Making and Breaking Streptococcal Rhamnose Polysaccharides

Samantha L. van der Beek

Making and Breaking Streptococcal Rhamnose Polysaccharides
PhD thesis, Utrecht University, The Netherlands

Cover image: Job Fermie and Samantha van der Beek
Cover design and layout thesis: Guus Gijben (Proefschrift-AIO) and Samantha van der Beek
Printing: ProefschriftMaken
Photograph of the author (page 178) by Matthijs Hoebe

About the cover and chapter pages: Scanning electron microscopy images of either *Streptococcus mutans* (cover, chapters 4 to 7) or *Streptococcus pyogenes* (chapters 1 to 3) showing defects in streptococcal rhamnose polysaccharide biosynthesis. The hexagons represent the cell wall carbohydrates.

ISBN: 978-90-393-7064-3

Printing of this thesis was financially supported by the Netherlands Society of Medical Microbiology (NVMM) and the Royal Netherlands Society for Microbiology (KNVM); Infection and Immunity Center Utrecht; the University Medical Center Utrecht (UMCU), Quorum Technologies and Thijs! Bedrijfsadministratie en optimalisatie.

Making and Breaking Streptococcal Rhamnose Polysaccharides

Het maken en kraken van rhamnose polysachariden van streptokokken
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op

donderdag 8 november 2018 des middags te 12.45 uur

door

Samantha Lisette van der Beek
geboren op 16 januari 1990
te Dordrecht

Promotor
Prof. Dr. J. A. G. van Strijp

Copromotor
Dr. N. M. van Sorge

Table of content


1

1

General introduction

Samantha L. van der Beek

INTRODUCTION ON STREPTOCOCCAL RHAMNOSE POLYSACCHARIDES

The streptococcal cell wall

The bacterial surface is highly glycosylated. Cell wall glycopolymers, such as cell wall polysaccharides, teichoic acids, and capsular polysaccharides, are critical for bacterial physiology and for interaction with the host, including defense against our immune system [(1)]. Classically, two classes of bacteria are distinguished based on their cell wall architecture: Gram-positive and Gram-negative bacteria. Gram-negative bacteria are characterized by a thin layer of peptidoglycan in between inner and outer membranes, the latter embedding lipopolysaccharides, which are major virulence factors for these bacteria. Gram-positive bacteria lack an outer membrane and instead express a thick layer of peptidoglycan decorated with wall teichoic acid (WTA), in addition to lipoteichoic acid (LTA), a structurally similar molecule that is embedded in the inner membrane. In addition, both classes of bacteria can express capsular polysaccharides, which forms a protective cloak around the bacteria.

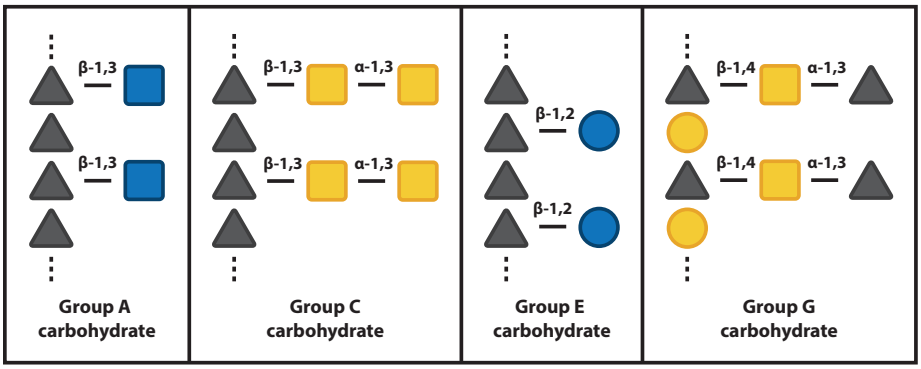
Textbooks describe WTA as a major cell wall component of Gram-positive bacteria representing as much as 60% of the total cell wall mass. However, many Gram-positive species from the genera *Streptococcus*, *Enterococcus* and *Lactococcus*, lack the expression of WTA and instead express glycopolymers, which contain large amounts of the sugar rhamnose [(2)]. Before genome sequencing was used to classify bacterial species, these bacteria were all designated to the genus *Streptococcus* due to their similar spherical or ovoid shaped cells arranged in chains or pairs [(3)]. Rebecca Lancefield developed a streptococcal classification system based immune reactivity of the so-called carbohydrate C [(4)]. Further research elucidated the structures of these C carbohydrates, which also became known as the Lancefield group carbohydrates [(5)](Fig 1A). However, this serological grouping system became confusing when it became apparent that different species can express the same Lancefield group carbohydrate or that different strains from the same species can express different Lancefield group carbohydrates. Also, the classification system was developed for hemolytic streptococci, excluding many other streptococcal species that are now known to express similar cell wall carbohydrates. These include *Streptococcus mutans*, whose cell wall polysaccharides are used for serotype determination [(6)] (Fig 1B) and *Streptococcus uberis* and *Streptococcus sobrinus*, whose cell wall carbohydrates currently have no application for classification or serotyping (Fig 1C). Regardless of their practical application in early diagnostics, we now know that many of these streptococcal polysaccharides share structural, compositional and functional characteristics. For this reason, I suggest to designate this family of cell wall glycopolymers Streptococcal Rhamnose Polysaccharides (SRPs).

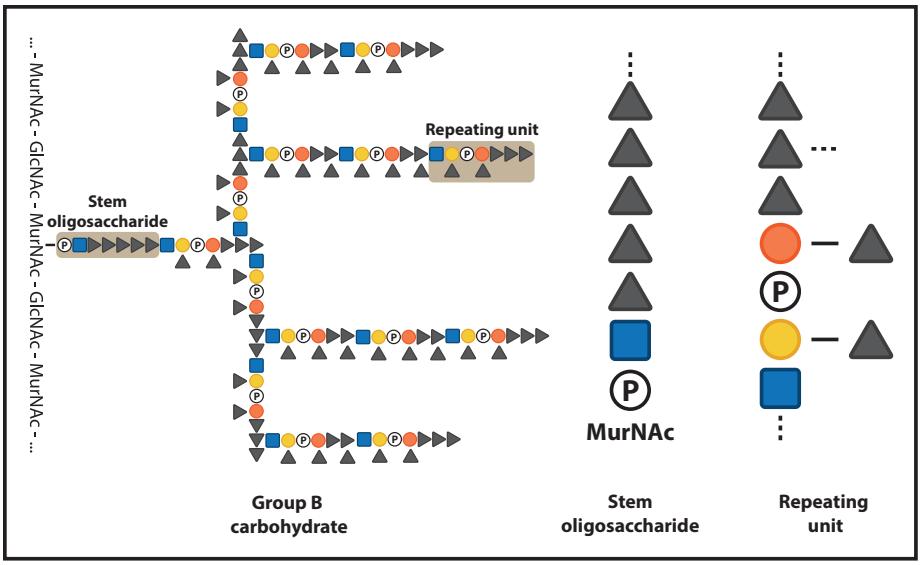
Some SRPs are known to contribute to virulence of human and animal pathogens similar to WTA [7-12]. Further similarities with WTAs include their role in bacteria-phage interactions and maintenance of cell wall shape. In addition to their importance for bacterial physiology and infection, their high density and surface exposed nature makes both SRPs and WTAs excellent therapeutic targets. However, in contrast to WTA, we currently have a limited knowledge of SRP biosynthesis and their role in pathogenesis. Functional studies on WTAs have focused on the Gram-positive model species *Bacillus subtilis* and *Staphylococcus aureus* due to their ease of genetic manipulation and relevance to human disease, respectively. For similar reasons, SRP biosynthesis and virulence studies have predominantly been carried out in *S. mutans, Streptococcus pyogenes* (Group A *Streptococcus*) and *Streptococcus agalactiae* (Group B *Streptococcus*). In this introduction, I aim to summarize the common themes in SRP biosynthesis and to discuss their potential as therapeutic targets through bioinformatics analysis and comparison with the extensively studied WTA biosynthesis pathway.

