities that were observed by NMR, it further allowed the visualization of an additional reactivity of PgfE as the enzyme showed high affinity for UDP-HexNAc glycans as well as lower affinity for UDP-Hex sugars. While both UDP-Glc and UDP-Gal substrates were epimerized by PgfE, the reactions did not reach equilibrium and less product was formed than in the other reactions, indicating only a weak PgfE affinity for UDP-Hex as a substrate.

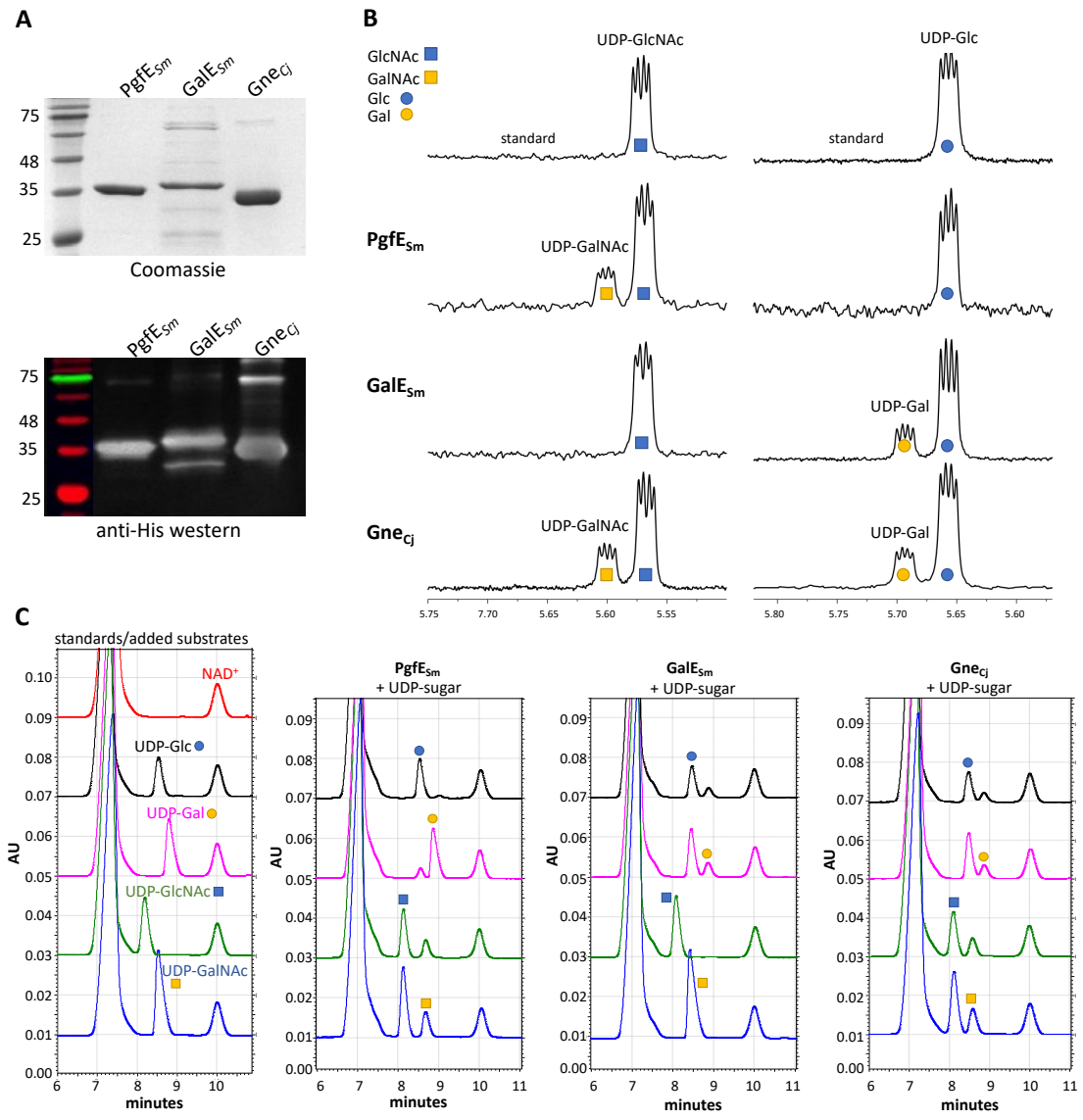


Figure 3.4: The substrate specificity of the epimerases was analyzed using NMR and CE. (A) PgfE and GalE from *S. mutans*, as well as Gne from *Campylobacter jejuni* were cloned into expression vectors including C-terminal hexa-His tags. The proteins were expressed in *E. coli* and purified using a nickel-NTA column. Shown are the eluted epimerases used in this study (Coomassie stain and anti-polyhistidine western blot). Molecular weights are indicated in kDa. **(B)** The epimerases were expressed and purified from *E. coli*. Then, 25 ng/µl protein, 1 mM NAD+ and 300 µM UDP-sugar in PBS were incubated for 1 h at 37 °C. Regions of 1D proton NMR spectra show the anomeric protons for UDP-sugars. The top trace shows starting material and lower traces the results after addition of the indicated enzymes. Spectra showing the epimerase products correspond to less than 15 minute reactions, whereas spectra without products were obtained after 60 minute reactions. Peaks are labeled with SNFG symbols for Glc, Gal, GlcNAc and GalNAc (35). **(C)** For the analysis of the substrate specificity with CE, 250 ng/µL, 0.5 mM NAD+ and 100 µM UDP-sugar in Tris buffer were incubated for 1 h at 37 °C. All spectra shown

in the same lane contain the same sugar substrate as indicated in the standards (left column), the substrate peak for each sample is additionally indicated by the labeling with the SNFG glycan symbol.

PgfE, but not GalE, is involved in protein glycosylation

The collected data on Cnm glycosylation suggested PgfE would be important for GlcNAc/GalNAc addition. To compare and contrast the roles of the two epimerases during protein O-glycosylation, a variety of assays were employed. We probed western blots with anti-Cnm antibodies to examine the molecular weight of Cnm in whole cell lysates derived from the *S. mutans* wildtype and *pgfE* and *galE* epimerase-deficient mutants (Fig. 3.5A). As described before, loss of *pgfE* or the entire *pgf* operon resulted in a reduction in Cnm protein size, suggesting the absence of glycosylation. The deletion of *galE* however had no effect on the migration pattern of Cnm, indicating that GalE plays no part in Cnm glycosylation. As expected, the double mutant, missing both *S. mutans* 4-epimerases showed the same change in molecular weight as the *ΔpgfE* single mutant (Fig. 3.5A). When combined with the SBA reactivities and NMR analyses, our data indicates that PgfE, but not GalE, plays a role in protein O-glycosylation in *S. mutans*.

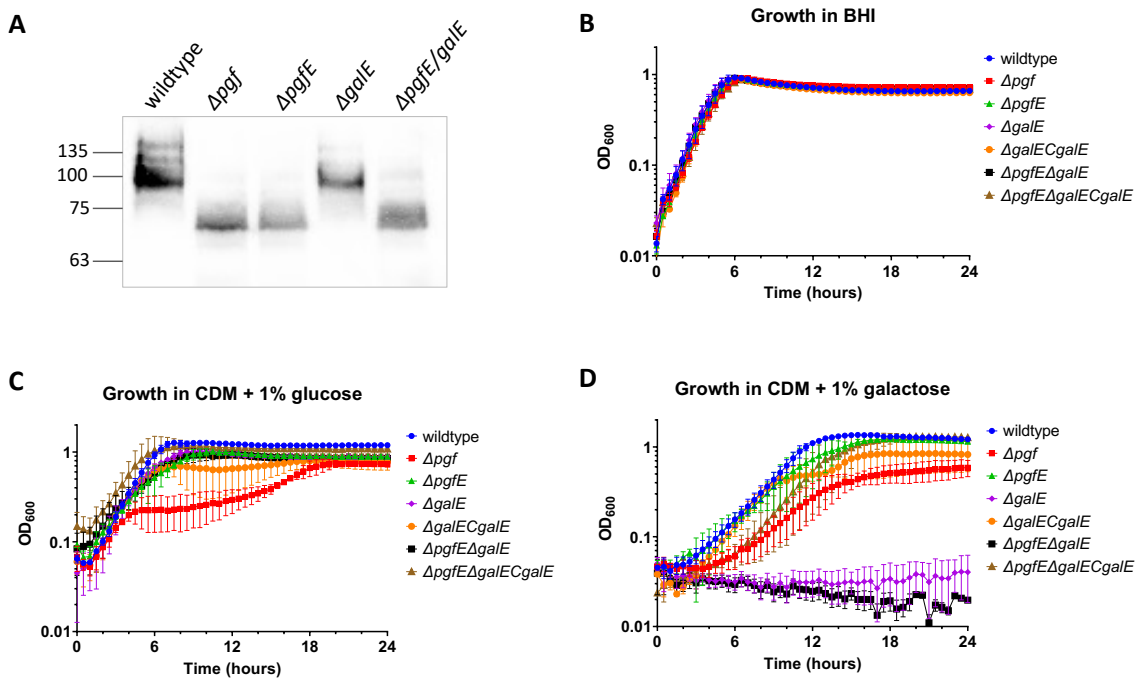


Figure 3.5: PgfE plays a role in protein glycosylation, while GalE is important for galactose metabolism in *S. mutans*. (A) Whole cell lysates of *S. mutans* strains were separated by SDS-PAGE, transferred onto a nitrocellulose membrane and probed with an anti-Cnm antibody. The bands corresponding to Cnm detection were excised from SDS-PAGE and analyzed via MS for the presence of Cnm and glycosylation with HexNAc2. **(B-D)** *S. mutans* wildtype and indicated mutant strains were grown on BHI or CDM + 1% sugar as carbon source as indicated.

GalE, but not PgfE, impacts growth on galactose

As an enzyme of the Leloir pathway, GalE has been implicated as a key factor in galactose metabolism since disruption of the Leloir pathway leads to the inability of *S. mutans* to grow on medium containing galactose as the only carbon source (shown for *ΔgalT* and *ΔgalTE*) (214, 268). Here, we aimed to investigate a possible secondary role of PgfE in galactose metabolism. To this end, single and double *galE* and *pgfE* mutants (*ΔgalE*, *ΔpgfE* and *ΔgalEΔpgfE*) were grown on either brain heart infusion (BHI) or in a chemically defined media (CDM) in the presence of different carbon sources (Fig. 3.5B-D). All strains showed lower final yield, but growth rates comparable to the *S. mutans*

wildtype on BHI (Fig. 3.5B). The *Δpgf* and *ΔpgfE* mutants phenocopied the wildtype strain when grown in either glucose or galactose, suggesting that PgfE is not important for glucose or galactose metabolism (Fig. 3.5C-D). In contrast, while the *ΔgalE* strain grew as well as the wildtype on glucose, it was unable to use galactose as a carbon source (Fig. 3.5C-D). The double epimerase mutant performed similarly to the *ΔgalE* mutant on both carbon sources ruling out a possible minor role of PgfE in galactose metabolism. Collectively, these results indicate that GalE is the only *S. mutans* UDP-Glc/Gal-4-epimerase essential for galactose metabolism.

Loss of GalE effects *S. mutans* surface glycosylation

In *E. coli*, GalE is implicated in cell wall biosynthesis due to the need for UDP-Glc and UDP-Gal as building blocks for certain bacterial surface glycans (214, 269). In *Lactococcus lactis*, deletion of *galE* led to a defect in bacterial separation and the formation of long chains when cells are grown using glucose as sole carbon source (270). We therefore aimed to investigate the effect of GalE and PgfE loss on cell wall biosynthesis in *S. mutans* strain OMZ175. Our previous growth analysis had shown a significant growth impairment for the *ΔgalE* mutant when given galactose as the sole carbon source. To circumvent this phenotype and allow for bacterial growth, wildtype and mutant strains were grown in low glucose (0.1%) medium supplemented with the addition of excess (1%) glucose or galactose as the main carbon source, before assessing their morphologies through bright field microscopy after crystal violet staining (Fig. 3.6A). While the use of a base medium with low glucose allowed for growth of all strains, the *galE* deletion strain and the epimerase double mutant displayed much lower cell densities after overnight

growth in 0.1% glucose + 1% galactose as was expected (data not shown). As shown in Fig. 3.6A, all strains showed comparable morphologies when grown in glucose consisting of short to medium length bacterial chains and small aggregates. The same arrangements could be observed for the wildtype and *ΔpgfE* strains when grown on galactose. However, the *ΔgalE* and double mutant strain formed long bacterial chains when the base media was supplemented with 1% galactose, indicating a possible defect in cell wall biogenesis (Fig. 3.6A).

To further analyze the cell morphologies observed during bright field microscopy, the wildtype, *ΔpgfE* and *ΔgalE* mutants were examined by transmission electron microscopy (TEM) after growth on low glucose media supplemented with 1% galactose. As shown in Fig. 3.6B, at low magnification, the strains revealed the same cell arrangements as observed by optical microscopy. While wildtype and *ΔpgfE* mutant formed small bacterial aggregates, the *ΔgalE* mutant formed long chains. Under higher magnification, both wildtype and *ΔpgfE* mutant were visibly enclosed by a thick surface structure (red arrows), which is likely to consist of polysaccharides such as the rhamnose glucose polysaccharide (RGP). For the *ΔgalE* strain however, striking phenotypes related to cell morphology could be observed (Fig. 3.6B). While the majority of cells in the *ΔgalE* strain exhibited the same surface composition as the wildtype, some demonstrated a loss of the thick surface stucutre (purple arrow). Secondly, the analysis of the *ΔgalE* mutant through TEM revealed a defect in cell separation since many "long cell units" with multiple septa could be observed (indicated by asterisks, lower 10,000x image Fig. 3.6B).

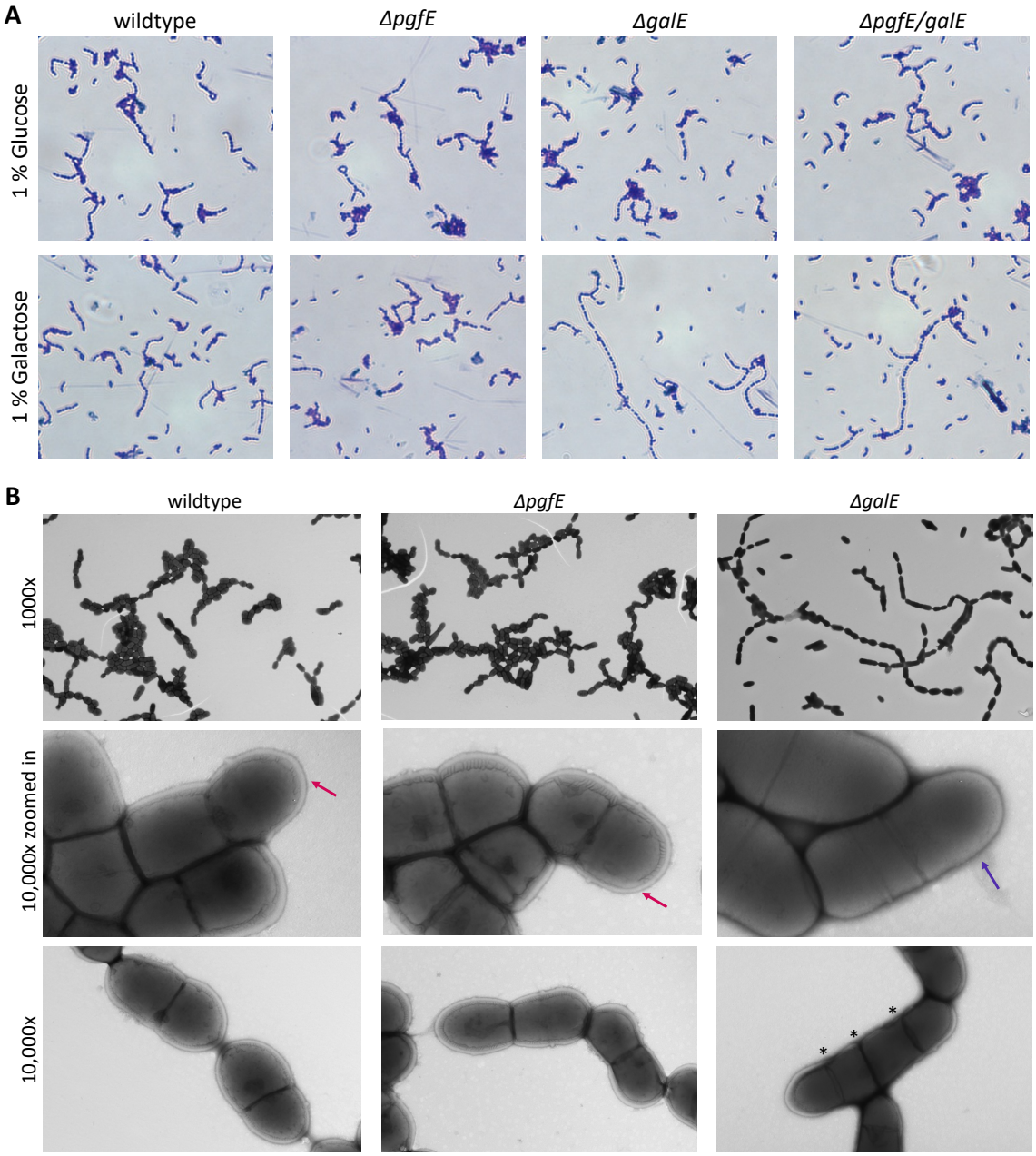


Figure 3.6: The loss of GalE causes an increase in bacterial chain length and changes in surface structure. (A) All *S. mutans* strains were grown on low glucose medium supplemented with 1 % glucose or galactose as the main carbon source overnight. Bacteria were applied to microscopy slides, stained with crystal violet, and imaged at 100x magnification using an Axio A1 light microscope. **(B)** *S. mutans* wildtype*, ΔpgfE and ΔgalE* were grown in 1 % galactose and visualized with TEM. Red arrows indicate the presence of a thick surface structure surrounding the bacteria and the purple arrows indicates the absence of a surface structure. Defects in septum formation are indicated by asterisks.

Discussion

Previous studies have shown that a significant proportion of the molecular weight of Cnm, which migrates on SDS-PAGE as multiple bands in excess of 100 kDa and has a predicted mass of 57.3 kDa, is attributed to glycan modifications added by the Pgf machinery. These glycans provide protection against proteolysis by a variety of proteases including ProK, trypsin and GluC, but also create challenges in the analysis of the sugar moieties (204). To overcome this barrier, we engineered a truncated version of Cnm, named tCnm, that contains only two threonine rich repeat units showing an average of 75-92 % glycan occupancy within the repeat region. Assuming this result translates to the full length Cnm, between 48 - 59 of the 64 threonine residues within the 21 threonine rich repeats would be glycosylated on average at any time. While this is just an extrapolation of occupancy, especially considering the heterogeneity of Cnm glycosylation, this amount of protein O-glycosylation is comparable to what is found in mucins. Muchins are known for their heavily glycosylated serine and threonine rich repeating units creating what is often referred as a bottle brush-like structure (262, 263).

Combining the presented work, as well as our previous studies, we have identified at least four proteins as likely core and non-core genome targets of the Pgf system, Cnm, WapA, SpaP, and GtfC (204). Notably, all four of these proteins are implicated in the initial steps of oral colonization by *S. mutans*. Both SpaP and Cnm are secreted and cell wall anchored proteins (LPXTG motif) that bind collagen and laminin, and in the case of SpaP also fibronectin (247, 250, 252). WapA, the wall associated protein, is also a surface adhesin and collagen binding protein (248, 250). Collagen, laminin and fibronectin are considered among the key components of the host extracellular matrix in the oral cavity,

making them ideal targets for adhesion during tooth colonization. Futhermore, SpaP, Cnm and WapA have been shown to form amyloid structures, which are often implicated in biofilm formation (271–273). While SpaP is considered an important factor during sucrose-independent biofilm formation, GtfC is employed during sucrose dependent surface adhesion (274, 275). Given the roles of these putative Pgf target proteins, it is not surprising that in the absence of Pgf glycosylation, *S. mutans* is impaired during collagen binding, host cell invasion, and *G. mellonella* infection (204). Future work will be directed at purifying these proteins to confirm that they are indeed modified.

The Pgf system modifies its target proteins with HexNAc2 moieties, predominantly comprised of GalNAc residues with minor additions of GlcNAc. While this study aimed to elucidate the role of PgfE as a UDP-GlcNAc/GalNAc 4-epimerase, less is known about the other proteins of the O-glycosylation pathway. PgfS is predicted to be a membrane-bound CAZy family 2 glycosyltransferase which transfers nucleotide activated sugars to lipid or sugar substrates. Both PgfM1 and PgfM2 are membrane proteins with putative undecaprenyl-phosphate (UndP)-sugar transferase activities. Thus, since Cnm exhibits the same mass shift in *ΔpgfS, ΔpgfE*, and *ΔpgfM1* backgrounds, our current hypothesis is that PgfS transfers UDP-GalNAc to UndP forming Und-P-P-GalNAc. Subsequently, PgfM1 transfers GalNAc from UndPP to threonine. For PgfM2, our previous studies showed a small decrease in Cnm mass in the *ΔpgfM2* strain when compared to the *ΔpgfS, ΔpgfE*, and *ΔpgfM1* strains coupled with loss of WGA (ie GlcNAc) reactivity (204), suggesting that PgfM2 may add GlcNAc to unmodified threonine sites (or to GalNAc residues) from the Und-P-P-GlcNAc precursor required for GlcNAc linker formation during RGP biosynthesis. Furthermore, SBA lectin blots of *ΔpgfM2* lysates in this study showed the

same protein recognition pattern as seen for the wildtype, indicating that PgfM2 is dispensable for GalNAc addition and further supporting our hypothesis. While this mechanism suggests a possible order of events, we are currently overexpressing the other enzymes in order to test their actual activities. In the Streptococcus genus, sequential O-glycosylation of serine-rich repeat proteins has been described in various strains, such as *S. pneumoniae* (for PsrP)*, S. parasanguinis* (for Fab1) and *S. gordonii* (for GspB) (36, 276, 277). In these organisms, O-glycosylation is achieved through sequential transfer of UDP-activated glycans to the target proteins, followed by export of the modified proteins through accessory SecA2 secretion systems (36, 211, 278). Further studies are required to elucidate the specific roles of the remaining enzymes in the *S. mutans* Pgf system.

Our current analyses of the two 4-epimerases in *S. mutans* demonstrated not only separate substrate specificities, but also a corresponding lack of functional redundancy. PgfE showed no involvement in galactose or glucose metabolism and its mutation had no effect on cell wall biogenesis. Instead, PgfE demonstrated major activity as a UDP-HexNAc epimerase during protein O-glycosylation. While UDP-GalNAc and UDP-GlcNAc constitute the main substrates of PgfE, we did observe a weak affinity for UDP-Hex sugars. This additional activity is likely determined by the sugar binding pocket of the epimerase with contributions from the gatekeeper amino acid.

In the case of GalE, the gatekeeping amino acid is leucine, which is a bulkier residue, and leads to the prediction that GalE is a group 1 epimerase with UDP-Glc/Gal specificity. As previously shown, this prediction (213, 214) could be verified from our *in vitro* analysis using NMR and CE. More surprising was the effect of GalE loss on bacterial physiology and cell wall biosynthesis. In contrast to most other gram-positive bacteria, *S.*

mutans lacks the ability to produce charged cell wall teichoic acids as part of its surface glycans, and instead is decorated with a neutral RGPs (197). RGPs are comprised of poly-rhamnose backbones that are assembled on undecaprenylphosphate (ie bactoprenylphosphate) in the bacterial cytoplasm and non-stoichiometrically decorated with glucose (279). The RGP is transported to the cell surface where it is then attached to the peptidoglycan. The synthesis of RGP is executed by enzymes encoded by the *rgp* operon in which the type of glucosyltransferase determines the linkage between glucose and rhamnose and separates *S. mutans* into 4 different serotypes (serotype c, α1,2-linkage; serotype e, β1,2-linkage; serotype f, α1,3-linkage; serotype k, no Glc) (197, 280, 281). The loss of the *rgp* genes has been shown to cause defects in morphogenesis and cell division in *S. mutans* (197). Similarly, the disruption of *galE* from the Leloir pathway in *L. lactis* led to the formation of long bacterial chains when grown on glucose, due to the lack of UDP-galactose as a building block for its lipoteichoic acids (270). In our hands, *S. mutans* displayed the same deficiency for cell separation when grown on low glucose base medium supplemented with 1 % galactose. Our data therefore may suggest that the loss of *galE* as a UDP-Glc/Gal epimerase caused a deficit in UDP-glucose needed for RGP synthesis, but this remains to be confirmed.

In conclusion, this study has demonstrated that Cnm is O-glycosylated with HexNAc2 moieties that are predominantly GalNAc, but with non-stoichiometric GlcNAc additions, and this is accomplished through the activity of the Pgf enzymes. The characterization of the two UDP-4-epimerases in *S. mutans* further showed that PgfE is a UDP-GlcNAc/GalNAc epimerase, solely involved in protein glycosylation, while GalE is a UDP-Glc/Gal epimerase not only involved in galactose metabolism, but also in cell wall

biogenesis. While more studies are needed to elucidate the roles of each enzyme within the Pgf glycosylation machinery, we were able to present the first insights into the enzymatic activities of the system and analyze the highly glycosylated protein Cnm as an exemplary target for Pgf O-glycosylation.

Material and Methods:

Strains and plasmids

For construction of the PgfE expression plasmid, primer sequences were based on the sequence information available at the time (Q8DRZ0_STRMU, UDP-glucose 4-epimerase *S. mutans* UA159 serotype c (ATCC 700610)). The primers for the GalE expression plasmid were designed based on the GeneBank sequence EMC56922.1 (UDP-glucose 4-epimerase of *S. mutans* OMZ175). The primer design for *galE* amplification included the upstream area of the gene (20 amino acids, beginning with Met) to ensure the complete transcription of *galE*, which had not previously been cloned for expression. The primers 1 and 2 (pgfE_NdeI_For and pgfE_XhoI_Rev) as well as 3 and 4 (galE_NdeI_For and galE_XhoI_Rev) were used for the amplification of *pgfE* and *galE* respectively, using genomic DNA from *S. mutans* OMZ175 (Table 3.1). Primers 5 and 6 (gne_NdeI_For and gne_XhoI_Rev) were employed for amplification of *gne* from *C. jejuni* genomic DNA (Q0P9C3_CAMJE, UDP-glucose 4-epimerase, *C. jejuni* subsp. *jejuni* serotype HS:2, strain ATCC 700819 / NCTC 11168). The cloning of the epimerase genes into pET30(a)+ involved standard molecular biology methods. Briefly, PCR amplification was executed using the Q5 High-Fidelity PCR kit (New England Biolabs), followed by restriction digests with NdeI and XhoI (high fidelity enzymes, Thermo Scientific) and ligation with T4 DNA

Ligase (New England Biolabs). This resulted in plasmids pET30_PgfE-His, pET30_GalE-His and pET30_Gne-His that were used to express C-terminally hexa-histidine tagged epimerases in *E. coli* BL21(DE). All expression plasmids were verified through sequencing.

All *S. mutans* strains were routinely grown on BHI at 37 °C in a 5% CO_2 atmosphere. The *S. mutans* OMZ175 *Δpgf*, *ΔpgfS*, *ΔpgfM1*, *ΔpgfE* and *ΔpgfM2* strains have been previously described (204, 210). The *ΔgalE* and *ΔgalEΔpgfE* strains were generated by replacing the *galE* gene with a non-polar erythromycin marker via allelic exchange, and the double mutant created on the background of the *ΔpgfE* single mutant. DNA fragments flanking the *galE* gene were amplified via PCR using the primers listed in Table 3.1. A PstI restriction site was inserted with primer *galE*-R1 and a BamHI restriction site was inserted with primer *galE*-F2. PCR products were digested with either PstI or BamHI and ligated to an erythromycin-resistance cassette previously digested with both enzymes using a T4 ligase. Transformation of OMZ175 and OMZ175*ΔpgfE* with the ligation product was carried out by growing cultures in BHI to an OD_{600} of 0.08 and then adding synthetic competence stimulating peptide to a final concentration of 1 μM, plus 250 ng of ligation product (282). Growth was maintained for four hours, and cultures were plated on BHI agar containing erythromycin (10 μg/mL). Deletion of *galE* was confirmed via Sanger sequencing. Cultures were grown at 37°C in a 5% (v/v) CO_2 incubator.

To construct the *S. mutans* strain expressing a truncated Cnm containing only the first two threonine rich repeats, we used a PCR ligation approach to replace the kanamycin resistance cassete (*cnm*::*kan*) in the *S. mutans* OMZ175 Δ*cnmΔvicK* strain with *tcnm* at its original locus (Fig. 3.8). In order to delete most of the highly repetitive sequences in *cnm*,

we used a sequential primer walking approach. Briefly, we first designed primers 11 and 12 (TCnmF1, TCnmR1A, Table 3.1) to amplify the approximately 1000 bp region containing the N-terminal portion and the first threonine rich repeat of Cnm (PCR1). Then, a second PCR reaction was performed using the purified PCR1 product as template, the same forward primer 11 (TCnmF1) and another reverse primer 13 (TCnmR1B) that contained the sequence of the second threonine repeat and also partially overlapped with first reverse primer. This second PCR product (PCR2) was gel purified and used as a template in a third PCR reaction using primer 11 (TCnmF1) and primer 14 (TCnmR1C) to generate a PCR product containing the first three threonine repeats of Cnm (PCR3) (Fig. 3.8). Lastly, to amplify the 3' sequence of Cnm containing the LPxTG cell wall anchoring domain, the primer set 15 (TCnmF2A) and 17 (TCnmR2) was used to amplify the approximately 1000 bp containing the 3' portion of the *cnm* gene that begins after the end of the threonine repeats and a portion of the flanking gene *pgfS* (PCR4). Again, the PCR4 product was purified and used as a template in a reaction containing primer 17 (TCnmR2) and forward primer 16 (TCnmF2B) containing 29 bp of the complementary region of 5' end of primer 14 (TCnmR1C) (PCR5). The purified products of PCR3 and PCR5 were used to perform overlapping extension PCR using primers 18 (TCnmF1) and 19 (TCnmR2) to amplify the desired sequence of *tcnm* (PCR6). The overlapping PCR product PCR6 was purified and used to transform the *S. mutans* OMZ175Δ*cnm* mutant. To screen for the tCnm recombinant strain, we took advantage of the antibiotic resistance of Δ*cnm* as a screening tool. Briefly, transformants were first grown on BHI and then patched onto both BHI and BHI containing kanamycin 1 mg/mL. Transformants that grew on BHI, but lost the ability to grow on BHI agar containing kanamycin were screened by PCR using the primers

cnm543F and cnm1514R with the insertion of the tCnm construct confirmed by Sanger sequencing.

Table 3.1: Primers used for the construction of epimerase expression plasmids and genetic manipulation of *S. mutans* OMZ175.

	Primer name	Sequence (5'-3')	Application
1	pgfE_NdeI_For	gcaataccatatgacgaaaattctagtaacaggg	Plasmid construction
2	pgfE_XhoI_Rev	caagtcctcgagctctctaaatccattagggtacttttttg	Plasmid construction
3	GalE_NdeI_For	gcaataccatatgCAGGTGTCTGCGTCAAGT	Plasmid construction
4	GalE_XhoI_Rev	caagtcctcgagGTCTCTGTCATCATAACCCTTTGGA	Plasmid construction
5	gne_NdeI_For	gcaataccatatgAAAATTCTTATTAGCGGTGGTGCAGG	Plasmid construction
6	gne_XhoI_Rev	caagtcctcgagACACTGTTTTTCCCAATCAAAAGCAGAT	Plasmid construction
7	*galE*-F1	GGGTGATGAGAAATGGGG	Deletion of *galE*
8	*galE*-R1	GAATAACTGCAGCCACATCAGG	Deletion of *galE*
9	*galE*-F2	TCAACAGGATCCTCAAATCTTCA	Deletion of *galE*
10	*galE*-R2	GGCAGCTAACTTGGACTT	Deletion of *galE*
11	TCnmF1	ATGAAAAGAAAAGGTTTACGAAG	tCnm strain: PCR1-3
12	TCnmR1A	GTGGTTGTTCTTCCGTCC	PCR1
13	TCnmR1B	GTCGTCGTTGTTGGCTTTTCTGTAGTAGTGGTTGTTCTTCCGTCC	PCR2
14	TCnmR1C	TGGAGCTTCCGTTGTTGTTGGAGCCTCTGTCGTCGTTGTTGGC	PCR3
15	TCnmF2A	CCTAATAAGCCATCAGGC	PCR4-5
16	TCnmF2B	CAGAGGCTCCAACAACAACGGAAGCTCCACCTAATAAGCCATCAGGC	PCR5
17	TCnmR2	CCAATCCATGAAAAAATTCC	PCR4
18	cnm543F	GCAGACAGATGACACCAATCATGTG	PCR6
19	cnm1514R	AGCTGCAGGAGATGCTGG	PCR6

Primers 1-6 were used to construct the epimerase expression plasmids for *E. coli*. Primers 7-10 were used for the deletion of *galE* in *S. mutans*. To better visualize the overlapping PCR approach for the creation of tCnm in *S. mutans* OMZ175 (primers 11-19), the overlapping primers were indicated through matching underlines. Colors in the primer sequences depict either the overlap or the reverse complement region (see Fig. 3.8 for more details).

Purification of Cnm and tCnm from *S. mutans*

A detailed purification protocol of full-length Cnm from *S. mutans* OMZ175 *ΔvicK* has previously been described (210). For the purification of tCnm, the same protocol was used, and a fresh affinity column was prepared to avoid cross-contamination with full-length Cnm. Briefly, *S. mutans* was grown in 10 L of BHI, and aliquots were pelleted, washed with PBS and lysed in the presence of Pierce Protease Inhibitor Tablets (ThermoScientific) as per the manufacturer's protocol. The soluble fraction of the lysate was bound to a custom NHS–anti-rCnmA overnight, followed by a wash with PBS and protein elution using 0.1 M glycine buffer (pH 2.5). The eluted protein was neutralized with 1 M Tris (pH 8.0), concentrated, and dialyzed against 100 mM ammonium bicarbonate. The protein concentration was estimated using the Pierce BCA Protein Assay Kit (ThermoScientific) and western blots were done to confirm integrity as detailed below.

LC-MS

Peptide and glycopeptide mapping by LC-MS was done for samples of purified Cnm and tCnm, as well as for proteins excised from SDS-PAGE. For the in-solution digestion of Cnm and tCnm, samples were buffered in 50mM ammonium bicarbonate buffer. Although there are no cysteines in Cnm to form disulfide bonds, DTT was used a low concentration (5mM) to disrupt any tertiary structure. To this, sequencing grade trypsin (Promega) was added, and the protein was digested overnight at 37˚C. The enzymatic digestion was stopped the next day by heating briefly to 100 ˚C and the sample was analyzed by LC-MS with fragmentation by HCD and ETD.

For in-gel digestion: Identification of proteins from gel slices were determined by mass spectrometry using methods described previously (283). Briefly, gel slices were cut into small pieces (~1mm), destained by washing the gels by alternately swelling/shrinking between ammonium bicarbonate buffer/acetonitrile, followed by reduction and alkylation with DTT and iodoacetamide within the gel. After removal of the chemicals, the protein within the gel was digested with trypsin (Sequencing Grade; Promega) overnight at 37˚C. The produced peptides were released from the gel with increasing proportions of acetonitrile acidified with formic acid. The samples were dried via SpeedVac then reconstituted with 0.1% formic acid before analysis with LC-MS.

For data acquisition of protein digests from solution or gels by LC-MS, samples were reconstituted in 0.1% formic acid for LC-MS/MS and pre-filtered through a 0.2 μm spin filter before injection. Protein and glycoprotein digests were analyzed on an Orbitrap Fusion Tribrid mass spectrometer equipped with a nanospray ion source and was connected in-line to a Thermo Ultimate RSLCnano chromatography system (ThermoFisher). A commercial nano-LC column (Cat. No. 164568, ThermoFisher) of 15 cm length with 75 μm internal diameter (id), filled with 3 μm C18 material (reverse phase), was used for chromatographic separation of samples. The precursor ion scan was acquired at 120,000 resolutions in the Orbitrap analyzer, and precursors at a time frame of 3 s were selected for subsequent MS/MS fragmentation in the Orbitrap analyzer at 15,000 resolution. The LC–MS/MS runs of each digest were conducted for 70 or 180 min. About 0.1% FA and 80% ACN–0.1% FA were used as mobile phases A and B, respectively, in order to separate the glycopeptides. The threshold for triggering an MS/MS event was set to 1000 counts, and monoisotopic precursor selection was enabled (only the monoisotopic peak of the overall

distribution is picked up and selected for MS/MS fragmentation, allowing for more accurate analysis). MS/MS fragmentation was conducted with stepped HCD product-triggered ETD program.

Compositional analysis of Cnm

For the glycosyl composition analysis of Cnm, the methods of Black et al. were followed (284). Briefly, 300 µg purified full length Cnm was dialyzed against DI water using a 3.5 kDa MWCO dialysis membrane (SpectraPor) for 36 hours. Per-O-trimethylsilyl (TMS) derivatives were generated from the methyl glycosides produced after hydrolysis of the sample as described previously (285).

The GC ramp temperature program was as follows. Starting with an initial temperature of 80 °C, the oven temperature was increased at a rate of 20 °C/min until it reached 140 °C, which was held for 2 min. Afterwards, the GC was ramped at a rate of 2 °C/min until it reached 200 °C. The final increase rate was 30 °C/min until the GC reached a temperature of 250 °C, where it was held for 5 min.

SBA blot and protein identification

Bacterial strains were harvested from overnight growth on agar and adjusted to an approximate OD_{600} of 2 in 1 ml PBS. The cells were spun down and resuspended in 100 µl B-Per (Thermo Fisher), supplemented with 200 µg lysozyme (Sigma) and 1x protease inhibitor (cOmplete Mini, EDTA-free, Roche) and incubated for 1 h for 37 °C. Then, 5x SDS loading dye was added to the samples for a final concentration of 1x before they were boiled at 95 °C for 12 minutes. Proteins were separated on a 12.5 % SDS-PAGE and either

stained with Coomassie or transferred to a nitrocellulose membrane. The membrane was blocked in 3 % BSA in PBST at 4 °C overnight. HRP-conjugated soybean agglutinin at 1:1000 in PBST was incubated with the membrane for 1 h at RT. The blots were developed using the Clarity Western ECL Substrate (BioRad) and visualized using the ChemiDoc XRS+ Gel Imaging System (BioRad).

The bands corresponding to SBA recognition were excised from the Coomassie stained SDS-PAGE and destained for several days at 4 °C before being subjected to digestion and LC-MS analysis.

Computational modeling

Models of PgfE and GalE from *S. mutans* were constructed using the AlphaFold CoLab v2.1.0 (267). These models and the crystal structure of Gne from *C. jejuni* (PDB 7K3P) were superimposed onto the crystal structure of WbpP from *P. aeruginosa* (PDB 1SB8) (266) using Chimera (286) and PyMOL (v2.5, Schrödinger LLC) to evaluate possible interactions with UDP-GalNAc.

Purification of recombinant PgfE, GalE and Gne

For the protein expression, *E. coli* BL21(DE) chemically competent cells were freshly transformed (heat-shock) with one of the three epimerase expression plasmids (pET30_PgfE-His, pET30_GalE-His and pET30_Gne-His). Then, 5 ml overnight starter cultures (2xYT with 16 g/L tryptone, 10g/L yeast extract, 5 g/L NaCl and Kan) were used to inoculate 1 L 2xYT medium containing 25 µg/ml kanamycin (0.5x) to a starting OD_{600} of 0.05. The cultures were grown shaking at 37 °C until an OD_{600} of 0.5, when they were

induced with 1mM IPTG. After 12-16 h of growth the cultures were harvested and resuspended in lysis buffer at 10 ml per gram cell pellet (100 mM HEPES, 150 mM NaCl, 10 % glycerol, 5 mM imidazole, pH 7.4, protease inhibitor cocktail (cOmplete Mini, EDTA-free, Roche, 1 tablet/50 ml) and Turbonucuclease (from *Serratia marcescens*, Sigma, 6.26 µl/50ml)). The bacteria were lysed by homogenization with an Emulsiflex-C5 Pressure homogenizer (ATA Scientific) according to the manufacturer's instructions. To remove cell debris, the lysate was centrifuged at 4255 x g for 30 min at 4 °C. The supernatant was used to purify the His-tagged proteins via nickel-NTA column (0.5 ml resin /1 L culture). After the cell lysate was passed through the column twice, the resin was washed with 10 column volumes (CV) of 100 mM HEPES, 150 mM NaCl, 10 % glycerol, 5 mM imidazole, pH 7.4, followed by elution in the same buffer with increasing imidazole concentrations (20 mM, 50 mM, 100 mM, 250 mM, 500 mM with 4 CV each). The collected elution fractions were assessed for the protein of interest via SDS-PAGE, followed by Coomassie staining and/or anti-His western blotting. For the western blot, the nitrocellulose membrane with the transferred proteins was blocked overnight at 4 °C in 5 % BSA in PBST, followed by a 1 h incubation at RT with 1:10,000 mouse anti-His antibody (BioRad) in blocking buffer and a 1 h incubation at RT with 1:20,000 goat anti-mouse secondary antibody. The blots were developed using the Clarity Western ECL Substrate (BioRad) and imaged with the ChemiDoc XRS+ Gel Imaging System (BioRad). Following visual assessment, the elution fractions with the purest and most highly concentrated protein of interest were combined and dialyzed into either 1x PBS or Tris buffer (50 mM Tris, 150 mM NaCl, pH 7.4). For storage, the proteins were aliquoted into smaller samples, lyophilized and kept at -20 °C.

NMR analysis

The substrate specificity of the epimerases was examined using proton spectra acquired at 37 °C on 800 and 900 MHz Bruker Neo NMR spectrometers. All reagents were lyophilized and stored at -20 °C in advance and reconstituted in D_2O on the day of the experiment. Within a total volume of 600 µl, 15 µg epimerase protein (25 µg/ml final concentration), 300 µM UDP-sugar and 1 mM NAD^+ were combined in 1xPBS. To obtain an experimental baseline for time point 0, UDP-sugar, NAD^+ and PBS were analyzed, before the enzyme was spiked in for further analysis. After addition of enzyme to the UDP-sugar, samples were returned to the magnet, equilibrated, and then separate proton spectra were collected for short intervals (1.5 or 3.75 minutes) over a period of hours. Data were analyzed using MestReNova software (Mestrelab Research).

Capillary electrophoresis

The epimerases PgfE, GalE and Gne were expressed in *E. coli*, purified, and stored in lyophilized form at -20 °C before analysis through CE. In 100 µl total volume, 2.5 µg epimerase (25 µg/ml), 100 µM UDP-sugar, 0.5 mM NAD^+ and 50 mM Tris, 150 mM NaCl, pH 7.4 were combined. Samples were incubated at 37 °C for 1 h before storage at 10 °C until endpoint analysis with CE. For the CE analysis, the method from Creuzenet et al. was modified (287). A 60 cm capillary with 50 µm i.d. (Beckman Coulter) was used on a Beckman Coulter PA 800plus capillary electropherograph. The samples were injected at 2 psi for 20 sec and run at 20 kV under regular polarity. Sample separation was detected at 262 nm and 25 mM sodium tetraborate at pH 9.4 served as running buffer with 0.2 mM NaOH and H_2O as column rinsing solutions.

Western blotting for detection of Cnm in lysates

OMZ175 wildtype, *Δpgf*, *ΔpgfE*, *ΔpgfECpgfE*, *ΔgalE* and *ΔpgfEΔgalE* cultures were grown overnight in 5 mL of BHI media. Bacteria were harvested and adjusted to an OD_{600} of 2 in 1 ml PBS. The cells were spun down and resuspended in 100 µl B-Per (Thermo Fisher), supplemented with 200 µg lysozyme (Sigma) and 1x protease inhibitor (cOmplete Mini, EDTA-free, Roche) and incubated for 2 h for 37 °C. Then, 5x SDS loading dye (4 % final SDS concentration) were added to the samples for a final concentration of 1x before boiling at 95 °C for 12 min. The proteins were separated using a 12.5 % SDS gel run for extended time to achieve better separation of high molecular weight proteins. The gel was either stained with Coomassie or used for transfer of the proteins to a nitrocellulose membrane. Cnm detection was performed using a rabbit anti-rCnmA (Cnm Collagen-Binding Domain) and an HRP-labeled anti-rabbit horseradish peroxidase secondary antibody (Sigma-Aldrich) as described elsewhere (204).

Growth curves

OMZ175 wildtype, *Δpgf*, *ΔpgfE*, *ΔpgfECpgfE*, *ΔgalE* and *ΔpgfEΔgalE* cultures were grown overnight in biological triplicates on BHI media. Cells were subcultured in 1:20 dilutions into 3 mL of fresh pre-warmed BHI to an OD_{600} of 0.4. Cells from 1 mL of each culture were pelleted via centrifugation at 10,000 x g for 2 minutes and washed twice with 1 mL of sterile PBS, before being resuspended in 1 mL of sterile PBS. Then, 10 µL of each cell suspension were added to 250 µL of media in wells of a Honeycomb plate (ThermoFischer Scientific) and covered with a 50 µL layer of sterile mineral oil in order to decrease aeration of cultures. Blanks for each medium were inoculated with 10 µL of

sterile PBS instead. Growth curves were done in either BHI or a Chemically Defined Medium (CDM) containing either 1% (m/v) of glucose or galactose as the carbon source (288). Growth was maintained at 37°C for 24 hours in an automated growth reader (Bioscreen c, Oy Growth Curves AB), with OD_{600} being monitored every 30 minutes after a 5-second plate agitation. Blank measurements were subtracted from each experimental reading.

Bright field microscopy

All *S. mutans* strains were grown on BHI plates containing their appropriate antibiotics. Bacteria were harvested into PBS and washed 3x before resuspension in MEM Alpha medium (Gibco). Then, 5 ml MEM alpha substituted with 25 mM iron (289) and 1 % D-glucose or 1 % D-galactose were inoculated and adjusted to an OD_{600} of 0.1 and grown overnight at 37 °C with 5 % CO_2. The next day, 100 µl of bacterial culture were transferred onto a clean microscopy slide and allowed to dry at RT. The samples were heat fixed and stained with crystal violet before analysis on a Axio A1 brightfield microscope (Zeiss) at 100x magnification.

Transmission Electron Microscopy

S. mutans wildtype, as well as the *ΔpgfE* and *ΔgalE* mutants were grown in MEM supplemented with 1 % galactose as described above (bright field microscopy). After overnight incubation, bacteria were applied to TEM grids directly (*ΔgalE*) or at a 1:1 dilution with PBS (wildtype and *ΔpgfE*) to achieve appropriate cell density for the visualization of cell morphology. Mesh 300 copper grids with Formvar coating and light

carbon coating were used. Then, 30 µl bacterial suspension were applied onto the Formvar coated side of the grids and incubated at RT for 45 min. All following steps were executed by floating the grids on 100 µl droplets of solution. After bacterial binding, the grids were washed for 5 min in PBS, followed by fixing in 5 % paraformaldehyde for 30 min. the grids were washed in PBS and H_2O 3x for five minutes each. Each grid was stained using 3% phosphotungstic acid (Electron Microscopy Sciences) for 5 sec, the stain was removed with a Kimwipe to stop the development. The microscopy was done on the JEOL JEM1011 TEM (JEOL Inc., Akishima, Tokyo, Japan).

Acknowledgements

The authors would like to thank Harald Nothaft for helpful discussions. S.A. was supported by NIGMS Training grant T32GM107004. This study was funded by NIDCR grant R01DE022559 awarded to J.A and J.L. This work was also supported by the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, U.S. Department of Energy grant DE-SC0015662 at the Complex Carbohydrate Research Center. The carbohydrate analysis was supported, in part, by NIH grant R24GM137782.

Abbreviations

Collagen binding proteins, CBP; post-translational modification, PTM; uridine diphosphate-glucose, UDP-Glc; uridine diphosphate- *N*-acetylglucosamine, UDP-GlcNAc; uridine diphosphate-galactose, UDP-Gal; uridine diphosphate- *N*-acetylgalactosamine, UDP-GalNAc; mass spectrometry, MS; nuclear magnetic resonance,

NMR; nicotinamide adenine dinucleotide, NAD; soybean agglutinin, SBA; wheat germ agglutinin, WGA; capillary electrophoresis, CE; procedure defined Units, pdU; transmission electron microscopy, TEM; brain-heart infusion, BHI; chemically defined media, CDM; rhamnose-glucose polysaccharides, RGP; optical density, OD; room temperature, RT; bovine serum albumin, BSA; phosphate buffered saline with triton, PBST.

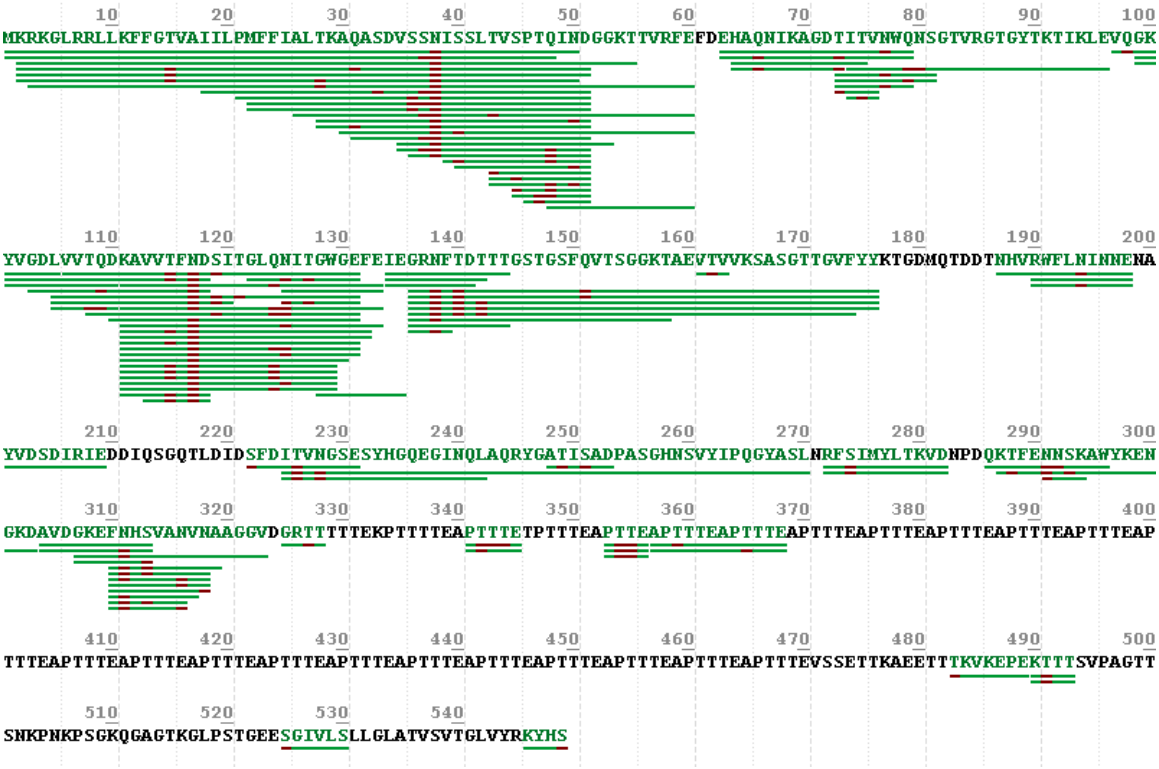


Figure 3.7: Coverage map for MS peptide analysis of purified Cnm form *S. mutans* wildtype. Green underlines indicate the identification/coverage of Cnm peptides. The lack of coverage in the threonine-rich repeat region is due to the heavy glycosylation.

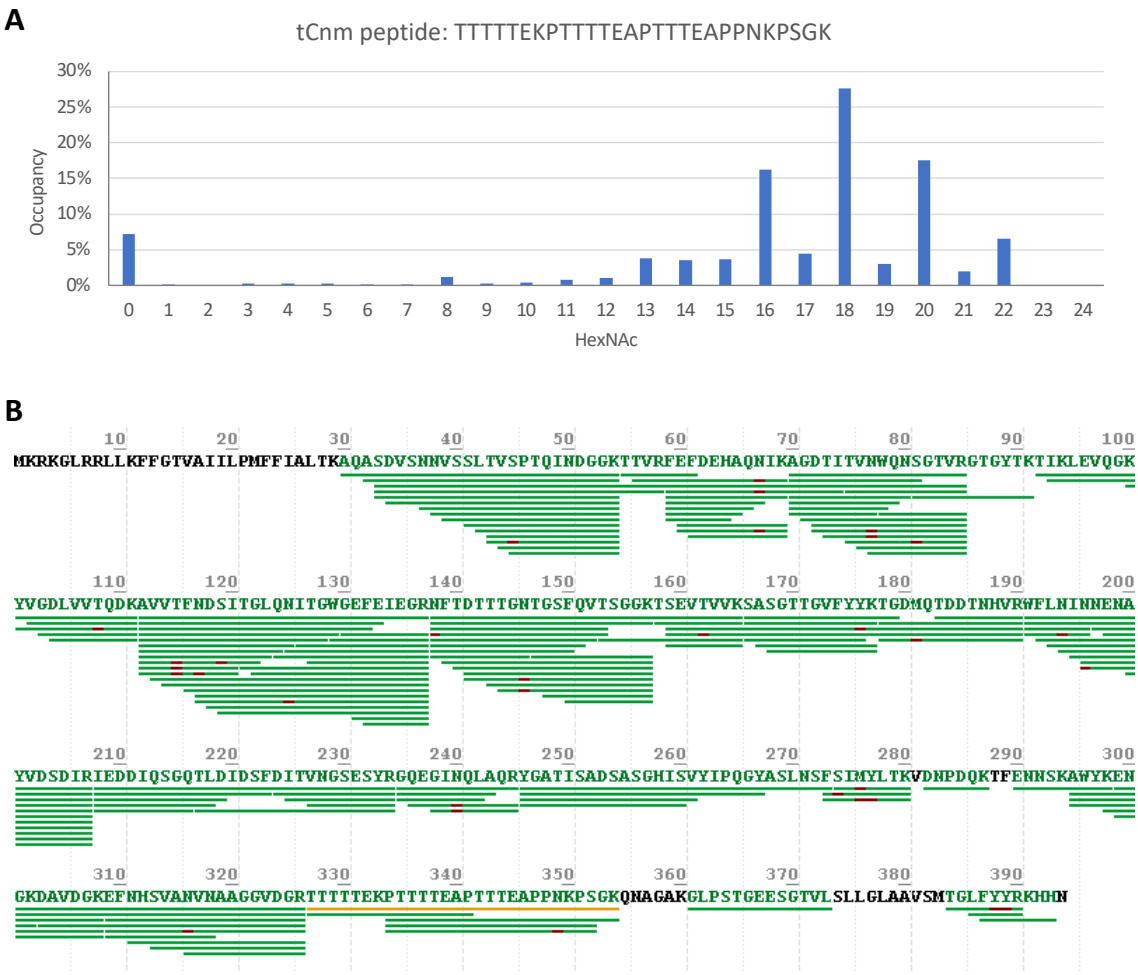
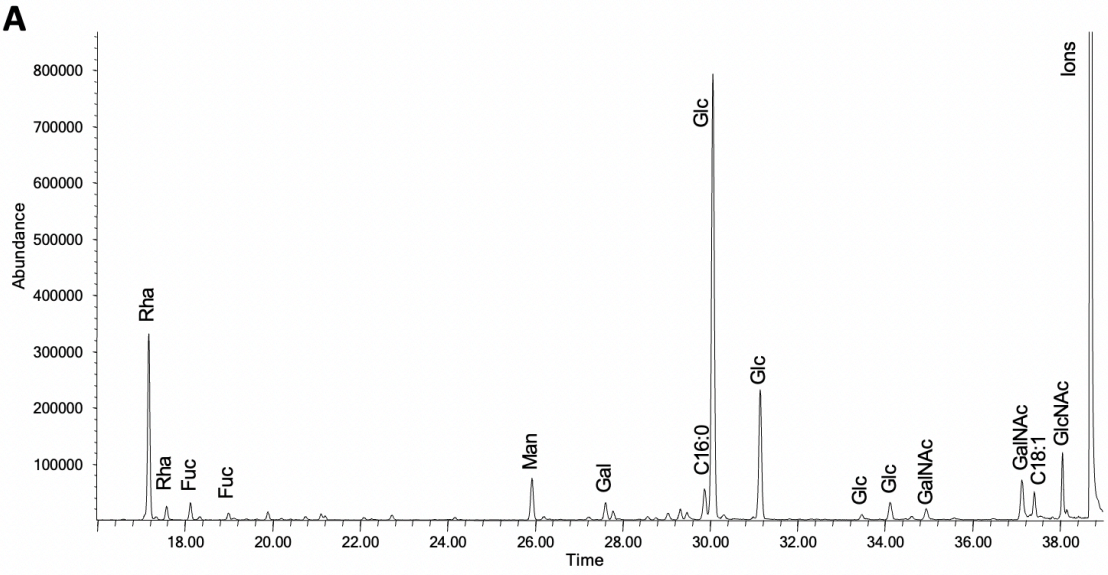


Figure 3.8: HexNAc occupancy and peptide coverage of tCnm MS analysis. tCnm protein was enzymatically digested by trypsin and analyzed via LC-MS/MS. The glycan occupancy of 12 threonine residues in the T-rich region of tCnm is shown as percent occupancy with HexNAc, which was manually determined by extracting ion abundances of each glycopeptide. (B) Coverage map for MS peptide analysis of purified tCnm form *S. mutans* wildtype. Green underlines indicate the identification/coverage of Cnm peptides. The lack of coverage in the threonine-rich repeat region is due to the heavy glycosylation.

A



B

Glycosyl residue	Mass (mg)	Mol %
Ribose (Rib)	n.d.	-
Arabinose (Ara)	n.d.	-
Rhamnose (Rha)	0.9	24.5
Fucose (Fuc)	0.1	3.4
Xylose (Xyl)	n.d.	-
Glucuronic Acid (GlcA)	n.d.	-
Galacturonic acid (GalA)	n.d.	-
Mannose (Man)	0.1	3.5
Galactose (Gal)	0.1	2.1
Glucose (Glc)	1.9	45.0
N-Acetyl Glucosamine (GlcNAc)	0.3	8.5
N-Acetyl Galactosamine (GalNAc)	0.6	12.9
N-Acetyl Mannosamine (ManNAc)	n.d.	-
S=	4.1	

Fig. 3.9: Composition analysis of purified full length Cnm from *S. mutans* OMZ175. Cnm was purified from its native host, dialyzed into water and analyzed via GC-MS for glycan composition of the sample. Shown is the MS spectrum of the Cnm composition analysis in **(A)** and the resulting percentages of glycans present in **(B)**. S is sum of masses in mg.

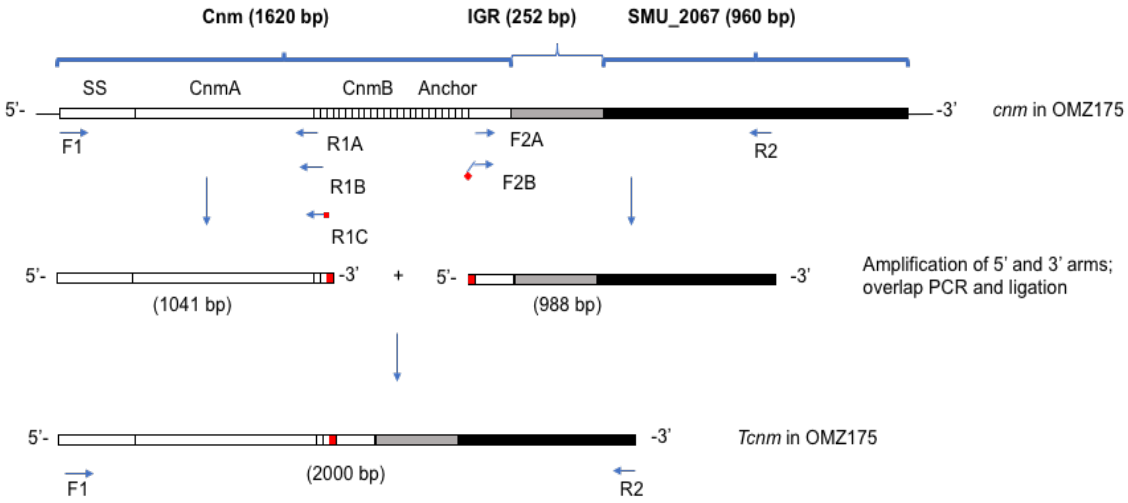


Fig. 3.10: Schematic showing overlapping PCR strategy for the creation of a truncated Cnm (tCnm) expressing strain in *S. mutans* OMZ175. A Cnm variant containing the first two threonine rich repeats (tCnm) was constructed by overlapping PCR to facilitate LC-MS and compositional analyses. In the tCnm construct, 19 out of the 21 threonine rich repeats were deleted, but the C-terminal portion of the gene containing the LPXTG cell wall anchoring domain was left intact. To facilitate annealing and extension, certain forward and reverse primers shared complementary sequences as indicated in red.

Table 3.2. Identification of putative Pgf target proteins from *S. mutans* whole cell lysates. *S. mutans* wildtype whole cell lysates were probed with SBA. Four SBA-reactive bands corresponding to GalNAc detection were excised from the Coomassie-stained SDS-PAGE (Fig. 2). Proteins present in the four excised gel pieces were identified by MS from the proteome of *S. mutans* OMZ175.

Gel piece 1_Hits against *S. mutans* (Filtered for Log Prob >5)

Prot. Rank	Protein Name	|Log Prob|	Best |Log Prob|	Best Score	# Spectra	# Uniq. Peps.
1 | 1	>sp|P23504|SPAP_STRMU Cell surface antigen I/II OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=spaP PE=1 SV=2	42.11	8.41	689.8	16	10
2 | 2	>sp|Q8DVP9|IF2_STRMU Translation initiation factor IF-2 OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=infB PE=3 SV=1	39.4	5.79	476.1	18	14
3 | 3	>sp|Q8CWY0|SYA_STRMU Alanine--tRNA ligase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=alaS PE=3 SV=1	34.15	9.39	739.4	11	9
4 | 4	>sp|Q8CWY5|SYGB_STRMU Glycine--tRNA ligase beta subunit OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=glyS PE=3 SV=1	26.2	9.46	730.5	6	5
5 | 5	>sp|Q8DSL1|SYV_STRMU Valine--tRNA ligase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=valS PE=3 SV=1	24.07	6.01	553.3	7	7
6 | 6	>sp|Q59934|PFL_STRMU Formate acetyltransferase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 /	23.61	6.41	568.3	14	6

	UA159) OX=210007 GN=pfl PE=3 SV=2					
7 | 7	>sp|Q8DV10|CAPP_STRMU Phosphoenolpyruvate carboxylase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=ppc PE=3 SV=1	21.24	4.77	594.5	10	8
8 | 8	>sp|P13470|GTFC_STRMU Glucosyltransferase-SI OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=gtfC PE=1 SV=2	17.15	5.79	456.5	6	6
9 | 9	>sp|Q8DSF0|SECA_STRMU Protein translocase subunit SecA OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=secA PE=3 SV=1	16.57	5.97	567.1	6	5
10 | 10	>sp|P11701|SACB_STRMU Levansucrase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=ftf PE=3 SV=2	16.25	6.12	539.5	6	5
11 | 11	>sp|P72483|EFTU_STRMU Elongation factor Tu OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=tuf PE=3 SV=2	15.79	12.17	855.8	2	2
12 | 12	>sp|Q8CWZ6|TIG_STRMU Trigger factor OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=tig PE=3 SV=1	13.78	5.86	518.1	4	3
13 | 13	>sp|Q59938|ACNA_STRMU Aconitate hydratase A OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=acn PE=3 SV=2	13.16	5.68	396	8	5

14 | 14	>sp|Q8DVV4|EFG_STRMU Elongation factor G OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=fusA PE=3 SV=1	12.3	6.62	580.1	6	6
15 | 15	>sp|P45595|PT1_STRMU Phosphoenolpyruvate-protein phosphotransferase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=ptsI PE=3 SV=2	9.17	4.66	494.5	3	3
16 | 16	>sp|Q8CWX8|FTSY_STRMU Signal recognition particle receptor FtsY OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=ftsY PE=3 SV=1	7.62	4.23	386.4	5	3
17 | 17	>sp|O06942|DNAK_STRMU Chaperone protein DnaK OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=dnaK PE=2 SV=2	5.85	6	443.6	1	1

Gel piece 2_Hits against *S. mutans* (Filtered for Log Prob >5)

Prot Rank	Protein Name	|Log Prob|	Best |Log Prob|	Best Score	# Spectra	# Uniq Peps *
1 | 1	>sp|O06942|DNAK_STRMU Chaperone protein DnaK OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=dnaK PE=2 SV=2	98.4	14.22	883.2	33	16
2 | 2	>sp|P45595|PT1_STRMU Phosphoenolpyruvate-protein phosphotransferase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=ptsI PE=3 SV=2	91.26	12.71	837.2	40	17

3 | 3	>sp|Q8CWW6|CH60_STRMU 60 kDa chaperonin OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=groL PE=3 SV=1	74.87	11.2	723.2	19	12
4 | 4	>sp|P72483|EFTU_STRMU Elongation factor Tu OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=tuf PE=3 SV=2	67.33	14.99	981.3	10	9
5 | 5	>sp|Q8DTQ3|EZRA_STRMU Septation ring formation regulator EzrA OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=ezrA PE=3 SV=1	58.47	9.81	725.1	21	14
6 | 6	>sp|Q8DTY0|GLMS_STRMU Glutamine--fructose-6-phosphate aminotransferase [isomerizing] OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=glmS PE=3 SV=3	45	14.27	924.4	10	6
7 | 7	>sp|P23504|SPAP_STRMU Cell surface antigen I/II OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=spaP PE=1 SV=2	29.84	6.91	571.1	10	8
8 | 8	>sp|Q8CWZ6|TIG_STRMU Trigger factor OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=tig PE=3 SV=1	27.84	11.69	758.5	11	6
9 | 9	>sp|Q8DRT7|ILVD_STRMU Dihydroxy-acid dehydratase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=ilvD PE=3 SV=1	16.09	8.74	628.3	5	4
10 | 10	>sp|P11000|WAPA_STRMU Wall-associated protein OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=wapA PE=3 SV=2	15.16	5.67	576.4	10	4

11 | 11	>sp|Q99040|DEXB_STRMU Glucan 1,6-alpha-glucosidase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=dexB PE=1 SV=2	12.64	6.37	472.2	4	4
12 | 12	>sp|Q8DSJ9|SYP_STRMU Proline--tRNA ligase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=proS PE=3 SV=1	10.2	5.21	462.7	2	2
13 | 13	>sp|Q8DSG3|SYDND_STRMU Aspartate--tRNA(Asp/Asn) ligase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=aspS1 PE=3 SV=1	9.56	6.41	415.5	4	3
14 | 14	>sp|Q8DTS9|ENO_STRMU Enolase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=eno PE=1 SV=1	8.18	5.78	515.9	3	3
15 | 15	>sp|Q59934|PFL_STRMU Formate acetyltransferase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=pfl PE=3 SV=2	7.04	5.32	365.1	2	2
16 | 16	>sp|Q8DWG1|PYRG_STRMU CTP synthase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=pyrG PE=3 SV=1	6.29	6.35	445.4	4	3
17 | 17	>sp|Q8DSY2|RRF_STRMU Ribosome-recycling factor OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=frr PE=3 SV=1	5.57	5.59	368.9	2	1
18 | 18	>sp|O70055|DLTC_STRMU D-alanyl carrier protein OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=dltC PE=3 SV=1	5.55	5.55	397	2	1
19 | 19	>sp|P45596|PTHP_STRMU Phosphocarrier protein HPr OS=*Streptococcus mutans* serotype	5.54	5.56	547.1	1	1

	c (strain ATCC 700610 / UA159) OX=210007 GN=ptsH PE=1 SV=3					
20 | 20	>sp|Q8DVV2|PGK_STRMU Phosphoglycerate kinase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=pgk PE=3 SV=1	5.54	5.59	377.2	2	2
21 | 21	>sp|Q9XB21|DBH_STRMU DNA-binding protein HU OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=hup PE=1 SV=1	5.19	5.21	494.1	1	1

Gel piece 3_Hits against *S. mutans* (Filtered for Log Prob >5)

Prot Rank	Protein Name	|Log Prob|	Best |Log Prob|	Best Score	# Spectra	# Uniq Peps
1 | 1	>sp|Q59934|PFL_STRMU Formate acetyltransferase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=pfl PE=3 SV=2	41.01	14.23	964.9	14	10
2 | 2	>sp|P23504|SPAP_STRMU Cell surface antigen I/II OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=spaP PE=1 SV=2	39.97	9.38	706	17	12
3 | 3	>sp|Q8CWY0|SYA_STRMU Alanine--tRNA ligase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=alaS PE=3 SV=1	36.25	8	624.3	9	7
4 | 4	>sp|Q8DVP9|IF2_STRMU Translation initiation factor IF-2 OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=infB PE=3 SV=1	30.23	5.71	503.3	14	11
5 | 5	>sp|P08987|GTFB_STRMU Glucosyltransferase-I OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=gtfB PE=3 SV=3	25.75	6.64	588.8	7	7

6 | 6	>sp|Q8DSF0|SECA_STRMU Protein translocase subunit SecA OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=secA PE=3 SV=1	25.03	5.64	494.8	9	8
7 | 7	>sp|Q8DSL1|SYV_STRMU Valine--tRNA ligase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=valS PE=3 SV=1	22.91	6.33	549	10	8
8 | 8	>sp|Q8DVV4|EFG_STRMU Elongation factor G OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=fusA PE=3 SV=1	21.45	5.83	606.6	8	6
9 | 9	>sp|Q8CWY5|SYGB_STRMU Glycine--tRNA ligase beta subunit OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=glyS PE=3 SV=1	21.41	8.89	686.1	6	5
10 | 10	>sp|Q8DV10|CAPP_STRMU Phosphoenolpyruvate carboxylase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=ppc PE=3 SV=1	19.05	4.85	569.1	5	5
11 | 11	>sp|P11701|SACB_STRMU Levansucrase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=ftf PE=3 SV=2	14.45	6.5	543.3	5	4
12 | 12	>sp|P45596|PTHP_STRMU Phosphocarrier protein HPr OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=ptsH PE=1 SV=3	12.17	6.18	483.8	4	2
13 | 13	>sp|Q8CWZ6|TIG_STRMU Trigger factor OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=tig PE=3 SV=1	11.9	6.21	548	4	2
14 | 14	>sp|Q8CWX8|FTSY_STRMU Signal recognition particle receptor FtsY OS=*Streptococcus mutans* serotype	8.43	4.77	410.5	3	3

	c (strain ATCC 700610 / UA159) OX=210007 GN=ftsY PE=3 SV=1					
15 | 15	>sp|Q9XB21|DBH_STRMU DNA-binding protein HU OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=hup PE=1 SV=1	8.41	5.22	374	3	2
16 | 16	>sp|Q8DVV2|PGK_STRMU Phosphoglycerate kinase OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=pgk PE=3 SV=1	7.88	4.89	368.9	2	2
17 | 17	>sp|Q8CWX2|SYFB_STRMU Phenylalanine--tRNA ligase beta subunit OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=pheT PE=3 SV=1	6.39	6.46	481.3	2	2
18 | 18	>sp|Q8DRW8|MUTS_STRMU DNA mismatch repair protein MutS OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=mutS PE=3 SV=1	6.03	4.77	381.9	3	3
19 | 19	>sp|Q59938|ACNA_STRMU Aconitate hydratase A OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=acn PE=3 SV=2	5.84	3.9	318.6	3	3
20 | 20	>sp|Q8DTC7|CLPB_STRMU Chaperone protein ClpB OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=clpB PE=3 SV=1	5.37	5.6	486.6	1	1
21 | 21	>sp|O70055|DLTC_STRMU D-alanyl carrier protein OS=*Streptococcus mutans* serotype c (strain ATCC 700610 / UA159) OX=210007 GN=dltC PE=3 SV=1	5.36	5.38	381.2	1	1

Gel piece 4_Hits against *S. mutans* (Filtered for Log Prob >5)

Prot Rank	Protein Name	|Log Prob|	Best |Log Prob|	Best Score	# Spectra	# Uniq Peps
1 | 1	>tr|J9V290|J9V290_STRMG AgI/II OS=*Streptococcus mutans* OX=1309 GN=spaP PE=3 SV=1	38.71	7.18	655	15	10
2 | 2	>tr|M4JBQ7|M4JBQ7_STRMG Dextransucrase OS=*Streptococcus mutans* OX=1309 GN=gtfC PE=3 SV=1	12.59	6.36	520.3	9	3
3 | 3	>tr|F5VX98|F5VX98_STROR Putative cross-wall-targeting lipoprotein signal OS=*Streptococcus oralis* SK255 OX=1005704 GN=HMPREF9968_1245 PE=3 SV=1	7.14	3.79	536.9	3	2

*Prot = protein; Prob = probability; Uniq = unique; Peps = peptides

CHAPTER 4

CONCLUSIONS AND FUTURE DIRECTIONS

Associations of hIntL-1 with bacteria and the immune system

It is widely appreciated that the glycans decorating bacterial cell surfaces are playing a key role in initial interactions between these microbes and their surroundings. This holds particularly true during microbial invasion of a host, where the immune system specializes in bacterial surface glycan recognition. Yet, many aspects and players during this process of interaction, evasion and elimination remain unknown. This study has provided insight into the specific effects of the binding of an under characterized member of the human innate immune system with a prevalent bacterial pathogen. We were able to demonstrate that serotype-specific binding of hIntL-1 to *S. pneumoniae* CPS results in agglutination. The bacterial recognition through the lectin causes neither direct killing nor complement dependent killing. Instead we noted increased bacterial adhesion to host cells, a mechanism that was previously described for *S. pneumoniae* mutants of increased chain length (227). It is yet to be determined if this phenomenon is a strategy evolved by the bacteria to use immune molecules to their own benefit or whether the increase in bacterial host attachment can be replicated in an *in vivo* model in the presence of mucus and immune proteins/cells. Furthermore, our results indicated that the primary method of pneumococcal elimination through hIntL-1 lies in the phagocytosis of aggregated bacteria by neutrophils,

which could occur through one of two distinct receptor mediated mechanisms; opsonic and non-opsonic (11). In the non-opsonic mechanism, target bacteria are directly recognized by membrane-bound neutrophil receptors. All characterized non-opsonic receptors to-date are C-type lectins, such as dentin-1 and Mincle. These receptors recognize specific glycan moieties within bacterial surface glycans, such as the LTA anchor of group A Streptococcus through the Mincle receptor (68). In contrast, during opsonin-mediated phagocytosis no direct contact between the microbe and phagocyte is established. Instead, the bacterium is recognized by one or multiple members of the humoral immune response, leading to the binding of an opsonin i.e. various host molecules that bind to antigens and are recognized by receptors on the phagocyte. During this interaction, the opsonins, such as antibodies or the MBL, can either be bound directly to the bacterial surface or indirectly as part of a complex with other molecules that do bind directly, as is the case for C3b within the complement cascade (8, 225, 231, 290). Our studies have shown that neutrophil killing of *S. pneumoniae* is only increased through hIntL-1 in the presence of complement-containing serum, providing a strong indication that hIntL-1-mediated killing by neutrophils is likely opsonin-dependent. To further investigate this, neutrophil cell lysates were probed with hIntL-1 in a far western blot showing no specific interactions between the lectin and any neutrophil components. An alternative approach to examine the possibility of interactions between hIntL-1 and neutrophils would be fluorescent microscopy. In this approach, not only could the neutrophil and hIntL-1 be specifically labeled with fluorophores, but different opsonins could be labeled as well. This allows visualization of any co-localization of hIntL-1 and neutrophils in the presence or absence of serum, as well as the involvement of any opsonins in the presence of serum.

Our studies suggest a role for hIntL-1 in bacterial surveillance, during which the lectin has the ability to bind a wide variety of microbes due to the commonality of its glycan ligand. This would allow for hIntL-1 to aid in the sequestration between host tissues and microbiota. This hypothesis is further supported by the widespread expression and secretion of hIntL-1 within various human tissues, allowing the lectin to act as a generalized control mechanism for the immune system. Additionally, hIntL-1 has demonstrated a lack of well-established regulatory mechanisms for its expression and an inability to kill bacteria even in the presence of serum proteins. Combined these features suggest that hIntL-1 is active in many areas of the body, recognizing a wide variety of bacteria, followed by their binding. The agglutination by the lectin causes an increase in bacterial killing by other components of the immune response. Future studies could explore not only the interplay between hIntL-1 and cells or molecules of the innate immune response, but also the binding specificity for other pathogens and commensals.

The Pgf glycosylation machinery of *S. mutans*

Our analysis of Cnm showed abundant post translational modification with HexNAc2 moieties of GlcNAc and GalNAc. Apart from the Cnm collagen binding protein, several additional targets for the Pgf glycosylation machinery were proposed (WapA, GtfC, SpaP), all of which are key proteins during the initial colonization of *S. mutans* in the oral cavity. This indicates the importance of protein O-glycosylation during pathogenesis, as Pgf deletion mutants demonstrate impairments in cell invasion and infectivity (204).

Within the Pgf machinery, the proteins PgfS, PgfM1 and PgfE have been shown to be directly involved in the addition of glycans to target proteins, while PgfM2 seems to

play a role during the later steps of glycosylation or to assist in glycosylation. While PgfS, PgfM1 and PgfM2 remain largely unstudied, we were able to characterize the 4-epimerase, PgfE. NMR and CE analysis of PgfE demonstrated its activity as a UDP-GlcNAc/GalNAc 4-epimerase with minor conversion of UDP-Glc/Gal. We further confirmed its key function during protein O-glycosylation in *S. mutans* while no involvement in galactose metabolism or cell division were detected. For the second epimerase in *S. mutans*, the UDP-Glc/Gal epimerase GalE, a novel connection to surface polysaccharide biosynthesis and cell division was also established. These findings open a new area of exploration investigating the interconnectedness between glycosyltransferases from different *S. mutans* pathways and surface glycosylation. Future experiments could utilize RGP mutants and focus on the role of GalE during RGP synthesis and septum formation.

Our studies on the two epimerases in *S. mutans* demonstrated the severe phenotypical effects associated with the loss of bacterial glycosylation. The deletion of the Pgf machinery led to deficits in pathogen colonization of its host, while the loss of the epimerase GalE influenced cell morphology and division. To gain a deeper insight into the mechanism and cellular effects of protein O-glycosylation in *S. mutans,* future studies will focus on the characterization of the remaining Pgf proteins.

Summary

Overall, the work described in this dissertation was directed at better understanding both protein and surface glycosylation in *Streptococci* and the interaction of these glycans with the host environment. This included the study of the Pgf O-glycosylation machinery in *S. mutans*, as well as the interactions of the *S. pneumoniae* CPS with human immune

molecules. We discovered that the serotype-specific recognition of *S. pneumoniae* serotype 43 by hIntL-1 leads not only to increased bacterial attachment to host cells, but also to enhanced killing through phagocytosis by neutrophils. We further carried out the first characterization of an enzyme in the Pgf glycosylation pathway of *S. mutans* as a UDP-GlcNAc/GalNAc 4-epimerase, and found a novel connection between the UDP-Glc/Gal 4-epimerase GalE and RGP formation, septum formation and cell division. In addition, we demonstrated the importance of protein glycosylation in the context of *S. mutans* oral colonization, utilizing a variety of phenotypical analyses, and discovered that protein O-glycosylation can be replaced by phosphorylation as an alternative post-translational modification in the absence of the glycosylation machinery. This work presents new information on all aspects of bacterial glycosylation and demonstrates the diversity with which glycans affect bacterial and host interactions.

REFERENCES

1. Amon R, Reuven EM, Leviatan Ben-Arye S, Padler-Karavani V. 2014. Glycans in immune recognition and response. Carbohydr Res 389:115–122.

2. Suresh R, Mosser DM. 2013. Pattern recognition receptors in innate immunity, host defense, and immunopathology. Adv Physiol Educ 37:284.

3. Turvey SE, Broide DH. 2010. Innate immunity. J Allergy Clin Immunol 125:S24–S32.

4. Ragland SA, Criss AK. 2017. From bacterial killing to immune modulation: Recent insights into the functions of lysozyme. PLoS Pathog 13.