Structural comparison of SRPs

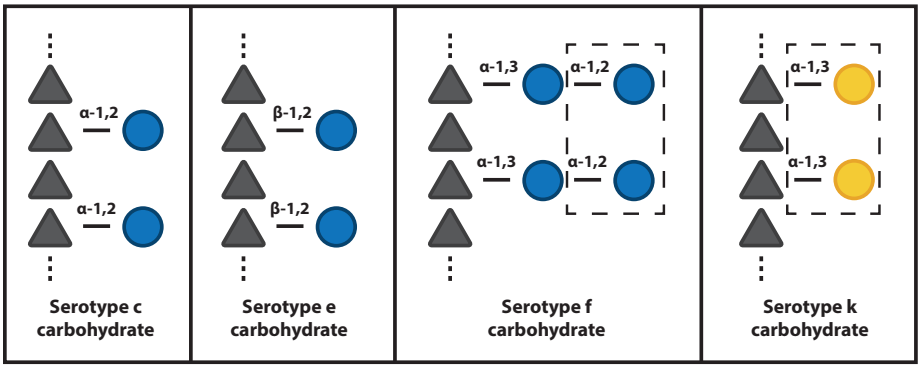
To date, SRP structures have been characterized from *S. pyogenes* [13], *S. agalactiae* [14-16], *Streptococcus porcinus* [18, 19], *Streptococcus dysgalactiae* subsp. *equisimilis* [13, 20], *Streptococcus dysgalactiae* subsp. *dysgalactiae* [17], *S. mutans* [21-25], *S. uberis* [26] and *S. sobrinus* (formerly known as *S. mutans* serotypes d and g) [27-29] and display several common features (Fig 1). In general, SRPs are uncharged glycopolymers that are mainly composed of the sugars rhamnose, glucose, *N*-acetylglucosamine (GlcNAc), galactose and *N*-acetylgalactosamine (GalNAc). However, rhamnose is the only monosaccharide that is common to all SRPs. Most SRPs are composed of an α-1,2/α-1,3 linked polyrhamnose (rhamnan) backbone, which is decorated with different arrangements of aforementioned glycans (Fig 1). An exception is the group G carbohydrate, which contains a backbone of $[\alpha\text{-}1,2\ \text{galactose}\ \alpha\text{-}1,3\ \text{rhamnose}]_n$ and a GalNAc-rhamnose side chain linked to the rhamnose backbone residue (Fig 1A) [20]. In contrast to the linear structures of aforementioned SRPs, the group B carbohydrate (GBC) is composed of two oligosaccharides, a stem oligosaccharide and a repeating unit that branches into a multiantennary structure (Fig 1A) [15, 16]. The stem oligosaccharide resembles the rhamnan backbone of other SRPs, although the rhamnose residues are connected through α-1,3 linkages [16]. So far, the GBC is the only SRP structure known to contain negatively charged phosphate residues, which has raised the suggestion that GBC and possibly other SRPs are homologs of WTAs.

A. Lancefield group carbohydrates from various streptococci





B. Serotype-specific carbohydrates from *S. mutans*



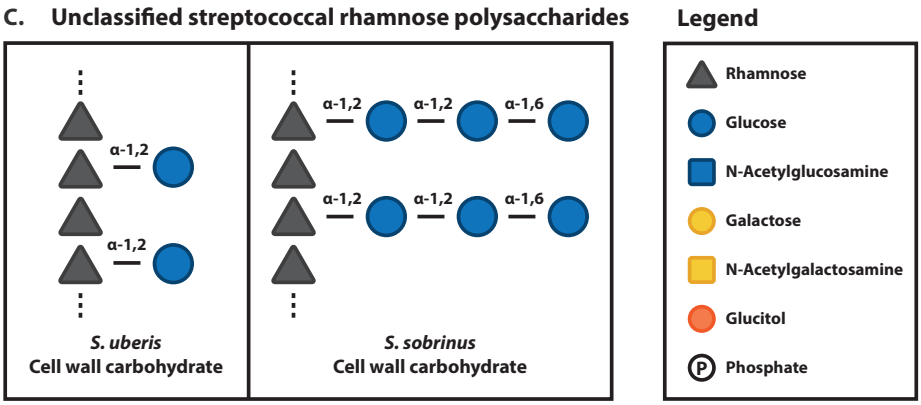


Figure 1. Schematic overview of streptococcal rhamnose polysaccharide (SRP) structures. Structural composition of **(A)** Lancefield group carbohydrates, **(B)** *S. mutans* serotype-specific carbohydrates, **(C)** unclassified SRPs. Except for the group B and G carbohydrates, these glycopolymers share a common α-1,2/α-1,3 linked polyrhamnose backbone and are uniquely distinguished by the composition and linkage of their side chains. Structures were determined in *Streptococcus pyogenes* J17A4/ATCC 12385 (group A) [13], *Streptococcus agalactiae* SS1073 [14] and 090/ATCC 12386 (group B) [15, 16], *Streptococcus dysgalactiae* subsp. *equisimilis* C74/ATCC 12388 and *Streptococcus dysgalactiae* subsp. *dysgalactiae* 2023/ATCC 43078 (group C) [13, 17], *Streptococcus porcinus* K129/ATCC 12390 (group E) [18, 19], *Streptococcus dysgalactiae* subsp. *equisimilis* D166B/ATCC 12394 (Group G) [20], *S. mutans* strains GS-5 [21] and 10449 (serotype c) [22], LM7 (serotype e) [23], OMZ175 (serotype f) [22, 24, 25], OM42X, OM88X and OM98X (serotype k) [22], *S. uberis* 233/ATCC 13387 [26] and *S. sobrinus* B13 and 6715-T_2 [27]. Sugar residues indicated by a dashed box were recently identified during structure redetermination [22].

WTAs are anionic glycopolymers that are covalently attached to peptidoglycan. Their exact chemical composition varies among organisms and even within species [30-32]. Most well studied are the Type I WTAs, which are composed of a poly-ribitolphosphate $(RboP)_n$ or poly-glycerolphosphate $(GroP)_n$ backbone with modifications such as glycosylation and *D*-alanylation [33, 34]. The presence of D-ala residues neutralizes the negative charge of the abundant phosphates in the backbone [33]. Depending on the strain and species, WTA glycopolymers size is quite uniform and can reach up to 50 polymer units in length. Several genes have been suggested to be involved in regulation of WTA polymer chain length, including *tarF*, *tarE*, *tarK* and *tarL*, although the mechanism are not completely resolved [35-39].

Function of SRPs in bacterial physiology

SRPs are linked to peptidoglycan and encompass about 20-60% of the cell wall mass depending on the strain, species and culture conditions [27, 29, 40-44]. Many genes in the

SRP biosynthesis pathway have been identified as essential or critical for survival as determined by generation and analysis of mutant transposon libraries for *S. pyogenes*, *S. mutans*, *S. equi*, *S. agalactiae* and *S. uberis* [45-50]. These essential genes are required for biosynthesis of sugar precursors, initiation of SRP production, elongation of the rhamnan backbone or transport of the SRP across the membrane. However, mutant transposon libraries present competitive conditions, resulting in quick loss of non-fit but viable mutants. Therefore, some 'essential' genes can be mutated in isolation or application of compensating growth conditions, such as the addition of sucrose or magnesium. Indeed, these mutants often display severely attenuated growth, cell separation and cell division defects, misplaced septa, resulting in morphological aberrancies [44, 51, 52]. In the case of *S. agalactiae*, loss of the GBC resulted in incomplete peptidoglycan polymerization and mislocalization of a peptidoglycan hydrolase important for cell division [44]. Deletion of genes that are important for side chain formation resulted only in minor morphological changes, generally elongation of streptococcal chains as a result of reduced cell separation [10, 53]. These results suggest that SRPs are involved in cell division possibly through the recruitment of cell division or cell separation proteins, such as peptidoglycan hydrolases or autolysins. In *S. mutans*, other roles for SRPs in bacterial physiology included maintenance of the proton gradient across the membrane, stress resistance, resistance to acid challenge, biofilm formation and adherence to hydroxyapatite (dental) surfaces [52, 54, 55].

Role of SRPs in streptococcal mediated disease

The importance of SRPs in bacterial physiology also impacts host colonization, immune evasion and virulence. *S. mutans* is considered the main etiological agent of dental caries through biofilm-mediated dysbiosis of the oral microbiome. In addition, *S. mutans* can also cause invasive disease, predominantly endocarditis [6, 56-61]. *S. mutans* host colonization of the oral cavity is affected by the serotype-specific SRPs through their impact on biofilm formation and adherence to dental tissue. Furthermore, reduced *in vitro* bacterial fitness, resulting from mutations in SRP biosynthesis genes, often enhances immune clearance. Indeed, *S. mutans* lacking the serotype c, e or f SRPs were more susceptible to phagocytosis by human neutrophils and killing by antimicrobial compounds [7, 8]. Consequently, lack of SRP expression resulted in reduced virulence of *S. mutans* in several *in vivo* models, including a mouse model of dental caries, an infective endocarditis rat model and a waxworm *Galleria mellonella* infection model [46, 54, 61]. The contribution of SRPs to the pathogenesis of infective endocarditis may be related to the observation that SRP-specific IgG antibodies induce platelet aggregation [62], which plays an important role in development of infective endocarditis [63].