5. Athman R, Philpott D. 2004. Innate immunity via Toll-like receptors and Nod proteins. Curr Opin Microbiol 7:25–32.

6. Wicherska-pawłowska K, Wróbel T, Rybka J. 2021. Toll-Like Receptors (TLRs), NOD-Like Receptors (NLRs), and RIG-I-Like Receptors (RLRs) in Innate Immunity. TLRs, NLRs, and RLRs Ligands as Immunotherapeutic Agents for Hematopoietic Diseases. Int J Mol Sci 22.

7. Parkin J, Cohen B. 2001. An overview of the immune system. Lancet 357:1777–1789.

8. Mathern DR, Heeger PS. 2015. Molecules Great and Small: The Complement System. Clin J Am Soc Nephrol 10:1636–50.

9. Brady AM, Calix JJ, Yu J, Geno KA, Cutter GR, Nahm MH. 2014. Low Invasiveness of Pneumococcal Serotype 11A Is Linked to Ficolin-2 Recognition of O-acetylated Capsule Epitopes and Lectin Complement Pathway Activation. J Infect Dis 210:1155–1165.

10. Roche AM, Richard AL, Rahkola JT, Janoff EN, Weiser JN. 2015. Antibody blocks acquisition of bacterial colonization through agglutination. Mucosal Immunol 8:176.

11. Uribe-Querol E, Rosales C. 2020. Phagocytosis: Our Current Understanding of a Universal Biological Process. Front Immunol 11:1066.

12. Riera Romo M, Pérez-Martínez D, Castillo Ferrer C. 2016. Innate immunity in vertebrates: an overview. Immunology 148:125–39.

13. Kobayashi SD, Malachowa N, DeLeo FR. 2018. Neutrophils and Bacterial Immune Evasion. J Innate Immun 10:432–441.

14. Kaplan MJ, Radic M. 2012. Neutrophil extracellular traps (NETs): Double-edged swords of innate immunity. J Immunol 189:2689.

15. Segal AW. 2005. How Neutrophils Kill Microbes. Annu Rev Immunol 23:197.

16. Martín-Orozco N, Isibasi A, Ortiz-Navarrete V. 2001. Macrophages present exogenous antigens by class I major histocompatibility complex molecules via a secretory pathway as a consequence of interferon-γ activation. Immunology 103:41.

17. ten Broeke T, Wubbolts R, Stoorvogel W. 2013. MHC Class II Antigen Presentation by Dendritic Cells Regulated through Endosomal Sorting. Cold Spring Harb Perspect Biol 5.

18. Topham NJ, Hewitt EW. 2009. Natural killer cell cytotoxicity: how do they pull the trigger? Immunology 128:7.

19. Vivier E, Tomasello E, Baratin M, Walzer T, Ugolini S. 2008. Functions of natural killer cells. Nat Immunol 2008 95 9:503–510.

20. Lunderius-Andersson C, Enoksson M, Nilsson G. 2012. Mast Cells Respond to Cell Injury through the Recognition of IL-33. Front Immunol 3.

21. Alberts, Bruce; Johnson, Alexander; Lewis, Julian; Raff, Martin; Roberts, Keith; Walter P. 2002. Molecular Biology of the Cell in Cell 4thFigure4 edition. Garland Science.

22. Lebien TW, Tedder TF. 2008. B lymphocytes: how they develop and function. Blood 112:1570–1580.

23. Petersone L, Edner NM, Ovcinnikovs V, Heuts F, Ross EM, Ntavli E, Wang CJ, Walker LSK. 2018. T Cell/B Cell Collaboration and Autoimmunity: An Intimate Relationship. Front Immunol 9:1941.

24. Hoffman W, Lakkis FG, Chalasani G. 2016. B Cells, Antibodies, and More. Clin J Am Soc Nephrol 11:137.

25. Cohen M, Varki A. 2014. Modulation of glycan recognition by clustered saccharide patches, p. 75–125. *In* International Review of Cell and Molecular Biology. Elsevier Inc.

26. Varki A, Gagneux P. 2017. Chapter 7 Biological Functions of Glycans. Essentials Glycobiol 3rd Ed Chapter 7.

27. Apweiler R, Hermjakob H, Sharon N. 1999. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim

Biophys Acta - Gen Subj 1473:4–8.

28. Guzman-Aranguez A, Argüeso P. 2010. Structure and biological roles of mucin-type O-glycans at the ocular surface. Ocul Surf 8:8–17.

29. Kowarik M, Young NM, Numao S, Schulz BL, Hug I, Callewaert N, Mills DC, Watson DC, Hernandez M, Kelly JF, Wacker M, Aebi M. 2006. Definition of the bacterial N-glycosylation site consensus sequence. EMBO J 25:1957–1966.

30. Szymanski CM, Yao R, Ewing CP, Trust TJ, Guerry P. 1999. Evidence for a system of general protein glycosylation in Campylobacter jejuni. Mol Microbiol 32:1022–30.

31. Nothaft H, Szymanski CM. 2010. Protein glycosylation in bacteria: sweeter than ever. Nat Rev Microbiol 8:765–778.

32. Gross J, Grass S, Davis AE, Gilmore-Erdmann P, Townsend RR, St. Geme JW. 2008. The Haemophilus influenzae HMW1 adhesin is a glycoprotein with an unusual N-linked carbohydrate modification. J Biol Chem 283:26010–26015.

33. Dell A, Galadari A, Sastre F, Hitchen P. 2010. Similarities and Differences in the Glycosylation Mechanisms in Prokaryotes and Eukaryotes. Int J Microbiol 2010.

34. Nothaft H, Szymanski CM. 2013. Bacterial protein n-glycosylation: New perspectives and applications. J Biol Chem 288:6912–6920.

35. Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH, Prestegard JH, Schnaar RL, Seeberger PH. 2015. Essentials of GlycobiologyEssentials of Glycobiology. Cold Spring Harbor Laboratory Press.

36. Jiangm YL, Jin H, Yang HB, Zhao RL, Wang S, Chen Y, Zhou CZ. 2017. Defining the enzymatic pathway for polymorphic O-glycosylation of the pneumococcal serine-rich repeat protein PsrP. J Biol Chem 292:6213–6224.

37. Filloux A, Whitfield C. 2016. Editorial: The many wonders of the bacterial cell surface. FEMS Microbiol Rev 40:161–163.

38. Zamyatina A. 2017. Aminosugar-based immunomodulator lipid A: Synthetic approaches. Beilstein J Org Chem. Beilstein-Institut Zur Forderung der Chemischen Wissenschaften.

39. Rietschel ET, Kirikae T, Schade FU, Mamat U, Schmidt G, Loppnow H, Ulmer AJ, Zähringer U, Seydel U, Di Padova F, Schreier M, Brade H. 1994. Bacterial endotoxin: molecular relationships of structure to activity and function. FASEB J 8:217–225.

40. Lodowska J, Wolny D, Weglarz L. 2013. The sugar 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) as a characteristic component of bacterial endotoxin - A review

of its biosynthesis, function, and placement in the lipopolysaccharide core. Can J Microbiol.

41. Kalynych S, Morona R, Cygler M. 2014. Progress in understanding the assembly process of bacterial O-antigen. FEMS Microbiol Rev 38:1048–1065.

42. Sperandeo P, Dehò G, Polissi A. 2009. The lipopolysaccharide transport system of Gram-negative bacteria. Biochim Biophys Acta - Mol Cell Biol Lipids. Elsevier.

43. Whitfield C, Williams DM, Kelly SD. 2020. Lipopolysaccharide O-antigens-bacterial glycans made to measure. J Biol Chem 295:10593–10609.

44. Brown S, Santa Maria JP, Walker S. 2013. Wall Teichoic Acids of Gram-Positive Bacteria. Annu Rev Microbiol 67:313–336.

45. Pereira MP, Brown ED. 2010. Biosynthesis of Cell Wall Teichoic Acid Polymers, p. 337–350. *In* Microbial Glycobiology. Elsevier Inc.

46. Percy MG, Gründling A. 2014. Lipoteichoic acid synthesis and function in gram-positive bacteria. Annu Rev Microbiol 68:81–100.

47. Van Der Es D, Hogendorf WFJ, Overkleeft HS, Van Der Marel GA, Codée JDC. 2017. Teichoic acids: Synthesis and applications. Chem Soc Rev. Royal Society of Chemistry.

48. Wenzel CQ, Mills DC, Dobruchowska JM, Vlach J, Nothaft H, Nation P, Azadi P, Melville SB, Carlson RW, Feldman M, Szymanski CM. 2020. An atypical lipoteichoic acid from Clostridium perfringens elicits a broadly cross-reactive and protective immune response. J Biol Chem 295:9513–9530.

49. Geno KA, Gilbert GL, Song JY, Skovsted IC, Klugman KP, Jones C, Konradsen HB, Nahm MH. 2015. Pneumococcal Capsules and Their Types: Past, Present, and Future. Clin Microbiol Rev 28:871–99.

50. Russell W, Herwald H. 2005. Concepts in bacterial virulence. Karger.

51. Woodward L, Naismith JH. 2016. Bacterial polysaccharide synthesis and export. Curr Opin Struct Biol. Elsevier Ltd.

52. Chan YGY, Kim HK, Schneewind O, Missiakas D. 2014. The capsular polysaccharide of Staphylococcus aureus is attached to peptidoglycan by the LytR-CpsA-Psr (LCP) family of enzymes. J Biol Chem 289:15680–15690.

53. Whitfield C, Wear SS, Sande C. 2020. Assembly of Bacterial Capsular Polysaccharides and Exopolysaccharides. https://doi.org/101146/annurev-micro-011420-075607 74:521–543.

54. Roberson EB, Firestone MK. 1992. Relationship between desiccation and

exopolysaccharide production in a soil Pseudomonas sp. Appl Environ Microbiol 58:1284–1291.

55. Ophir T, Gutnick DL. 1994. A role for exopolysaccharides in the protection of microorganisms from desiccation. Appl Environ Microbiol.

56. Cywes C, Stamenkovic I, Wessels MR. 2000. CD44 as a receptor for colonization of the pharynx by group A Streptococcus. J Clin Invest 106:995–1002.

57. Roberts IS, Saunders FK, Boulnois GJ. 1989. Bacterial capsules and interactions with complement and phagocytes. Biochem Soc Trans 17:462–464.

58. Hyams C, Yuste J, Bax K, Camberlein E, Weiser JN, Brown JS. 2010. Streptococcus pneumoniae resistance to complement-mediated immunity is dependent on the capsular serotype. Infect Immun 78:716–25.

59. Howard CJ, Glynn AA. 1971. The virulence for mice of strains of Escherichia coli related to the effects of K antigens on their resistance to phagocytosis and killing by complement. Immunology 20:767–77.

60. Kumar Kk, Reddy Gs, Reddy B, Shekar Pc, Sumanthi J, Chandra KlP. 2012. Biological role of lectins: A review. J Orofac Sci 4:20.

61. Lord JM, Roberts LM. 2004. Ricin: structure, synthesis, and mode of action, p. 215–233. *In* . Springer, Berlin, Heidelberg.

62. Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH, Prestegard JH, Schnaar RL, Seeberger PH. 2015. Essentials of GlycobiologyEssentials of Glycobiology. Cold Spring Harbor Laboratory Press.

63. Zelensky AN, Gready JE. 2005. The C-type lectin-like domain superfamily. FEBS J 272:6179–6217.

64. Brown GD, Willment JA, Whitehead L. 2018. C-type lectins in immunity and homeostasis. Nat Rev Immunol 18:374–389.

65. Flornes LM, Bryceson YT, Spurkland A, Lorentzen JC, Dissen E, Fossum S. 2004. Identification of lectin-like receptors expressed by antigen presenting cells and neutrophils and their mapping to a novel gene complex. Immunogenetics 56:506–517.

66. Behler-Janbeck F, Takano T, Maus R, Stolper J, Jonigk D, Tort Tarré M, Fuehner T, Prasse A, Welte T, M Timmer MS, Stocker BL, Nakanishi Y, Miyamoto T, Yamasaki S, Maus UA. 2016. C-type Lectin Mincle Recognizes Glucosyl-diacylglycerol of Streptococcus pneumoniae and Plays a Protective Role in Pneumococcal Pneumonia https://doi.org/10.1371/journal.ppat.1006038.

67. Patin EC, Orr SJ, Schaible UE. 2017. Macrophage inducible C-type lectin as a

multifunctional player in immunity. Front Immunol 8:861.

68. Imai T, Matsumura T, Mayer-Lambertz S, Wells CA, Ishikawa E, Butcher SK, Barnett TC, Walker MJ, Imamura A, Ishida H, Ikebe T, Miyamoto T, Ato M, Ohga S, Lepenies B, van Sorge NM, Yamasaki S. 2018. Lipoteichoic acid anchor triggers Mincle to drive protective immunity against invasive group A Streptococcus infection. Proc Natl Acad Sci U S A 115:E10662–E10671.

69. Takahashi K, Ip WE, Michelow IC, Ezekowitz RAB. 2006. The mannose-binding lectin: a prototypic pattern recognition molecule. Curr Opin Immunol 18:16–23.

70. Mathern DR, Heeger PS. 2015. Molecules great and small: The complement system. Clin J Am Soc Nephrol 10:1636–1650.

71. Juul-Madsen HR, Viertlböeck B, Härtle S, Smith AL, Göbel TW. 2013. Innate Immune Responses, p. 121–147. *In* Avian Immunology: Second Edition. Elsevier Inc.

72. Wallis R. 2007. Interactions between mannose-binding lectin and MASPs during complement activation by the lectin pathway. Immunobiology 212:289–299.

73. Hanage WP, Kaijalainen TH, Syrjänen RK, Auranen K, Leinonen M, Mäkelä PH, Spratt BG. 2005. Invasiveness of serotypes and clones of Streptococcus pneumoniae among children in Finland. Infect Immun 73:431–5.

74. Sahagún-Ruiz A, Breda LCD, Castiblanco Valencia MM, Elias WP, Munthe-Fog L, Garred P, Barbosa AS, Isaac L. 2015. Studies of the binding of ficolin-2 and ficolin-3 from the complement lectin pathway to Leptospira biflexa, Pasteurella pneumotropica and Diarrheagenic Escherichia coli. Immunobiology 220:1177–1185.

75. Luo F, Sun X, Wang Y, Wang Q, Wu Y, Pan Q, Fang C, Zhang XL. 2013. Ficolin-2 Defends against Virulent Mycobacteria Tuberculosis Infection In Vivo, and Its Insufficiency Is Associated with Infection in Humans. PLoS One 8:e73859.

76. Stowell SR, Arthur CM, Dias-Baruffi M, Rodrigues LC, Gourdine J-P, Heimburg-Molinaro J, Ju T, Molinaro RJ, Rivera-Marrero C, Xia B, Smith DF, Cummings RD. 2010. Innate immune lectins kill bacteria expressing blood group antigen. Nat Med 16:295–301.

77. Springer GF, Horton RE. 1969. Blood group isoantibody stimulation in man by feeding blood group-active bacteria. J Clin Invest 48:1280–91.

78. Springer GF, Williamson P, Brandes WC. 1961. Blood group activity of Gram-negative bacteria. J Exp Med 113:1077–1093.

79. Chen X, Khambu B, Zhang H, Gao W, Li M, Chen X, Yoshimori T, Yin XM. 2014. Autophagy induced by calcium phosphate precipitates targets damaged endosomes.

J Biol Chem 289:11162–11174.

80. Thurston TLM, Wandel MP, Von Muhlinen N, Foeglein Á, Randow F. 2012. Galectin 8 targets damaged vesicles for autophagy to defend cells against bacterial invasion. Nature 482:414–418.

81. Cash HL, Whitham C V., Behrendt CL, Hooper L V. 2006. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. Science (80-) 313:1126–1130.

82. Mukherjee S, Zheng H, Derebe MG, Callenberg KM, Partch CL, Rollins D, Propheter DC, Rizo J, Grabe M, Jiang QX, Hooper L V. 2014. Antibacterial membrane attack by a pore-forming intestinal C-type lectin. Nature 505:103–107.

83. Gallo RL, Hooper L V. 2012. Epithelial antimicrobial defence of the skin and intestine. Nat Rev Immunol 12:503–16.

84. Angata T, Brinkman-Van der Linden ECM. 2002. I-type lectins. Biochim Biophys Acta - Gen Subj 1572:294–316.

85. Powell LD, Varki A. 1995. I-type lectins. J Biol Chem 270:14243–14246.

86. Crocker PR, Paulson JC, Varki A. 2007. Siglecs and their roles in the immune system. Nat Rev Immunol 2007 74 7:255–266.

87. Powell$ LD, Sgroiq D, Sjoberg ER, Stamenkovic$ I, Varkill A. 1993. Natural ligands of the B cell adhesion molecule CD22 beta carry N-linked oligosaccharides with alpha-2,6-linked sialic acids that are required for recognition. J Biol Chem 268:7019–7027.

88. Kelm S, Schauer R, Manuguerra JC, Gross HJ, Crocker PR. 1994. Modifications of cell surface sialic acids modulate cell adhesion mediated by sialoadhesin and CD22. Glycoconjugate J 1994 116 11:576–585.

89. Crocker PR, Varki A. 2001. Siglecs, sialic acids and innate immunity. Trends Immunol 22:337–342.

90. Nitschke L. 2005. The role of CD22 and other inhibitory co-receptors in B-cell activation. Curr Opin Immunol 17:290–297.

91. Erickson-Miller CL, Freeman SD, Hopson CB, D'Alessio KJ, Fischer EI, Kikly KK, Abrahamson JA, Holmes SD, King AG. 2003. Characterization of Siglec-5 (CD170) expression and functional activity of anti-Siglec-5 antibodies on human phagocytes. Exp Hematol 31:382–388.

92. Lee JK, Buckhaults P, Wilkes C, Teilhet M, King M Lou, Moremen KW, Pierce M. 1997. Cloning and expression of a *Xenopus laevis* oocyte lectin and characterization of its mRNA levels during early development. Glycobiology 7:367–372.

93. Wyrick RE, Nishihara T, Hedrick JL. 1974. Agglutination of jelly coat and cortical granule components and the block to polyspermy in the amphibian Xenopus laevis. Proc Natl Acad Sci U S A 71:2067–71.

94. Nishihara T, Wyrick RE, Working PK, Chen YH, Hedrick JL. 1986. Isolation and characterization of a lectin from the cortical granules of Xenopus laevis eggs. Biochemistry 25:6013–6020.

95. Roberson4 MM, Barondesg SH. 1982. Lectin from Embryos and Oocytes of Xenopus laevisTHE JOURNAL OF BIOLWICAL CHEMISTRY.

96. Grey RD, Wolf DP, Hedrick JL. 1974. Formation and structure of the fertilization envelope in Xenopus laevis. Dev Biol 36:44–61.

97. Grey RD, Working PK, Hedrick JL. 1976. Evidence that the fertilization envelope blocks sperm entry in eggs of Xenopus laevis: Interaction of sperm with isolated envelopes. Dev Biol 54:52–60.

98. Monk BC, Hedrick JL. 1986. The cortical reaction in Xenopus laevis eggs--cortical granule lectin release as determined by radioimmunoassay. Zoolog Sci 3:p459-466.

99. Lee JK, Baum LG, Moremen K, Pierce M. 2004. The X-lectins: A new family with homology to the Xenopus laevis oocyte lectin XL-35. Glycoconj J 21:443–450.

100. Tsuji S, Uehori J, Matsumoto M, Suzuki Y, Matsuhisa A, Toyoshima K, Seya T. 2001. Human intelectin is a novel soluble lectin that recognizes galactofuranose in carbohydrate chains of bacterial cell wall. J Biol Chem 276:23456–63.

101. Lee JK, Schnee J, Pang M, Wolfert M, Baum LG, Moremen KW, Pierce M. 2001. Human homologs of the Xenopus oocyte cortical granule lectin XL35. Glycobiology 11:65–73.

102. Wesener DA, Wangkanont K, McBride R, Song X, Kraft MB, Hodges HL, Zarling LC, Splain RA, Smith DF, Cummings RD, Paulson JC, Forest KT, Kiessling LL. 2015. Recognition of microbial glycans by human intelectin-1. Nat Struct Mol Biol 22:603–610.

103. Wesener DA, Dugan A, Kiessling LL. 2017. Recognition of microbial glycans by soluble human lectins. Curr Opin Struct Biol 44:168–178.

104. McMahon CM, Isabella CR, Windsor IW, Kosma P, Raines RT, Kiessling LL. 2020. Stereoelectronic Effects Impact Glycan Recognition. J Am Chem Soc jacs.9b11699.

105. Hausdorff WP, Bryant J, Kloek C, Paradiso PR, Siber GR. 2000. The Contribution of Specific Pneumococcal Serogroups to Different Disease Manifestations: Implications for Conjugate Vaccine Formulation and Use, Part II. Clin Infect Dis 30:122–140.

106. Jochems SP, Weiser JN, Malley R, Ferreira DM. 2017. The immunological mechanisms that control pneumococcal carriage. PLOS Pathog 13:e1006665.

107. Hatzios SK, Abel S, Martell J, Hubbard T, Sasabe J, Munera D, Clark L, Bachovchin DA, Qadri F, Ryan ET, Davis BM, Weerapana E, Waldor MK. 2016. Chemoproteomic profiling of host and pathogen enzymes active in cholera. Nat Chem Biol 12:268–274.

108. Sigal M, Reinés M del M, Müllerke S, Fischer C, Kapalczynska M, Berger H, Bakker ERM, Mollenkopf H-J, Rothenberg ME, Wiedenmann B, Sauer S, Meyer TF. 2019. R-spondin-3 induces secretory, antimicrobial Lgr5+ cells in the stomach. Nat Cell Biol 21:812–823.

109. de Souza Batista CM, Yang R-Z, Lee M-J, Glynn NM, Yu D-Z, Pray J, Ndubuizu K, Patil S, Schwartz A, Kligman M, Fried SK, Gong D-W, Shuldiner AR, Pollin TI, McLenithan JC. 2007. Omentin plasma levels and gene expression are decreased in obesity. Diabetes 56:1655–61.

110. Washimi K, Yokose T, Yamashita M, Kageyama T, Suzuki K, Yoshihara M, Miyagi Y, Hayashi H, Tsuji S. 2012. Specific Expression of Human Intelectin-1 in Malignant Pleural Mesothelioma and Gastrointestinal Goblet Cells. PLoS One 7:e39889.

111. Wrackmeyer U, Hansen GH, Seya T, Danielsen EM. 2006. Intelectin: a novel lipid raft-associated protein in the enterocyte brush border. Biochemistry 45:9188–97.

112. Tsuji S, Uehori J, Matsumoto M, Suzuki Y, Matsuhisa A, Toyoshima K, Seya T. 2001. Human Intelectin Is a Novel Soluble Lectin That Recognizes Galactofuranose in Carbohydrate Chains of Bacterial Cell Wall. J Biol Chem 276:23456–23463.

113. Lee JK, Schnee J, Pang M, Wolfert M, Baum LG, Moremen KW, Pierce M. 2001. Human homologs of the Xenopus oocyte cortical granule lectin XL35. Glycobiology 11:65–73.

114. Yang R-Z, Lee M-J, Hu H, Pray J, Wu H-B, Hansen BC, Shuldiner AR, Fried SK, Mclenithan JC, Gong D-W. 2006. Identification of omentin as a novel depot-specific adipokine in human adipose tissue: possible role in modulating insulin action. Am J Physiol Endocrinol Metab 290:1253–1261.

115. Kuperman DA, Lewis CC, Woodruff PG, Rodriguez MW, Yang YH, Dolganov GM, Fahy J V., Erle DJ. 2005. Dissecting asthma using focused transgenic modeling and functional genomics. J Allergy Clin Immunol 116:305–311.

116. Voehringer D, Stanley SA, Cox JS, Completo GC, Lowary TL, Locksley RM. 2007. Nippostrongylus brasiliensis: Identification of intelectin-1 and -2 as Stat6-dependent genes expressed in lung and intestine during infection. Exp Parasitol 116:458–466.

117. French AT, Bethune JA, Knight PA, McNeilly TN, Wattegedera S, Rhind S, Miller

HRP, Pemberton AD. 2007. The expression of intelectin in sheep goblet cells and upregulation by interleukin-4. Vet Immunol Immunopathol 120:41–46.

118. French AT, Knight PA, Smith WD, Brown JK, Craig NM, Pate JA, Miller HRP, Pemberton AD. 2008. Up-regulation of intelectin in sheep after infection with Teladorsagia circumcincta. Int J Parasitol 38:467–475.

119. Pemberton AD, Knight PA, Gamble J, Colledge WH, Lee J-K, Pierce M, Miller HRP. 2004. Innate BALB/c Enteric Epithelial Responses to Trichinella spiralis: Inducible Expression of a Novel Goblet Cell Lectin, Intelectin-2, and Its Natural Deletion in C57BL/10 Mice. J Immunol 173:1894–1901.

120. Yang R-Z, Lee M-J, Hu H, Pray J, Wu H-B, Hansen BC, Shuldiner AR, Fried SK, McLenithan JC, Gong D-W. 2006. Identification of omentin as a novel depot-specific adipokine in human adipose tissue: possible role in modulating insulin action. Am J Physiol Metab 290:E1253–E1261.

121. Batista CM de S, Yang R-Z, Lee M-J, Glynn NM, Yu D-Z, Pray J, Ndubuizu K, Patil S, Schwartz A, Kligman M, Fried SK, Gong D-W, Shuldiner AR, Pollin TI, McLenithan JC. 2007. Omentin Plasma Levels and Gene Expression Are Decreased in Obesity. Diabetes 56:1655–1661.

122. Tan Y-L, Zheng X-L, Tang C-K. 2015. The protective functions of omentin in cardiovascular diseases. Clin Chim Acta 448:98–106.

123. Ruffie P, Feld R, Minkin S, Cormier Y, Boutan-Laroze A, Ginsberg R, Ayoub J, Shepherd FA, Evans WK, Figueredo A. 1989. Diffuse malignant mesothelioma of the pleura in Ontario and Quebec: a retrospective study of 332 patients. J Clin Oncol 7:1157–1168.

124. Sorensen M, Sorensen SPL. 1939. The proteins in whey. Compt Rend Trav Lab Carlsb 23:55–59.

125. Schanbacher FL, Goodman RE, Talhouk RS. 1993. Bovine Mammary Lactoferrin: Implications from Messenger Ribonucleic Acid (mRNA) Sequence and Regulation Contrary to Other Milk Proteins. J Dairy Sci 76:3812–3831.

126. Masson PL, Heremans JF, Dive CH. 1966. An iron-binding protein common to many external secretions. Clin Chim Acta 14:735–739.

127. Baker EN, Baker HM. 2005. Molecular structure, binding properties and dynamics of lactoferrin. Cell Mol Life Sci 62:2531–2539.

128. Anderson BF, Baker HM, Norris GE, Rice DW, Baker EN. 1989. Structure of human lactoferrin: crystallographic structure analysis and refinement at 2.8 A resolution. J Mol Biol 209:711–34.

129. García-Montoya IA, Cendón TS, Arévalo-Gallegos S, Rascón-Cruz Q. 2012.

Lactoferrin a multiple bioactive protein: An overview. Biochim Biophys Acta - Gen Subj 1820:226–236.

130. Conneely OM. 2001. Antiinflammatory activities of lactoferrin. J Am Coll Nutr 20:389S-395S; discussion 396S-397S.

131. Arnold RR, Cole MF, McGhee JR. 1977. A bactericidal effect for human lactoferrin. Science 197:263–5.

132. Rainard P. 1986. Bacteriostatic activity of bovine milk lactoferrin against mastitic bacteria. Vet Microbiol 11:387–92.

133. Ellison RT, Giehl TJ, LaForce FM. 1988. Damage of the outer membrane of enteric gram-negative bacteria by lactoferrin and transferrin. Infect Immun 56:2774–81.

134. Leitch EC, Willcox MDP. 1999. Elucidation of the antistaphylococcal action of lactoferrin and lysozyme. J Med Microbiol 48:867–871.

135. Suzuki YA, Shin K, Lönnerdal B. 2001. Molecular cloning and functional expression of a human intestinal lactoferrin receptor. Biochemistry 40:15771–9.

136. Shin K, Wakabayashi H, Yamauchi K, Yaeshima T, Iwatsuki K. 2008. Recombinant human intelectin binds bovine lactoferrin and its peptides. Biol Pharm Bull 31:1605–8.

137. Kerr SC, Carrington SD, Oscarson S, Gallagher ME, Solon M, Yuan S, Ahn JN, Dougherty RH, Finkbeiner WE, Peters MC, Fahy J V. 2014. Intelectin-1 is a prominent protein constituent of pathologic mucus associated with eosinophilic airway inflammation in asthma. Am J Respir Crit Care Med 189:1005–7.

138. Pemberton AD, Knight PA, Gamble J, Colledge WH, Lee J-K, Pierce M, Miller HRP. 2004. Innate BALB/c Enteric Epithelial Responses to Trichinella spiralis: Inducible Expression of a Novel Goblet Cell Lectin, Intelectin-2, and Its Natural Deletion in C57BL/10 Mice. J Immunol 173:1894–1901.

139. Fischetti VA, Ryan P. 1996. Streptococcus. Pract Handb Microbiol Third Ed 411–428.

140. Hardie JM, Whiley RA. 1995. The genus Streptococcus. Genera Lact Acid Bact 55–124.

141. Couvigny B, De Wouters T, Kaci G, Jacouton E, Delorme C, Doré J, Renault P, Blottière HM, Guédon E, Lapaque N. 2015. Commensal Streptococcus salivarius Modulates PPARγ Transcriptional Activity in Human Intestinal Epithelial Cells. PLoS One 10.

142. Pérez-Trallero E, Vicente D, Montes M, Marimon JM, Piñeiro L. 2001. High proportion of pharyngeal carriers of commensal streptococci resistant to

erythromycin in Spanish adults. J Antimicrob Chemother 48:225–229.

143. Krzyściak W, Pluskwa KK, Jurczak A, Kościelniak D. 2013. The pathogenicity of the Streptococcus genus. Eur J Clin Microbiol Infect Dis 32:1361.

144. Denapaite D, Brückner R, Hakenbeck R, Vollmer W. 2012. Biosynthesis of teichoic acids in Streptococcus pneumoniae and closely related species: lessons from genomes. Microb Drug Resist 18:344–358.

145. Zorzoli A, Meyer BH, Adair E, Torgov VI, Veselovsky V V., Danilov LL, Uhrin D, Dorfmueller HC. 2019. Group A, B, C, and G Streptococcus Lancefield antigen biosynthesis is initiated by a conserved α-D-GlcNAc-β-1,4-L-rhamnosyltransferase. J Biol Chem 294:15237–15256.

146. Lancefield RC. 1933. A SEROLOGICAL DIFFERENTIATION OF HUMAN AND OTHER GROUPS OF HEMOLYTIC STREPTOCOCCI. J Exp Med 57:571.

147. Heidelberger M, Avery OT. 1923. The soluble specific substance of pneumococcus. J Exp Med 38:73.

148. Slater J. 2007. Bacterial Infections of the Equine Respiratory Tract. Equine Respir Med Surg 327–353.

149. Zhu F, Zhang H, Yang T, Haslam SM, Dell A, Wu H. 2016. Engineering and Dissecting the Glycosylation Pathway of a Streptococcal Serine-rich Repeat Adhesin. J Biol Chem 291:27354–27363.

150. Latousakis D, Juge N. 2018. How sweet are our gut beneficial bacteria? A focus on protein glycosylation in Lactobacillus. Int J Mol Sci. MDPI AG.

151. Huang SS, Johnson KM, Ray GT, Wroe P, Lieu TA, Moore MR, Zell ER, Linder JA, Grijalva CG, Metlay JP, Finkelstein JA. 2011. Healthcare utilization and cost of pneumococcal disease in the United States. Vaccine 29:3398–3412.

152. Loughran AJ, Orihuela CJ, Tuomanen EI. 2019. Streptococcus pneumoniae: Invasion and Inflammation. Microbiol Spectr 7.

153. Weiser JN, Ferreira DM, Paton JC. 2018. Streptococcus pneumoniae: Transmission, colonization and invasion. Nat Rev Microbiol. Nature Publishing Group.

154. Hyams C, Camberlein E, Cohen JM, Bax K, Brown JS. 2010. The Streptococcus pneumoniae capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. Infect Immun 78:704–15.

155. Rep BE-PH, 1944 undefined. Cross reactions between the several pneumococcic types and their significance in the preparation of polyvalent antiserum. ncbi.nlm.nih.gov.

156. Sørensen UB. 1993. Typing of pneumococci by using 12 pooled antisera. J Clin Microbiol 31:2097–100.

157. Synytsya A, Novak M. 2014. Structural analysis of glucans. Ann Transl Med 2:17.

158. Batt SL, Charalambous BM, McHugh TD, Martin S, Gillespie SH. 2005. Novel PCR-restriction fragment length polymorphism method for determining serotypes or serogroups of Streptococcus pneumoniae isolates. J Clin Microbiol 43:2656–61.

159. Brito DA, Ramirez M, de Lencastre H. 2003. Serotyping Streptococcus pneumoniae by multiplex PCR. J Clin Microbiol 41:2378–84.

160. Calix JJ, Porambo RJ, Brady AM, Larson TR, Yother J, Abeygunwardana C, Nahm MH. 2012. Biochemical, genetic, and serological characterization of two capsule subtypes among Streptococcus pneumoniae Serotype 20 strains: discovery of a new pneumococcal serotype. J Biol Chem 287:27885–94.

161. Cartee RT, Forsee WT, Schutzbach JS, Yother J. 2000. Mechanism of type 3 capsular polysaccharide synthesis in Streptococcus pneumoniae. J Biol Chem 275:3907–14.

162. Guidolin A, Morona JK, Morona R, Hansman D, Paton JC. 1994. Nucleotide sequence analysis of genes essential for capsular polysaccharide biosynthesis in Streptococcus pneumoniae type 19F. Infect Immun 62:5384–96.

163. Raetz CRH, Whitfield C. 2002. Lipopolysaccharide Endotoxins. Annu Rev Biochem 71:635.

164. Kim JO, Weiser JN. 1998. Association of Intrastrain Phase Variation in Quantity of Capsular Polysaccharide and Teichoic Acid with the Virulence of *Streptococcus pneumoniae*. J Infect Dis 177:368–377.

165. Cundell DR, Weiser JN, Shen J, Young A, Tuomanen EI. 1995. Relationship between colonial morphology and adherence of Streptococcus pneumoniae. Infect Immun 63:757–761.

166. Weiser JN, Austrian R, Sreenivasan PK, Masure HR. 1994. Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization. Infect Immun 62:2582–9.

167. Gilley RP, Orihuela CJ. 2014. Pneumococci in biofilms are non-invasive: implications on nasopharyngeal colonization. Front Cell Infect Microbiol 4:163.

168. Mavroidi A, Aanensen DM, Godoy D, Skovsted IC, Kaltoft MS, Reeves PR, Bentley SD, Spratt BG. 2007. Genetic relatedness of the Streptococcus pneumoniae capsular biosynthetic loci. J Bacteriol 189:7841–55.

169. Dalia AB, Weiser JN. 2011. Minimization of bacterial size Allows for complement

evasion and Is overcome by the agglutinating effect of antibody. Cell Host Microbe 10:486–496.

170. Jarva H, Jokiranta TS, Würzner R, Meri S. 2003. Complement resistance mechanisms of streptococci. Mol Immunol 40:95–107.

171. Brown EJ, Hosea SW, Frank MM. 1983. The Role of Antibody and Complement in the Reticuloendothelial Clearance of Pneumococci from the Bloodstream. Rev Infect Dis 5:S797–S805.

172. Paterson GK, Mitchell TJ. 2006. Innate immunity and the pneumococcus. Microbiology. Microbiology Society.

173. Andre GO, Converso TR, Politano WR, Ferraz LFC, Ribeiro ML, Leite LCC, Darrieux M. 2017. Role of Streptococcus pneumoniae Proteins in Evasion of Complement-Mediated Immunity. Front Microbiol 8:224.

174. Mukerji R, Mirza S, Roche AM, Widener RW, Croney CM, Rhee D-K, Weiser JN, Szalai AJ, Briles DE. 2012. Pneumococcal Surface Protein A Inhibits Complement Deposition on the Pneumococcal Surface by Competing with the Binding of C-Reactive Protein to Cell-Surface Phosphocholine. J Immunol 189:5327–5335.

175. Tu AHT, Fulgham RL, Mccrory MA, Briles DE, Szalai AJ. 1999. Pneumococcal surface protein A inhibits complement activation by Streptococcus pneumoniae. Infect Immun 67:4720–4724.

176. Yother J, Briles DE. 1992. Structural properties and evolutionary relationships of PspA, a surface protein of Streptococcus pneumoniae, as revealed by sequence analysis. J Bacteriol 174:601–609.

177. Dave S, Carmicle S, Hammerschmidt S, Pangburn MK, McDaniel LS. 2004. Dual Roles of PspC, a Surface Protein of Streptococcus pneumoniae , in Binding Human Secretory IgA and Factor H . J Immunol 173:471–477.

178. Dave S, Brooks-Walter A, Pangburn MK, McDaniel LS. 2001. PspC, a pneumococcal surface protein, binds human factor H. Infect Immun 69:3435–3437.

179. O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, Lee E, Mulholland K, Levine OS, Cherian T. 2009. Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. Lancet 374:893–902.

180. Daniels CC, Rogers PD, Shelton CM. 2016. A review of pneumococcal vaccines: Current polysaccharide vaccine recommendations and future protein antigens. J Pediatr Pharmacol Ther. Pediatric Pharmacy Advocacy Group, Inc.

181. Sun L, Middleton DR, Wantuch PL, Ozdilek A, Avci FY. 2016. Carbohydrates as T-cell antigens with implications in health and disease. Glycobiology 26:1029–

1040.

182. Wantuch PL, Avci FY. 2018. Current status and future directions of invasive pneumococcal diseases and prophylactic approaches to control them. Hum Vaccin Immunother 14:2303–2309.

183. Aliberti S, Mantero M, Mirsaeidi M, Blasi F. 2014. The role of vaccination in preventing pneumococcal disease in adults. Clin Microbiol Infect 20:52–58.

184. Moore MR, Link-Gelles R, Schaffner W, Lynfield R, Lexau C, Bennett NM, Petit S, Zansky SM, Harrison LH, Reingold A, Miller L, Scherzinger K, Thomas A, Farley MM, Zell ER, Taylor TH, Pondo T, Rodgers L, McGee L, Beall B, Jorgensen JH, Whitney CG. 2015. Effect of use of 13-valent pneumococcal conjugate vaccine in children on invasive pneumococcal disease in children and adults in the USA: Analysis of multisite, population-based surveillance. Lancet Infect Dis 15:301–309.

185. Nuorti J, CG W. 2010. Prevention of pneumococcal disease among infants and children—use of 13-valent pneumococcal conjugate vaccine and 23-valent pneumococcal polysaccharide vaccine—recommendations of the Advisory Committee on Immunization Practices (ACIP). Morb Mortal Wkly Rep 59:1–18.

186. Tomczyk S, Bennett NM, Stoecker C, Gierke R, Moore MR, Whitney CG, Hadler S, Pilishvili T. 2014. Use of 13-Valent Pneumococcal Conjugate Vaccine and 23-Valent Pneumococcal Polysaccharide Vaccine Among Adults Aged ≥65 Years: Recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Morb Mortal Wkly Rep 63:822.

187. Dagan R, Poolman J, Siegrist C-A. 2010. Glycoconjugate vaccines and immune interference: A review. Vaccine 28:5513–5523.

188. Keller LE, Robinson DA, McDaniel LS. 2016. Nonencapsulated *Streptococcus pneumoniae* : Emergence and Pathogenesis. MBio 7:e01792.

189. Lemos JA, Palmer SR, Zeng L, Wen ZT, Kajfasz JK, Freires IA, Abranches J, Brady LJ. 2019. The Biology of Streptococcus mutans. Microbiol Spectr 7.

190. Centers for Disease Control and Prevention U. 2001. Promoting oral health: interventions for preventing dental caries, oral and pharyngeal cancers, and sports-related craniofacial injuries. a report on recommendations of the task force on community preventive services. MMWR Recomm reports Morb Mortal Wkly report Recomm reports / Centers Dis Control 50:1–13.

191. Abranches J, Miller JH, Martinez AR, Simpson-Haidaris PJ, Burne RA, Lemos JA. 2011. The collagen-binding protein Cnm is required for Streptococcus mutans adherence to and intracellular invasion of human coronary artery endothelial cells. Infect Immun 79:2277–2284.

192. Drangsholt MT. 1998. A New Causal Model of Dental Diseases Associated With

Endocarditis. Ann Periodontol 3:184–196.

193. Nakano K, Ooshima T. 2009. Serotype classification of Streptococcus mutans and its detection outside the oral cavity. http://dx.doi.org/102217/fmb0964 4:891–902.

194. Mistou MY, Sutcliffe IC, Van Sorge NM. 2016. Bacterial glycobiology: rhamnose-containing cell wall polysaccharides in Gram-positive bacteria. FEMS Microbiol Rev 40:464–479.

195. Kovacs CJ, Faustoferri RC, Quivey RG. 2017. RgpF is required for maintenance of stress tolerance and virulence in Streptococcus mutans. J Bacteriol 199.

196. Shibata Y, Yamashita Y, Ozaki K, Nakano Y, Koga T. 2002. Expression and characterization of streptococcal rgp genes required for rhamnan synthesis in Escherichia coli. Infect Immun 70:2891–2898.

197. Kovacs CJ, Faustoferri RC, Bischer AP, Quivey RG. 2019. Streptococcus mutans requires mature rhamnose-glucose polysaccharides for proper pathophysiology, morphogenesis and cellular division. Mol Microbiol 112:944–959.

198. Zamakhaeva S, Chaton CT, Rush JS, Ajay Castro S, Kenner CW, Yarawsky AE, Herr AB, van Sorge NM, Dorfmueller HC, Frolenkov GI, Korotkov K V., Korotkova N. 2021. Modification of cell wall polysaccharide guides cell division in Streptococcus mutans. Nat Chem Biol 2021 178 17:878–887.

199. van der Beek SL, Le Breton Y, Ferenbach AT, Chapman RN, van Aalten DMF, Navratilova I, Boons GJ, Mciver KS, van Sorge NM, Dorfmueller HC. 2015. GacA is essential for Group A Streptococcus and defines a new class of monomeric dTDP-4-dehydrorhamnose reductases (RmlD). Mol Microbiol 98:946–962.

200. Glim JE, Everts V, Niessen FB, Ulrich MM, Beelen RHJ. 2014. Extracellular matrix components of oral mucosa differ from skin and resemble that of foetal skin. Arch Oral Biol 59:1048–1055.

201. Avilés-Reyes A, Miller JH, Lemos JA, Abranches J. 2017. Collagen-binding proteins of Streptococcus mutans and related streptococci. Mol Oral Microbiol 32:89–106.

202. Sciotti M, Yamodo I, … JK-F microbiology, 1997 undefined. The N-terminal half part of the oral streptococcal antigen I/IIf contains two distinct binding domains. academic.oup.com.

203. Han T, Zhang C, research MD-B and biophysical, 2006 undefined. Identification and characterization of collagen-binding activity in Streptococcus mutans wall-associated protein: a possible implication in dental root caries and. Elsevier.

204. Avilés-Reyes A, Freires IA, Besingi R, Purushotham S, Deivanayagam C, Brady LJ, Abranches J, Lemos JA. 2018. Characterization of the pgf operon involved in the

posttranslational modification of Streptococcus mutans surface proteins. Sci Rep 8.

205. Nakano K, Lapirattanakul J, Nomura R, Nemoto H, Alaluusua S, Grönroos L, Vaara M, Hamada S, Ooshima T, Nakagawa I. 2007. Streptococcus mutans clonal variation revealed by multilocus sequence typing. J Clin Microbiol 45:2616–2625.

206. Nomura R, Nakano K, Taniguchi N, Lapirattanakul J, Nemoto H, Grönroos L, Alaluusua S, Ooshima T. 2009. Molecular and clinical analyses of the gene encoding the collagen-binding adhesin of Streptococcus mutans. J Med Microbiol 58:469–475.

207. Nomura R, Nakano K, Naka S, Nemoto H, Masuda K, Lapirattanakul J, Alaluusua S, Matsumoto M, Kawabata S, Ooshima T. 2012. Identification and characterization of a collagen-binding protein, Cbm, in Streptococcus mutans. Mol Oral Microbiol 27:308–323.

208. Sato Y, Okamoto K, Kagami A, Yamamoto Y, Igarashi T, Kizaki H. 2004. Streptococcus mutans strains harboring collagen-binding adhesin. J Dent Res 83:534–539.

209. Miller JH, Avilés-Reyes A, Scott-Anne K, Gregoire S, Watson GE, Sampson E, Progulske-Fox A, Koo H, Bowen WH, Lemos JA, Abranches J. 2015. The collagen binding protein Cnm contributes to oral colonization and cariogenicity of Streptococcus mutans OMZ175. Infect Immun 83:2001–2010.

210. Avilés-Reyes A, Miller JH, Simpson-Haidaris PJ, Hagen FK, Abranches J, Lemos JA. 2014. Modification of Streptococcus mutans Cnm by PgfS contributes to adhesion, endothelial cell invasion, and virulence. J Bacteriol 196:2789–2797.

211. Chen Y, Seepersaud R, Bensing BA, Sullam PM, Rapoport TA. 2016. Mechanism of a cytosolic O-glycosyltransferase essential for the synthesis of a bacterial adhesion protein. Proc Natl Acad Sci U S A 113:E1190–E1199.

212. Hyams C, Camberlein E, Cohen JM, Bax K, Brown JS. 2010. The Streptococcus pneumoniae capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. Infect Immun 78:704–715.

213. Abranches J, Chen YYM, Burne RA. 2004. Galactose Metabolism by Streptococcus mutans. Appl Environ Microbiol 70:6047.

214. Zeng L, Das S, Burne RA. 2010. Utilization of lactose and galactose by Streptococcus mutans: Transport, toxicity, and carbon catabolite repression. J Bacteriol 192:2434–2444.

215. Matsushita M, Endo Y, Fujita T. 2000. Cutting edge: complement-activating complex of ficolin and mannose-binding lectin-associated serine protease. J Immunol 164:2281–4.

216. Calix JJ, Nahm MH, Zartler ER. 2013. Elucidation of Structural and Antigenic Properties of Pneumococcal Serotype 11A, 11B, 11C, and 11F Polysaccharide Capsules. J Bacteriol 195:4552–4552.

217. Wyrick RE, Nishihara T, Hedrick JL. 1974. Agglutination of jelly coat and cortical granule components and the block to polyspermy in the amphibian Xenopus laevis. Proc Natl Acad Sci U S A 71:2067–71.

218. Uta Wrackmeyer ‡, Gert H. Hansen ‡, Tsukasa Seya § and, E. Michael Danielsen* ‡. 2006. Intelectin: A Novel Lipid Raft-Associated Protein in the Enterocyte Brush Border† https://doi.org/10.1021/BI060570X.

219. Carolan BJ, Harvey B-G, De BP, Vanni H, Crystal RG. 2008. Decreased Expression of Intelectin 1 in the Human Airway Epithelium of Smokers Compared to Nonsmokers. J Immunol 181:5760–5767.

220. Kamm K, Schierwater B, DeSalle R. 2019. Innate immunity in the simplest animals - placozoans. BMC Genomics 20:5.

221. Dalia AB, Weiser JN. 2011. Article Minimization of Bacterial Size Allows for Complement Evasion and Is Overcome by the Agglutinating Effect of Antibody. Cell Host Microbe 10:486–496.

222. Takei H, Araki A, Watanabe H, Lchinose A, Sendo F. 1996. Rapid killing of human neutrophils by the potent activator phorbol 12-myristate 13-acetate (PMA) accompanied by changes different from typical apoptosis or necrosis. Wiley Online Libr 59:229.

223. Brinkmann V, Reichard U, Goosmann C, Fauler B, Uhlemann Y, Weiss DS, Weinrauch Y, Zychlinsky A. 2004. Neutrophil Extracellular Traps Kill Bacteria. Science (80-) 303:1532–1535.

224. Mukherjee S, Zheng H, Derebe MG, Callenberg KM, Partch CL, Rollins D, Propheter DC, Rizo J, Grabe M, Jiang Q-X, Hooper L V. 2014. Antibacterial membrane attack by a pore-forming intestinal C-type lectin. Nature 505:103–107.

225. Tietze C, Schlesinger P, Stahl P. 1982. Mannose-specific endocytosis receptor of alveolar macrophages: demonstration of two functionally distinct intracellular pools of receptor and their roles in receptor recycling. J Cell Biol 92:417–424.

226. Stahl PD, Ezekowitz RAB. 1998. The mannose receptor is a pattern recognition receptor involved in host defense. Curr Opin Immunol 10:50–55.

227. Rodriguez JL, Dalia AB, Weiser JN. 2012. Increased Chain Length Promotes Pneumococcal Adherence and Colonization. Infect Immun 80:3454.

228. Brouwer N, Dolman KM, van Houdt M, Sta M, Roos D, Kuijpers TW. 2008. Mannose-binding lectin (MBL) facilitates opsonophagocytosis of yeasts but not of

bacteria despite MBL binding. J Immunol 180:4124–4132.

229. Caron E, Hall A. 1998. Identification of two distinct mechanisms of phagocytosis controlled by different Rho GTPases. Science 282:1717–1721.

230. Gordon DL, Johnson GM, Hostetter MK, Gordon DL, Johnson GM, Hostetter MK. 1986. Ligand-receptor interactions in the phagocytosis of virulent Streptococcus pneumoniae by polymorphonuclear leukocytes. J Infect Dis 154:619–626.

231. Le Cabec V, Carréno S, Moisand A, Bordier C, Maridonneau-Parini I. 2002. Complement receptor 3 (CD11b/CD18) mediates type I and type II phagocytosis during nonopsonic and opsonic phagocytosis, respectively. J Immunol 169:2003–2009.

232. Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, Kuo A, Mitros T, Salamov A, Carpenter ML, Signorovitch AY, Moreno MA, Kamm K, Grimwood J, Schmutz J, Shapiro H, Grigoriev I V., Buss LW, Schierwater B, Dellaporta SL, Rokhsar DS. 2008. The Trichoplax genome and the nature of placozoans. Nature 454:955–960.

233. SCHULZE, E. F. 1883. Trichoplax adhaerens, nov. gen., nov. spec. Zool Anz 6:92–97.

234. Ruthmann A, Behrendt G, Wahl R. 1986. The ventral epithelium of Trichoplax adhaerens (Placozoa): Cytoskeletal structures, cell contacts and endocytosis. Zoomorphology 106:115–122.

235. Kamm K, Osigus H-J, Stadler PF, DeSalle R, Schierwater B. 2018. Trichoplax genomes reveal profound admixture and suggest stable wild populations without bisexual reproduction. Sci Rep 8:11168.

236. O'Toole GA. 2011. Microtiter Dish Biofilm Formation Assay. J Vis Exp https://doi.org/10.3791/2437.

237. Novick S, Shagan M, Blau K, Lifshitz S, Givon-Lavi N, Grossman N, Bodner L, Dagan R, Mizrachi Nebenzahl Y. 2017. Adhesion and invasion of Streptococcus pneumoniae to primary and secondary respiratory epithelial cells. Mol Med Rep 15:65–74.

238. Crippen CS, Glushka J, Vinogradov E, Szymanski CM. 2021. Trehalose-deficient Acinetobacter baumannii exhibits reduced virulence by losing capsular polysaccharide and altering membrane integrity. Glycobiology https://doi.org/10.1093/GLYCOB/CWAB096.

239. Lees-Miller RG, Iwashkiw JA, Scott NE, Seper A, Vinogradov E, Schild S, Feldman MF. 2013. A common pathway for O-linked protein-glycosylation and synthesis of capsule in Acinetobacter baumannii. Mol Microbiol 89:816–830.

240. Fantone K, Tucker SL, Miller A, Yadav R, Bernardy EE, Fricker R, Stecenko AA, Goldberg JB, Rada B. 2021. Cystic Fibrosis Sputum Impairs the Ability of Neutrophils to Kill Staphylococcus aureus. Pathog (Basel, Switzerland) 10.

241. Rada BK, Geiszt M, Káldi K, Timár C, Ligeti E. 2004. Dual role of phagocytic NADPH oxidase in bacterial killing. Blood 104:2947–2953.

242. Sil P, Yoo DG, Floyd M, Gingerich A, Rada B. 2016. High Throughput Measurement of Extracellular DNA Release and Quantitative NET Formation in Human Neutrophils In Vitro. JoVE (Journal Vis Exp 2016:e52779.

243. Rada B, Jendrysik MA, Pang L, Hayes CP, Yoo DG, Park JJ, Moskowitz SM, Malech HL, Leto TL. 2013. Pyocyanin-Enhanced Neutrophil Extracellular Trap Formation Requires the NADPH Oxidase. PLoS One 8:e54205.

244. Yoo D, Winn M, Pang L, Moskowitz SM, Malech HL, Leto TL, Rada B. 2014. Release of cystic fibrosis airway inflammatory markers from Pseudomonas aeruginosa-stimulated human neutrophils involves NADPH oxidase-dependent extracellular DNA trap formation. J Immunol 192:4728–4738.

245. Pitts NB, Zero DT, Marsh PD, Ekstrand K, Weintraub JA, Ramos-Gomez F, Tagami J, Twetman S, Tsakos G, Ismail A. 2017. Dental caries. Nat Rev Dis Prim 2017 31 3:1–16.

246. Bowen WH, Burne RA, Wu H, Koo H. 2018. Oral Biofilms: Pathogens, Matrix, and Polymicrobial Interactions in Microenvironments. Trends Microbiol 26:229–242.

247. Sciotti MA, Chatenay-Rivauday C, Yamodo I, Ogier J. 1997. The N-terminal half part of the oral streptococcal antigen I/IIf contains two distinct functional domains. Adv Exp Med Biol 418:699–701.

248. Han TK, Zhang C, Dao ML. 2006. Identification and characterization of collagen-binding activity in Streptococcus mutans wall-associated protein: A possible implication in dental root caries and endocarditis. Biochem Biophys Res Commun 343:787–792.

249. Nomura R, Ogaya Y, Nakano K. 2016. Contribution of the collagen-binding proteins of Streptococcus mutans to bacterial colonization of inflamed dental pulp. PLoS One 11.

250. Avilés-Reyes A, Miller JH, Lemos JA, Abranches J. 2017. Collagen-binding proteins of Streptococcus mutans and related streptococci. Mol Oral Microbiol 32:89–106.

251. Nakano K, Nomura R, Taniguchi N, Lapirattanakul J, Kojima A, Naka S, Senawongse P, Srisatjaluk R, Grönroos L, Alaluusua S, Matsumoto M, Ooshima T. 2010. Molecular characterization of Streptococcus mutans strains containing the cnm gene encoding a collagen-binding adhesin. Arch Oral Biol 55:34–39.

252. Sato Y, Okamoto K, Kagami A, Yamamoto Y, Igarashi T, Kizaki H. 2004. Streptococcus mutans strains harboring collagen-binding adhesin. J Dent Res 83:534–539.

253. Naka S, Wato K, Misaki T, Ito S, Matsuoka D, Nagasawa Y, Nomura R, Matsumoto-Nakano M, Nakano K. 2021. Streptococcus mutans induces IgA nephropathy-like glomerulonephritis in rats with severe dental caries. Sci Rep 11:5784.

254. Esberg A, Sheng N, Mårell L, Claesson R, Persson K, Borén T, Strömberg N. 2017. Streptococcus mutans adhesin biotypes that match and predict individual caries development. EBioMedicine 24:205–215.

255. Garcia BA, Acosta NC, Tomar SL, Roesch LFW, Lemos JA, Mugayar LRF, Abranches J. 2021. Association of Candida albicans and Cbp+ Streptococcus mutans with early childhood caries recurrence. Sci Rep 11:1–11.

256. Inenaga C, Hokamura K, Nakano K, Nomura R, Naka S, Ohashi T, Ooshima T, Kuriyama N, Hamasaki T, Wada K, Umemura K, Tanaka T. 2018. A Potential New Risk Factor for Stroke: Streptococcus Mutans With Collagen-Binding Protein. World Neurosurg 113:e77–e81.

257. Hosoki S, Saito S, Tonomura S, Ishiyama H, Yoshimoto T, Ikeda S, Ikenouchi H, Yamamoto Y, Hattori Y, Miwa K, Friedland RP, Carare RO, Nakahara J, Suzuki N, Koga M, Toyoda K, Nomura R, Nakano K, Takegami M, Ihara M. 2020. Oral Carriage of Streptococcus mutans Harboring the cnm Gene Relates to an Increased Incidence of Cerebral Microbleeds. Stroke 3632–3639.

258. Nomura R, Otsugu M, Hamada M, Matayoshi S, Teramoto N, Iwashita N, Naka S, Matsumoto-Nakano M, Nakano K. 2020. Potential involvement of Streptococcus mutans possessing collagen binding protein Cnm in infective endocarditis. Sci Rep 10:1–14.

259. Freires IA, Avilés-Reyes A, Kitten T, Simpson-Haidaris PJ, Swartz M, Knight PA, Rosalen PL, Lemos JA, Abranches J. 2017. Heterologous expression of Streptococcus mutans Cnm in Lactococcus lactis promotes intracellular invasion, adhesion to human cardiac tissues and virulence. Virulence. Taylor and Francis Inc.

260. Avilés-Reyes A, Miller JH, Simpson-Haidaris PJ, Lemos JA, Abranches J. 2014. Cnm is a major virulence factor of invasive Streptococcus mutans and part of a conserved three-gene locus. Mol Oral Microbiol 29:11–23.

261. Szymanski CM, Wren BW. 2005. Protein glycosylation in bacterial mucosal pathogens. Nat Rev Microbiol 2005 33 3:225–237.

262. Bergstrom KSB, Xia L. 2013. Mucin-type O-glycans and their roles in intestinal homeostasis. Glycobiology 23:1026–1037.

263. Ince D, Lucas TM, Malaker SA. 2022. Current strategies for characterization of

mucin-domain glycoproteins. Curr Opin Chem Biol 69:102174.

264. Beerens K, Soetaert W, Desmet T. 2015. UDP-hexose 4-epimerases: a view on structure, mechanism and substrate specificity. Carbohydr Res 414:8–14.

265. Bernatchez S, Szymanski CM, Ishiyama N, Li J, Jarrell HC, Lau PC, Berghuis AM, Young NM, Wakarchuk WW. 2005. A Single Bifunctional UDP-GlcNAc/Glc 4-Epimerase Supports the Synthesis of Three Cell Surface Glycoconjugates in Campylobacter jejuni*. J Biol Chem 280:4792–4802.

266. Ishiyama N, Creuzenet C, Lam JS, Berghuis AM. 2004. Crystal structure of WbpP, a genuine UDP-N-acetylglucosamine 4-epimerase from Pseudomonas aeruginosa: substrate specificity in udp-hexose 4-epimerases. J Biol Chem 279:22635–22642.

267. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589.

268. Merritt J, Tsang P, Zheng L, Shi W, Qi F. 2007. Construction of a counterselection-based in-frame deletion system for genetic studies of Streptococcus mutans. Oral Microbiol Immunol 22:95–102.

269. Neidhardt FC, Curtiss R. 1996. Escherichia coli and Salmonella typhimurium: : cellular and molecular biologyTranslation Mechanisms2nd ed. ASM Press, Washington D.C.

270. Grossiord BP, Luesink EJ, Vaughan EE, Arnaud A, De Vos WM. 2003. Characterization, expression, and mutation of the Lactococcus lactis galPMKTE genes, involved in galactose utilization via the Leloir pathway. J Bacteriol 185:870–878.

271. Besingi RN, Wenderska IB, Senadheera DB, Cvitkovitch DG, Long JR, Wen ZT, Brady LJ. 2017. Functional amyloids in Streptococcus mutans, their use as targets of biofilm inhibition and initial characterization of SMU_63c. Microbiology 163:488.

272. Oli MW, Otoo HN, Crowley PJ, Heim KP, Nascimento MM, Ramsook CB, Lipke PN, Brady LJ. 2012. Functional amyloid formation by Streptococcus mutans. Microbiology 158:2903.

273. di Cologna NM, Samaddar S, Valle CA, Vargas J, Aviles-Reyes A, Morales J, Ganguly T, Pileggi R, Brady LJ, Lemos JA, Abranches J. 2021. Amyloid Aggregation of Streptococcus mutans Cnm Influences Its Collagen-Binding Activity. Appl Environ Microbiol 87:e0114921.

274. Nakano K, Nomura R, Nemoto H, Lapirattanakul J, Taniguchi N, Grönroos L, Alaluusua S, Ooshima T. 2008. Protein antigen in serotype k streptococcus mutans clinical isolates. J Dent Res 87:964–968.

275. Tsumori H, Kuramitsu H. 1997. The role of the Streptococcus mutans glucosyltransferases in the sucrose-dependent attachment to smooth surfaces: essential role of the gtfC enzyme. Oral Microbiol Immunol 12:274–280.

276. Zhu F, Zhang H, Yang T, Haslam SM, Dell A, Wu H. 2016. Engineering and Dissecting the Glycosylation Pathway of a Streptococcal Serine-rich Repeat Adhesin. J Biol Chem 291:27354–27363.

277. Michael F St., Szymanski CM, Li J, Chan KH, Khieu NH, Larocque S, Wakarchuk WW, Brisson J-R, Monteiro MA. 2002. The structures of the lipooligosaccharide and capsule polysaccharide of Campylobacter jejuni genome sequenced strain NCTC 11168. Eur J Biochem 269:5119–5136.

278. Sequeira S, Kavanaugh D, MacKenzie DA, Šuligoj T, Walpole S, Leclaire C, Gunning AP, Latousakis D, Willats WGT, Angulo J, Dong C, Juge N. 2018. Structural basis for the role of serine-rich repeat proteins from Lactobacillus reuteri in gut microbe-host interactions. Proc Natl Acad Sci U S A 115:E2706–E2715.

279. Yamashita Y, Shibata Y, Nakano Y, Tsuda H, Kido N, Ohta M, Koga T. 1999. A novel gene required for rhamnose-glucose polysaccharide synthesis in Streptococcus mutans. J Bacteriol 181:6556–6559.

280. mutans RL and immunology of S, 1986 undefined. Immunochemical aspects of serotype carbohydrate antigens of Streptococcus mutans. cir.nii.ac.jp.

281. Shibata Y, Ozaki K, Seki M, Kawato T, Tanaka H, Nakano Y, Yamashita Y. 2003. Analysis of loci required for determination of serotype antigenicity in Streptococcus mutans and its clinical utilization. J Clin Microbiol 41:4107–4112.

282. Lau PCY, Sung CK, Lee JH, Morrison DA, Cvitkovitch DG. 2002. PCR ligation mutagenesis in transformable streptococci: Application and efficiency. J Microbiol Methods 49:193–205.

283. Shajahan A, Heiss C, Ishihara M, Azadi P. 2017. Glycomic and glycoproteomic analysis of glycoproteins-a tutorial. Anal Bioanal Chem 409:4483–4505.

284. Black IM, Heiss C, Jain M, Muszyński A, Carlson RW, Gabriel DW, Azadi P. 2021. Structure of Lipopolysaccharide from Liberibacter crescens Is Low Molecular Weight and Offers Insight into Candidatus Liberibacter Biology. Int J Mol Sci 22.

285. Santander J, Martin T, Loh A, Pohlenz C, Gatlin DM, Curtiss R. 2013. Mechanisms of intrinsic resistance to antimicrobial peptides of Edwardsiella ictaluri and its influence on fish gut inflammation and virulence. Microbiology 159:1471–1486.

286. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612.

287. Creuzenet C, Belanger M, Wakarchuk WW, Lam JS. 2000. Expression, purification, and biochemical characterization of WbpP, a new UDP-GlcNAc C4 epimerase from Pseudomonas aeruginosa serotype O6. J Biol Chem 275:19060–19067.

288. Chang JC, LaSarre B, Jimenez JC, Aggarwal C, Federle MJ. 2011. Two group a streptococcal peptide pheromones act through opposing rgg regulators to control biofilm development. PLoS Pathog 7.

289. Bian X, Garber JM, Cooper KK, Huynh S, Jones J, Mills MK, Rafala D, Nasrin D, Kotloff KL, Parker CT, Tennant SM, Miller WG, Szymanski CM. 2020. Campylobacter Abundance in Breastfed Infants and Identification of a New Species in the Global Enterics Multicenter Study . mSphere 5.

290. Eddie Ip WK, Takahashi K, Alan Ezekowitz R, Stuart LM. 2009. Mannose-binding lectin and innate immunity. Immunol Rev 230:9–21.

291. Eradi P, Ghosh S, Andreana PR. 2018. Total Synthesis of Zwitterionic Tetrasaccharide Repeating Unit from *Bacteroides fragilis* ATCC 25285/NCTC 9343 Capsular Polysaccharide PS A1 with Alternating Charges on Adjacent Monosaccharides. Org Lett 20:4526–4530.

APPENDIX A

POST-TRANSLATIONAL MODIFICATION BY THE PGF GLYCOSYLAITON MACHINERY MODULATES STREPTOCOCCUS MUTANS PHYSIOLOGY AND VIRULENCE[3]

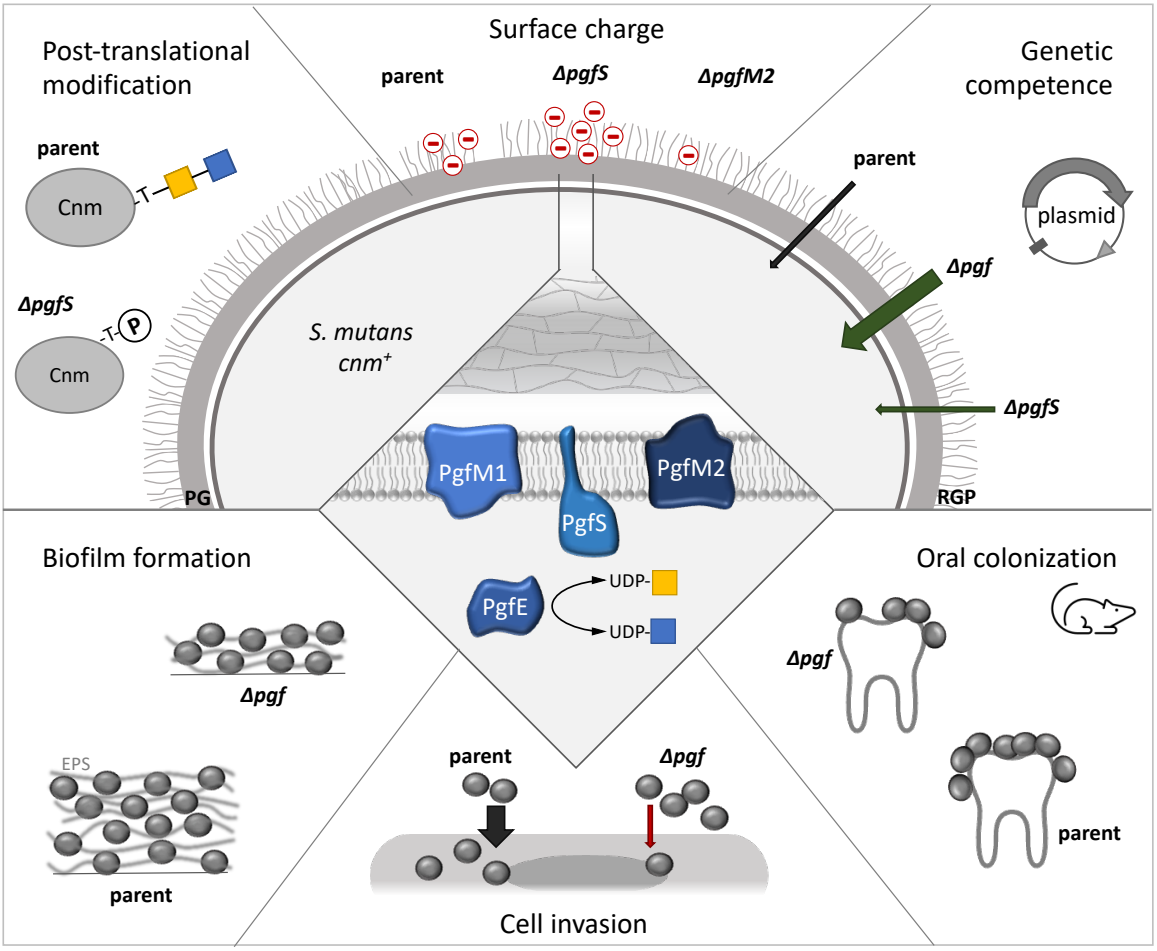
[3] Nicholas de Mojana di Cologna, Silke Andresen, Sandip Samaddar, Stephanie Archer-Hartmann, Tridib Ganguly, Jessica K. Kajfasz, Bruna A. Garcia, Irene Saengpet, Alexandra M. Peterson, Parastoo Azadi, Christine M. Szymanski, José A. Lemos and Jacqueline Abranches, 2022. *Submitted to Molecular Microbiology.*

NMC, JAL and JA conceptualized and designed this study with input from CMS. NMC conducted part of the computational modeling, biofilm formation analysis, survival assays with saliva, MICs for SDS and tunicamycin, the measurement of the zeta potential, confocal image acquisition and analysis and analysis for genetic competence. NMC, TG and JK designed and performed the rat oral colonization experiment. BG designed and assisted the confocal image acquisition and analysis and assisted with all experiments' statistical analysis. AP performed protein structure modeling with AlphaFold. SA executed the survival assays with serum and phagocytes, the measurement of membrane permeability and antibiotic resistance and the design of graphical abstract and summary image. SS constructed the S. mutans parental and pgf mutant strains for tCnm expression. SS, TG and IS expressed and purified Cnm and tCnm from S. mutans. SA-H and PA conducted MS analysis of Cnm and tCnm.

Abstract

Streptococcus mutans is a keystone pathogen of dental caries, and the ability to form biofilms is essential for its pathogenicity. We identified a glycosylation machinery (Pgf) in *S. mutans* that post-translationally modifies two surface-associated adhesins, Cnm and WapA. The four *pgf* genes (*pgfS*, *pgfM1*, *pgfE,* and *pgfM2*) are part of *S. mutans* core genome and we hypothesized that the scope of Pgf goes beyond Cnm and WapA. By inactivating each *pgf* gene individually or creating a quadruple *pgf* mutant in *S. mutans* OMZ175, we showed that the Pgf machinery is important for biofilm formation. Compared to OMZ175, differences in surface charge, membrane stability, and genetic competence were also observed for most mutants. Importantly, *in silico* analyses and tunicamycin MIC assays suggest a functional redundancy between the Pgf machinery and the rhamnose-glucose polysaccharide synthesis pathway. Using a rat oral colonization model, we showed a 10-fold reduction in recovered CFUs for the *pgf* quadruple mutant compared to OMZ175. Finally, using Cnm as a model, we showed by glycoproteomics analyses that Cnm is heavily modified with N-acetyl hexosamine in OMZ175 whereas phosphorylations were observed for the *pgfS* mutant. Our findings indicate that the Pgf machinery participates in important aspects of *S. mutans* pathobiology.

Graphical Abstract



Abbreviated summary

In this study, we demonstrate that the Pgf glycosylation machinery of *Streptococcus mutans*, a keystone pathogen of dental caries, regulates several aspects of bacterial pathophysiology that ultimately contribute to *S. mutans* fitness in oral colonization experiments. Using the heavily glycosylated Cnm adhesin as a model, we found that inactivation of the glycosyltransferase PgfS results in loss of Cnm glycosylation, but instead, Cnm became heavily phosphorylated, suggesting a crosstalk/competition between these two post-translational modification mechanisms.

Introduction

Living organisms require fine regulation of their biological processes by post-translational modification of proteins, which allows for rapid and often reversible modulation of protein stability and protein-substrate interactions (Prabakaran *et al.*, 2012; Beltrao *et al.*, 2013). Phosphorylation, glycosylation, ubiquitination, acetylation, lipidation, and many other protein modifications add useful diversity to living organisms' toolkits (Beltrao *et al.*, 2013). Glycosylation, for example, modulates protein structure, solubility, and stability, and therefore influences a myriad of cellular processes. Glycosylation can also mediate interactions with other cells and with the extracellular matrix (Spiro, 2002; Nothaft and Szymanski, 2010; Lu *et al.*, 2015; Schäffer and Messner, 2017; Eichler, 2019). Not surprisingly, glycosylation is a critical post-translational modification that modulates bacterial traits associated with pathogenesis such as adhesion, motility, glycosyltransferase-based toxicity (toxic glycosylation of host proteins), host-pathogen interaction, and evasion of the immune system (Valguarnera *et al.*, 2016; Bhat *et al.*, 2019; Lin *et al.*, 2020).

Widely present among bacteria, N- and O-glycosylations are the most common forms of protein glycosylation. The N- or -O- classification indicates the atom in the side chain of the protein that is linked to the glycan (nitrogen in asparagine and oxygen in serine and threonine). (Spiro, 2002; Nothaft and Szymanski, 2010; Nothaft and Szymanski, 2013). Both N- and O-glycosylation machineries are found in Gram-positive and Gram-negative bacteria and the addition of sugars can occur *en bloc* with the assistance of an oligosaccharyltransferase or sequentially with the assistance of glycosyltransferases (Tan *et al.*, 2015). Importantly, protein glycosylation exponentially increases the complexity of

the proteome, as individual proteins can be modified by oligosaccharide chains bearing different permutations of monomeric and oligomeric sugars of various sizes, charges, isomeric forms, and branching patterns (Eichler, 2019). Although these mono- and oligosaccharides are of major relevance to protein biochemistry, they are secondary gene products that are not coded in the genome, but rather covalently linked after translation by enzymes, mainly glycosyltransferases (Schäffer and Messner, 2017). This subclass of enzymes (E. C. 2.4) catalyzes the formation of glycosidic bonds between a sugar residue from a donor (most commonly a nucleotide-sugar conjugate) to an acceptor molecule such as other carbohydrates, proteins, or lipids (Schmid *et al.*, 2016). There are currently 115 families of glycosyltransferases listed on CAZY (Carbohydrate-Active enZYme Database, at http://www.cazy.org) covering a wide range of biochemical processes (Lombard *et al.*, 2014).

In *Streptococcus mutans*, a keystone pathogen of dental caries, glycosyltransferases (GTs), especially glycosyltransferase (Gtf) -B, -C, and -D enzymes, play a central role in pathogenicity. These GTs are part of the sucrose-dependent mechanism of colonization of *S. mutans* and are responsible for the production of the extracellular polysaccharides (EPS), a major component of the biofilm extracellular matrix (ECM). Using sucrose as a substrate, GtfB produces insoluble α(1-3)-glucans, while GtfD produces soluble α(1-6)-glucans and GtfC produces a mixture of both (Bowen and Koo, 2011; Simón-Soro and Mira, 2015; Pitts *et al.*, 2017). Furthermore, Gtf enzymes can bind to microbial cell surfaces, converting other organisms into *de facto* glucan-producers, and therefore contributing to the development of thick multi-species biofilms (Krzyściak *et al.*, 2014; Koo *et al.*, 2018; Lemos *et al.*, 2019). For instance, the binding of *S. mutans* GtfB to mannans present in the

cell wall of the fungus *Candida albicans* promotes the development of a highly cariogenic biofilm (Falsetta *et al.*, 2014; Hwang *et al.*, 2017; Koo *et al.*, 2018). Importantly, the enzymatic activity of GtfB, -C, and -D is based on the formation and breakage of glycosidic bonds between carbohydrate moieties, and not on protein glycosylation (Bowen and Koo, 2011; Krzyściak *et al.*, 2014).

In addition to the aforementioned sucrose-dependent colonization mechanisms, *S. mutans* also possesses sucrose-independent mechanisms of colonization that are mainly involved in initial attachment and biofilm development (Avilés-Reyes *et al.*, 2017; Arora *et al.*, 2021). The surface adhesin P1, also known as Antigen I/II or SpaP, mediates *S. mutans* binding to constituents of the salivary pellicle such as the salivary agglutinin glycoprotein 340 (Gp340) and also promotes interactions between bacterial cells (Krzyściak *et al.*, 2014; Sullan *et al.*, 2015). The wall-associated protein A (WapA) has also been demonstrated to be important for cell-cell interaction, binding to the salivary pellicle, and organization of biofilm structure (Zhu *et al.*, 2006). Both P1 and WapA were also shown to mediate collagen binding *in vitro* (Sciotti *et al.*, 1997; Han *et al.*, 2006). Moreover, approximately 20% of *S. mutans* clinical isolates express the collagen- and laminin-binding adhesin Cnm (Sato *et al.*, 2004; Nomura *et al.*, 2009; Nakano *et al.*, 2010; Abranches *et al.*, 2011; Avilés-Reyes *et al.*, 2014b), an important virulence factor associated with dental caries severity and recurrence (Miller *et al.*, 2015; Esberg *et al.*, 2017; Garcia *et al.*, 2021) that also contributes to extra-oral infections such as cerebral microbleeds, hemorrhagic stroke and infective endocarditis (Abranches *et al.*, 2011; Nomura *et al.*, 2013; Miller *et al.*, 2015; Avilés-Reyes *et al.*, 2017; Inenaga *et al.*, 2018; Hosoki *et al.*, 2020; Nomura *et al.*, 2020; Arora *et al.*, 2021). In addition to binding to

collagen, Cnm, WapA and P1 have also been characterized as functional amyloids that help model the biofilm structure (Oli *et al.*, 2012; Besingi *et al.*, 2017; di Cologna *et al.*, 2021; Yarmola *et al.*, 2022). In addition to sharing functional and structural similarities, Cnm and WapA were both shown to be post-translationally modified through glycosylation (Avilés-Reyes *et al.*, 2017; Avilés-Reyes *et al.*, 2018).

Post-translational modifications are increasingly implicated as having relevant functional roles in the pathogenesis of oral bacteria (Ma *et al.*, 2021). In *S. mutans* OMZ175, the *cnm* gene is co-transcribed with the four-gene *pgf* operon (*pgfS*, *pgfM1*, *pgfE,* and *pgfM2*) that is part of the *S. mutans* core genome and is required to glycosylate both core (WapA) and non-core (Cnm) LPXTG-anchored surface proteins (Avilés-Reyes *et al.*, 2018). Like *cnm,* the *pgf* operon is negatively regulated by VicRKS and positively regulated by CovR, two signal transduction systems that control the expression of many virulence-associated genes of *S. mutans* (Stipp *et al.*, 2013; Avilés-Reyes *et al.*, 2018; Alves *et al.*, 2018). In *S. mutans* OMZ175, deletion of the *pgf* genes, either individually or the entire operon, resulted in a reduction of the observed mass of Cnm and WapA in polyacrylamide gels, and both proteins demonstrated reduced stability and resistance to proteolysis. Also, deletion of the *pgf* genes led to a significant decrease in expression of Cnm-associated virulence traits including collagen-binding, invasion of endothelial and epithelial cells, and killing of the invertebrate *Galleria mellonella* model of systemic infection (Avilés-Reyes *et al.*, 2014a; Avilés-Reyes *et al.*, 2018).

To gain further insights into post-translational glycosylation mechanisms employed by the Pgf machinery and the contribution of each Pgf protein to *S. mutans* fitness, we performed a series of phenotypic assays to study important traits associated with *S. mutans*

pathophysiology using a panel of *pgf* mutants created in the *cnm*$^{+}$ strain OMZ175. Additionally, using the glycoprotein Cnm as a Pgf substrate model, we performed glycoproteomic analyses of a series of Cnm constructs expressed in OMZ175 and the *pgf* mutants.

Results and Discussion

The Pgf glycosylation system is responsible for many aspects of bacterial pathobiology

Our previous studies focused on the role of the Pgf glycosylation system on the expression of Cnm-dependent phenotypes such as collagen- and laminin-binding and intracellular invasion (Avilés-Reyes et al., 2014a; Avilés-Reyes et al., 2018). However, considering that Pgf is also present in *cnm*-negative strains and that it modifies the core adhesin WapA, we anticipated that the scope of the Pgf system goes beyond Cnm and WapA modification (Avilés-Reyes *et al.*, 2014a; Avilés-Reyes *et al.*, 2018). Current and past phenotypical characterization of mutants for the Pgf machinery revealed a wide spectrum of changes in phenotype as compared to the parental strain and complemented strains (Avilés-Reyes *et al.*, 2014a; Avilés-Reyes *et al.*, 2018), summarized in Table A1.