Genes implicated in SRP side chain formation are not essential under laboratory growth conditions, however these genes are important for certain aspects of host infectivity. In *S. mutans*, mutants with a transposon insertion in side chain biosynthesis genes are lost from the library in aforementioned *in vivo* mouse model of dental caries [46]. In contrast, a mutant with reduced SRP glucose side chains did not alter *S. mutans* virulence in a systemic infection model, i.e. an *in vivo* infective endocarditis rat model [61]. *S. pyogenes* mutants carrying a transposon insertion in genes involved in GAC GlcNAc side chain biosynthesis show reduced fitness in human saliva and a mouse model for skin and soft tissue infection, but not in human blood [9, 64, 65]. The last observation does not correspond to findings with a defined knockout in *gacI*, encoding a glycosyltransferase critical for GAC GlcNAc side chain formation. In this study, the *gacI* mutant was shown to be less fit in human blood compared to WT GAS [10]. An explanation for the discrepancy between the two studies could be the anticoagulant used to collect human blood. Overall, the *gacI* mutant was less virulent in both a mouse and a rabbit infection model and displayed increased sensitivity to neutrophil killing, platelet-derived antimicrobials in serum, and the cathelicidin antimicrobial peptide LL-37 [10, 53]. However, the effect of *gacI* mutation depends on the specific *S. pyogenes* background, since different serotypes carrying the same mutation did not behave similar in both *in vitro* and *in vivo* assays. This most likely reflects the interaction of the GAC with other cell wall components, which influence the overall susceptibility to immune clearance and infection [53]. Indeed, *S. pyogenes* isolates, which are passed numerous times through mice, occasionally lose the expression of the GAC GlcNAc side chain and are referred to as A-variant strains. This implies that for mouse infection loss of the GlcNAc side chain can occur without any detrimental consequences [66-68]. Overall, it is evident that SRPs play a major role in host colonization, immune evasion and pathogenesis, even though the precise molecular mechanisms of their virulence potential remains to be elucidated.

MAKING SRPs

The SRP biosynthesis pathway

Even though the cell wall of Gram-positive bacteria harbors many structurally diverse cell wall glycopolymers, such as teichoic acids, SRPs, capsular polysaccharides and peptidoglycan, common themes in their biosynthesis mechanisms can be discerned (reviewed in [1]). In general, glycopolymer biosynthesis proceeds in six basic steps (Fig. 2); 1) **Biosynthesis of activated sugar precursors**, which are needed as building blocks for glycopolymer biosynthesis, 2) **Initiation** of glycopolymer biosynthesis through activation of a polyprenoid carrier, often undecaprenyl-phosphate (Und-P), on the cytoplasmic side of the membrane, 3) **Elongation** by sequential addition of activated

sugar precursors to form the glycopolymer on the polyprenoid linker, 4) **Transport** of polyprenoid-linked precursors, either repeating units or the complete glycopolymer, across the membrane by ABC transporters or 'flippases' [69], 5) **Additional modifications** to the glycopolymer that can occur after translocation, and 6) **Linkage** of the glycopolymer to peptidoglycan [70, 71]. In some cases, polymerization of translocated repeating units into a mature glycopolymer occurs between steps 4 and 6 by a dedicated polymerase. It is currently unknown, whether linkage to peptidoglycan occurs before or after or even during additional modifications to the glycopolymer (step 5).

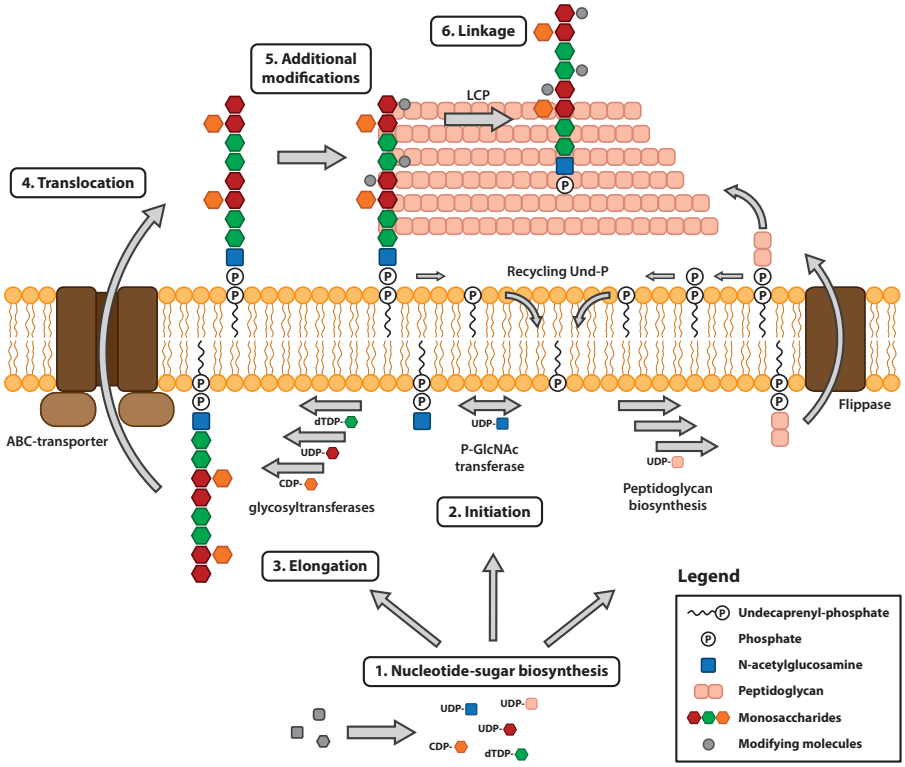


Figure 2. General steps in glycopolymer and peptidoglycan biosynthesis. Despite structural and compositional diversity, glycopolymer biosynthesis is quite conserved and proceeds in six basic steps: **(1)** Biosynthesis of activated sugar precursors, which are the building blocks of the glycopolymer, **(2)** initiation of biosynthesis through activation of a lipid carrier on the cytoplasmic side of the membrane, **(3)** elongation of the glycopolymer on the polyprenoid linker by sequential addition of activated monosaccharides by glycosyltransferases, **(4)** translocation of polyprenoid-linked precursors across the membrane by ABC-transporters, **(5)** additional modifications of the glycopolymer, for example *D*-alanylation of wall teichoic acid, **(6)** linkage of the glycopolymer to peptidoglycan by members of the LytR/CpsA/Psr (LCP) protein family. It is uncertain whether additional modifications mentioned in step five are added before or after anchoring to peptidoglycan. GlcNAc, *N*-acetylglucosamine; Und-P, undecaprenyl-phosphate.

For Gram-positive bacteria, the WTA biosynthesis pathway is arguably the best studied and well characterized glycopolymer biosynthesis pathway. Indeed, many of the WTA biosynthesis genes have been identified in *S. aureus* and *B. subtilis* permitting dissection of the different biosynthetic steps by genetic mutagenesis [72, 73]. In addition, the intracellular steps of WTA biosynthesis have even been successfully reconstituted *in vitro*, highlighting interesting differences between *S. aureus* and *B. subtilis* WTA biosynthesis [74, 75]. Compared to WTA biosynthesis, research on SRP biosynthesis has been limited. Below, I will discuss the recent insight into the different steps of SRP biosynthesis and I will draw comparisons with the well-studied WTA biosynthesis pathway. For the purpose of this introduction, the different steps of WTA biosynthesis will only be illustrated through the biosynthesis of $(RboP)_n$ WTA in *S. aureus.*

Step 1. Production of nucleotide sugars

An essential step in the biosynthesis of cell wall glycopolymers is the production of activated sugar precursors; monosaccharides linked to phosphate-nucleotides such as uridine diphosphate (UDP) or deoxythymidine diphosphate (dTDP). Glycosyltransferases use the high energetic state of these nucleotide-sugars to catalyze the transfer of glycans to an acceptor, such as Und-P (step 2: Initiation) or the growing glycan chain (step 3: Elongation). In bacteria, specific nucleotide-sugar combinations are identified and their biosynthesis pathways are highly conserved (Fig. 3A). The most common nucleotide-sugars used for SRP and WTA biosynthesis include dTDP-L-rhamnose, UDP-glucose, UDP-GlcNAc, UDP-galactose, UDP-GalNAc, UDP-*N*-acetylmannosamine (UDP-ManNAc), CDP-ribitol and CDP-glycerol. Often, genes involved in nucleotide-sugar biosynthesis, especially those involved in glycans incorporation into the core glycopolymer structure are essential, as is discussed for select nucleotide sugars in more detail below.