Carbohydrates are essential for *S. mutans* pathobiology, as this organism relies mostly on sugars for energy generation, biofilm formation, and building the cell wall and its associated polysaccharides such as the rhamnose-glucose polysaccharide (RGP) layer (Abranches *et al.*, 2018; Lemos *et al.*, 2019). There are several predicted glycosyltransferases in the genome of *S. mutans* including those from the RGP synthesis pathway (RgpA, RgpB, RgpE, RgpF, RgpH, and RgpI) (Shibata *et al.*, 2009; Mistou *et al.*,

2016; Rainey *et al.*, 2018; Kovacs *et al.*, 2019; Bischer *et al.*, 2020). Other uncharacterized putative glycosyltransferases are annotated in the *S. mutans'* genome including Smu.834, located immediately downstream of the *rgp* operon (Lara Vasquez *et al.*, 2021), Smu1039c, Smu1434c, and Smu1806. Therefore, the Pgf proteins may not be the only designated protein glycosylation machinery in *S. mutans*, raising the possibility of functional redundancy between pathways and highlighting the importance of these systems for bacterial pathobiology.

In silico analyses of the Pgf proteins offer hints related to their structure and function

Our first step in understanding the function of the Pgf machinery was by expanding the bioinformatics analysis initiated in our previous study (Avilés-Reyes *et al.*, 2018). In addition to domain prediction using InterPro (Blum *et al.*, 2021) and domain localization of transmembrane proteins using TMHMM 2.0 (Krogh *et al.*, 2001), the current state-of-the-art predictor of protein structure AlphaFold2 (Jumper *et al.*, 2021) was used in HiPerGator 3.0, the University of Florida's supercomputer. AlphaFold2 has been demonstrated as the best currently available tool for the prediction of the 3D structure of proteins for which no homologous structure is known (the "protein folding problem") in the latest Critical Assessment of protein Structure Prediction (CASP, 14th edition), and it operates utilizing a neural network-based machine learning algorithm that interprets evolutionary, biophysical, and steric constraints starting from a protein sequence input (Jumper *et al.*, 2021). Together, these analyses indicate that PgfS, PgfM1, and PgfM2 are likely membrane proteins, while PgfE is an intracellular globular protein (Figure A.1A). Moreover, PgfS contains an intracellular nucleotide-diphospho-sugar-transferase domain,

while PgfM1 and PgfM2 have an extracellular globular domain of no predicted function. PgfE is predicted to contain a Rossmann Fold, a highly conserved motif consisting of up to seven strands of alternating β-sheets and α-helixes which interact via hydrogen bonds (Hanukoglu, 2015). This domain is well known for binding NAD(P), and FAD dinucleotides (Hanukoglu, 2015), which suggests the possibility of these dinucleotides being coenzymes of PgfE. GT-A nucleotide-sugar-dependent glycosyltransferases such as PgfS (Avilés-Reyes *et al.*, 2014a) and the GalT proteins of oral streptococci are known to contain a single Rossmann fold (Zhu *et al.*, 2015), and the presence of a Rossmann Fold motif is suggested by AlphaFold2 modeling of PgfS. Initial analysis by InterPro revealed that the homologous membrane proteins PgfM1 and PgfM2 (88% coverage, with 38.5% identity and 54% positives when aligned) are not predicted to contain a Rossmann fold. However, AlphaFold2 structure predictions of these proteins also display alternating β-sheets and α-helixes in their extracellular globular domains (Figure A.1A), suggesting that the entire Pgf machinery could be relying on NAD(P) and/or FADH as coenzymes. An *in silico* homo-oligomerization analysis using the AlphaFold2 predicted structures of Pgf proteins in the GalaxyHomomer platform (Baek *et al.*, 2017) predicts that PgfS likely forms a tetramer, while PgfE likely forms a dimer (Figure A.1B). The template-based prediction of the PgfS tetramer was based on the GtrB polyisoprenyl-phosphate glycosyltransferase from *Synechocystis* sp recombinantly expressed in *Escherichia coli* (PDB ID 5EKP), with an interface area of 7662.2 $Å^2$ between monomers. GtrB and PgfS share a 34.7% sequence identity, and a 76.2% structural similarity. The crystal structure of the GtrB transmembrane tetramer was previously determined via X-ray crystallography with 3.19 Å of resolution (Ardiccioni *et al.*, 2016). The dimer prediction of PgfE was based on the UDP-4-galactose

epimerase GalE structure from *Homo sapiens* (PDB ID 1I3K), an enzyme of the Leloir pathway for galactose catabolism that also exists in *S. mutans*, with 920.3 $Å^2$ of interface area between monomers. The PgfE and the human GalE proteins share 53.6% of sequence identity and 98.26% of structural similarity. The crystal structure of the human GalE was also determined via X-ray crystallography with 1.50 Å of resolution (Thoden *et al.*, 2001). A GalaxyHomomer analysis of the AlphaFold2 structures from PgfM1 and PgfM2 did not return any expected oligomers.

Glycosylation by the Pgf machinery is required for optimal biofilm formation and possible crosstalk with other glycosylation systems

To gain insight into the importance of the Pgf machinery in *S. mutans'* key virulence traits, we analyzed the biofilm formation ability of our panel of mutant strains (*ΔpgfS, ΔpgfE, ΔpgfM1, or ΔpgfM2* strains and *Δpgf* which designates a strain with all four *pgf* genes deleted). Figure A.10 displays a schematic diagram of the mutant strains utilized in our experiments and of the *cnm-pgf* locus in OMZ175. When biofilms were allowed to grow for 48 hours in a chemically defined biofilm medium (CDM) supplemented with 1% sucrose, all of the mutant strains showed a defect in biofilm biomass when compared to the parent strain, except for *ΔpgfS* (Figure A.2A). Previously, we showed that the inactivation of *cnm* in OMZ175 resulted in decreased biofilm formation in media containing sucrose (Abranches *et al.*, 2011). Thus, the observed phenotype for the mutants could be, at least in part, related to Cnm function and stability, since this adhesin is a known target of the Pgf glycosylation machinery. However, the *ΔpgfS* mutant phenocopied the parent strain in the biofilm biomass assay, and because PgfS is required for full Cnm glycosylation

(Avilés-Reyes *et al.*, 2018), it is likely that the lack of proper protein glycosylation also affects biofilm biomass in a Cnm-independent manner. A similar trend was observed for the recovery of viable cells from these biofilms (Figure A.2B), with comparable CFUs recovered from the parent strain and *ΔpgfS* mutant biofilms and significantly lower cells recovered from the other *pgf* mutant biofilms. Thus, our findings indicate that proper protein glycosylation is important for the ability of *S. mutans* to form robust sucrose-dependent biofilms.

The architecture of *S. mutans* biofilms is largely dependent upon the production of a matrix whose components act as scaffolds that guide the three-dimensional distribution of microenvironments within the biofilm (Klein *et al.*, 2015). To further dissect the effects of the *pgf* deletions in biofilm formation, confocal laser microscopy scanning (CLMS) of all strains was performed using labeling fluorophores for *S. mutans* cells and extracellular polysaccharides. To quantitatively analyze the characteristic parameters of these biofilms, each z-stack was converted into a binary layer and the data were used for an in-depth quantitative analysis of the biofilms (Figure A.2C, D, and Figure A.11). When compared to the parent strain, the overall biofilm architecture was affected in each *pgf* mutant tested (Figure A.3). Although all quantitative metrics for the comparison between the parent strain and the *ΔpgfS* mutant were comparable as indicated above, when the complex 3D structure is visualized, it is noticeable that the parent strain's biofilm is more homogeneously distributed over the area in each layer, while some degree of patchiness and larger microcolonies are observed for *ΔpgfS, ΔpgfE,* and *ΔpgfM1*. The *ΔpgfM2* biofilm, however, was homogeneously distributed, similar to the parent strain. Finally, the *Δpgf* quadruple mutant showed a thinner, more compact, and more homogeneous biofilm when

compared to the parent and other mutant strains. Changes in biofilm thickness (distance between bottom and top of biofilms) and volume of components (cells and EPS) were the most striking phenotypes observed for these mutants. The parent strain biofilm was thicker than all mutants' biofilms except for *ΔpgfS*, and the *Δpgf* mutant biofilm appears to be the thinnest of the biofilms as seen in Figure A.2C. The volume of the biofilms (cells, EPS, and sum) was also estimated (Figure A.2D). All mutants presented an overall lower volume of cells, EPS, and their sum than the parental strain, except for the *ΔpgfS* mutant. Thus, all *pgf* genes are important for robust biofilm development whereas *pgfS* appears to contribute to a lesser extent. Despite the differences in architecture, the volumetric proportion of each analyzed component remained at approximately a 3:2 ratio (*S. mutans*:EPS) for all analyzed biofilms (Figure A.11A). Lastly, biofilm compactness was estimated, indicating high compactness of the *Δpgf* mutant (Figure A.11B). The observation that the *ΔpgfS* mutant phenocopies the parental strain is intriguing, considering that it is the only member of this glycosylation machinery with a traditional GT domain. This observation raises the possibility of functional redundancy of PgfS with other *S. mutans* glycosyltransferases. A protein BLAST search against an *S. mutans* database using the PgfS amino acid sequence revealed only two significant hits: glucosyltransferase RgpI (95% coverage with 41% identities and 61% positives) and rhamnosyltransferase RgpB (36% coverage with 28% identities and 47% positives), both enzymes from the RGP synthesis machinery (Kovacs *et al.*, 2019). The RGP-derived modifications of the *S. mutans* cell wall are important in guaranteeing proper septation and location of cell division complexes, and also in preventing autolysis (Kovacs *et al.*, 2019; Zamakhaeva *et al.*, 2021). RgpI has been demonstrated as having a role in modulating the matrix of biofilms in *S. mutans*, mainly

by decreasing glucan components and increasing the abundance of extracellular DNA (eDNA), without altering the overall biomass (Rainey *et al.*, 2018). Thus, a similar effect may have occurred for the *ΔpgfS* mutant, where a change in biofilm architecture, but not biofilm biomass is observed (Figures A.2 and A.3). The AlphaFold2 predicted structure for RgpI is also very similar to that of PgfS (Figure A.14). Therefore, we cannot rule out that PgfS and RgpI may have overlapping roles in promoting appropriate biofilm development. Taken altogether, our findings show an important role of the Pgf glycosylation machinery in biofilm accumulation and architecture, and a possible functional redundancy between PgfS and RgpI as further discussed below.

Adequate glycosylation by the Pgf (and possibly RGP) system interferes with *S. mutans* survival in human saliva, surface charge, and membrane stability.

In order to establish a biofilm in the oral cavity, microorganisms must be able to utilize salivary components as nutrients as well as tolerate salivary secretory immunity, comprised of soluble immunoglobulins (sIgA and, to a lesser extent, sIgM), proteases, and salivary antimicrobial proteins and peptides (Amorim *et al.*, 2011; Fábián *et al.*, 2012; Brandtzaeg, 2013; Lynge Pedersen and Belstrøm, 2019). When we measured the survival in saliva (supplemented with glucose to prevent starvation-related cell death) of our panel of strains and compared it to that of the parent strain, we observed that the absence of Pgf proteins, except for PgfM2, significantly diminished survival in saliva (Figure A.4). Despite the high homology between PgfM1 and PgfM2, the *ΔpgfM1* strain was almost entirely cleared after 20 hours of incubation (≈0.1% survival, our limit of detection), whereas the *ΔpgfM2* mutant only crossed the 0.1% survival threshold sometime between

28-48 hours of incubation. Interestingly, the *ΔpgfE* mutant was the most susceptible and was below the detection limit after 20 hours of incubation, indicating an important role of the regulation of the availability of GlcNAc and GalNAc for the bacterial ability to survive in saliva. Also, the quadruple Δ*pgf* mutant showed a significant reduction in survival after 28 hours of incubation in saliva. Post-translational modification of bacterial surface proteins by phosphorylation, lipidation, glycosylation, and others is used by bacteria to evade the immune system. For example, bacteria can be cloaked in hydrophobic, hydrophilic, or charged molecules that repel antimicrobial peptides and proteins, while molecular mimicry of human protein-associated glycans avoids recognition by antibodies (Reddick and Alto, 2014; Brockhausen, 2014; Macek *et al.*, 2019). Since at least two of the known targets of the Pgf machinery are surface-anchored proteins, it is not surprising that the mutants behave differently than the parent strain when in the presence of saliva. Deletion of PgfM2 still results in partially glycosylated Cnm and WapA, while the deletion of any other Pgf protein resulted in fully unglycosylated Cnm and WapA (Avilés-Reyes *et al.*, 2018), and this may be the reason why the survival in saliva of the *ΔpgfM2* strain was not compromised as it was for the other mutants. UDP-4-galactose epimerases have been demonstrated as crucial agents in cell surface glycocomponent generation in both Prokaryotes and Eukaryotes (Fry *et al.*, 2000; Roper *et al.*, 2005; Lee *et al.*, 2014; Schäper *et al.*, 2019; Broussard *et al.*, 2020). Thus, exposed cell envelope-linked glycans are likely important for the recognition of this bacteria by salivary antimicrobial components.

We next directed our efforts to the characterization of cell envelope homeostasis in the *pgf* mutants. First, we estimated the cell surface charge via zeta potential measurement (Figure A.5A), which serves as an indirect measurement of the electrical potential between

the bacterial surface and the surrounding water pellicle (Wilson *et al.*, 2001). Of note, proper surface charge is pivotal to the maintenance of a surface hydrophobicity that allows for transmembrane transport of nutrients and waste in an aqueous environment (Wilson *et al.*, 2001). Gram-positive lactic acid bacteria present negatively charged surfaces under physiological pH due to the presence of lipoteichoic acids (LTAs) and wall teichoic acids (WTAs) imbued within the peptidoglycan layer (Wilson *et al.*, 2001; Chapot-Chartier and Kulakauskas, 2014). In *S. mutans*, there are no WTAs and the RGP layer is considered to fulfill their roles, although it does not contain the negative charge associated with WTA (Kovacs *et al.*, 2019; Bischer *et al.*, 2020). When compared to the parental strain, the deletion of *pgfS* significantly increased the negative charge by ≈44%, while the deletion of *pgfM2* significantly decreased the negative charge by ≈26%. On the other hand, the deletion of *pgfE* and *pgfM1* did not significantly change the bacterial zeta potential. Of note, the deletion of the entire *pgf* operon (*Δpgf*) did not alter the cell surface charge. Therefore, the simultaneous deletion of *pgfS* and *pgfM2* in *Δpgf* appears to erase the opposing effects of each individual mutation. We hypothesized that this surface charge variation was due to a modification of the glycosylation pattern of surface proteins or, perhaps, other constituents of the cell envelope like lipids and the RGP. In addition to PgfS being nearly identical to the glycosyltransferase RgpI of the RGP synthesis pathway (Figure A.14), the PgfS and PgfM1/M2 proteins are also highly homologous at both structural and sequence levels to LTA and WTA glycosylation enzymes characterized in other Gram-positive bacteria, such as *Listeria monocytogenes* (GtlA/Lmo2550 as PgfS-like and GtlB/Lmo1079 as PgfM1/M2-like proteins, respectively) and *Bacillus subtilis* (CsbB and YfhO as PgfS-like and PgfM1/M2-like proteins, respectively) (Rismondo *et al.*,

2018; Rismondo *et al.*, 2020; Rismondo *et al.*, 2021). Figure A.15 depicts and compares the predicted structures for PgfS and GtlA, Lmo2550 and CsbB, and Figure A.16 depicts the same comparison for PgfM1, PgfM2, Lmo1079, GtlB, and YfhO. Also, as previously shown in Figure A.1B, the PgfS predicted tetramer is structurally similar to the GtrB tetramer from *Synechocystis* sp., a well-characterized and conserved polyisoprenyl-phosphate glycosyltransferase (PGT) (Ardiccioni *et al.*, 2016). PGTs are integral membrane proteins that catalyze the transfer of a sugar moiety from a nucleotide-sugar to undecaprenyl-phosphate (in Bacteria) or dolichol-phosphate (Eukarya and Archaea), which act as transmembrane carriers of sugar to allow for the formation of membrane glycoconjugates (Ardiccioni *et al.*, 2016; Allen and Imperiali, 2019). The surface charge of teichoic acids can be modulated by replacing free hydroxyl groups of their alditol-phosphate with carbohydrates or D-alanine residues (Chapot-Chartier and Kulakauskas, 2014). The presence or absence of neutral or charged carbohydrates in the membrane, as well as the relative abundance of exposed peptidoglycan/LTA (negatively charged) when compared to the glycosylated peptidoglycan/LTA, may alter the cell's zeta potential and overall membrane stability. Further research is required to determine whether the Pgf machinery can also glycosylate membrane lipids since the speculations made here are solely based on *in silico* analysis. Importantly, the characterized LTA/WTA glycosylation machineries do not possess a designated sugar epimerase like PgfE, but, surprisingly, PgfE is 38.85% identical (56% positives, 99% coverage) to the glucose-galactose epimerase GalE of the Leloir pathway in *S. mutans*. Our group recently characterized and compared the functions of PgfE and GalE and found that despite their structural similarities, PgfE mostly recognizes N-acetylated sugars as substrates (GlcNAc and GalNAc), while GalE

only recognizes non-acetylated sugars (Glc and Gal) (Andresen *et al.*, 2022a). When we tested our panel of strains for resistance to detergents, significant differences in phenotypes were seen only for the negatively charged detergent sodium dodecyl sulfate (SDS). Interestingly, the Δ*pgfS*, Δ*pgfE,* and the quadruple Δ*pgf* mutant presented a significantly increased sensitivity to SDS as measured by minimum inhibitory concentration (MIC) determination (Figure A.5B, top and Figure A.12) whereas the *ΔpgfM1* and *ΔpgfM2* mutants phenocopied the parent strain. We speculate that the fact that the *ΔpgfM2* mutant (and possibly *ΔpgfM1* as well) surface is less negatively charged than the parental strain's, this difference in surface charge allows for a partial neutralization of the disturbance that SDS inflicts upon the cellular membrane. No differences in MIC were observed for non-ionic detergents Tween-20 and Triton X-100 (data not shown). To test the hypothesis that the *pgf* operon is also acting on RGP glycosylation, we utilized the RGP/peptidoglycan synthesis inhibitor tunicamycin (Zhu *et al.*, 2018; Kovacs *et al.*, 2019) for MIC determination (Figure A.5B, bottom and Figure A.13). The high sensitivity of most *pgf* mutants, especially the quadruple mutant *Δpgf*, is suggestive of the participation of the Pgf machinery in RGP synthesis. Thus, our findings may indicate that the scope of the Pgf machinery might go beyond glycosylation of surface adhesins, also contributing to cell wall synthesis and membrane homeostasis. Future studies including an *ΔrgpI* mutant and a double *ΔpgfSΔrgpI* mutant could help us understand how redundant those two enzymes can be, and perhaps explain why the *ΔpgfS* biofilms were similar to the parental strain biofilms. Other phenotypic traits of interest including antibiotic disc inhibition, ethidium bromide permeability, survival in serum, and opsonophagocytosis susceptibility were analyzed, but all mutants phenocopied the parental strain (Figure A.17).

Transformation ability via the ComABCDE and ComRS pathways is suppressed by the Pgf machinery.

The phenotypes observed for membrane homeostasis prompted us to also analyze genetic competence, a process dependent on the ability of surface-associated proteins to bind and take up DNA (Dubnau and Blokesch, 2019). The ability to share genetic material through competence is a keystone event in streptococcal pathogenesis (Sitkiewicz, 2018; Kaspar and Walker, 2019). *S. mutans* is a naturally competent bacterium, with the competent state triggered by activation of the alternative sigma factor ComX via the interconnected ComABCDE and ComRS pathways (Son *et al.*, 2015). These pathways are activated by the competence-stimulating peptide (CSP, encoded by *comC*) and by the *comX*-inducing peptide (XIP, encoded by *comS*) in response to stresses and environmental conditions (Son *et al.*, 2015; Underhill *et al.*, 2018; Underhill *et al.*, 2019). The regulation of genetic competence in *S. mutans* is complex and there are still gaps in knowledge. Transformation assays were performed using CSP or XIP to assess whether the Pgf glycosylation machinery can influence these competence pathways in *S. mutans* (Figure A.6). Our findings revealed that when compared to the parent strain, a 3-log (CSP) and 2-log (XIP) increase in transformability for all mutants, except for *ΔpgfS*, was observed, suggesting that the optimal glycosylation state conferred by the Pgf machinery dampens competence. Genome-wide screening of a transposon library in *S. mutans* identified several genes that, when deleted, induce an aberrant expression of *comX* (Shields *et al.*, 2018). In addition to the known competence-related genes, other genes involved in cell division and cell envelope homeostasis were identified (resistance to stresses and biogenesis),

suggesting that they are likely important for the activation of early competence (Shields *et al.*, 2018). Interestingly, *rgpI* was among the genes identified in the transposon library with increased competence, which due to its high homology with *pgfS,* raises the possibility of the latter also influencing DNA uptake. However, our data revealed only a modest, non-significant increase in competence for the *ΔpgfS* mutant, once again suggesting a possible functional redundancy between PgfS and RgpI in the expression of some bacterial traits. Moreover, RGP glycosylation has been recently shown to be a crucial step for the recruitment of the cell-division machinery (Zamakhaeva *et al.*, 2021) and cell growth (Bischer *et al.*, 2020), strengthening the link between protein glycosylation, cell division, and genetic competence. An alternative explanation for the small increase in transformability observed for the *ΔpgfS* mutant when compared to other mutants is that the more negative surface charge of this mutant (Figure A.5A) could be repelling the negatively charged DNA, inhibiting its intake.

Deletion of the *pgf* operon significantly diminishes *S. mutans* fitness in the rat model of oral colonization

The role of the Pgf machinery in *S. mutans* OMZ175 fitness *in vivo* was analyzed using an established rat oral colonization model by infecting the animals with the parent strain or its isogenic Δ*pgf* mutant (quadruple mutant). Daily infection inocula for both strains were comparable (approximately 7.50 x 10^8 CFU/mL), and rats from both groups had comparable weight gain throughout the entire experiment (data not shown). Cell homogenates from dissected jaws were plated on blood-agar to estimate the total oral microbiota, and on BHI supplemented with appropriate antibiotics to estimate *S. mutans*

numbers. The CFU counts for blood-agar were comparable between strains, indicating that the total culturable microbiota was not affected by the infection with either type of bacteria (Figure A.7A). Interestingly, a 10-fold decrease in counts recovered from the rats' jaws infected with the *Δpgf* mutant was observed in comparison to the parent strain. When the *S. mutans* counts were represented as a percentage of the total microbiota, a significant decrease from 41.1 ± 8.7% (parental) to 16.8 ± 13.9% (*Δpgf*) in *S. mutans* percentage of total oral microbiota was observed in the *Δpgf* mutant (Figure A.7B). Thus, our findings suggest that protein glycosylation by the Pgf machinery is important for oral colonization and, therefore, the overall fitness of *S. mutans* in its natural habitat. Taken altogether, the oral colonization defect of the *Δpgf* mutant is likely related to its severe biofilm defect, reduced ability to survive in saliva, and altered cell envelope homeostasis.

Glycoproteomics analysis of the Cnm adhesin reveals $HexNAc_2$ glycosylation in the parent strain and suggests competition between the Pgf machinery with the serine/threonine kinase PknB for common amino acid residues.

We have recently reported that Cnm is glycosylated by the Pgf glycosylation machinery with predominantly $HexNAc_2$ residues in its threonine-rich repeat region (TRRR) (Avilés-Reyes *et al.*, 2018; Andresen *et al.*, 2022a). Surprisingly, we discovered that the TRRR region of Cnm is phosphorylated in the absence of Pgf-mediated glycosylation (Figure A.8). Figure A.8A summarizes the findings from our mass spectrometry (MS) analysis. In an attempt to overcome the limitations of obtaining peptide cleavage from the highly glycosylated Cnm, we constructed a mutant harboring a truncated variant of the *cnm* gene, coding for a Cnm protein containing only the first three threonine-

rich repeats (tCnm) (Andresen *et al.*, 2022a). This truncated gene was inserted in the parental OMZ175 and also in the mutants *ΔpgfS* (no Cnm glycosylation) and Δ*pgfM2* (partial Cnm glycosylation). Thus, we found that the engineered tCnm obtained from the parent strain was modified with HexNAc2 in the TRRR (Figure A.8B). Examination of purified native full-length Cnm from the *ΔpgfS* mutant strain showed extensive signs of phosphorylation by using the Byonic software analysis (Figure A.18 and Figure A.19). MS/MS fragmentation of one of those peptides confirmed extensive phosphorylation of the TRRR (Figure A.8C). Similar Byonic analysis of full-length Cnm from the parent strain showed no evidence of phosphorylation (Figure A.19), but the TRRR could not be appropriately detected due to extensive glycosylation. In contrast, tCnm from the *ΔpgfM2* mutant showed less glycosylation compared to the parent strain, so we tested whether Cnm from this mutant could be undergoing glycosylation and phosphorylation at the same time. Analysis of tCnm from the *ΔpgfM2* mutant (Figure A.8D) demonstrated that peptides exist with one less $HexNAc_2$ modification than parent strain peptides, and some do indeed possess an additional phosphate modification suggesting that both post-translational modifications can be present in this Cnm variant. We attempted to purify tCnm from the *ΔpgfS* strain for an MS analysis, but it appears to be too unstable as it quickly degrades during purification and sample processing. Future studies will examine the glycosylation and phosphorylation profiles of Cnm (and tCnm) in the other two *pgf* mutants, *ΔpgfE* and *ΔpgM1*. We acknowledge that these analyses remain challenging due to the heterogeneity observed in post-translational modifications and the instability of the truncated variants. In mammalian cells, O-linked glycosylations and O-linked phosphorylations were shown to crosstalk and compete for the same amino acid residues (threonines and serines), and

changes in the balance between these two post-translational modifications may be implicated in an array of human pathologies like cancer and Alzheimer's disease (Wang *et al.*, 2012; Takahashi *et al.*, 2022; Yi *et al.*, 2022; Li *et al.*, 2022). In bacteria, the crosstalk between protein phosphorylation and glycosylation remains largely underexplored. Interestingly, *S. mutans* harbors only one serine/threonine kinase, namely PknB (Banu *et al.*, 2010), and future studies will include mutants for this gene in parental and *pgf* mutant backgrounds to further characterize the post-translational modification profile of the model protein Cnm. A better understanding of the Pgf substrates and specificities will be crucial for a deeper understanding of the phenotypes observed in our panel of mutant strains. For instance, the presence of heavily phosphorylated Cnm in the cell surface of *ΔpgfS* may contribute to its significantly more negative surface charge when compared to the parent strain (Figure A.5A).

Conclusions

The Pgf machinery regulates, through post-translational protein glycosylation and possibly by modifying other targets like the RGP, a wide range of complex biological processes in *S. mutans* that are essential for its pathobiology (Table A2, with a visual representation in Figure A.9) such as proper biofilm development, survival in saliva, cell envelope homeostasis, competence and, most importantly, oral colonization *in vivo*. Our findings suggest functional redundancies between the Pgf machinery, especially PgfS, and the RGP synthesis pathways. We also found that in the absence of proper glycosylation through the Pgf machinery, Cnm becomes phosphorylated in the TRRR, suggesting that

Pgf machinery and the only serine-threonine kinase present in *S. mutans* compete for the same sites in Cnm.

Material and Methods

Bacterial strains and growth conditions

The strains used in this study are listed in Table A3. All *pgf* mutants have been previously complemented (Avilés-Reyes *et al.*, 2014a; Avilés-Reyes *et al.*, 2018). *S. mutans* was routinely grown in Brain Heart Infusion (BHI) media and Chemically Defined Media (CDM) (van de Rijn and Kessler, 1980) with 1% (m/v) glucose at 37°C and 5% CO_2. When required, Kanamycin or Erythromycin was added to the media at 1 $mg.mL^{-1}$ or 0.3 $mg.mL^{-1}$, respectively. Biofilms were grown in CDM with 1% (m/v) sucrose as a carbon source.

Bacterial mutagenesis

An OMZ175 mutant harboring a truncated *cnm* gene that generates a protein containing only the first three threonine-rich repeats in its B domain was generated as described elsewhere (Andresen *et al.*, 2022a). This mutant strain (*tcnm*) had either the *pgfS* or the *pgfM2* genes deleted by the same strategy used to create the *ΔpgfS* and *ΔpgfM2* mutants (Avilés-Reyes *et al.*, 2018), generating the double mutants *tcnmΔpgfS* and *tcnmΔpgfM2*.

Biofilm biomass assays

Quantification of biofilm biomass and colony-forming unit counts (CFUs) were performed as previously described with some modifications (Lemos *et al.*, 2010). Briefly, overnight cultures of *S. mutans* strains grown in BHI were diluted 1:100 in CDM containing 1% sucrose as the sole carbon source in 96-well plates. Four biological replicates were grown in technical triplicates on duplicate plates for 48 hours. Media was carefully removed, and wells were washed twice with sterile PBS to remove non-adherent cells. Biofilms were scraped from one of the plates, resuspended in 20 µL of PBS, and plated on BHI-agar plates for determination of CFUs. Biofilms of the second plate were stained for 15 minutes with a 0.05% crystal violet solution. Wells were washed twice with PBS and stained biofilms were resuspended in 7% acetic acid (v/v) and absorbance was read at 595 nm in a Synergy H1 hybrid multimode reader (BioTek).

Confocal Laser Scanning Microscopy

Overnight cultures were diluted 1:100 in CDM + 1% sucrose in µ-Slide 8 Well polymer coverslip chambers (Ibidi) and grown for 48 hours. Spent media and planktonic cells were carefully removed by pipetting and biofilms were gently washed twice with a sterile 0.85% (v/v) NaCl solution. Then, a 0.85% NaCl solution containing 5 µM Syto9 (ThermoScientific) to identify live *S. mutans* cells and 1 µM Dextran Alexa Fluor 647 (ThermoScientific) to identify extracellular polysaccharides was added to stain biofilms for 30 minutes. Biofilms were again washed to remove unbound dyes, and images were acquired using a Nikon Ti2 Confocal Microscope with a Plan Apo λ 60x oil objective coupled with a Nikon C2 Plus camera. NIS-Elements software (Nikon) was used to

analyze three-dimensional images with a 1 μm height difference between z-stacks for three spots of each biofilm using at least 4 biological replicates. Thresholds were determined for the creation of binary layers for *S. mutans* cells and extracellular polysaccharides, allowing for quantitative analysis of biofilms (Paula *et al.*, 2020).

Cnm purification and digestion

Cnm from OMZ175 and the *ΔpgfS* mutant, and tCnm from OMZ175 parental background and the *tcnmΔpgfS* and *tcnmΔpgfM2* mutant backgrounds were purified by affinity chromatography using custom columns coupled with anti-Cnm collagen-binding domain (CBD) antibodies, as previously published (Avilés-Reyes *et al.*, 2014a; Avilés-Reyes *et al.*, 2014b). Briefly, the OMZ175*ΔvicK* strain, known to overexpresses Cnm (Alves *et al.*, 2018), and the *ΔpgfS* mutant were grown overnight in liquid cultures and then lysed by bead beating in the presence of Halt protease inhibitor single-use cocktail (Thermo Fischer Scientific) and ethylenediaminetetraacetic acid (EDTA). The soluble protein fractions were then bound to custom N-hydroxysuccinimide (NHS)-anti-CBD columns with overnight rocking at 4°C. Different columns were used for proteins from different strains to avoid cross-contamination. Then, columns were washed with 15 mL of 1x PBS at pH 7.2, and protein was eluted by incubation in a 0.1 M glycine buffer (pH 2.5) for 5 min. Elutions were immediately neutralized with 1/10 volume of a basic 1 M Tris buffer (pH 8.0). Purified native proteins were later concentrated and dialyzed against a 50 mM ammonium bicarbonate solution. Proteins were then digested in solution as described elsewhere (Shajahan *et al.*, 2017). Although there are no cysteines in Cnm to form disulfide bonds, dithiothreitol (DTT) was used at a low concentration (5mM) to help disrupt the

protein tertiary structure. To this, sequencing grade trypsin (Promega) was added at a protease:protein mass ratio of 1:20 and digested overnight at 37˚C. The enzymatic digestion was halted the next day by briefly heating at 100˚C, and the digested peptides were cleaned-up and dried by C18 solid phase extraction cartridges (Hawach Scientific) and dried on a vacuum centrifuge.

Mass spectrometry analysis

Tandem liquid chromatography mass spectrometry (LC-MS/MS) was performed on an Orbitrap Eclipse Tribrid mass spectrometer equipped with a nanospray ion source coupled with a Thermo Ultimate RSLCnano chromatography system (ThermoFisher). Digested peptides were reconstituted in 0.1% (v/v) formic acid (FA) and pre-filtered through a 0.2 µm spin filter before injection into the system. A commercial nano-LC column (ThermoFisher) of 15 cm length with 75 µm internal diameter and filled with 3 µm C18 material (reverse phase) was used for the chromatographic separation of samples. The precursor ion scans were acquired at 120,000 resolution in the Orbitrap analyzer, and precursors at a time frame of 3 seconds were selected for subsequent MS/MS higher-energy C-trap dissociation (HCD) fragmentation in the Orbitrap analyzer at 15,000 resolution. Runs of each digest were conducted for 70 or 180 min for phosphorylation and glycosylation analysis, respectively. Solutions of 0.1% FA and 80% acetonitrile–0.1% FA were used as mobile phases to separate the glycopeptides. The threshold for triggering an MS/MS event was set to 1,000 counts, and monoisotopic precursor selection was enabled. Data were processed with Byonic (Protein Metrics v4.0.12) software using the full-length sequence of Cnm as a reference. Precursor mass tolerance was set to 5 ppm and fragment

mass tolerance was set to 20 ppm. Modifications included oxidation and deamidation on methionine and asparagine respectively, and phosphorylation on serine, threonine, and tyrosine. Detection of multiple phosphates (e.g., 2X phosphorylation = 159.9327) was also programmed for analysis.

Zeta potential measurement

Estimation of *S. mutans* envelope charge via zeta potential in liquid culture was performed as previously described with some modifications (Wilson *et al.*, 2001; Soni *et al.*, 2008; Ng and Ting, 2016). Cultures were grown in BHI to an OD_{600} of 0.15, then cells were harvested by centrifugation (10,000 x g for 2 minutes) and washed three times with 0.1X PBS to remove loosely bound ions at the bacterial surface. Pellets were resuspended in 0.1X PBS to prevent osmotic stress and cell clumping. Then, 800 µL of each cell suspension was added to a 12 mm glass cuvette (Malvern Panalytical) which was closed with a Dip Cell with palladium electrodes with 2 mm spacing (Malvern Panalytical), inserted into a Malvern Zetasizer Ultra particle analysis system (Malvern Panalytical), and equilibrated to 25°C for 60 seconds. Readings were performed using 30 individual measurements with self-refinements, resulting in sharp and stable peaks for each sample.

Minimum Inhibitory Concentration determination

To determine the MIC of SDS and tunicamycin, overnight cultures were used to respectively inoculate BHI supplemented with up to 0-0.01% (m/v) Sodium Dodecyl Sulphide or up to 20 µg/mL of tunicamycin from *Streptomyces lysosuperficus* (Millipore Corp) solubilized in Dimethyl Sulfoxide, in quadruplicate. After 24h of incubation at 37°C

in 5% CO_2, the OD_{600} was read using a Synergy H1 hybrid multimode reader (BioTek). MIC values were considered as the concentration in which >90% of inhibition occurred.

Ethidium Bromide permeability assays

The assay was performed as previously described, with minor modifications (Patry *et al.*, 2019). Briefly, *S. mutans* strains were grown to mid-logarithmic phase (OD_{600} of ≈0.5) in BHI, harvested by centrifugation, and resuspended in PBS to an OD_{600} of 0.2. The bacterial suspensions were mixed with 20 µM ethidium bromide (EtBr) solution to a final concentration of 1 µM. Immediately after the addition of the EtBr, spectrophotometric monitoring at 530 nm extinction and 600 nm emission was started using a Synergy H1 hybrid multimode reader (BioTek) with measurements at 20-second intervals. At least four independent biological replicates were performed for each strain, with two technical replicates.

Antibiotic resistance disc assays

Antibiotic discs were prepared using sterile Whatman paper discs (7 mm in diameter) saturated with 20 µL of antibiotic stock solutions (Penicillin 1.5 $mg.mL^{-1}$, Vancomycin 1.5 $mg.mL^{-1}$, Gentamycin 1.5 $mg.mL^{-1}$, Streptomycin 1.5 $mg.mL^{-1}$, Chloramphenicol 1.5 $mg.mL^{-1}$, Tetracycline 1.5 $mg.mL^{-1}$, Ampicillin 0.5 $mg.mL^{-1}$). The discs were placed on BHI agar plates which had been freshly coated with 150 µL of mid-logarithmic phase *S. mutans* cultures. Diameters of zones of growth inhibition around each antibiotic disc were measured after 24 hours of incubation. Assays were performed with three biological replicates per strain.

Opsonophagocytosis assays

The method for the opsonophagocytosis assay was adapted from a previously published method (Andresen *et al.*, 2022b). Blood was drawn from healthy, adult, human volunteers at the Clinical Translational Research Unit of the University of Georgia by venipuncture under informed consent. Heparin was used as an anticoagulant. Briefly, 5 mL of whole human blood was layered onto 5 mL of Histopaque-1077 Hybri-MaxTM (Sigma Life Science, H8889-100 mL) in a column, followed by centrifugation at 350 x g for 30 min to induce the formation of a density gradient. The peripheral blood mononuclear cells (PBMCs) were collected from the corresponding layer, washed in PBS, and resuspended in Dulbecco's Modified Eagle Medium (DMEM). PBMCs were stained with Methylene Blue and counted by visualization of intact cells in a hemocytometer. Briefly, *S. mutans* cultures were grown to mid-exponential phase (OD_{600} of ≈0.5) and adjusted to an OD_{600} of 0.01 in DMEM ($8x10^6$ CFU/mL). Then, $1.25x10^5$ CFU of bacteria were combined with $5x10^5$ PBMCs for a final multiplicity of infection (MOI) of about 1:4 bacteria to PBMCs. Pooled normal human complement serum (NHS; a source of complement) (Innovative Research) or heat-inactivated NHS (HIS) were added to the bacterial:PBMC mixture in DMEM to a final volume of 200 µL. As a control condition, the complement components were heat inactivated by heating to 56 °C for 1 h. The components were incubated for 2 hours, then serially diluted and plated on BHI for CFU quantifications of viable bacterial cells. The percentage of survival was calculated relative to controls free of PBMCs (= 100% survival).

Serum survival assay

Pooled human complement serum was purchased from Innovative Research. The serum survival assay was done as previously described (Lees-Miller *et al.*, 2013) with some modifications. *S. mutans* strains grown to mid-exponential phase were harvested by centrifugation and adjusted to an OD_{600} of 2. In a 96-well plate, 5 µL of bacterial suspension were combined with 20 µl (20% v/v final concentration) of baby rabbit complement (BRC) and with heat-inactivated (1 h at 56 °C) normal human serum (Innovative Research) in a final volume of 100 µL. The mixtures were incubated for 24 or 48 hours before plating serial dilutions onto BHI plates. The survival of *S. mutans* in the presence of BRC was calculated as a percentage of survival in heat-inactivated BRC (1 h at 56 °C). Four biological replicates of the experiment were performed for each strain. Controls for the assay were performed using the *Acinetobacter baumannii* 5075 WT (robust survival) and *A. baumannii* 5075*ΔpglC* (complete killing) (Crippen *et al.*, 2021).

Saliva survival assay

Saliva was pooled from healthy donors and clarified by filtration (0.2 µm cutoff filter). Stationary phase *S. mutans* cultures were harvested by centrifugation and resuspended in PBS to one-tenth of the original volume. Twenty microliters of each concentrated culture were added to 200 µL of pooled human saliva supplemented with 20 µM of glucose and incubated for 48 hours. Samples were removed at various time points to allow for the counting of viable CFUs.

Genetic competence

Biological replicates (three for each strain) of *S. mutans* were subcultured into BHI or CDM + 1% glucose and incubated until the cultures reached an OD_{600} of approximately 0.08. The BHI-grown cultures were then supplemented with 1 μM Competence Stimulating Peptide (CSP; SGSLSTFFRLFNRSFTQALGK) and the CDM-growth cultures were supplemented with *comX*-Inducing Peptide (XIP; GLDWWSL) for the stimulation of early competence via the *comABCDE* and *comRS* pathways, respectively. Five hundred nanograms of purified spectinomycin-resistance conferring plasmid pDL278 (Addgene, #46882) were also added to each culture. Incubation was continued for four hours, followed by tenfold serial dilutions plated on BHI-agar plates with or without spectinomycin (1 mg/mL). After 48 hours, CFUs were enumerated, and transformation efficiency was calculated by dividing the CFUs counted from the plates with antibiotic by the CFUs counted from the plates without antibiotic.