A common residue of both SRPs and WTAs is GlcNAc, required for the initiation of biosynthesis on Und-P. Furthermore, GlcNAc is added as side chain modifications to both the GAC backbone in *S. pyogenes* (Fig 1A)[10, 53] as well as the WTA RboP backbone in *S. aureus* (Fig 3C). In addition, GlcNAc is one of the two major components of peptidoglycan together with *N*-acetylmuramic acid (MurNAc), rendering GlcNAc an essential monosaccharide for bacterial viability and maintenance of proper cell wall architecture. Indeed, all three UDP-GlcNAc biosynthesis genes, GlmS, GlmM and GlmU (Fig 3A), were found to be essential under competitive circumstances in several streptococcal transposon libraries [45, 46, 48, 49, 76]. In some instances, inactivation can be overcome by compensating measures, such as inactivation of *glmS* in *S. mutans* in the presence of exogenous GlcNAc [77]. Similarly, a *gmlS*-deficient *S. equi* subsp.

zooepidemicus strain was only viable in the presence of glucosamine and displayed a mucoid colony morphology [78]. For the second enzyme in the UDP-GlcNAc biosynthesis pathway GlmM, the situation is less straight forward. For example, *S. equi* subsp. *zooepidemicus* expresses two GlmM homologs, HasD and GcaD, since the bacterium also requires expression of UDP-GlcNAc for hyaluronic acid capsule biosynthesis. Although deletion of the individual genes was possible, it was not possible to obtain a *hasD-gcaD* double mutant [78]. Surprisingly, a *glmM*-deficient *S. mutans* strain could also be generated and displayed longer streptococcal chains, reduced growth rate, increased autolysis and attenuated biofilm formation [79]. Similar to *S. equi* subsp. *zooepidemicus*, a homolog might be present elsewhere on the genome, allowing deletion of the *glmM*. Although no functional studies have been performed on GlmU in any streptococcal species, it is noteworthy that this enzyme is bifunctional, catalyzing the last two steps in UDP-GlcNAc biosynthesis (Fig. 3A)[80]. Besides GlcNAc, the WTA anchor in *S. aureus* also contains ManNAc and GroP (Fig 3C). The enzyme UDP-GlcNAc:UDP-ManNAc 2-epimerase, encoded by *mnaA*, is essential for the incorporation of ManNAc in the WTA anchor (Fig 3A, C). However, *mnaA* is not essential for the viability of *S. aureus* under laboratory conditions and an *mnaA* mutant completely lacks WTA in its cell wall [81].

In contrast to MnaA, enzymes needed for biosynthesis of the nucleotide sugars CDP-glycerol (TarD) and CDP-ribitol (TarI and TarJ), which are incorporated in the Type I WTA backbone (Fig 3A), are essential for *S. aureus* viability [82]. Similarly, all four dTDP-L-rhamnose biosynthesis genes *rmlABCD* (Fig. 3A), required for the construction of the rhamnan backbone of many SRPs (Fig 3B), are essential in many streptococcal transposon libraries [45, 46, 48-50]. Under non-competitive conditions however, it is possible to inactivate *rmlA, rmlB, rmlC or rmlD* in *S. mutans* Xc although this has a dramatic impact on growth and morphology [83-85]. The reason why it is possible to mutate rhamnose biosynthesis genes in some streptococcal species but not others is currently not elucidated, but likely relates to overall composition of the cell wall.

Step 2. Initiation

As mentioned, initiation of glycopolymer biosynthesis, including WTAs and SRPs, generally starts with activation of the same lipid precursor Und-P [86]. WTA biosynthesis is initiated by TarO, an UDP-GlcNAc:Und-P *N*-acetylglucosaminyl 1-P transferase, which catalyzes the transfer of GlcNAc to Und-P (UDP-GlcNAc + Und-P → UMP + Und-P-P-GlcNAc) (Fig 3C)[82]. Homologs of TarO have been identified in *S. mutans* (RgpG) (Fig 3B), *S. agalactiae* (GbcO) and *S. pyogenes* (GacO) and have been implicated or proven to be required for SRP biosynthesis [10, 44, 51, 52, 87, 88]. Deletion of *S. mutans rgpG* or *S. agalactiae gbcO* reduced SRP expression, resulting in aberrant

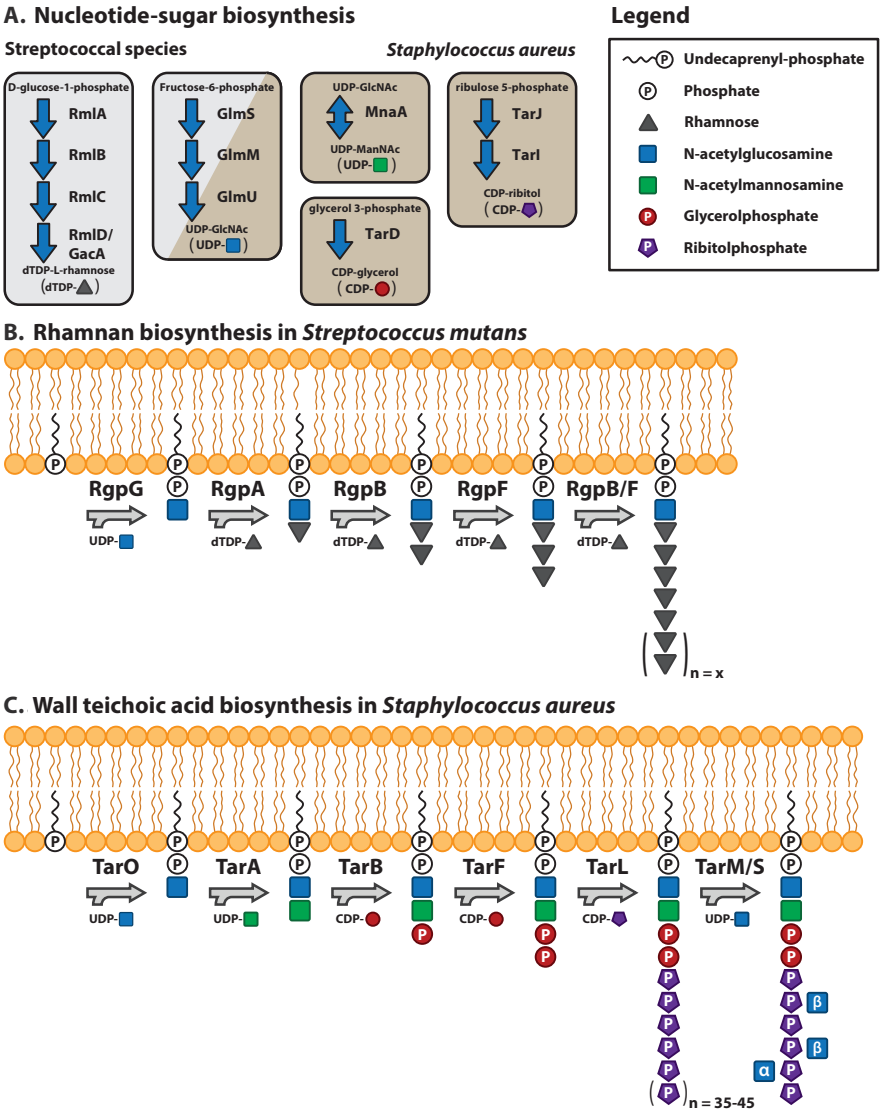


Figure 3. Schematic overview of specific steps in nucleotide biosynthesis, initiation and elongation for rhamnan and WTA biosynthesis. (A) Biosynthesis pathways of essential activated sugar precursors in streptococci (grey) and *S. aureus* (beige). Overview of initiation and elongation steps in **(B)** rhamnan biosynthesis on undecaprenyl-phosphate in *S. mutans* and **(C)** WTA biosynthesis on undecaprenyl-phosphate in *Staphylococcus aureus*.

bacterial morphology, slower growth rates and reduced biofilm formation [44, 51, 52, 87]. Mutant bacteria aggregate at the bottom of culture tubes and appear as long chains of swollen cocci with misplaced septa, suggesting SRPs play important roles in both cell separation and cell division. In addition, the *gbcO*-deficient mutant display reduced peptidoglycan crosslinking, resulting in increased capsule production to cope with the inferred cell wall stress [44, 51]. The *gbcO* mutant phenotype can be replicated by treatment of *S. agalactiae* with tunicamycin, an UDP-GlcNAc:Und-P transferase inhibitor. Similarly, treatment of *S. pyogenes* with tunicamycin resulted in loss of rhamnose in the cell wall, aberrant morphology and attenuated growth, suggesting GacO activates biosynthesis of the GAC through the transfer of P-GlcNAc to Und-P [10]. Indeed, *in vitro* reconstitution of GacO verified the function of this protein as a UDP-GlcNAc:Und-P transferase [88]. It is currently unknown whether all streptococcal species use GlcNAc to initiate SRP biosynthesis on Und-P.

Step 3. Elongation

After initiation of glycopolymer biosynthesis through activation of Und-P, various glycosyltransferases create a glycopolymer by step-wise attachment of monosaccharides. Since SRP biosynthesis genes involved in backbone formation are essential in most streptococci [45-50], extensive studies on the SRP elongation step are currently lacking. To circumvent gene essentiality, Shibata et al. established a heterologous expression system in *Escherichia coli* to study SRP backbone biosynthesis genes of *S. mutans*. In this *E. coli* heterologous expression system, the polyrhamnose backbone is attached as O-antigen to lipopolysaccharide on the outer membrane [89]. Results from these studies suggest that the rhamnan backbone is formed by three rhamnosyltransferases, designated RgpA, RgpB and RgpF, using dTDP-L-rhamnose as nucleotide-sugar precursor (Fig 3B). First, RgpA links rhamnose to the GlcNAc moiety of the Und-P-P-GlcNAc precursor [89]. Most likely, alternate activity of RgpB and RgpF subsequently add rhamnose in an α-1,2 or α-1,3 configuration (Fig 3B) [89]. Homologs of all three rhamnosyltransferases can be identified in streptococci expressing α-1,2/α-1,3-rhamnan containing SRPs with 60-75% protein sequence identity compared to *S. mutans* RgpA, RgpB or RgpF. It was previously suggested that the GAC from *S. pyogenes* must be at least 20 rhamnose residues long [13, 90]. However, it is currently unknown whether chain length of the growing glycopolymer is regulated, for example through addition of a terminating group.

Until recently, it was assumed that the complete SRP, including side chains, was synthesized intracellularly before transport across the membrane. However, Rush et al. discovered that only the rhamnan backbone of the GAC is transported and

the GlcNAc side chains are produced by a separate biosynthesis pathway, which is finally attached extracellularly to the rhamnan backbone by GacL [88]. Unsubstituted polyrhamnose SRP has been identified alongside the side chain substituted SRP in the cell walls of several *S. mutans* strains, *S. dysgalactiae* subsp. *dysgalactiae* and *S. uberis*, suggesting that the two-step biosynthesis system is likely widespread among streptococcal species [17, 22, 26]. In addition, homologs of most genes involved in *S. pyogenes* GlcNAc side chain formation and attachment can be identified in these streptococcal species.

In contrast to SRP biosynthesis, WTA elongation in *S. aureus* has been well studied both structurally and enzymatically. WTA contains an anchor consisting of four non-repeating residues, Und-P-P-GlcNAc-ManNAc-$(GroP)_2$ (Fig 3C), which are subsequently added by the activity of TarA, TarB and TarF [82, 91, 92]. The full length WTA chain of approximately 30-50 RboP is subsequently synthesized by the activity of a single enzyme TarL, which has both priming and polymerase activity (Fig 3C) [74]. In contrast to *S. pyogenes*, *S. aureus* does add side chains intracellularly to the WTA backbone [69, 93]. This strain-dependent modification with α-O-GlcNAc or β-O-GlcNAc, initially noted in the 1960s [94], occurs at the at the C4 hydroxyl group of RboP units through the activity of TarM and TarS, respectively (Fig 3C) [31, 95-97]. It is suggested that glycosylation of the WTA backbone prevents binding of the polymerase protein and that polymer chain length is regulated in this manner [36].

Step 4. Translocation

Transport of WTA across the membrane to the extracellular site occurs via an ABC-transporter consisting of permease (TarG) and ATPase (TarH) proteins (Fig 4B) and constitutes the rate-limiting step in glycopolymer biosynthesis [39]. Interestingly, the TarGH transporter is promiscuous with regard to substrate specificity; WTA composed of either a polyRboP or a polyGroP backbone is processed by the *S. aureus* transporter, suggesting that substrate recognition may be determined by the shared lipid-linked disaccharide unit [93]. This would concur with recent experimental evidence that the Gram-negative homodimeric ABC transporter PglK, which can 'flip' oligosaccharides without apparent substrate selectivity, recognizes the conserved pyrophosphate-linker on which the oligosaccharide is synthesized [98]. Whether transport occurs after or concomitant with WTA backbone polymerization is currently not known.

Very little is currently known regarding SRP transport. However, genes encoding ABC-transporter proteins are found in all SRP biosynthesis gene clusters, suggesting SRP translocation resembles WTA transport (Fig 4A). Two different approaches

were used to identify RgpC and RgpD as the SRP permease and ATPase transporter proteins in *S. mutans*, respectively. First, insertional inactivation of *rgpC* or *rgpD* drastically decreased serotype c carbohydrate in the *S. mutans* cell wall and colonies appeared smaller and more grainy. [85, 89]. Second, deletion of *rgpC* and *rgpD* in the heterologous *E. coli* expression system mentioned previously [89], prevented rhamnan export. Interestingly, the substrate specificity of the transporter seems to be side chain independent as these biosynthesis genes were not expressed in *E. coli*. Heterologous expression of SRP transporters in other *S. mutans* serotypes or even in different streptococcal species could provide more information about substrate specificity. A remarkable difference between WTA TarH and all SRP ATPase proteins is the presence of an extended approximately 150 amino acid C-terminal domain of unknown function (Fig 4A). Using bioinformatics analysis, Cuthbertson et al. identified two phylogenetic groups of ATPase proteins involved in glycopolymer transport harboring such extended C-terminal domains [99]. While ATPase activity is restricted to the highly conserved N-terminal region, more diverse functions have been attributed to the C-terminal domain. Some of these domains could recognize a terminal so-called 'capping residue' on the glycopolymer backbone, while catalytic domains were identified in others. However, blasting of SRP ATPase C-terminal domains, did not provide any clues regarding a possible function for these domains.

Step 5. Additional decorations

Assuming only the SRP rhamnan backbone is transported across the membrane, as is the case for the GAC from *S. pyogenes*, side chain attachment must occur extracellularly, either before or after coupling to peptidoglycan. Considering that most proteins involved in additional modification are membrane bound, it can be assumed that these modifications occur close to the membrane, most likely before coupling to peptidoglycan. In *S. pyogenes*, GlcNAc side chain addition involves the activity of four proteins, GacI-L, which are located at the end of the GAC biosynthesis gene cluster (Fig. 4A) [10, 53, 88]. First, GacI and GacJ catalyze the formation of Und-P-GlcNAc from Und-P and UDP-GlcNAc on the intracellular site [88]. Then, after translocation possibly by the potential flippase GacK, Und-P-GlcNAc is used as the substrate for GacL, which transfers GlcNAc to a rhamnose subunit of the GAC rhamnan backbone in an alternating fashion [88]. GacI homologs and related bactoprenol glycosyl transferase paralogs, which catalyze the transfer of i.e. glucose, galactose or GalNAc to Und-P, can be identified in SRP biosynthesis gene clusters of most streptococci using blast searches, suggesting the side chain biosynthesis mechanism is highly conserved among streptococcal species. In *S. mutans* Xc, a serotype c strain, RgpE, RgpH and RgpI have been identified as side chain glucosylating proteins [85, 89, 100]. However, their precise functions in side chain attachment was never unraveled. With this recent

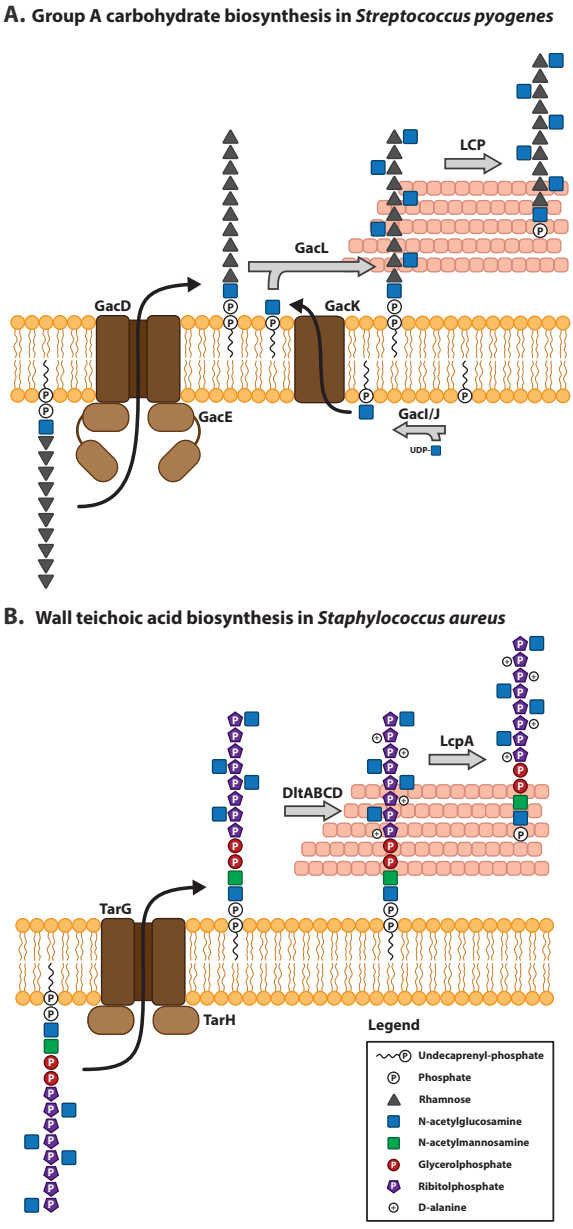


Figure 4. Schematic overview of SRP and WTA translocation, additional modifications and linkage to peptidoglycan. **(A)** Translocation of the GAC rhamnan backbone by ABC-transporter, decoration of rhamnan with GlcNAc side chains by GacIJKL proteins [88] and attachment of the GAC to peptidoglycan in *S. pyogenes*. Translocation and anchoring are likely to occur through the activity of proteins GacDE and specific members of the LytR/CpsA/Psr (LCP) protein family, respectively. **(B)** Translocation of fully glycosylated WTA by ABC-transporter TarGH, further decoration of ribitolphosphate with positively-charged *D*-alanine by DltABCD proteins and attachment of WTA to the peptidoglycan MurNAc residue by LcpA in *S. aureus*.

insight into GlcNAc side chain biosynthesis, it is likely that *rgpI* encodes a bactoprenol glucosyltransferase that transfers glucose from UDP-glucose to Und-P and *rgpH* likely encodes a *gacK* paralog, transporting Und-P-glucose across the membrane.