Rat oral colonization assay

An established rat oral colonization model (Miller *et al.*, 2015; Galvão *et al.*, 2017) was performed with minor modifications using the parental OMZ175 strain and a *Δpgf* mutant. An *S. mutans* OMZ175 strain harboring an Erythromycin resistance cassette integrated at the *gtfA* locus was used as the parent strain in this study (Miller *et al.*, 2015). This gene has been demonstrated not to interfere with *S. mutans* virulence in rat oral colonization models (Barletta *et al.*, 1988), and therefore this mutant is herein referred to as the parental strain. Upon arrival, Sprague-Dawley female rats were provided with sulfamethoxazole (0.8 mg/mL) and trimethoprim (0.16 mg/mL) in their drinking water to

suppress their endogenous microbiota, facilitating infection. Antibiotics were removed three days before the animals were infected for four consecutive days by inoculating the oral cavities with cotton swabs saturated with overnight cultures of *S. mutans*. At the start of the infection period, animals were given a balanced rodent diet containing 12% sucrose (Envigo), provided for the duration of the experiment. Fourteen days after the initial oral infection, animals were sacrificed, and the lower jaws were removed for sonication in sterile PBS to remove biofilms from the tooth surfaces and to homogenize the cells. The sonicate was then tenfold serially diluted and plated on blood-agar (total bacteria), and BHI containing either Kanamycin (*Δpgf* only) or Erythromycin (Parental only). The number of total bacteria and *S. mutans* recovered is expressed as CFU per mL of sonicate on each media, and the percentage of *S. mutans* was estimated based on the number of bacteria in BHI + Antibiotic plates relative to blood-agar CFUs.

Bioinformatics

Protein structure predictions were performed using AlphaFold2 (Jumper et al., 2021) in the University of Florida's supercomputer HiPerGator 3.0. Domain function predictions were performed using InterPro (Blum et al., 2021), and the position of domains in transmembrane proteins was predicted using TMHMM 2.0 (Krogh et al., 2001). Homomer predictions were performed by template-based modeling using GalaxyHomomer (Baek et al., 2017) and the predicted AlphaFold2 structures of PgfS and PgfE. Protein alignments were performed using the Basic Local Alignment Search Tool (BLAST), available at the National Center for Biotechnology Information (NCBI) website (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

Statistical Analyses

Statistical analyses were performed using GraphPad Prism 9.4 (GraphPad Software, La Jolla, CA, USA). Quantitative analysis of mutant strains' phenotypes was performed using an analysis of variance (ANOVA) with a Dunnett post hoc test, using the parental strain as a control. Comparisons between strain pairs were performed using a Student's t-test when appropriate. A p-value of $< .05$ was considered to indicate a significant difference in all tests.

Acknowledgments

We thank Dr. Balazs Rada for providing the human blood samples, and also thank the healthy volunteers for the donation of their blood and the staff of the UGA Clinical and Translational Research Unit for drawing blood used in studies, supported by the National Institutes of Health (5R01HL136707 and R21AI147097 to Balazs Rada). The authors also acknowledge the University of Florida Research Computing for providing computational resources and support that have contributed to the research results reported in this publication. URL: http://researchcomputing.ufl.edu.

Ethics statement

Rat oral colonization protocol has been approved by the University of Florida Institutional Animal Care and Use Committee (Protocol 201810421). Human saliva was collected from healthy subjects with written consent, with approval from the University of Florida's Internal Review Board (Protocol 01600877). Whole fresh human blood was

drawn from adult volunteers at the Health Center or the Clinical Translational Research Unit of the University of Georgia under informed consent according to procedures approved by the Institutional Review Boards at the University of Georgia (Protocol 20121076906).

Funding statement

This work was supported by NIH/NIDCR R01DE022559 to J.A. and J.A.L.

Tables

Table A2. Summary of phenotypical differences observed for the *pgf* mutants as compared to the parental strain OMZ175 in current and previous studies.

Analyzed Phenotype	Mutant					Reference
	Δpgf	***ΔpgfS***	***ΔpgfE***	***ΔpgfM1***	***ΔpgfM2***	
Biofilm biomass	↓	=	↓	↓	↓	This paper
Recovered viable cells from biofilms	↓	=	↓	↓	↓	
Biofilm thickness	↓↓	=	↓	↓	↓	
Volume of cells and EPS in biofilms	↓	=	↓	↓	↓	
Patchiness of biofilm	=	↑	↑↑	↑	=	
Biofilm compactness	↑↑	=	↓	↓	↓	
Survival in saliva	=	=	↓↓	↓	↑	
Surface charge	=	↓	=	=	↑	

Resistance to SDS	↓↓	↓↓	↓↓	↓	=	
Genetic competence	↑↑↑	↑	↑↑↑	↑↑↑	↑↑↑	
Rat oral colonization ability	↓↓	N/A	N/A	N/A	N/A	
Human coronary artery cell invasion	↓↓	↓	↓↓	↓↓	↓	(Avilés-Reyes *et al.*, 2014a; Avilés-Reyes *et al.*, 2018)
Killing in *Galleria mellonella*	↓	↓	↓	↓	↓	
Collagen-binding ability	↓	↓	↓	↓	↓	
Glycosylation of LPXTG-anchored proteins Cnm and WapA	↓↓	↓↓	↓↓	↓↓	↓	

= sign used for no significant differences observed; up/down arrows indicate an increase or decrease in phenotype; one arrow is used for modest phenotype, two for moderate, and three for robust.

Table A3. Bacterial strains utilized in this study.

Strain	Genotype	Reference
Streptococcus mutans OMZ175	Parental strain	B. Guggenheim
	ΔgtfA Erm^R	(Miller *et al.*, 2015)
	ΔpgfS Kan^R	(Avilés-Reyes *et al.*, 2014a)
	Δpgf Kan^R	(Avilés-Reyes *et al.*, 2018)
	ΔpgfE Kan^R	
	ΔpgfM1 Kan^R	
	ΔpgfM2 Kan^R	
	ΔvicK Kan^R	(Alves *et al.*, 2018)
	tcnm (no resistances)	(Andresen *et al.*, 2022a)
	tcnm ΔpgfS Kan^R	This study
	tcnm ΔpgfM2 Kan^R	This study

Figures

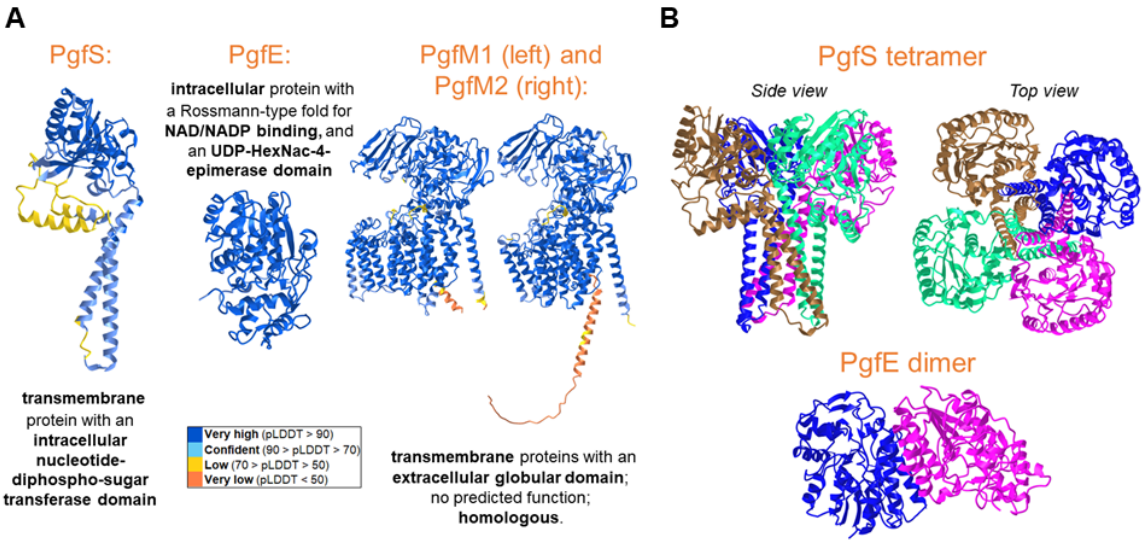


Figure A.1: Bioinformatics prediction for structural features of Pgf proteins. (A) Monomer analysis via Alpha Fold (structure prediction), InterPro (Domains), and TMHMM 2.0 (prediction of domain positions in transmembrane proteins). Colors indicate AlphaFold per-residue estimate of confidence (pLDDT). **(B)** GalaxyHomomer template-based modeling for Pgf proteins predicts PgfS as a tetramer (template PDB - 5EKP) and PgfE as a dimer (template PDB - 1I3K). Each color indicates a different monomer.

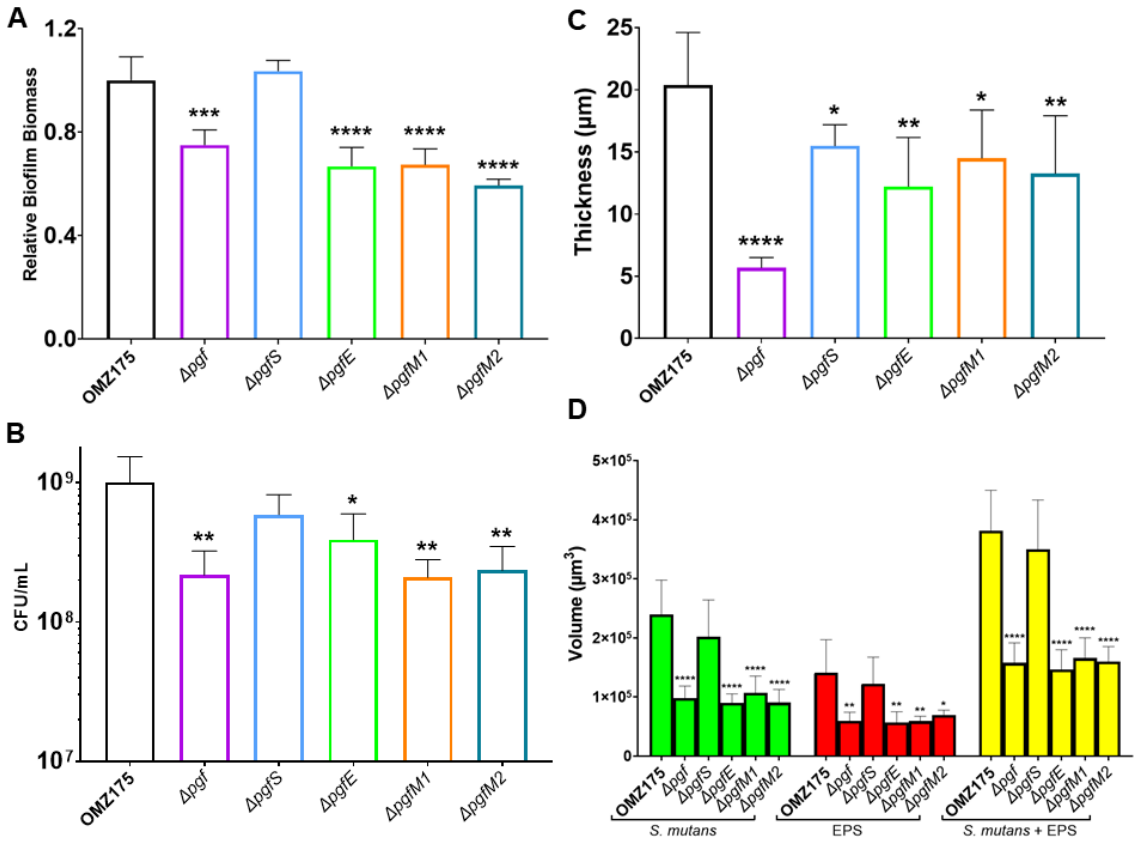


Figure A.2: The Pgf glycosylation machinery is important for proper biofilm formation. Deletion of the entire *pgf* operon or each gene individually, except for *pgfS*, results in biofilm biomass defects. **(A)** Crystal violet staining of biofilms relative to the parental strain, measured by absorbance at 595 nm. **(B)** Viable cells recovered from biofilms. **(C)** Biofilm thickness as measured on a volumetric view via Confocal Laser Microscopy Scanning ($p < .001$). **(D)** The estimated volume of biofilms was calculated from Confocal Laser Microscopy Scanning by multiplying each area by the distance between layers. ($p < .0001$ for *S. mutans*; $p = .0003$ for EPS; $p < .0001$ for sum). Each group was analyzed via one-way ANOVA. Asterisks denote post-hoc comparison with the parental strain. * = $p < .05$; ** = $p < .01$; *** = $p < .001$; **** = $p < .0001$. Bars indicate mean values and error bars indicate standard deviations. N = 4.

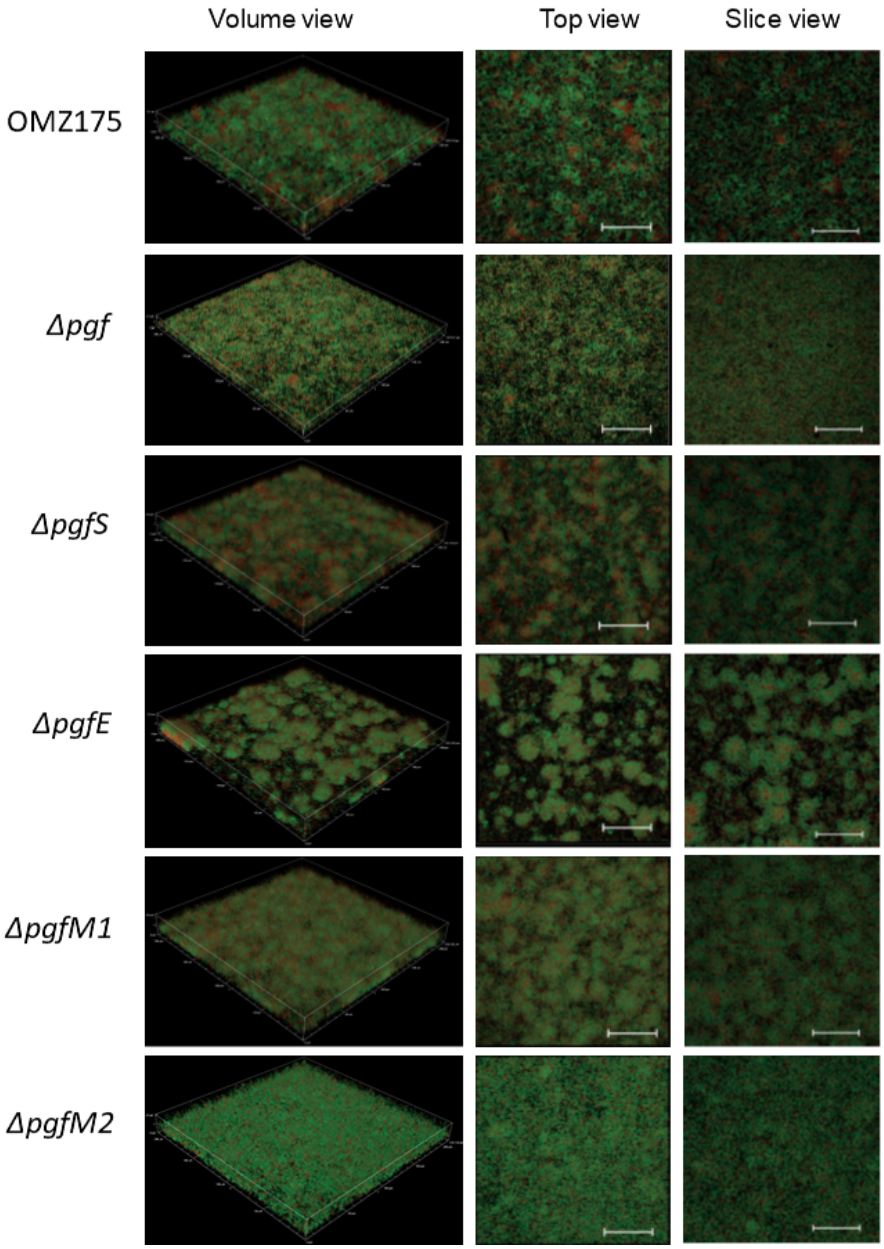


Figure A.3: A proper glycosylation state is important for biofilm assembly and architecture. Confocal Laser Scanning Microscopy of the panel of *S. mutans* strains in volumetric view displays architectural differences for all *pgf* mutants when compared to the parental strain OMZ175. A top (from above) and a slice view (middle of biofilm) are also represented. Representative images were chosen from four biological replicates, with three images collected for each. Green = *S. mutans* cells, Red = Extracellular Polysaccharides (EPS). Each side of the square base = 210 µm. Scale bar = 50 µm.

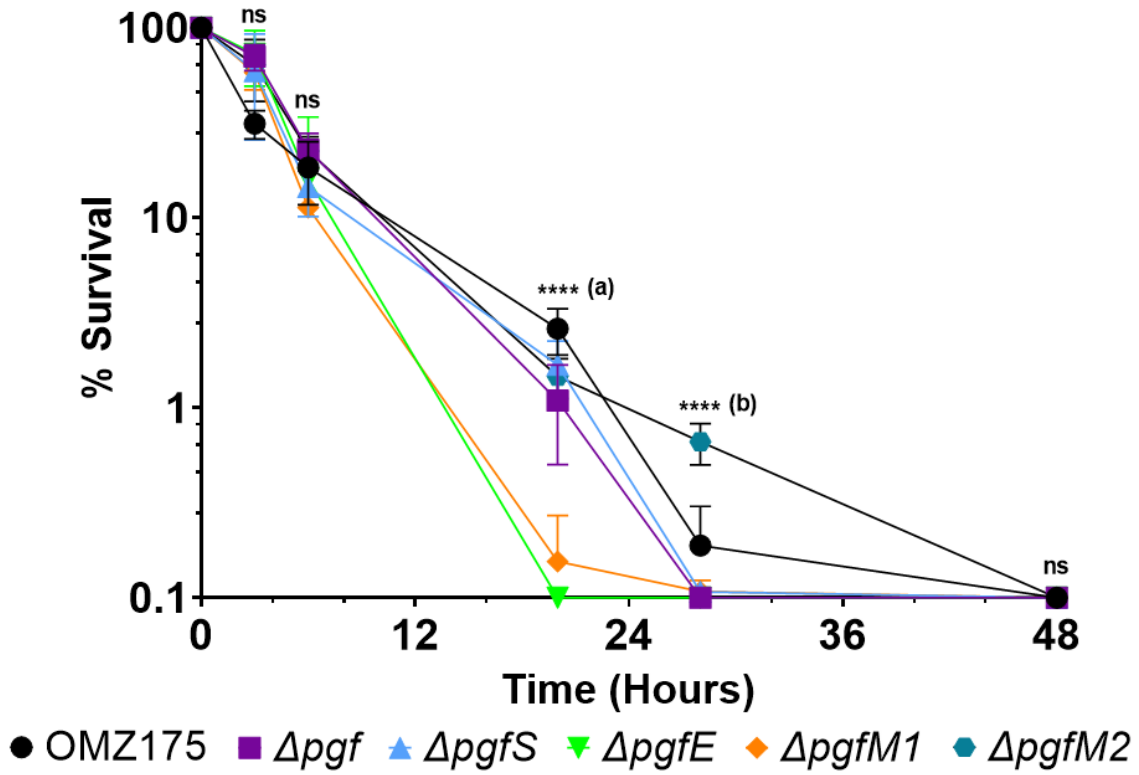


Figure A.4: The Pgf glycosylation machinery contributes to *S. mutans* survival in human saliva supplemented with 20 µM of glucose. The deletion of any *pgf* gene or all of them decreased the survival of bacteria in human saliva, except for the deletion of only *pgfM2*. 0.1% survival was the limit of detection. Connected dots indicate means, and error bars indicate standard deviations. One-way ANOVA was performed on each timepoint to determine differences between each mutant and the parental strain. Asterisks denote post-hoc comparison with the parental strain. (a) denotes differences between parental and *ΔpgfS/ΔpgfM2* ($p < .05$) and between parental and *Δpgf/ΔpgfE/ΔpgfM1* ($p < .0001$). (b) denotes differences between parental and ΔpgfM2 only ($p < .0001$). N = 4. ns = non-significant; * = $p < .05$; ** = $p < .01$; *** = $p < .001$; **** = $p < .0001$.

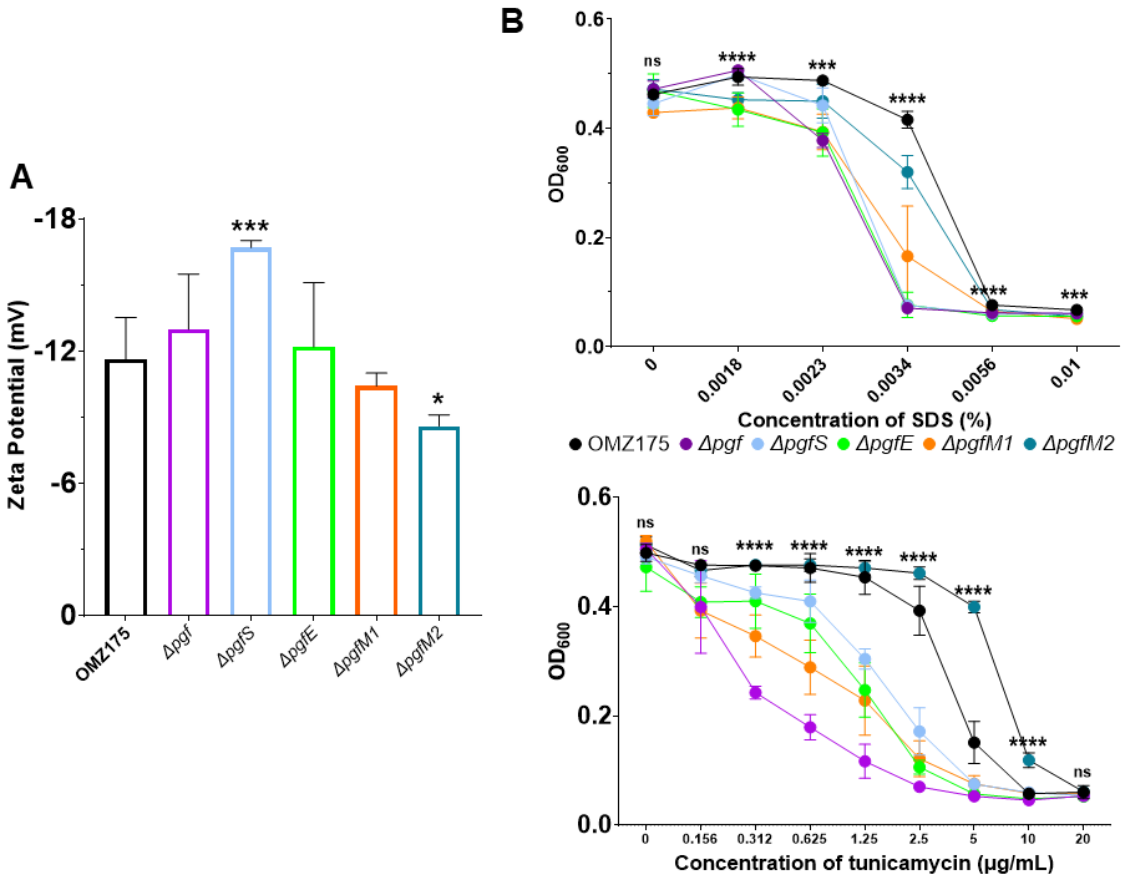


Figure A.5: The Pgf machinery contributes to bacterial surface charge and stability. **(A)** Zeta potential of cells grown to mid-log phase. **(B)** MIC determination for the negatively charged detergent SDS (top) and the rhamnose-glucose polysaccharide synthesis inhibitor tunicamycin (bottom). Bars and dots indicate mean values, and error bars indicate standard deviations. One-way ANOVA was performed on each timepoint to determine differences between each mutant and the parental strain. Asterisks denote post-hoc comparison with the parental strain. Pairwise comparisons for growth in SDS and tunicamycin can be found in Figure A.12 and Figure A.13, respectively. N = 4. ns = non-significant; * = $p < .05$; ** = $p < .01$; *** = $p < .001$; **** = $p < .0001$.

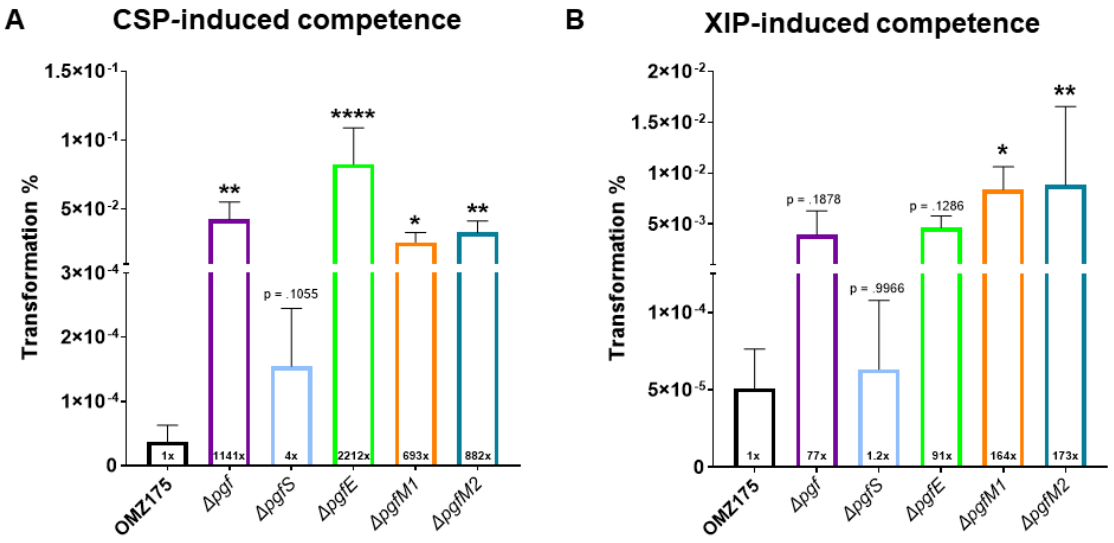


Figure A.6: The lack of proper glycosylation status affects *S. mutans* OMZ175 competence. Transformation assays via stimulation of either *comABCDE* **(A)** or *comRS* **(B)** pathways with CSP and XIP, as indicated. The number inside each bar indicates the average fold increase in competence for each mutant in comparison to the parental strain. One-way ANOVA was utilized for comparisons, with a post-hoc test to compare differences between mutants and the parental strain. * = $p < .05$; ** = $p < .01$; **** = $p < .0001$. Non-significand comparisons have their p values shown above the bar. Bars indicate mean values, and error bars indicate standard deviations. N = 4.

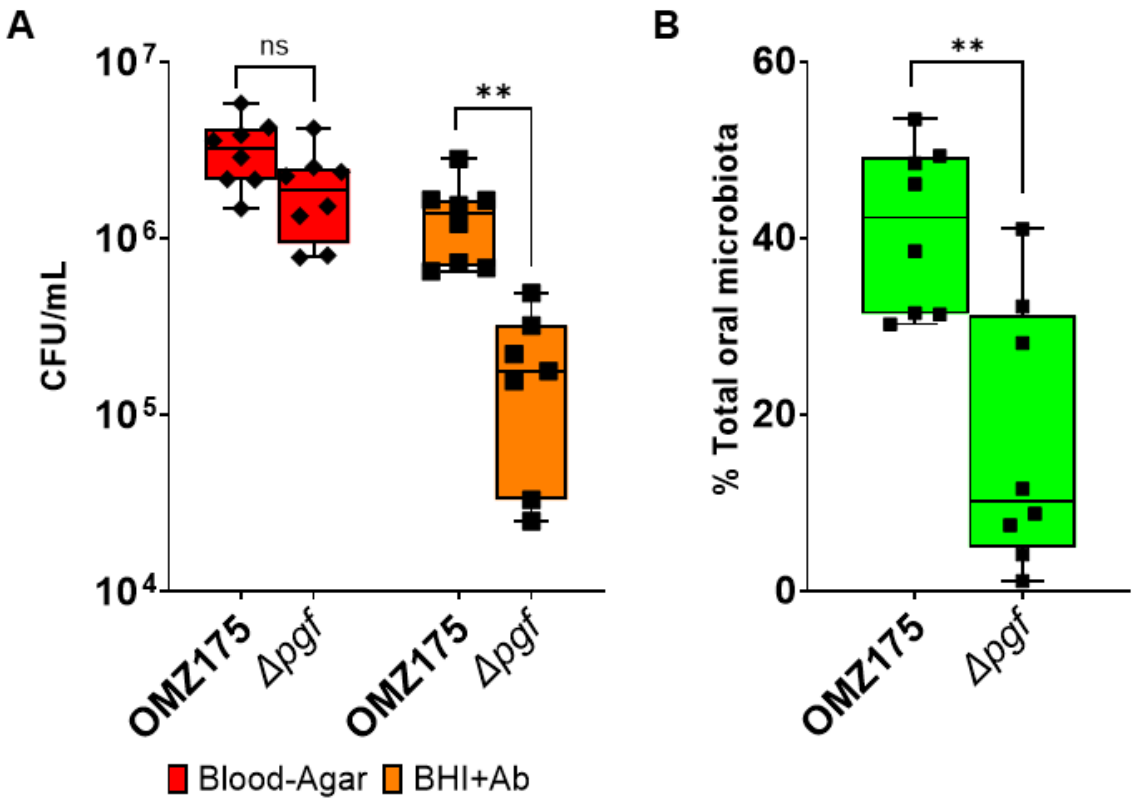
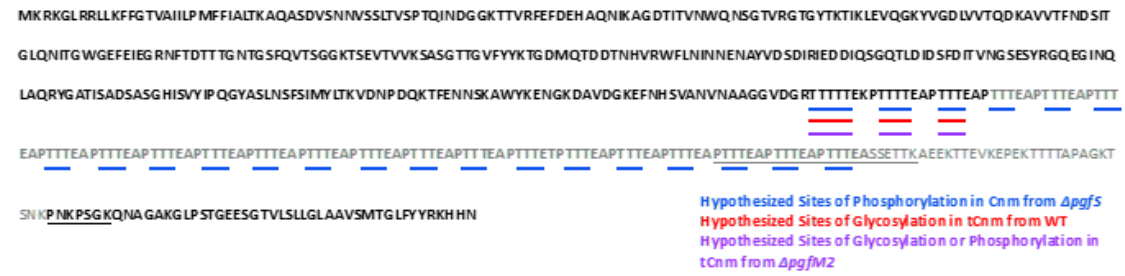


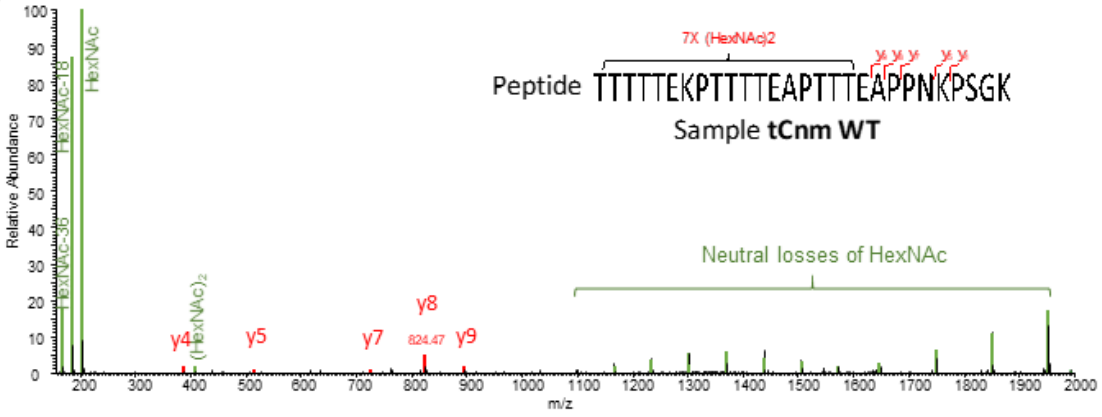
Figure A.7: The Pgf glycosylation machinery contributes to *S. mutans* fitness *in vivo*. Rat oral colonization with *S. mutans* OMZ175 and the quadruple *Δpgf* strain. **(A)** Direct comparison of recovered CFUs between strains for each type of media. Blood-agar indicates CFU counts of total cultivable microbiota, and BHI+antibiotics (BHI+Ab) reveals *S. mutans* CFU counts. **(B)** Comparison between the percentage of *S. mutans* relative to the total oral microbiota for each strain. T-tests were performed for each pair. ns = not significant; ** = $p < .01$. Bars indicate mean values, and error bars indicate standard deviations. N = 8.

A



Hypothesized Sites of Phosphorylation in Cnm from *Δpgf5*
Hypothesized Sites of Glycosylation in tCnm from WT
Hypothesized Sites of Glycosylation or Phosphorylation in tCnm from *ΔpgfM2*

B



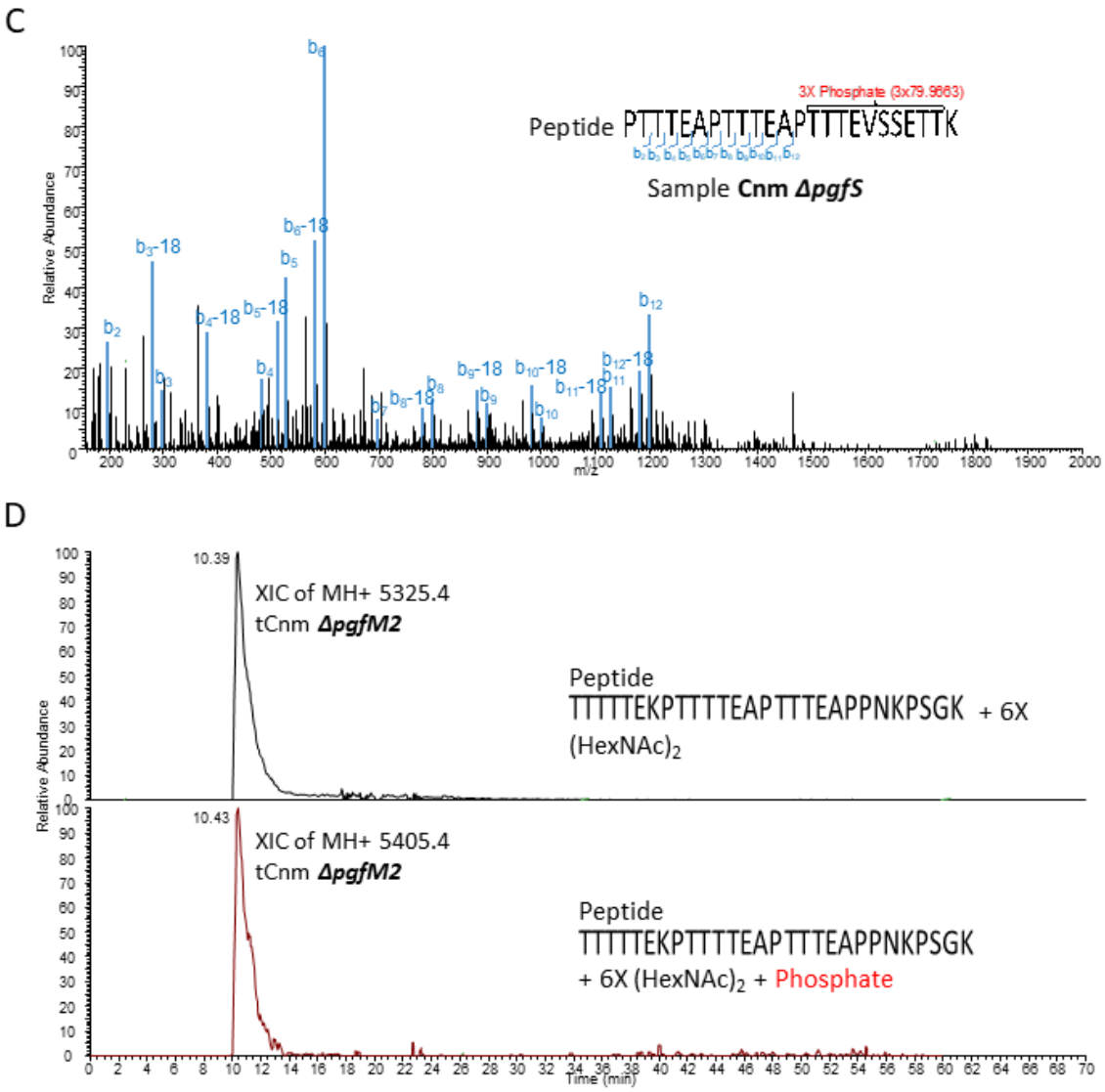


Figure A.8: Cnm and the truncated Cnm (tCnm) are glycosylated with HexNAc in the threonine-rich repeats whereas the unglycosylated Cnm and tCnm from the *ΔpgfS* mutant are phosphorylated. Mass spectrometric analysis of glycosylation and phosphorylation status of Cnm and tCnm purified from parent and *ΔpgfS and ΔpgfM2* mutants. **(A)** Sequence of Cnm (full sequence shown in black and grey) and tCnm (sequence shown in black only) with proposed sites of post-translational modifications within the threonine-rich repeat region. The peptides shown in panels B-D are indicated as underlined. **(B)** MS/MS fragmentation (HCD) of a phosphorylated peptide from Cnm purified from the *ΔpgfS* mutant**. (C)** MS/MS fragmentation (HCD) of a glycosylated peptide from parent tCnm. **(D)** Extracted ion chromatographs (XICs) of peptides from tCnm purified from *ΔpgfM2* indicating partial glycosylation and partial phosphorylation.

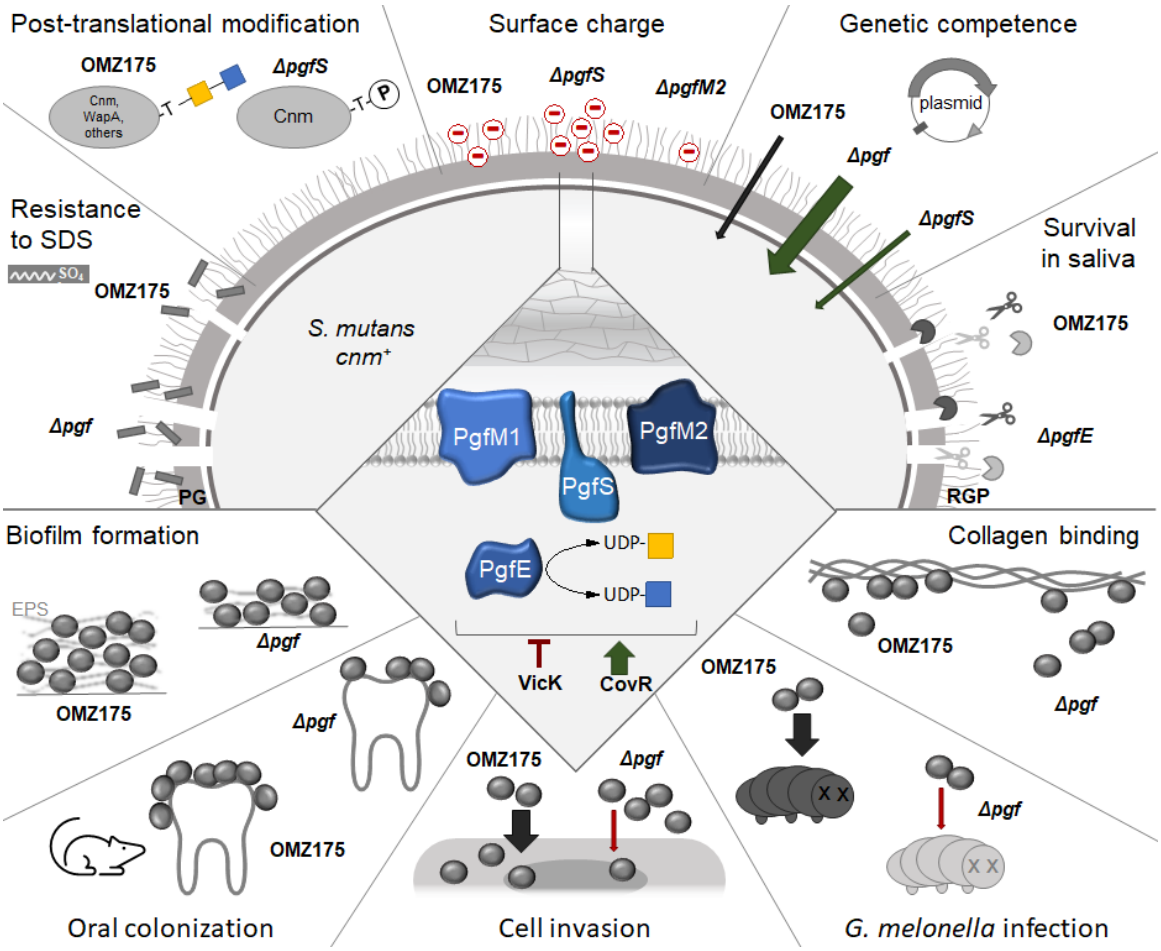


Figure A.9: Summary of findings from our current and previous studies. The Pgf glycosylation machinery is directly involved in the glycosylation of surface adhesins and possibly of the rhamnose-glucose polysaccharide layer. Glycosylation appears to be favored over phosphorylation, suggesting crosstalk or competition between post-translational modification pathways. Glycosylation also modulates surface charge, membrane homeostasis, genetic competence, biofilm formation, saliva survival, and fitness in an oral colonization model. Previously studied phenotypes associated with proper glycosylation of Cnm are endothelial and epithelial cell invasion, systemic infection in the *Galleria mellonella* model, and collagen binding. Expression of *pgf* genes is under positive regulation via CovR and negative regulation via VicRKS. PG = peptidoglycan; RGP = rhamnose-glycose polysaccharide; P = phosphate; EPS = extracellular polysaccharides. Following the Symbol Nomenclature for Glycans, GlcNAc (N-acetylglucosamine) is depicted as a blue square and GalNAc (N-acetylgalactosamine) is depicted as a yellow square.

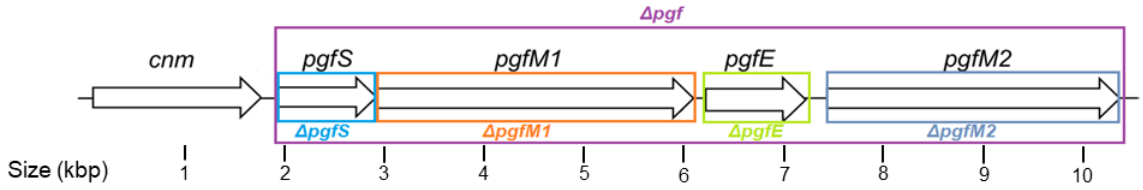


Figure A.10: Organization of the *cnm-pgf* locus in *S. mutans* strain OMZ175, with each mutant in the panel of strains indicated in colored rectangles.

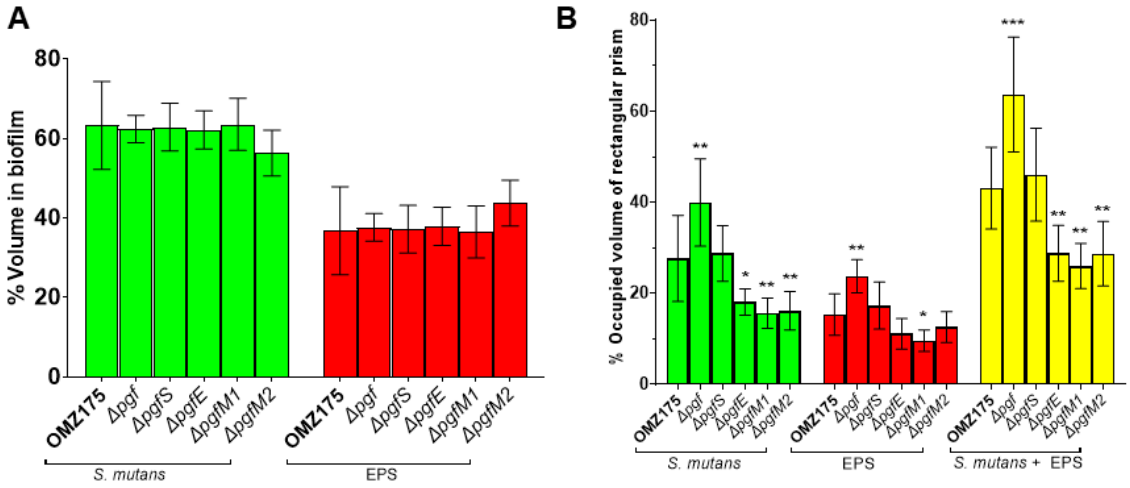


Figure A.11: Quantitative analysis of Confocal Lases Microscopy Scanning images. **(A)** Distribution of *S. mutans* cells and EPS as a percentage of the volume of the sum ($p = .7506$). **(B)** Biofilm compactness as estimated by calculating the volumetric occupancy of each biofilm component relative to the rectangular prism that encases the entire biofilm (base area = 210 x 210 μm^2, height = distance between bottom and top of biofilm) ($p < .0001$ for *S. mutans*; $p = .0002$ for EPS; $p < .0001$ for sum). Bars indicate means, and error bars indicate standard deviation.

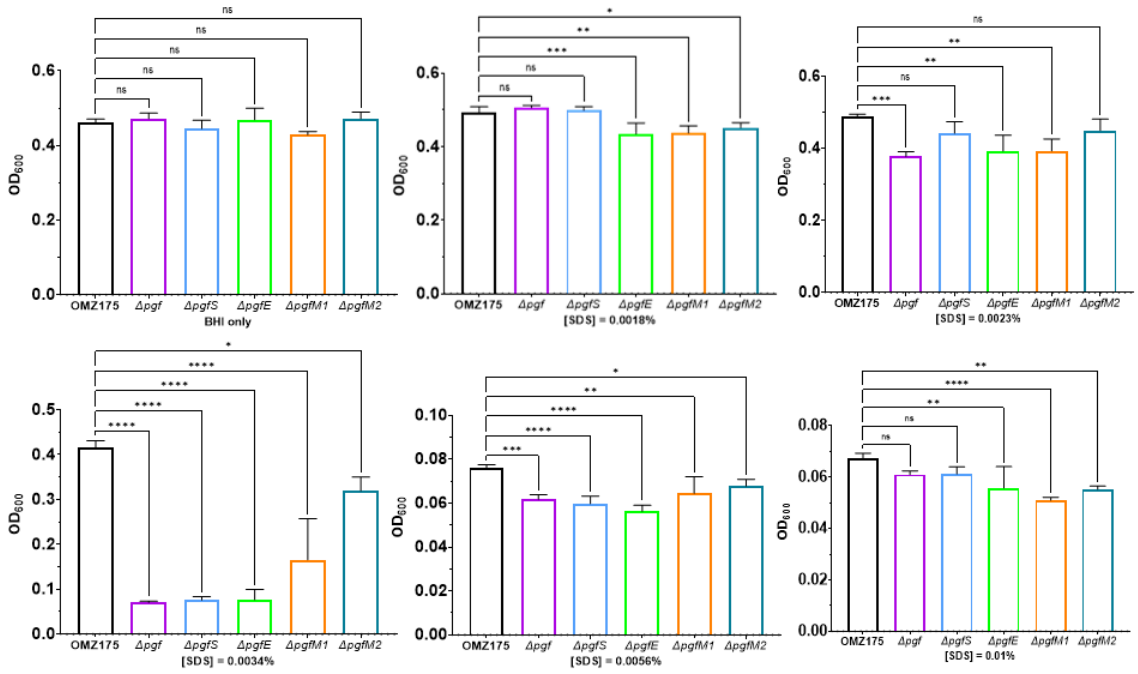


Figure A.12: One-way ANOVA comparisons between the growth of the parental strain and each mutant in different concentrations of SDS. N = 4. ns = non-significant; * = p < .05; ** = p < .01; *** = p < .001; **** = p < .0001.

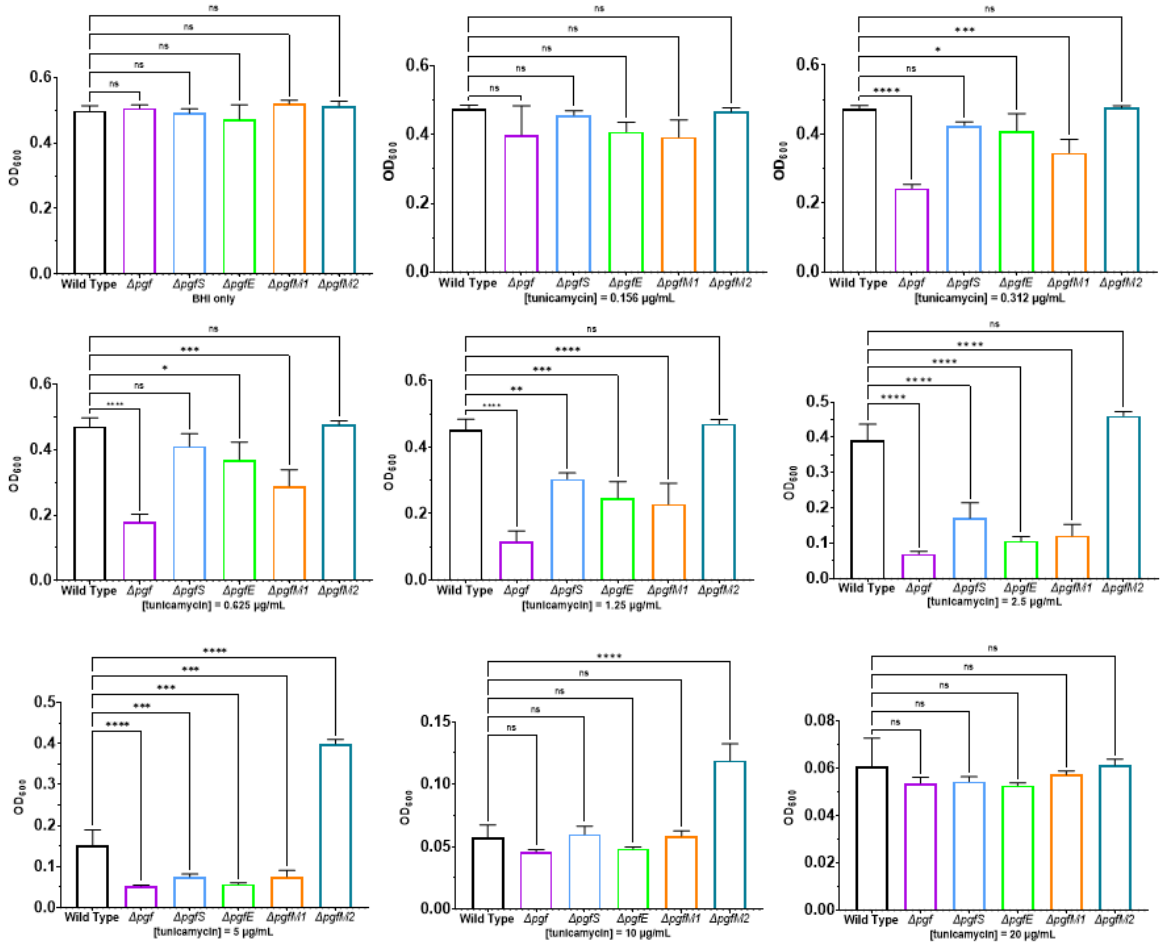


Figure A.13: One-way ANOVA comparisons between the growth of the parental strain and each mutant in different concentrations of tunicamycin. N = 4. ns = non-significant; * = p < .05; ** = p < .01; *** = p < .001; **** = p < .0001.

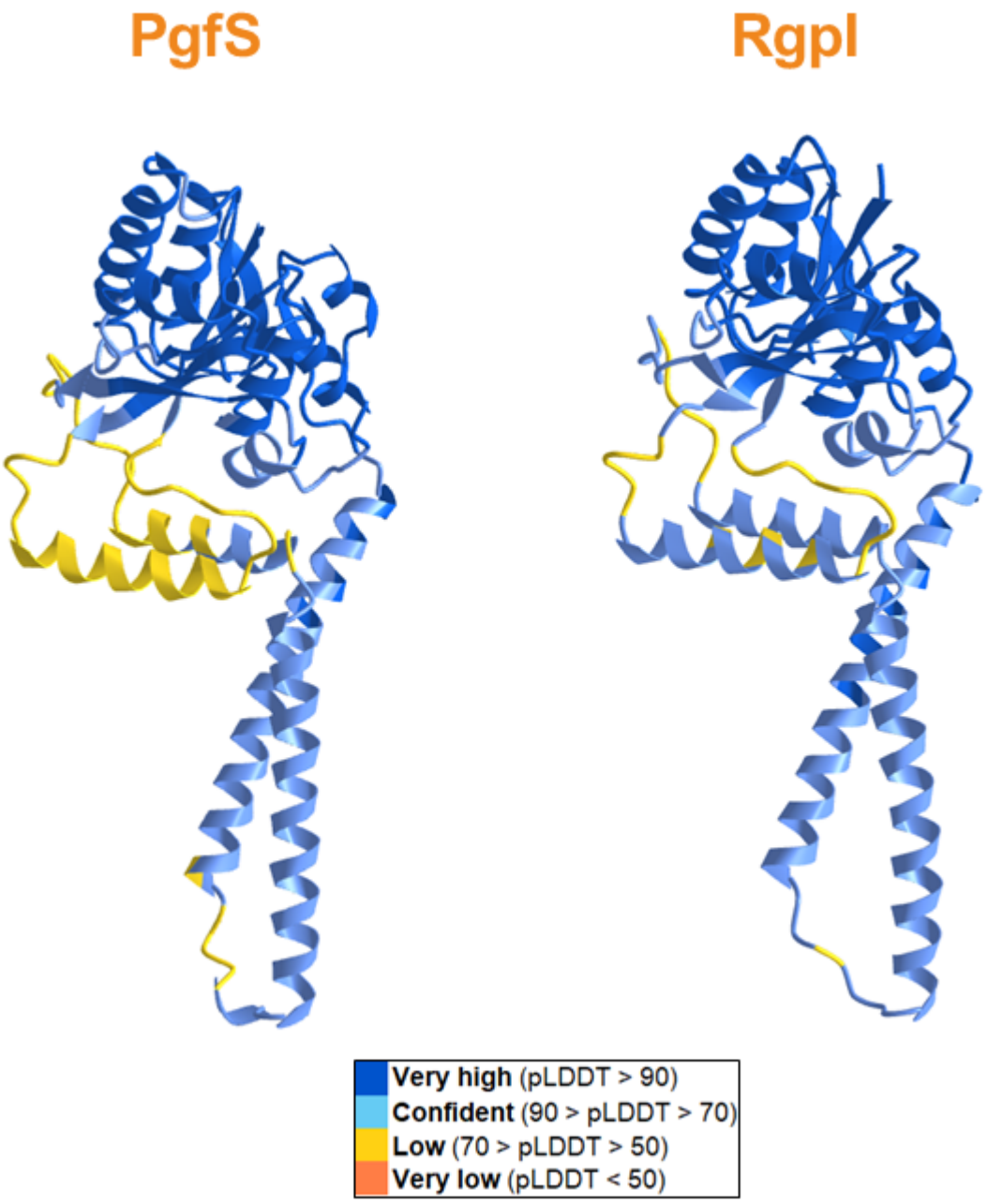


Figure A.14: AlphaFold2 structure predictions for PgfS and RgpI from *S. mutans* indicate a high level of structure homology between the two proteins.

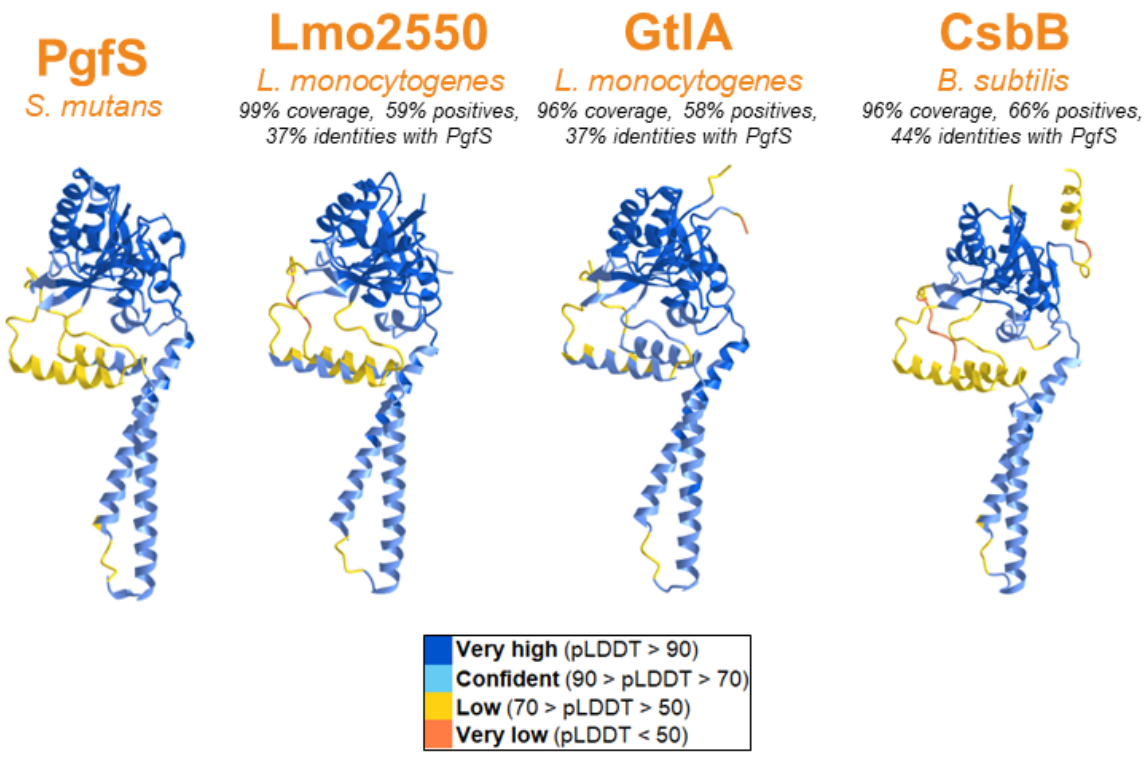


Figure A.15: AlphaFold2 structure predictions for PgfS from *S. mutans* and PgfS-like enzymes involved in lipid glycosylation in *L. monocytogenes* and *B. subtilis*.

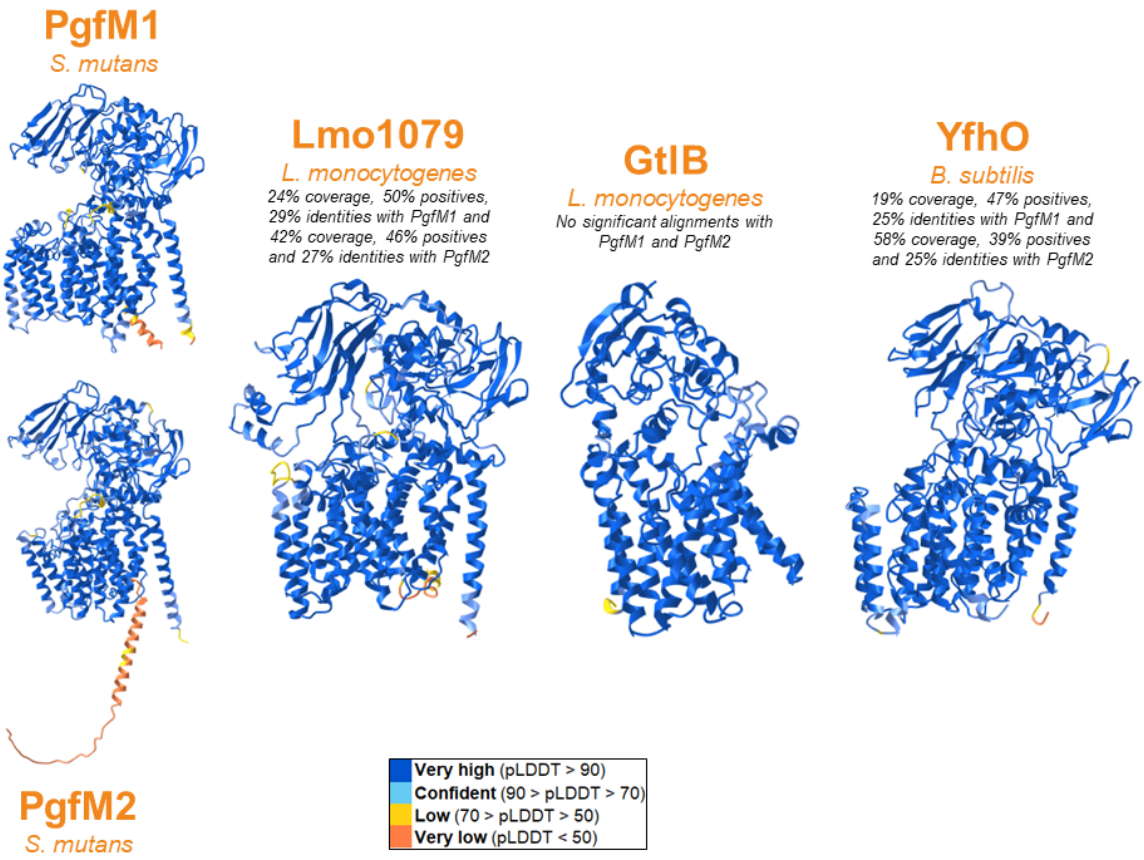


Figure A.16: AlphaFold2 structure predictions for PgfM1 and PgfM2 from *S. mutans* and PgfM1/M2-like enzymes involved in lipid glycosylation in *L. monocytogenes* and *B. subtilis*.

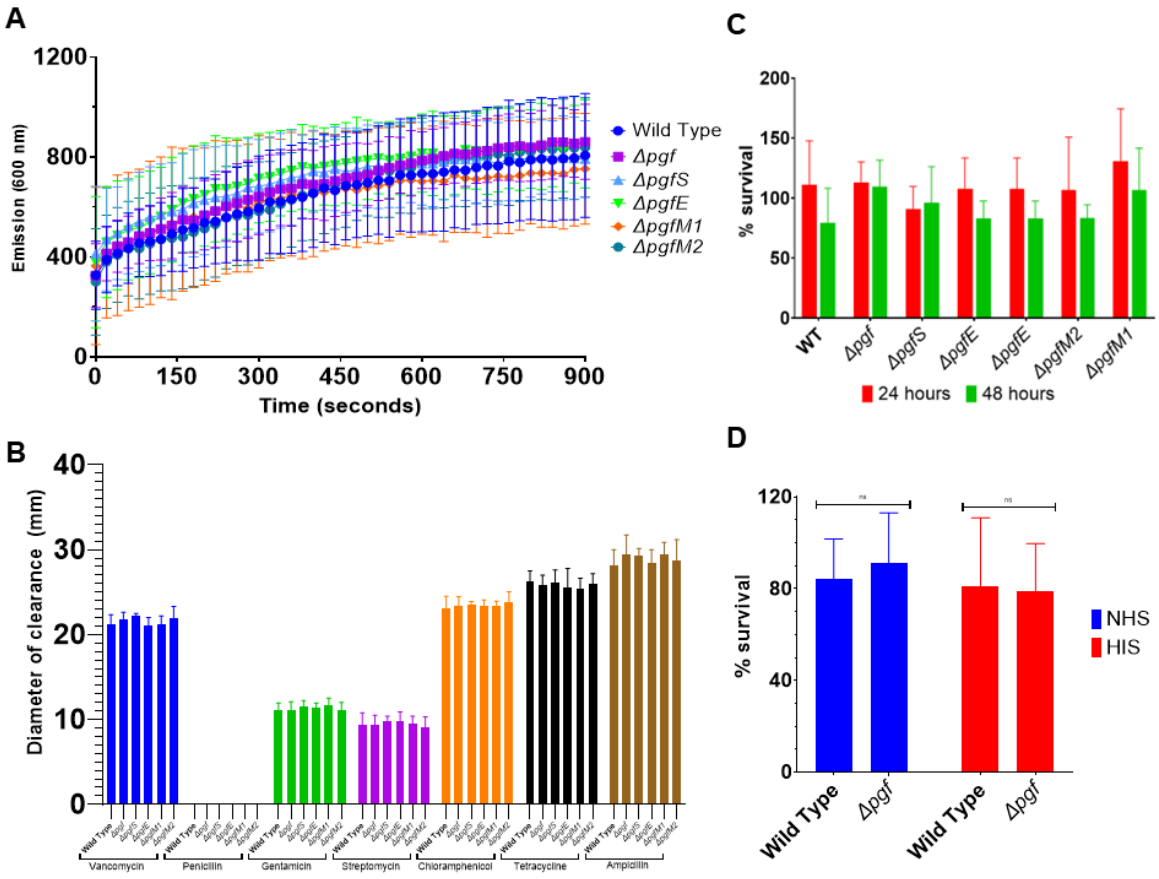
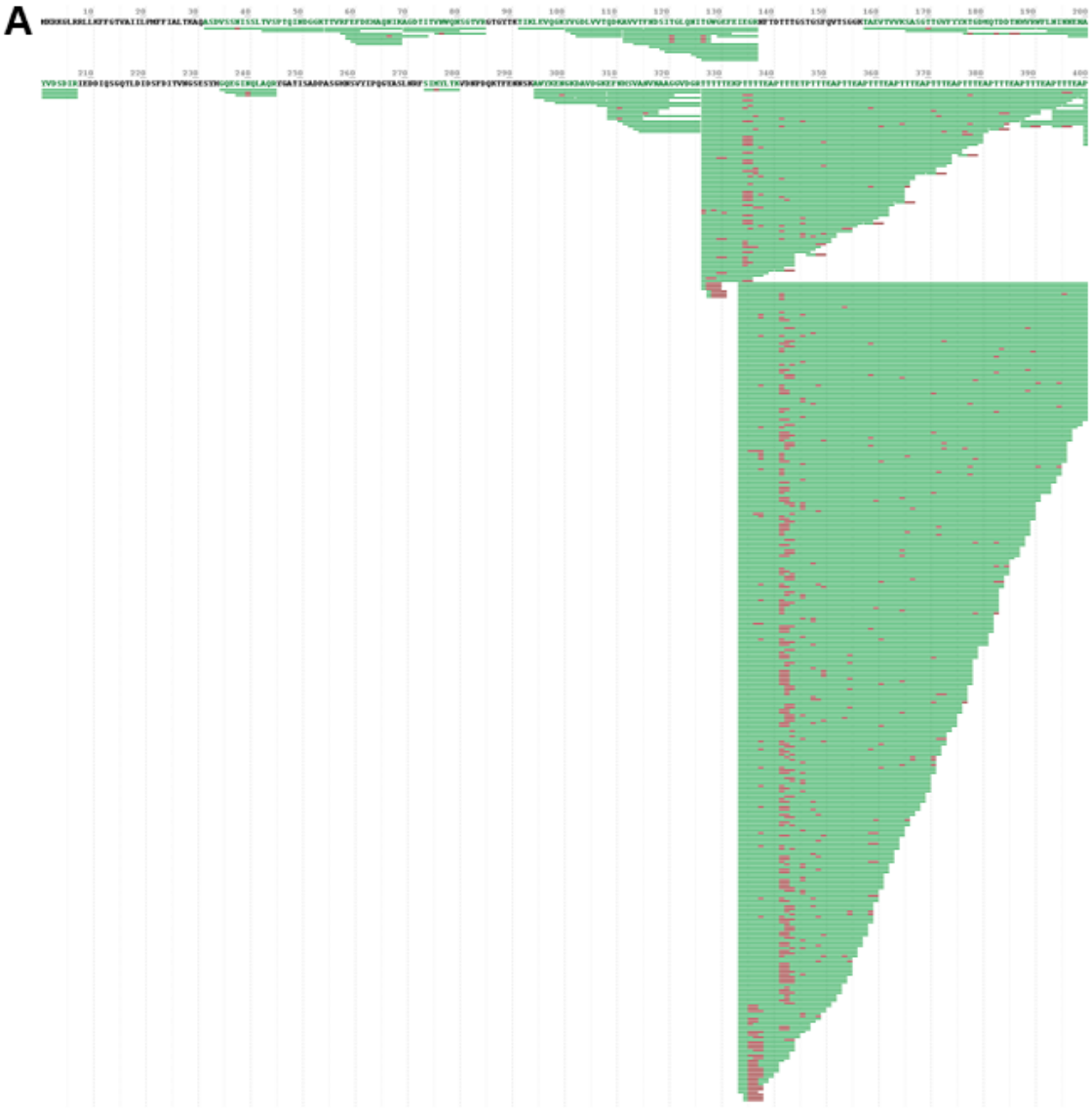


Figure A.17: Ethidium bromide permeability (A), serum survival (B), antibiotic disc resistance (C) and opsonophagocytosis (D) assays did not reveal different phenotypes between the parental strain and the *pgf* mutants.


A

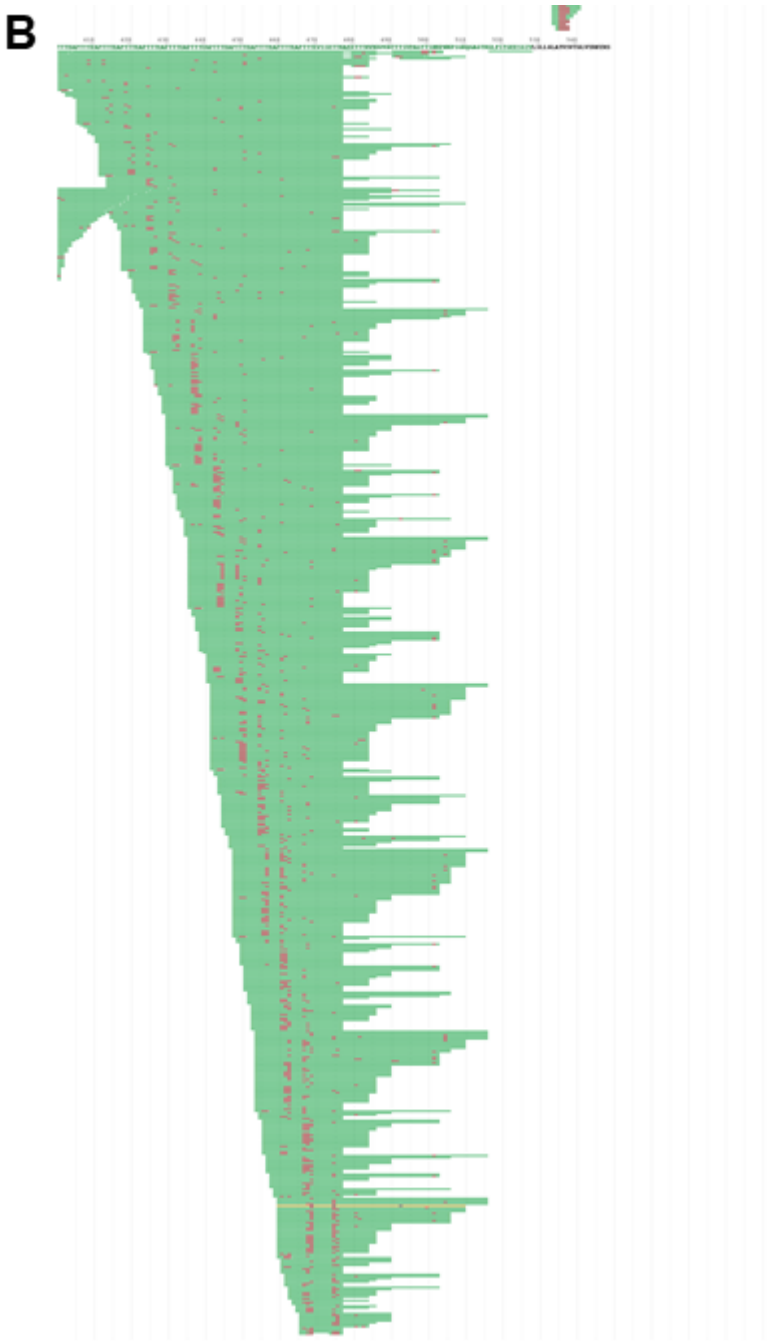


Figure A.18: Phosphate coverage of Cnm from the *ΔpgfS* mutant. (A) N-terminal portion of full-length Cnm from the *ΔpgfS* mutant was subjected to phosphorylation mapping with Byonic. **(B)** C-terminal portion of Cnm, same map as panel A. Red markings at threonines are indications of phosphorylation. Other posttranslational modifications from processing, including oxidation on methionines and deamidation on asparagines, are also indicated. Partial cleavage (see methods) was used to obtain fragments in the threonine-rich repeat region yielding peptides with jagged ends. One of these phosphopeptides is illustrated in Figure A.7.

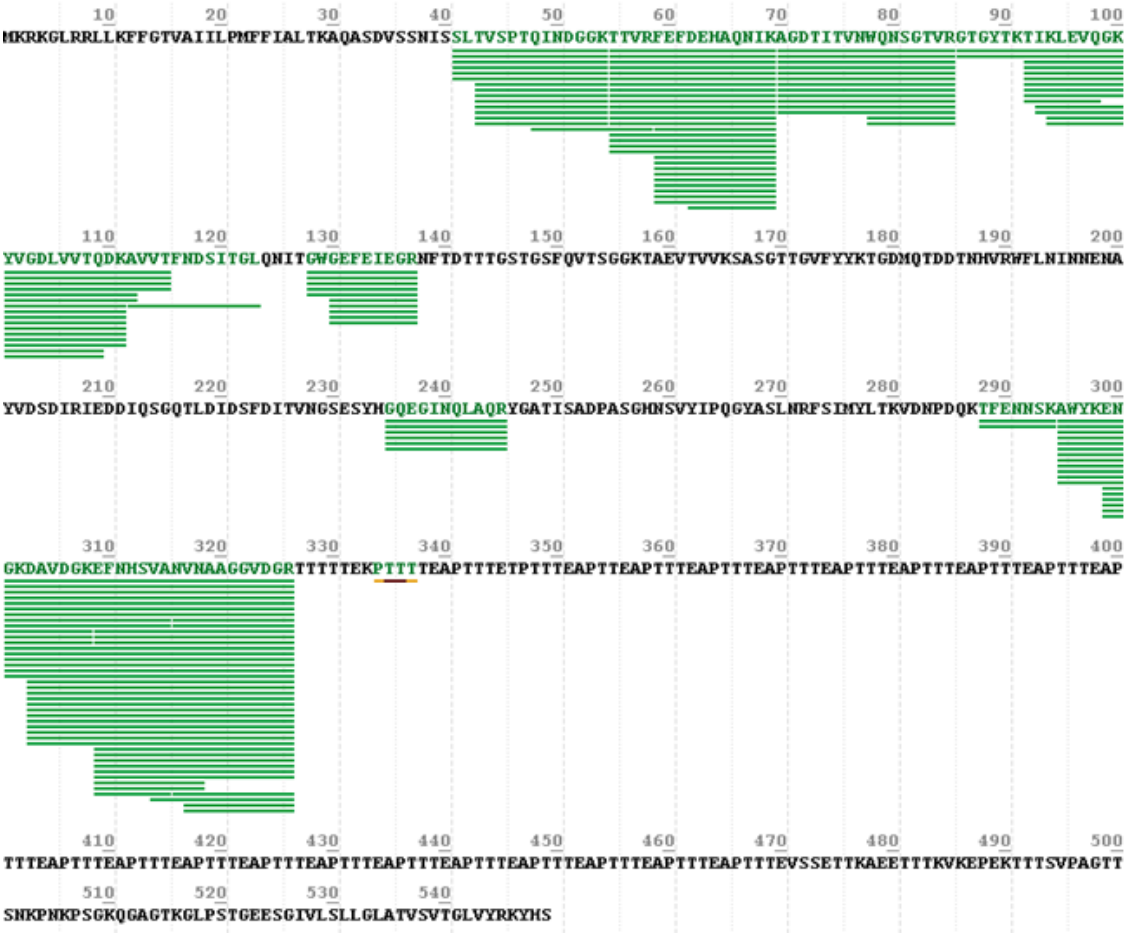


Figure A.19: Phosphate coverage of Cnm from parental strain OMZ175. Mass spectrometry data were subjected to the same search with Byonic as Cnm derived from the *ΔpgfS* mutant. No indications of phosphorylation were found in this sample (as evidenced by the lack of red markings). The lack of coverage in the threonine-rich repeat region is due to the heavy glycosylation we observed in another study (Andresen *et al.*, 2022a).

References

Abranches, J., Miller, J.H., Martinez, A.R., Simpson-Haidaris, P.J., Burne, R.A., and Lemos, J.A. (2011) The collagen-binding protein Cnm is required for Streptococcus mutans adherence to and intracellular invasion of human coronary artery endothelial cells. *Infect Immun* **79**: 2277–2284.

Abranches, J., Zeng, L., Kajfasz, J.K., Palmer, S.R., Chakraborty, B., Wen, Z.T., *et al.* (2018) Biology of Oral Streptococci. *Microbiol Spectr* **6**.

Allen, K.N., and Imperiali, B. (2019) Structural and mechanistic themes in glycoconjugate biosynthesis at membrane interfaces. *Curr Opin Struct Biol* **59**: 81–90.

Alves, L.A., Ganguly, T., Mattos-Graner, R.O., Kajfasz, J., Harth-Chu, E.N., Lemos, J.A., *et al.* (2018) CovR and VicRKX regulate transcription of the collagen binding protein Cnm of Streptococcus mutans. *J Bacteriol* **200**: 1–14 https://journals.asm.org/doi/10.1128/JB.00141-18.

Amorim, H. V., Lopes, M.L., Castro Oliveira, J.V. de, Buckeridge, M.S., Goldman, G.H., Aro, N., *et al.* (2011) Production and characterization of hemicellulase activities from Trichoderma harzianum strain T4. *Appl Microbiol Biotechnol* **40**: 145–155.

Andresen, S., Cologna, N. de M. di, Archer-Hartman, S., Rogers, A.M., Samaddar, S., Black, I.M., *et al.* (2022a) Involvement of the Streptococcus mutans UDP-4-epimerases PgfE and GalE in protein glycosylation, carbon metabolism and cell surface glycosylation. *Glycobiology, conditionally accepted*.

Andresen, S., Fantone, K., Chapla, D., Rada, B., Moremen, K.W., Pierce, M., and Szymanski, C.M. (2022b) Human Intelectin-1 Promotes Cellular Attachment and Neutrophil Killing of Streptococcus pneumoniae in a Serotype-Dependent Manner. *Infect Immun* **90**.

Ardiccioni, C., Clarke, O.B., Tomasek, D., Issa, H.A., Alpen, D.C. von, Pond, H.L., *et al.* (2016) Structure of the polyisoprenyl-phosphate glycosyltransferase GtrB and insights into the mechanism of catalysis. *Nat Commun* **7**: 10175.

Arora, S., Gordon, J., and Hook, M. (2021) Collagen Binding Proteins of Gram-Positive Pathogens. *Front Microbiol* **12**: 1–16.

Avilés-Reyes, A., Freires, I.A., Besingi, R., Purushotham, S., Deivanayagam, C., Brady, L.J., *et al.* (2018) Characterization of the pgf operon involved in the posttranslational modification of Streptococcus mutans surface proteins. *Sci Rep* **8**: 1–12.

Avilés-Reyes, A., Miller, J.H., Simpson-Haidaris, P.J., Hagen, F.K., Abranches, J., and Lemos, J.A. (2014a) Modification of Streptococcus mutans Cnm by PgfS contributes to adhesion, endothelial cell invasion, and virulence. *J Bacteriol* **196**: 2789–2797.

Avilés-Reyes, A., Miller, J.H., Simpson-Haidaris, P.J., Lemos, J.A., and Abranches, J. (2014b) Cnm is a major virulence factor of invasive Streptococcus mutans and part of a conserved three-gene locus. *Mol Oral Microbiol* **29**: 11–23.

Avilés-Reyes, A., Miller, J.H.H., Lemos, J.A.A., and Abranches, J. (2017) The collagen-binding proteins of Streptococcus mutans and related streptococci. *Mol Oral Microbiol* **32**: 89–106.

Baek, M., Park, T., Heo, L., Park, C., and Seok, C. (2017) GalaxyHomomer: a web server for protein homo-oligomer structure prediction from a monomer sequence or structure. *Nucleic Acids Res* **45**: W320–W324.

Banu, L.D., Conrads, G., Rehrauer, H., Hussain, H., Allan, E., and Ploeg, J.R. van der (2010) The Streptococcus mutans Serine/Threonine Kinase, PknB, Regulates Competence Development, Bacteriocin Production, and Cell Wall Metabolism. *Infect Immun* **78**: 2209–2220.