An enigmatic gene that is highly conserved within SRP gene clusters is annotated as a sulfatase or glycerolphosphate transferase, which is a function homologous to the lipoteichoic acid synthase (*ltaS*). The LtaS membrane protein is responsible for the polymerization of the GroP LTA backbone and transfers GroP from the membrane lipid phosphatidylglycerol to the LTA glycolipid anchor [101, 102]. The presence of this gene in SRP biosynthesis gene clusters is surprising since most SRPs, except for the GBC, have been structurally characterized as uncharged molecules with an apparent lack of phosphate (Fig 1). However, GroP was identified as a part of the GAC, then called C-polysaccharide, as early as 1967. It was proposed to be covalently linked to rhamnose either to attach GAC to peptidoglycan or as part of the interior of the GAC, although this was never verified or followed up [103, 104]. Similarly, GroP was identified in the rhamnose-containing cell wall polysaccharide of *Streptococcus sanguis* and small amounts of phosphorous were detected in *S. mutans* and *S. sobrinus* [23, 29, 105]. A likely explanation that GroP modifications were often undetected in SRP structural analysis is the harsh chemical treatment used to isolate the polysaccharides from the cell wall, which break phosphodiester bonds. In other cases, fractions containing GroP were considered to be contaminated with LTA and therefore not analyzed further. Again, these examples highlight how much is still unknown about SRP biosynthesis. Redetermination of SRP structures using alternative and especially mild isolation procedures is advised to investigate the possible presence of GroP. Considering that GacH is predicted to contain a large extracellular domain, it is expected that GroP modifications on SRPs occur on the extracellular site, similar to the mechanism of action of LtaS [101, 102].

In contrast to SRPs, the *S. aureus* WTA $(RboP)_n$ backbone is highly negatively charged and requires modification with positively charged *D*-alanine residues at the C2 hydroxyl group (Fig 4B) [106, 107]. Proteins required for *D*-alanylation are encoded by the *dltABCD* operon and disruption of any gene in this cluster abrogates addition of *D*-alanine to both WTA and structurally similar LTA [108]. Even though several models currently exist on how the DltABCD complex operates [109], one model is very similar to the GacI-L mechanism and involves the formation of Und-P-*D*-alanine. It is also suggested that *D*-alanine is first attached to LTA and subsequently transferred to WTA [110, 111], but proteins catalyzing this reaction have not yet been identified.

Step 6. Attachment to peptidoglycan

Once across the cytoplasmic membrane, both SRPs and WTAs are covalently linked to peptidoglycan. Attachment of glycopolymers to peptidoglycan is generally catalyzed by LytR-CpsA-Psr (LCP) family proteins (Fig 2). Based on database searches, this family of enzymes has been identified in bacteria of at least eight different phyla, which are hugely diverse with regard to cell envelope architecture and morphology, metabolic pathways and niche preference [112]. Given the diversity of cell wall glycopolymers, it is not surprising that LCP family members are not specific for SRPs or WTAs, but also anchor other structures such as capsular polysaccharides to peptidoglycan [113]. WTA is covalently linked to the peptidoglycan MurNAc moiety by enzymes belonging to the LCP family, renamed TagTUV or LcpABC (Fig 4B) [70, 71, 114]. Although all three LCP proteins in *S. aureus* were able to attach WTA to peptidoglycan *in vitro,* only deletion of *lcpA* resulted in a phenotype similar to a WTA null-mutant [114]. These results suggest that LCP proteins do have a preference for certain glycoconjugate structures in the context of the intact bacterium. Furthermore, *in vitro* reconstitution of WTA transfer to peptidoglycan in *B. subtilis*, revealed that all three LCP protein (TagTUV) were capable of transferring WTA to nascent (non-cross-linked) or mature (cross-linked) peptidoglycan, but not to lipid II (membrane bound biosynthetic peptidoglycan intermediate) [115]. Since LCP proteins localize near the septum, it is suggested that coupling of WTAs and other glycopolymers occurs at the division site, where cross-linking of newly synthesized peptidoglycan is highest [116-119].

S. mutans possesses two different LCP paralogs, BrpA and Psr, which are likely candidate proteins for the transfer of the serotype specific SRP to peptidoglycan. Deletion of either gene results in increased antimicrobial susceptibility and autolysis, reduced growth rates, acid and oxidative stress tolerance and defects in cell morphology, cell division and biofilm formation [120-122]. In addition, both proteins are regulators of gene expression [122, 123]. Deletion of both *brpA* and *psr* is lethal. In these double mutants, SRPs would accumulate on the outer membrane, thereby hampering recycling of Und-P, which in turn impedes the biosynthesis of new peptidoglycan. This lethal phenotype can be prevented in two ways: creation of a *psr*-deficient and *brpA*-down double mutant, which still has very low expression of *brpA* [123] or by additional deletion of *rgpG*, which prevents flux into the SRP biosynthesis pathway and leaves Und-P available for peptidoglycan biosynthesis [52].

Unfortunately, no biochemical studies have been performed on LCP proteins from *S. mutans* or other streptococci and direct evidence of SRP linkage to peptidoglycan by these proteins is currently lacking. However, structural studies have revealed that SRPs from *S. pyogenes* and *S. agalactiae* are linked to the MurNAc residue of

peptidoglycan, similar to WTA, although the nature of this linkage remains to be elucidated [42, 103, 104, 124]. Many glycoploymer structures are covalently linked to peptidoglycan through a phosphate containing linker, similar to the phosphodiester linkage that couples WTA to peptidoglycan MurNAc or capsular polysaccharides from *S. agalactiae* to peptidoglycan GlcNAc [30, 70, 124]. Considering the promiscuous nature of LCP proteins towards transferred substrates and the widespread use of phosphodiester linkage to covalently attach glycopolymers to peptidoglycan, it is likely that LCP proteins confer phosphotransferase activity, predominantly recognizing the conserved phosphate group. This would suggest that SRPs are coupled to MurNAc by a phophodiester linkage as well (Fig 4A). Indeed, there is some evidence that the GAC from *S. pyogenes,* the cell wall carbohydrate from *S. sobrinus* and the serotype e carbohydrate from *S. mutans* are linked to peptidoglycan by a phosphate containing group, because small amounts of phosphorous (<1%) were detected after SRP isolation for structure determination [23, 29, 40, 42, 103].

BREAKING SRPs

SRPs as therapeutic targets

Most of the streptococcal species described in this introduction are human or veterinary pathogens causing a wide range of disease [58, 125-129]. Despite the availability of antibiotics to treat streptococcal infections, streptococci still cause a shocking number of deaths annually. *S. pyogenes* even ranks in the top ten of infection-associated mortality worldwide [129, 130]. Efforts to produce a vaccine against these bacteria have proven fruitless so far and with the evermore increasing emergence of antibiotic-resistant pathogens, there is an urging need for the development of vaccines and new antibiotics. SRPs are abundant molecular hallmarks in the streptococcal cell wall and have long been used to identify streptococcal species for therapeutic diagnostics [4, 5, 131, 132]. However, SRPs have also proven to be essential for bacterial viability and virulence alike, making their biosynthesis pathways excellent therapeutic targets for the development of both vaccines and new antimicrobials.

Antimicrobials

In analogy with WTA biosynthesis, not all SRP biosynthesis steps are essential for bacterial viability and might therefore be less attractive as antibiotic targets. For example, in *S. aureus* expression of WTA is not required for cell viability and can be achieved by mutation of the early synthesis genes *tarO* or *tarA* [91, 133, 134] or by the antibiotic tunicamycin, which selectively blocks TarO at low concentrations [135]. Since tunicamycin does not kill the bacterium directly, it has been suggested to

work as an antivirulence agent since loss of WTA expression attenuates colonization and development of endocarditis [134, 136, 137]. Unfortunately, direct use of tunicamycin in animals and humans is precluded due to its off-target effects on eukaryotic phosphosugar transferases. Likewise, it is possible to create deletion mutants of the SRP initiation genes *gbcO* in *S. agalactiae* and *rgpG* in *S. mutans* [44, 51, 52]. However, efforts to make a *gacO* deletion mutant in *S. pyogenes* were unsuccessful, suggesting a more essential role for the GAC than the GBC or *S. mutans* SRPs.

Comparable results are expected with regard to essentiality when targeting the dTDP-L-rhamnose biosynthesis pathway since deletion or inhibition of one of the four dTDP-L-rhamnose biosynthesis enzymes, RmlABCD, would result in the inability to initiate SRP rhamnan backbone elongation. Indeed, rhamnose is essential for the viability and/or virulence of all SRP-expressing streptococcal pathogens and rhamnose is not expressed in humans [138]. Therefore, the rhamnose biosynthesis pathway is an interesting and relative selective antibiotic target. In addition, rhamnose is also essential for the viability or virulence of many other human pathogens including *Enterococcus faecalis* [139, 140], *Mycobacterium spp.* [141, 142], *Pseudomonas spp.* [143] and *Salmonella enterica* serovar *Typhimurium* [144]. Extensive studies on the structural characteristics as well as the mechanisms of action of the four dTDP-L-rhamnose biosynthesis enzymes, RmlABCD, in both Gram-positive and Gram-negative bacteria [145-154], have directed the development of preliminary inhibitors of which some could inhibit growth of *Mycobacterium tuberculosis in vitro* [155-159].

Alternatively, the accumulating intermediate products of SRP or WTA biosynthesis may be directly toxic to cells. Indeed, in contrast to non-essential mutation of the early steps of WTA biosynthesis in *S. aureus*, inhibition of steps downstream of TarA involved in WTA backbone formation are lethal to bacteria [82]. Similarly, inhibition of enzymes involved in SRP rhamnan backbone elongation and translocation across the membrane are expected to be lethal to streptococci even under non-competitive circumstances. No inhibitors that target the SRP or WTA elongation pathway have yet been identified. Nevertheless, the essentiality of the WTA transporter TarGH has been exploited for the development of antibiotics, resulting in the discovery of the TarG inhibitor Targocil that synergizes with beta-lactam antibiotics that target peptidoglycan biosynthesis [135, 160]. The recently identified compound teixobactin combines these two activities in one molecule, targeting both lipidII (Und-P-P-MurNac-GlcNAc; peptidoglycan biosynthesis) and lipidIII (Und-P-P-GlcNAc; WTA biosynthesis) [161]. In addition, inhibitors that would work on the non-essential SRP side chain biosynthesis pathway, might also act as anti-virulence agents, but further research on the exact role of these side chains in pathogenesis is required.

Vaccines

SRPs have been proposed as vaccine candidates as early as the 1960s due to their high abundance in the streptococcal cell wall, exposed nature and contribution to virulence [22, 162-167]. In addition, elevated anti-SRP antibody levels were detected in humans after streptococcal infection [167-170]. *S. pyogenes* antibodies detected in humans were found to be predominantly directed against the GAC GlcNAc, but also against the group A-variant carbohydrate, which lacks the GlcNAc side chains, suggesting both the rhamnan backbone and the GlcNAc side chains are immunogenic structures [167, 171-173]. However, there was no direct correlation between the detection of anti-rhamnan antibodies and other anti-*S. pyogenes* antibodies. Therefore, anti-rhamnan antibodies may also arise from colonization or infection with other streptococcal species harboring the same rhamnan backbone in their SRP [171, 173]. Furthermore, GAC-antibodies have GlcNAc-dependent opsonophagocytic capabilities and were able to passively protect mice against *S. pyogenes* [174, 175]. However, patients with rheumatic heart disease maintain elevated levels of GAC-antibodies over prolonged time [169, 176, 177]. Moreover, GAC-antibodies were found to cross-react with GlcNAc antigens on heart valve tissue, which has caused concern regarding its use for vaccine development [178-181]. Recent studies have shown that antibodies directed against only the rhamnan backbone also promoted opsonophagocytic killing of multiple *S. pyogenes* serotypes and protected against systemic *S. pyogenes* challenge after passive immunization [10]. Conversely, a vaccine based on only the rhamnan backbone is expected to be less specific since many (pathogenic) streptococci are known to express the rhamnan structure in their cell wall as part of their SRP (Fig 1). In addition, rhamnan might also be expressed in the cell walls of bacteria that have not yet been identified and promote a healthy microbiome. Precaution should therefore be taken when pursuing this vaccine strategy.

In case of *S. agalactiae*, it is generally accepted that serotype-specific capsule antibodies, but not GBC-specific antibodies, can provide immunity against this pathogen [162, 182-186]. Even though the GBC is no longer considered a suitable vaccine candidate, in one study monoclonal GBC antibodies could offer protection against serotype Ia and type III *S. agalactiae* in a neonatal rat infection model in a rhamnose dependent manner [187].

Vaccines have not only been considered to prevent life threatening infections, but have also been suggested for prevention of dental caries. Dental caries is a result of biofilm-mediated dysbiosis of acid-producing and acid-tolerant bacteria [59]. Elevated levels of mutans streptococci, predominantly *S. mutans* and *S. sobrinus*, have been associated with the development of dental caries and it is believed that targeted

therapy against these bacteria will help restore the healthy oral microbiome [58, 188]. Early vaccination studies in the 1970's and 1980's with fixed whole cells or protein antigens of *S. mutans* in rodents and primates have shown the feasibility of a dental caries vaccine directed against *S. mutans* [22, 189-192]. In addition, active immunization of rats with cell walls of *S. mutans* was able to prevent oral colonization and reduced the occurrence of dental caries [193]. However, similar to *S. pyogenes*, certain *S. mutans* antigens are cross reactive with human heart tissue and the use of a whole cell vaccine is therefore not advised [194]. More recently, Fukuizumi et al. showed that tonsillar application of killed *S. mutans* and *S. sobrinus* in rabbits does not induce antibodies that cross reacts with cardiac muscle tissue [195]. Salivary and blood antibodies derived from these experiments was partly directed against the SRPs from several mutans streptococci [196]. Also monoclonal antibodies isolated from mice immunized with *S. mutans* serotype e, specifically interacted with the glucose side chain of the serotype e SRP [197]. More importantly, also serum from humans infected with *S. mutans* contained IgG and IgA against *S. mutans* serotype c SRPs and did not cross-react with heart tissue [198, 199]. The immunogenicity of a serotype f *S. mutans* SRP could be further increased by the coupling of purified SRP to a peptide containing both T- and B-cell epitopes or to a saliva receptor protein conjugated to liposomes and even induced an IgA memory response [200-203]. Additionally, serum derived from mice immunized with serotype f or k *S. mutans* SRP conjugated to human serum albumin facilitated opsonophagocytic killing of homologous (f or k) and heterologous (f, k or c) *S. mutans* serotypes [22]. All together, these data suggest that SRPs from *S. mutans*, but possibly also from other streptococci, could be developed into efficient vaccine candidates [22, 204].

Bacteriophage therapy

Recently, bacteriophage therapy has gained increased attention in the media [205]. Bacteriophages are viruses that use bacteria as a host to replicate and kill them in the process. These phages can be species and even strain specific, allowing targeted antimicrobial therapy, thereby leaving the healthy microbiome undisturbed. Phage resistance has been readily observed, but sometimes led to reduced fitness in animal models creating alternative strategies for the use of phage therapy [206]. In light of these recent developments, it is noteworthy that SRPs are known targets for phage binding [207-211]. More specifically, several phage resistant *S. equisimilis* strains lacked GalNAc in their group C carbohydrate (C-variants), suggesting a role for the GalNAc side chain in phage absorption [212]. A role in phage absorption has also been identified for the GlcNAc side chain of the GAC [210].

Scope of this thesis

SRPs have proven to be important molecular structures in the cell wall of many (pathogenic) streptococcal species, contributing to bacterial physiology and virulence. Their biosynthesis pathway would provide an excellent target for antimicrobial therapy and SRPs have potential as vaccine targets. SRPs share many functional similarities with WTAs and extensive knowledge on WTA biosynthesis has advanced antibiotic development targeting the WTA biosynthesis pathway. However, knowledge on SRP biosynthesis is currently limited and besides many similarities, many differences with WTA biosynthesis have been identified as well, highlighting the need for more research on the topic. In this thesis, I investigated two different steps in the SRP biosynthesis pathway at a functional and structural level; the biosynthesis of the rhamnose nucleotide precursor, dTDP-L-rhamnose (**Chapter 2 + 3**) and the transport of the SRP (backbone) across the membrane by an ABC-transporter (**Chapter 4**). I limited my studies to the biosynthesis of the GAC from *S. pyogenes* and the serotype c carbohydrate from *S. mutans*. The GAC and the serotype c carbohydrate share many structural similarities and have identical rhamnan backbones. Combined with the ease of genetic manipulation of *S. mutans*, this allowed me to study the highly essential GAC biosynthesis genes by heterologous expression in *S. mutans*. In addition, I have identified inhibitory fragments against the dTDP-L-rhamnose biosynthesis enzymes from *S. pyogenes*, which could be further developed into antibiotic compounds (**Chapter 3**). Finally, my genetic studies using *S. mutans* has uncovered the unexpected occurrence of merodiploidy as a survival mechanism in strain Xc upon deletion of critical genes encoding the SRP ABC-transporter (**Chapter 5**). This chapter contains a cautionary note with regard to dismissal of unexpected results when generating mutants in essential or crucial genes.