Barletta, R.G., Michalek, S.M., and Curtiss, R. (1988) Analysis of the virulence of Streptococcus mutans serotype c gtfA mutants in the rat model system. *Infect Immun* **56**: 322–330.

Beltrao, P., Bork, P., Krogan, N.J., and Noort, V. van (2013) Evolution and functional cross-talk of protein post-translational modifications. *Mol Syst Biol* **9**: 714 https://onlinelibrary.wiley.com/doi/full/10.1002/msb.201304521. Accessed September 27, 2022.

Besingi, R.N., Wenderska, I.B., Senadheera, D.B., Cvitkovitch, D.G., Long, J.R., Wen, Z.T., and Brady, L.J. (2017) Functional amyloids in streptococcus mutans, their use as targets of biofilm inhibition and initial characterization of SMU_63c. *Microbiology (United Kingdom)* **163**: 488–501.

Bhat, A.H., Maity, S., Giri, K., and Ambatipudi, K. (2019) Protein glycosylation: Sweet or bitter for bacterial pathogens? *Crit Rev Microbiol* **45**: 82–102.

Bischer, A.P., Kovacs, C.J., Faustoferri, R.C., and Quivey, R.G. (2020) Disruption of L-Rhamnose Biosynthesis Results in Severe Growth Defects in Streptococcus mutans. *J Bacteriol* **202** https://journals.asm.org/doi/10.1128/JB.00728-19.

Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., *et al.* (2021) The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res* **49**: D344–D354.

Bowen, W.H.H., and Koo, H. (2011) Biology of Streptococcus mutans-Derived Glucosyltransferases: Role in Extracellular Matrix Formation of Cariogenic Biofilms. *Caries Res* **45**: 69–86.

Brandtzaeg, P. (2013) Secretory immunity with special reference to the oral cavity. *J Oral Microbiol* **5**.

Brockhausen, I. (2014) Crossroads between Bacterial and Mammalian Glycosyltransferases. *Front Immunol* **5**.

Broussard, A., Florwick, A., Desbiens, C., Nischan, N., Robertson, C., Guan, Z., *et al.* (2020) Human UDP-galactose 4′-epimerase (GALE) is required for cell-surface glycome structure and function. *Journal of Biological Chemistry* **295**: 1225–1239 https://linkinghub.elsevier.com/retrieve/pii/S0021925817498826.

Chapot-Chartier, M.P., and Kulakauskas, S. (2014) Cell wall structure and function in lactic acid bacteria. *Microb Cell Fact* **13**: 1–23.

Cologna, N.M. di, Samaddar, S., Valle, C.A., Vargas, J., Aviles-Reyes, A., Morales, J., *et al.* (2021) Amyloid Aggregation of Streptococcus mutans Cnm Influences Its Collagen-Binding Activity. *Appl Environ Microbiol* **87**: e0114921.

Crippen, C.S., Glushka, J., Vinogradov, E., and Szymanski, C.M. (2021) Trehalose-deficient Acinetobacter baumannii exhibits reduced virulence by losing capsular polysaccharide and altering membrane integrity. *Glycobiology* **31**: 1520–1530.

Dubnau, D., and Blokesch, M. (2019) Mechanisms of DNA Uptake by Naturally Competent Bacteria. *Annu Rev Genet* **53**: 217–237.

Eichler, J. (2019) Protein glycosylation. *Current biology* **29**: R229–R231.

Esberg, A., Sheng, N., Mårell, L., Claesson, R., Persson, K., Borén, T., and Strömberg, N. (2017) Streptococcus mutans adhesin biotypes that match and predict individual caries development. *EBioMedicine* **24**: 205–215.

Fábián, T.K., Hermann, P., Beck, A., Fejérdy, P., and Fábián, G. (2012) Salivary defense proteins: their network and role in innate and acquired oral immunity. *Int J Mol Sci* **13**: 4295–320.

Falsetta, M.L., Klein, M.I., Colonne, P.M., Scott-Anne, K., Gregoire, S., Pai, C.H., *et al.* (2014) Symbiotic relationship between Streptococcus mutans and Candida albicans synergizes virulence of plaque biofilms in vivo. *Infect Immun* **82**: 1968–1981.

Fry, B.N., Feng, S., Chen, Y.-Y., Newell, D.G., Coloe, P.J., and Korolik, V. (2000) The galE Gene of Campylobacter jejuni Is Involved in Lipopolysaccharide Synthesis and Virulence. *Infect Immun* **68**: 2594–2601.

Galvão, L.C.C., Rosalen, P.L., Rivera-Ramos, I., Franco, G.C.N., Kajfasz, J.K., Abranches, J., *et al.* (2017) Inactivation of the spxA1 or spxA2 gene of Streptococcus mutans decreases virulence in the rat caries model. *Mol Oral Microbiol* **32**: 142–153.

Garcia, B.A., Acosta, N.C., Tomar, S.L., Roesch, L.F.W., Lemos, J.A., Mugayar, L.R.F., and Abranches, J. (2021) Association of Candida albicans and Cbp+ Streptococcus mutans with early childhood caries recurrence. *Sci Rep* **11**: 1–11.

Han, T.K., Zhang, C., and Dao, M.L. (2006) Identification and characterization of collagen-binding activity in Streptococcus mutans wall-associated protein: A possible implication in dental root caries and endocarditis. *Biochem Biophys Res Commun* **343**: 787–792.

Hanukoglu, I. (2015) Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites. *Biochemistry and Molecular Biology Education* **43**: 206–209.

Hosoki, S., Saito, S., Tonomura, S., Ishiyama, H., Yoshimoto, T., Ikeda, S., *et al.* (2020) Oral Carriage of Streptococcus mutans Harboring the cnm Gene Relates to an Increased Incidence of Cerebral Microbleeds. *Stroke* 3632–3639.

Hwang, G., Liu, Y., Kim, D., Li, Y., Krysan, D.J., and Koo, H. (2017) Candida albicans mannans mediate Streptococcus mutans exoenzyme GtfB binding to modulate cross-kingdom biofilm development in vivo. *PLoS Pathog* **13**: 1–25.

Inenaga, C., Hokamura, K., Nakano, K., Nomura, R., Naka, S., Ohashi, T., *et al.* (2018) A Potential New Risk Factor for Stroke: Streptococcus Mutans With Collagen-Binding Protein. *World Neurosurg* **113**: e77–e81.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**: 583–589.

Kaspar, J.R., and Walker, A.R. (2019) Expanding the Vocabulary of Peptide Signals in Streptococcus mutans. *Front Cell Infect Microbiol* **9**.

Klein, M.I., Hwang, G., Santos, P.H.S., Campanella, O.H., and Koo, H. (2015) Streptococcus mutans-derived extracellular matrix in cariogenic oral biofilms. *Front Cell Infect Microbiol* **5**.

Koo, H., Andes, D.R., and Krysanid, D.J. (2018) Candida-streptococcal interactions in biofilm-associated oral diseases Bacterial-fungal interactions and oral diseases. 1–7.

Kovacs, C.J., Faustoferri, R.C., Bischer, A.P., and Quivey, R.G. (2019) Streptococcus mutans requires mature rhamnose-glucose polysaccharides for proper pathophysiology, morphogenesis and cellular division. *Mol Microbiol* **112**: 944–959 https://onlinelibrary.wiley.com/doi/10.1111/mmi.14330.

Krogh, A., Larsson, B., Heijne, G. von, and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**: 567–80.

Krzyściak, W., Jurczak, A., Kościelniak, D., Bystrowska, B., and Skalniak, A. (2014) The virulence of Streptococcus mutans and the ability to form biofilms. *European Journal of Clinical Microbiology & Infectious Diseases* **33**: 499–515.

Lara Vasquez, P., Mishra, S., Kuppuswamy, S.K., Crowley, P.J., and Brady, L.J. (2021) Protein Interactomes of Streptococcus mutans YidC1 and YidC2 Membrane Protein Insertases Suggest SRP Pathway-Independent- and -Dependent Functions, Respectively. *mSphere* **6**.

Lee, M.J., Gravelat, F.N., Cerone, R.P., Baptista, S.D., Campoli, P. v, Choe, S.-I., *et al.* (2014) Overlapping and distinct roles of Aspergillus fumigatus UDP-glucose 4-epimerases in galactose metabolism and the synthesis of galactose-containing cell wall polysaccharides. *J Biol Chem* **289**: 1243–56 https://www.jbc.org/article/S0021-9258(20)33607-3/fulltext.

Lees-Miller, R.G., Iwashkiw, J.A., Scott, N.E., Seper, A., Vinogradov, E., Schild, S., and Feldman, M.F. (2013) A common pathway for O-linked protein-glycosylation and synthesis of capsule in Acinetobacter baumannii. *Mol Microbiol* **89**: 816–830.

Lemos, J.A., Abranches, J., Koo, H., Marquis, R.E., and Burne, R.A. (2010) Protocols to study the physiology of oral biofilms. *Methods Mol Biol* **666**: 87–102.

Lemos, J.A., Palmer, S.R., Zeng, L., Wen, Z.T., Kajfasz, J.K., Freires, I.A., *et al.* (2019) The Biology of Streptococcus mutans. *Microbiol Spectr* **7**: 1–18.

Li, J., Dong, X., Cui, Y., Li, S., Chen, C., Zhang, X., *et al.* (2022) Simultaneous enrichment and sequential separation of O-linked glycopeptides and phosphopeptides with immobilized titanium (IV) ion affinity chromatography materials. *J Chromatogr A* **1681**: 463462.

Lin, B., Qing, X., Liao, J., and Zhuo, K. (2020) Role of Protein Glycosylation in Host-Pathogen Interaction. *Cells* **9**: 1022.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**: 490–495.

Lu, Q., Li, S., and Shao, F. (2015) Sweet Talk: Protein Glycosylation in Bacterial Interaction With the Host. *Trends Microbiol* **23**: 630–641.

Lynge Pedersen, A.M., and Belstrøm, D. (2019) The role of natural salivary defences in maintaining a healthy oral microbiota. *J Dent* **80**: S3–S12 https://doi.org/10.1016/j.jdent.2018.08.010.

Ma, Q., Zhang, Q., Chen, Y., Yu, S., Huang, J., Liu, Y., *et al.* (2021) Post-translational Modifications in Oral Bacteria and Their Functional Impact. *Front Microbiol* **12**.

Macek, B., Forchhammer, K., Hardouin, J., Weber-Ban, E., Grangeasse, C., and Mijakovic, I. (2019) Protein post-translational modifications in bacteria. *Nat Rev Microbiol* **17**: 651–664.

Miller, J.H., Avilés-Reyes, A., Scott-Anne, K., Gregoire, S., Watson, G.E., Sampson, E., *et al.* (2015) The collagen binding protein Cnm contributes to oral colonization and cariogenicity of Streptococcus mutans OMZ175. *Infect Immun* **83**: 2001–2010.

Mistou, M.-Y., Sutcliffe, I.C., and Sorge, N.M. van (2016) Bacterial glycobiology: rhamnose-containing cell wall polysaccharides in Gram-positive bacteria. *FEMS Microbiol Rev* **40**: 464–479.

Nakano, K., Nomura, R., Taniguchi, N., Lapirattanakul, J., Kojima, A., Naka, S., *et al.* (2010) Molecular characterization of Streptococcus mutans strains containing the cnm gene encoding a collagen-binding adhesin. *Arch Oral Biol* **55**: 34–39.

Ng, W., and Ting, Y.-P. (2016) Zeta potential of bacterial cells: Effect of wash buffers. *PeerJ Prepr* **Open Acces**: 2–3.

Nomura, R., Naka, S., Nemoto, H., Otsugu, M., Nakamura, S., Ooshima, T., and Nakano, K. (2013) Potential high virulence for infective endocarditis in Streptococcus mutans strains with collagen-binding proteins but lacking PA expression. *Arch Oral Biol* **58**: 1627–1634.

Nomura, R., Nakano, K., Taniguchi, N., Lapirattanakul, J., Nemoto, H., Grönroos, L., *et al.* (2009) Molecular and clinical analyses of the gene encoding the collagen-binding adhesin of Streptococcus mutans. *J Med Microbiol* **58**: 469–475.

Nomura, R., Otsugu, M., Hamada, M., Matayoshi, S., Teramoto, N., Iwashita, N., *et al.* (2020) Potential involvement of Streptococcus mutans possessing collagen binding protein Cnm in infective endocarditis. *Sci Rep* **10**: 1–14.

Nothaft, H., and Szymanski, C.M. (2010) Protein glycosylation in bacteria: Sweeter than ever. *Nat Rev Microbiol* **8**: 765–778.

Nothaft, H., and Szymanski, C.M. (2013) Bacterial protein n-glycosylation: New perspectives and applications. *Journal of Biological Chemistry* **288**: 6912–6920.

Oli, M.W., Otoo, H.N., Crowley, P.J., Heim, K.P., Nascimento, M.M., Ramsook, C.B., *et al.* (2012) Functional amyloid formation by Streptococcus mutans. *Microbiology (N Y)* **158**: 2903–2916.

Patry, R.T., Stahl, M., Perez-Munoz, M.E., Nothaft, H., Wenzel, C.Q., Sacher, J.C., *et al.* (2019) Bacterial AB 5 toxins inhibit the growth of gut bacteria by targeting ganglioside-like glycoconjugates. *Nat Commun* **10**.

Paula, A.J., Hwang, G., and Koo, H. (2020) Dynamics of bacterial population growth in biofilms resemble spatial and structural aspects of urbanization. *Nat Commun* **11**: 1–14.

Pitts, N.B., Zero, D.T., Marsh, P.D., Ekstrand, K., Weintraub, J.A., Ramos-Gomez, F., *et al.* (2017) Dental caries. *Nat Rev Dis Primers* **3**.

Prabakaran, S., Lippens, G., Steen, H., and Gunawardena, J. (2012) Post-translational modification: Nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip Rev Syst Biol Med* **4**: 565–583.

Rainey, K., Michalek, S.M., Wen, Z.T., and Wu, H. (2018) Glycosyltransferase-mediated biofilm matrix dynamics and virulence of Streptococcus mutans. *Appl Environ Microbiol* **85**: 1–15.

Reddick, L.E., and Alto, N.M. (2014) Bacteria fighting back: how pathogens target and subvert the host innate immune system. *Mol Cell* **54**: 321–8.

Rijn, I. van de, and Kessler, R.E. (1980) Growth characteristics of group A streptococci in a new chemically defined medium. *Infect Immun* **27**: 444–8.

Rismondo, J., Gillis, A., and Gründling, A. (2021) Modifications of cell wall polymers in Gram-positive bacteria by multi-component transmembrane glycosylation systems. *Curr Opin Microbiol* **60**: 24–33.

Rismondo, J., Haddad, T.F.M., Shen, Y., Loessner, M.J., and Gründling, A. (2020) GtcA is required for LTA glycosylation in Listeria monocytogenes serovar 1/2a and Bacillus subtilis. *Cell Surface* **6**: 100038.

Rismondo, J., Percy, M.G., and Gründling, A. (2018) Discovery of genes required for lipoteichoic acid glycosylation predicts two distinct mechanisms for wall teichoic acid glycosylation. *Journal of Biological Chemistry* **293**: 3293–3306.

Roper, J.R., Güther, M.L.S., MacRae, J.I., Prescott, A.R., Hallyburton, I., Acosta-Serrano, A., and Ferguson, M.A.J. (2005) The Suppression of Galactose Metabolism in Procylic Form Trypanosoma brucei Causes Cessation of Cell Growth and Alters Procyclin Glycoprotein Structure and Copy Number. *Journal of Biological Chemistry* **280**: 19728–19736 https://linkinghub.elsevier.com/retrieve/pii/S0021925820617574.

Sato, Y., Okamoto, K., Kagami, A., Yamamoto, Y., Igarashi, T., and Kizaki, H. (2004) Streptococcus mutans Strains Harboring Collagen-binding Adhesin. *J Dent Res* **83**: 534–539.

Schäffer, C., and Messner, P. (2017) Emerging facets of prokaryotic glycosylation. *FEMS Microbiol Rev* **41**: 49–91.

Schäper, S., Wendt, H., Bamberger, J., Sieber, V., Schmid, J., and Becker, A. (2019) A Bifunctional UDP-Sugar 4-Epimerase Supports Biosynthesis of Multiple Cell Surface Polysaccharides in Sinorhizobium meliloti. *J Bacteriol* **201** https://journals.asm.org/doi/10.1128/JB.00801-18.

Schmid, J., Heider, D., Wendel, N.J., Sperl, N., and Sieber, V. (2016) Bacterial glycosyltransferases: Challenges and Opportunities of a Highly Diverse Enzyme Class Toward Tailoring Natural Products. *Front Microbiol* **7**: 1–7.

Sciotti, M.A., Chatenay-Rivauday, C., Yamodo, I., and Ogier, J. (1997) The N-terminal half part of the oral streptococcal antigen I/IIf contains two distinct functional domains. *Adv Exp Med Biol* **418**: 699–701.

Shajahan, A., Heiss, C., Ishihara, M., and Azadi, P. (2017) Glycomic and glycoproteomic analysis of glycoproteins—a tutorial. *Anal Bioanal Chem* **409**: 4483–4505

http://link.springer.com/10.1007/s00216-017-0406-7. Accessed October 2, 2022.

Shibata, Y., Yamashita, Y., and Ploeg, J.R. van der (2009) The serotype-specific glucose side chain of rhamnose-glucose polysaccharides is essential for adsorption of bacteriophage M102 to Streptococcus mutans. *FEMS Microbiol Lett* **294**: 68–73.

Shields, R.C., O'Brien, G., Maricic, N., Kesterson, A., Grace, M., Hagen, S.J., and Burne, R.A. (2018) Genome-Wide Screens Reveal New Gene Products That Influence Genetic Competence in Streptococcus mutans. *J Bacteriol* **200**.

Simón-Soro, A., and Mira, A. (2015) Solving the etiology of dental caries. *Trends Microbiol* **23**: 76–82.

Sitkiewicz, I. (2018) How to become a killer, or is it all accidental? Virulence strategies in oral streptococci. *Mol Oral Microbiol* **33**: 1–12.

Son, M., Shields, R.C., Ahn, S.-J.J., Burne, R.A., and Hagen, S.J. (2015) Bidirectional signaling in the competence regulatory pathway of Streptococcus mutans. *FEMS Microbiol Lett* **362**: fnv159.

Soni, K.A., Balasubramanian, A.K., Beskok, A., and Pillai, S.D. (2008) Zeta potential of selected bacteria in drinking water when dead, starved, or exposed to minimal and rich culture media. *Curr Microbiol* **56**: 93–97.

Spiro, R.G. (2002) Protein glycosylation: Nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**.

Stipp, R.N., Boisvert, H., Smith, D.J., Höfling, J.F., Duncan, M.J., and Mattos-Graner, R.O. (2013) CovR and VicRK Regulate Cell Surface Biogenesis Genes Required for Biofilm Formation in Streptococcus mutans. *PLoS One* **8**: e58271 https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0058271. Accessed October 1, 2022.

Sullan, R.M.A., Li, J.K., Crowley, P.J., Brady, L.J., and Dufrêne, Y.F. (2015) Binding forces of Streptococcus mutans P1 adhesin. *ACS Nano* **9**: 1448–1460.

Takahashi, M., Hasegawa, Y., Maeda, K., Kitano, M., and Taniguchi, N. (2022) Role of glycosyltransferases in carcinogenesis; growth factor signaling and EMT/MET programs. *Glycoconj J* **39**: 167–176.

Tan, F.Y.Y., Tang, C.M., and Exley, R.M. (2015) Sugar coating: Bacterial protein glycosylation and host-microbe interactions. *Trends Biochem Sci* **40**: 342–350.

Thoden, J.B., Wohlers, T.M., Fridovich-Keil, J.L., and Holden, H.M. (2001) Molecular Basis for Severe Epimerase Deficiency Galactosemia. *Journal of Biological Chemistry* **276**: 20617–20623.

Underhill, S.A.M., Shields, R.C., Burne, R.A., and Hagen, S.J. (2019) Carbohydrate and PepO control bimodality in competence development by Streptococcus mutans. *Mol Microbiol* **112**: 1388–1402.

Underhill, S.A.M., Shields, R.C., Kaspar, J.R., Haider, M., Burne, R.A., and Hagen, S.J. (2018) Intracellular Signaling by the comRS System in Streptococcus mutans Genetic Competence . *mSphere* **3**: 1–21.

Valguarnera, E., Kinsella, R.L., and Feldman, M.F. (2016) Sugar and Spice Make Bacteria Not Nice: Protein Glycosylation and Its Influence in Pathogenesis. *J Mol Biol* **428**: 3206–3220.

Wang, S., Huang, X., Sun, D., Xin, X., Pan, Q., Peng, S., *et al.* (2012) Extensive Crosstalk between O-GlcNAcylation and Phosphorylation Regulates Akt Signaling. *PLoS One* **7**: e37427.

Wilson, W.W., Wade, M.M., Holman, S.C., and Champlin, F.R. (2001) Status of methods for assessing bacterial cell surface charge properties based on zeta potential measurements. *J Microbiol Methods* **43**: 153–164.

Yarmola, E., Ishkov, I.P., Cologna, N.M. di, Menashe, M., Whitener, R.L., Long, J.R., *et al.* (2022) Amyloid Aggregates Are Localized to the Nonadherent Detached Fraction of Aging Streptococcus mutans Biofilms. *Microbiol Spectr* e0166122.

Yi, L., Fu, M., Shao, Y., Tang, K., Yan, Y., and Ding, C.-F. (2022) Bifunctional super-hydrophilic mesoporous nanocomposite: a novel nanoprobe for investigation of glycosylation and phosphorylation in Alzheimer's disease. *J Chromatogr A* **1676**: 463236.

Zamakhaeva, S., Chaton, C.T., Rush, J.S., Ajay Castro, S., Kenner, C.W., Yarawsky, A.E., *et al.* (2021) Modification of cell wall polysaccharide guides cell division in Streptococcus mutans. *Nat Chem Biol* **17**: 878–887 https://www.nature.com/articles/s41589-021-00803-9.

Zhu, F., Zhang, H., and Wu, H. (2015) Glycosyltransferase-mediated Sweet Modification in Oral Streptococci. *J Dent Res* **94**: 659–65.

Zhu, L., Kreth, J., Cross, S.E., Gimzewski, J.K., Shi, W., and Qi, F. (2006) Functional characterization of cell-wall-associated protein WapA in Streptococcus mutans. *Microbiology (N Y)* **152**: 2395–2404.

Zhu, X., Liu, D., Singh, A.K., Drolia, R., Bai, X., Tenguria, S., and Bhunia, A.K. (2018) Tunicamycin Mediated Inhibition of Wall Teichoic Acid Affects Staphylococcus aureus and Listeria monocytogenes Cell Morphology, Biofilm Formation and Virulence. *Front Microbiol* **9**.

APPENDIX B

ADDITIONAL DATA AND FUTURE DIRECTIONS

Interactions between hIntL-1 and other bacteria

To-date, a number of bacterial species have been shown to be recognized by hIntL-1 (*Streptococcus, Vibrio, Helicobacter, Escherichia, Listeria, Staphylococcus*, etc.), the vast majority of which are considered to be human pathogens. The characteristics of this lectin, both in the literature and detailed in this study, suggest a more generalized role for hIntL-1. Lectins typically associated with roles in pathogen surveillance are regulated in their release and expression, and often perform highly specialized methods of direct bacterial killing. However, hIntL-1 demonstrates a broad binding specificity, absence of a controlled regulatory mechanism and an indirect method of bacterial killing. hIntL-1 recognizes excocyclic-1,2-diols, which are commonly found in bacterial surface glycans (including β-galactofuranose, glycerol-1-phosphate, KO and KDO (102)), and the lectin is constitutively expressed in a variety of human tissues with no known detailed mechanism of regulation (110, 115) (although expression levels have, in some cases, been shown to be affected by inducers such as IL-13). In addition, hIntL-1 binding causes agglutination that leads to increased phagocytic uptake, rather than direct killing. Combined, these features suggest a role for hIntL-1 in the sequestration of microbes and host without displaying specificity for pathogens over commensals. We therefore investigated the binding of hIntL-

1 to one of the major human gut commensals, *Bacteroides fragilis*. The CPS of *B. fragilis* contains the ligand for hIntL-1 in the form of a β-galactofuranose moiety, while the capsule of *B. vulgatus* not display exocyclic-1,2 diols and served as a control (Figure B.1A) (291). In a round bottom 96-well assay, the presence of hIntL-1 leads to the agglutination of *B. fragilis*, but not *B. vulgatus* (Figure B.1B). Binding, as well as agglutination of *B. fragilis* could also be visualized by fluorescent microscopy (Figure B.1C). This finding verifies the recognition of commensal species by hIntL-1 in addition the binding of human pathogens. When assessed for killing by PBMCs, the presence of hIntL-1 did not affect the survival of either *B. fragilis* or *B. vulgatus* (Figure B.1D). To further investigate the effect of hIntL-1 on commensals like *B. fragilis*, a phagocytosis assay with neutrophils could be done, since those immune cells showed a major killing effect when *S. pneumoniae* was bound by hIntL-1. While no killing mechanism for *B. fragilis* in the presence of hIntL-1 was observed, the binding of this commensal supports our hypothesis that hIntL-1 functions as a lectin with universal microbial surveillance potential. While *B. fragilis* is a commensal of the human intestine, commensals of other mucosal niches, such as the oral cavity or the stomach, could also be evaluated in their interaction with hIntL-1 to gain a deeper insight into the role of hIntL-1.

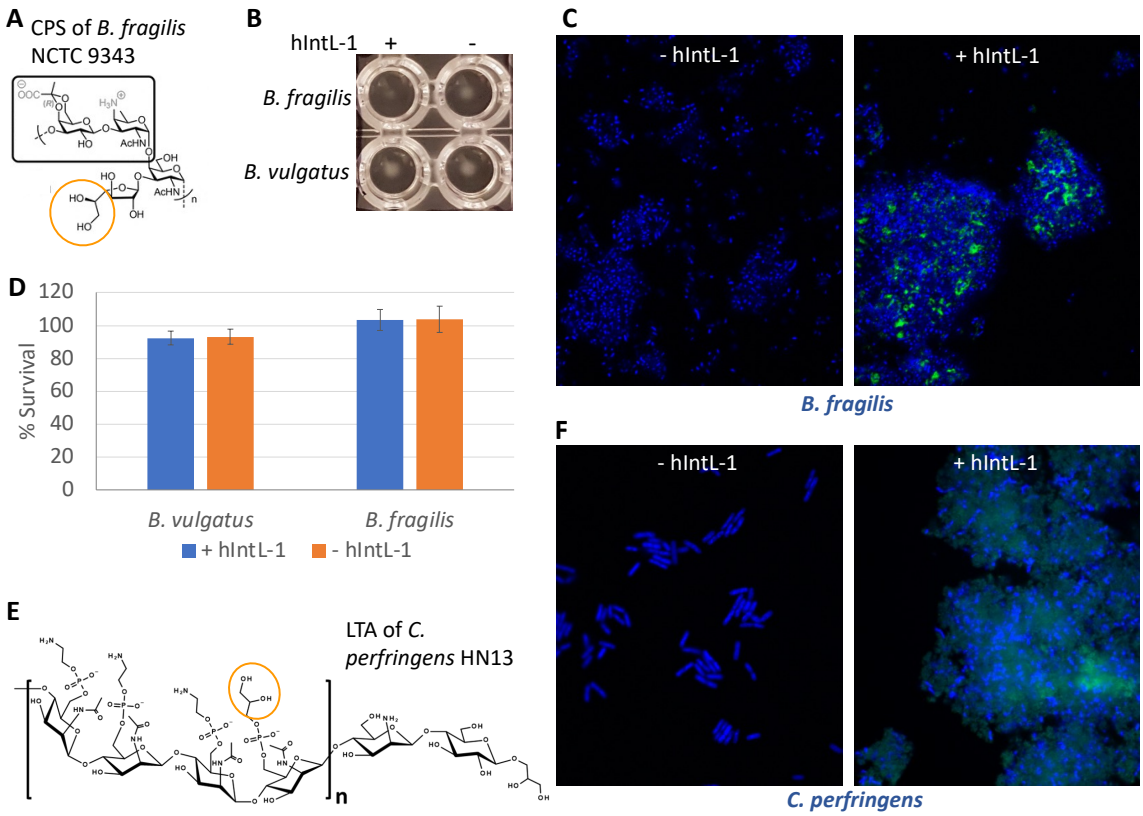


Figure B.1: hIntL-1 binds to *B. fragilis* and *C. perfringens*. The methods for all experimental procedures are described in Chapter 2. **(A)** CPS structure of *B. fragilis* NCTC 9343. **(B)** Agglutination assay of *B. fragilis* and *B. vulgatus* in the absence and presence of hIntL-1 (45 µg/ml). **(C)** Fluorescent microscopy of Alexa Fluor (green) labeled hIntL-1 binding to *B. fragilis*. **(D)** Opsonophagocytosis assay of *Bacteroides* with PBMCs in the absence and presence of (45 µg/ml) hIntL-1. **(E)** LTA structure of *C. perfringens* HN13. **(F)** Fluorescent microscopy of Alexa Fluor labeled hIntL-1 (15 µg/ml) binding to *C. perfringens* H13.

While both *S. pneumoniae* and *B. fragilis* display their hIntL-1 target epitopes within their CPS, many bacteria display other surface glycans that contain exocyclic-1,2-diols, such as teichoic acids or LPS. As a proof of principle, we showed binding of hIntL-1 to *C. perfringens*, which contains a glycerol-1-phosphate moiety in its LTA (Figure B.1 E-F), or other exopolysaccharides. While this result suggests that hIntL-1 can recognize its ligand in various carbohydrate structures, future studies could focus on further defining limitations to exocyclic-1,2-diol recognition. This binding epitope is not only found in

different surface glycans but is present at variable distances from the cell surface, and these factors are likely to impact the ability of the lectin to establish contact, due to steric interference by other sugars. For example, the presence of the target glycan on a capping structure might not be sufficient for full bacterial agglutination vs a ligand that is present in each repeating unit. A study in this area could be approached by designing and testing various surface glycan biosynthesis/assembly mutants within a single organism. *B. fragilis* could serve as the model for mutants within the CPS, which displays a β-galactofuranose, as well as within the LPS, which contains KDO as the anchor to lipid-A. Due to the three dimensional and multilayered envelop of surface polysaccharides that surround every bacterium, binding studies with such mutants would provide information on the binding limitations of hIntL-1 that cannot be achieved with glycan arrays.

Further characterization of the Pgf machinery

While the studies presented here characterized the UDP-GlcNAc/GalNAc 4-epimerase PgfE, little is known about the function of the remaining Pgf proteins. PgfM1 and PgfM2 have been predicted to be membrane proteins and this has been verified in the case when heterologously expressed in *E. coli.* In the case of PgfM2, we can hypothesize certain aspects of its function from previous observations. In the absence of PgfM2, Cnm was shown to possess a molecular weight similar to that of WT Cnm, indicating either an indirect role of PgfM2 during glycosylation or the involvement of PgfM2 in the later steps in the biosynthesis of this post-translational modification. In accordance with this, MS analysis of Cnm from a *ΔpgfM2* background showed a decrease in glycosylated residues, but the presence of additional phosphorylation. Further, the use of SBA demonstrated the

presence of GalNAc residues in both WT and *ΔpgfM2* lysates, but not in the other *pgf* mutants. Together, this suggests that PgfM2 is not involved in the attachment of GalNAc, and more likely to assist in the addition of GlcNAc, which constitutes the second glycan moiety present on Pgf target proteins based on some preliminary compositional analysis data. For both PgfM1 and PgfM2, a more detailed characterization of their substrates and functions during protein glycosylation is still needed.

PgfS is predicted to be a glycosyltransferase and shows high similarity with the enzyme involved in glucose transfer, RgpI, in *S. mutans.* This prediction, however, does not encompass the glycan moiety that is being transferred (i.e., GlcNAc vs. GalNAc), or the preference for lipid- or UDP-activated sugars. We have executed preliminary analysis for the glycan specificity of PgfS using the purified recombinant enzyme in a UDP-Glo assay. In the assay, a glycosyltransferase is combined with a glycan acceptor and UDP-sugar substrate, after the reaction UDP-detection reagent is added to the sample and UDP release is measured spectrophotometrically, correlating to glycan transfer. The ability of PgfS to transfer glycan moieties onto either free sugars or recombinant truncated Cnm was assessed and showed a preference towards Cnm as a protein acceptor over free sugars (Figure B.2A-C). The specificity of PgfS towards different UDP-activated glycans was less clear: while UDP-Glc demonstrated the lowest UDP-production, the produced luminescence in all UDP-sugar-containing reactions approached the detection limit and does not allow for a clear interpretation of the data. The same was true for overnight reactions using higher enzyme and acceptor concentrations, and assessment of these by Coomassie stained SDS-PAGE and anti-Cnm western blot did not reveal any noticeable shifts in the molecular weight of Cnm (Figure B.2D-F).

For future characterizations of PgfS, as well as PgfM1 and PgfM2, the UDP-Glo assay remains a valuable tool for the analysis of substrate specificity. The major limitation of the assay, however, is that it only allows for the use of UDP-activated sugars, and it has yet to be assessed whether the Pgf system utilizes lipid-activated substrates.

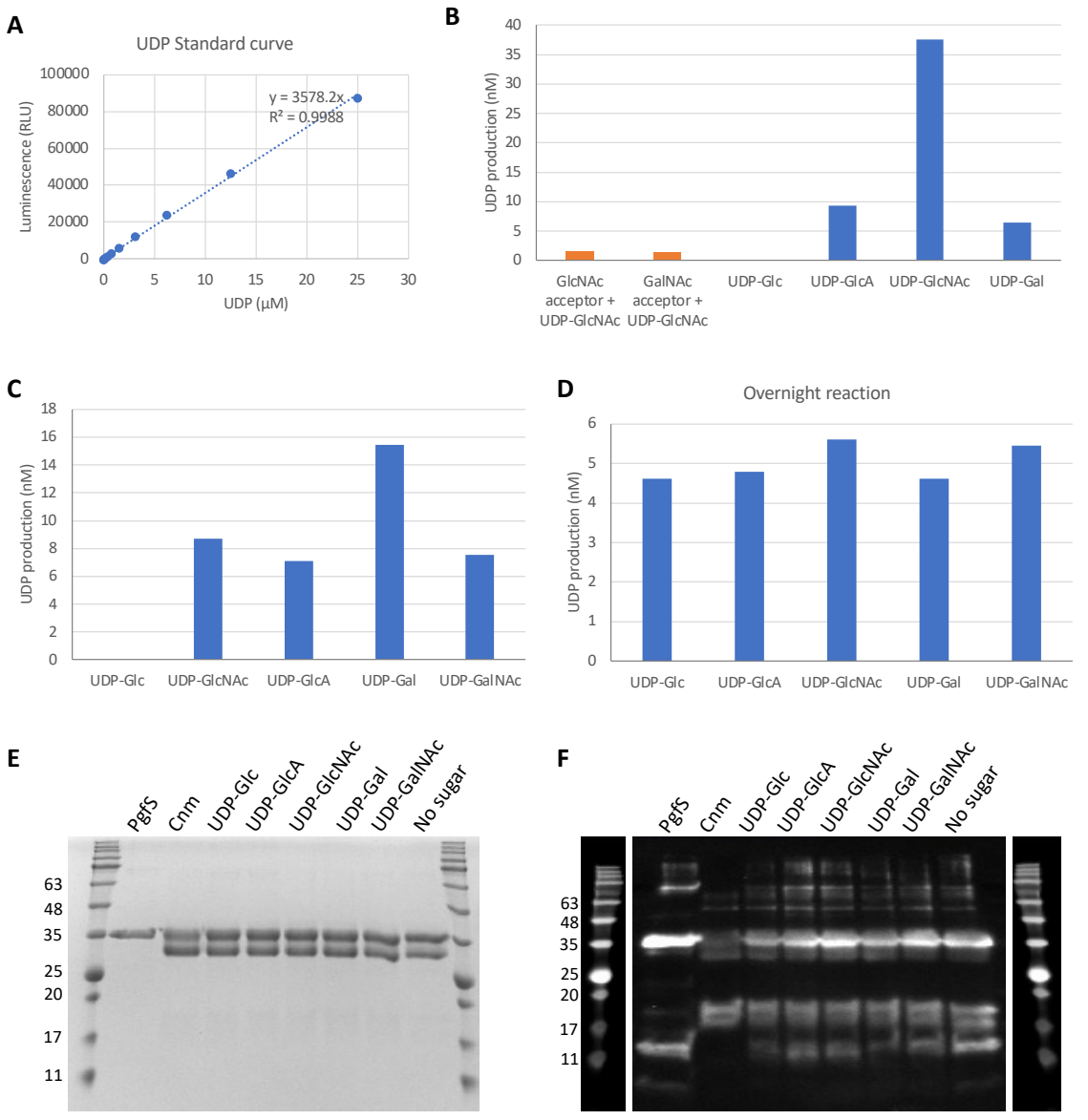


Figure B.2: Assessment of the substrate specificity of PgfS using the UDP-Glo assay. PgfS and truncated Cnm were recombinantly expressed in *E. coli* and purified through Nickel-NTA. **(A)** Exemplary UDP standard curve achieved with UDP-Glo assay. **(B)** 200 nm PgfS (2 days post purification) were combined with 1 µg Cnm or monosaccharides (GlcNAc or GalNAc, shown in orange) as acceptor. 100 mM HEPES buffer supplemented with 150 mM NaCl and 10 % glycerol at pH 7.5 was used as the assay buffer and different

UDP-sugar (100 mM) substrates were added. In all UDP-Glo assays the enzymatic reaction was run for 1 h at 37 °C in a total volume of 17.5 µl, followed by a 20 min cool down of the samples to RT before the UDP-Glo assay was performed, according to manufacturer's instructions. **(C)** Enzyme concentration of PgfS (3 days post-purification) was increased to 400 ng and 10 mM $MgCl_2$ were added. **(D-F)** Double enzymatic reactions were prepared at 4 µg PgfS (2 weeks post purification) and 12 µg Cnm to assess the glycosylation reaction simultaneously by UDP-Glo assay and western blot. The reactions were run overnight at 37 °C in the presence of 10 mM MgCl, 10 mM MnCl, 10 mM $CaCl_2$ before the samples were split for further assessment. The UDP-Glo assay was done as usual with a total of 1 µg Cnm per sample **(D)**. The remainder of the samples were separated by SDS-PAGE (5 µg Cnm/well), transferred onto a nitrocellulose membrane, and probed with an anti-Histidine antibody to detect a potential shift in molecular weight of the Cnm **(E-F)**.

Another interesting direction for future studies, is the similarity that was found between PgfS and RgpI through computational analysis. RgpI is one of the glucosyltransferases/regulators that attaches glucose residues to the poly-rhamnose backbone in a serotype-specific manner. Phenotypical studies on the *pgf* deletion mutants have repeatedly shown that the *ΔpgfS* strain behaves in a distinctly different/opposing manner to the other *pgf* mutants. Namely, deletion of *pgfS* had less of an impact on biofilm formation and genetic competence and resulted in increased negative surface charge compared to other *pgf* mutants. These effects could be explained through two observations that have been made for PgfS during our studies: (1) PgfS and RgpI execute simmilar functions within the cell and one enzyme has the potential to compensate for the loss of the other. (2) Preliminary data has shown the absence of glycosylation, but presence of phosphorylation on Cnm in a *ΔpgfS* background, and this switch in post-translational modification could be the cause for the observed phenotypic changes in this mutant. These two rationales, however, do not need to be mutually exclusive. To begin examining these mechanisms, a *ΔpgfSΔrgpI* double mutant could be evaluated for biofilm formation, genetic competence, and surface charge. If the loss of both glycosyltransferases causes a phenocopying of the other *pgf* machinery mutants, it is likely that RgpI is able to

compensate for the loss of PgfS. To assess the effects of phosphorylation, Cnm from all *pgf* mutants should be analyzed for the presence of phosphate groups in place of glycosylation via MS. This will show whether the phosphorylation of Pgf target proteins is specific to the loss of PgfS, or present in all *pgf* mutants. The latter can further be verified through anti-phosphothreonine western immunoblotting. Regardless of the outcome of the described preliminary experiments, the connection between PgfS and RGP, as well as the switching between two post-translational modifications, promises to be an interesting area of future exploration.