References

1. Tytgat, H.L. and Lebeer, S.; *The sweet tooth of bacteria: common themes in bacterial glycoconjugates.* Microbiol Mol Biol Rev, 2014. **78**(3): p. 372-417.
2. Mistou, M.Y., Sutcliffe, I.C., and van Sorge, N.M.; *Bacterial glycobiology: rhamnose-containing cell wall polysaccharides in Gram-positive bacteria.* FEMS Microbiol Rev, 2016. **40**(4): p. 464-79.
3. Hardie, J.M. and Whiley, R.A.; *Classification and overview of the genera Streptococcus and Enterococcus.* J Appl Microbiol, 1997. **83**(S1): p. 1S-11S.
4. Lancefield, R.C.; *A Serological Differentiation of Human and Other Groups of Hemolytic Streptococci.* J Exp Med, 1933. **57**(4): p. 571-95.
5. Slade, H.D. and Slamp, W.C.; *Cell-wall composition and the grouping antigens of Streptococci.* J Bacteriol, 1962. **84**: p. 345-51.
6. Nakano, K. and Ooshima, T.; *Serotype classification of Streptococcus mutans and its detection outside the oral cavity.* Future Microbiol, 2009. **4**(7): p. 891-902.
7. Tsuda, H., et al.; *Role of serotype-specific polysaccharide in the resistance of Streptococcus mutans to phagocytosis by human polymorphonuclear leukocytes.* Infect Immun, 2000. **68**(2): p. 644-50.
8. Saputo, S., Faustoferri, R.C., and Quivey, R.G., Jr.; *Vitamin D Compounds Are Bactericidal against Streptococcus mutans and Target the Bacitracin-Associated Efflux System.* Antimicrob Agents Chemother, 2018. **62**(1).
9. Le Breton, Y., et al.; *Genome-wide discovery of novel M1T1 group A streptococcal determinants important for fitness and virulence during soft-tissue infection.* PLoS Pathog, 2017. **13**(8): p. e1006584.
10. van Sorge, N.M., et al.; *The classical lancefield antigen of group a streptococcus is a virulence determinant with implications for vaccine design.* Cell Host Microbe, 2014. **15**(6): p. 729-40.
11. Voyich, J.M., et al.; *Genome-wide protective response used by group A Streptococcus to evade destruction by human polymorphonuclear leukocytes.* Proc Natl Acad Sci U S A, 2003. **100**(4): p. 1996-2001.
12. Voyich, J.M., et al.; *Engagement of the pathogen survival response used by group A Streptococcus to avert destruction by innate host defense.* J Immunol, 2004. **173**(2): p. 1194-201.
13. Coligan, J.E., Kindt, T.J., and Krause, R.M.; *Structure of the streptococcal groups A, A-variant and C carbohydrates.* Immunochemistry, 1978. **15**(10-11): p. 755-60.
14. Pritchard, D.G., Gray, B.M., and Dillon, H.C., Jr.; *Characterization of the group-specific polysaccharide of group B Streptococcus.* Archives of biochemistry and biophysics, 1984. **235**(2): p. 385-92.
15. Michon, F., et al.; *Structure of the complex group-specific polysaccharide of group B Streptococcus.* Biochemistry, 1987. **26**(2): p. 476-86.
16. Michon, F., et al.; *Multiantennary group-specific polysaccharide of group B Streptococcus.* Biochemistry, 1988. **27**(14): p. 5341-51.
17. Neiwert, O., Holst, O., and Duda, K.A.; *Structural investigation of rhamnose-rich polysaccharides from Streptococcus dysgalactiae bovine mastitis isolate.* Carbohydr Res, 2014. **389**: p. 192-5.
18. Soprey, P. and Slade, H.D.; *Chemical structure and immunological specificity of the streptococcal group e cell wall polysaccharide antigen.* Infect Immun, 1971. **3**(5): p. 653-8.
19. Pritchard, D.G. and Furner, R.L.; *Structure of the group-specific polysaccharide of group E Streptococcus.* Carbohydr Res, 1985. **144**(2): p. 289-96.
20. Pritchard, D.G., et al.; *Structure of the group G streptococcal polysaccharide.* Carbohydr Res, 1988. **173**(2): p. 255-62.
21. Wetherell, J.R., Jr. and Bleiweis, A.S.; *Antigens of Streptococcus mutans: characterization of a polysaccharide antigen from walls of strain GS-5.* Infect Immun, 1975. **12**(6): p. 1341-8.
22. St Michael, F., et al.; *Investigating the candidacy of the serotype specific rhamnan polysaccharide based glycoconjugates to prevent disease caused by the dental pathogen Streptococcus mutans.* Glycoconj J, 2017.

23. Pritchard, D.G., et al.; *Characterization of the serotype e polysaccharide antigen of Streptococcus mutans.* Mol Immunol, 1986. **23**(2): p. 141-5.
24. Linzer, R., Reddy, M.S., and Levine, M.J.; *Structural studies of the serotype-f polysaccharide antigen from Streptococcus mutans OMZ175.* Infect Immun, 1987. **55**(12): p. 3006-10.
25. Pritchard, D.G., et al.; *Structure of the serotype f polysaccharide antigen of Streptococcus mutans.* Carbohydr Res, 1987. **166**(1): p. 123-31.
26. Czabanska, A., Holst, O., and Duda, K.A.; *Chemical structures of the secondary cell wall polymers (SCWPs) isolated from bovine mastitis Streptococcus uberis.* Carbohydr Res, 2013. **377**: p. 58-62.
27. Linzer, R., Reddy, M.S., and Levine, M.J.; *Structural studies of the rhamnose-glucose polysaccharide antigen from Streptococcus sobrinus B13 and 6715-T2.* Infect Immun, 1985. **50**(2): p. 583-5.
28. Prakobphol, A. and Linzer, R.; *Purification and immunological characterization of a rhamnose-glucose antigen from Streptococcus mutans 6517-T2 (serotype g).* Infect Immun, 1980. **30**(1): p. 140-6.
29. Prakobphol, A., Linzer, R., and Genco, R.J.; *Purification and characterization of a rhamnose-containing cell wall antigen of Streptococcus mutans B13 (serotype d).* Infect Immun, 1980. **27**(1): p. 150-7.
30. Neuhaus, F.C. and Baddiley, J.; *A continuum of anionic charge: structures and functions of D-alanyl-teichoic acids in gram-positive bacteria.* Microbiol Mol Biol Rev, 2003. **67**(4): p. 686-723.
31. Winstel, V., Xia, G., and Peschel, A.; *Pathways and roles of wall teichoic acid glycosylation in Staphylococcus aureus.* Int J Med Microbiol, 2014. **304**(3-4): p. 215-21.
32. Weidenmaier, C. and Peschel, A.; *Teichoic acids and related cell-wall glycopolymers in Gram-positive physiology and host interactions.* Nat Rev Microbiol, 2008. **6**(4): p. 276-87.
33. Brown, S., Santa Maria, J.P., Jr., and Walker, S.; *Wall teichoic acids of gram-positive bacteria.* Annu Rev Microbiol, 2013. **67**: p. 313-36.
34. van der Es, D., et al.; *Teichoic acids: synthesis and applications.* Chem Soc Rev, 2017. **46**(5): p. 1464-1482.
35. Pereira, M.P., et al.; *The wall teichoic acid polymerase TagF efficiently synthesizes poly(glycerol phosphate) on the TagB product lipid III.* Chembiochem, 2008. **9**(9): p. 1385-90.
36. Allison, S.E., et al.; *Studies of the genetics, function, and kinetic mechanism of TagE, the wall teichoic acid glycosyltransferase in Bacillus subtilis 168.* J Biol Chem, 2011. **286**(27): p. 23708-16.
37. Schertzer, J.W. and Brown, E.D.; *Use of CDP-glycerol as an alternate acceptor for the teichoic acid polymerase reveals that membrane association regulates polymer length.* J Bacteriol, 2008. **190**(21): p. 6940-7.
38. Meredith, T.C., Swoboda, J.G., and Walker, S.; *Late-stage polyribitol phosphate wall teichoic acid biosynthesis in Staphylococcus aureus.* J Bacteriol, 2008. **190**(8): p. 3046-56.
39. Wanner, S., et al.; *Wall teichoic acids mediate increased virulence in Staphylococcus aureus.* Nat Microbiol, 2017. **2**: p. 16257.
40. McCarty, M.; *The lysis of group A hemolytic streptococci by extracellular enzymes of Streptomyces albus. II. Nature of the cellular substrate attacked by the lytic enzymes.* J Exp Med, 1952. **96**(6): p. 569-80.
41. Doran, T.I. and Mattingly, S.J.; *Association of type- and group-specific antigens with the cell wall of serotype III group B streptococcus.* Infect Immun, 1982. **36**(3): p. 1115-22.
42. De Cueninck, B.J., Shockman, G.D., and Swenson, R.M.; *Group B, type III streptococcal cell wall: composition and structural aspects revealed through endo-N-acetylmuramidase-catalyzed hydrolysis.* Infect Immun, 1982. **35**(2): p. 572-81.
43. Krause, R.M. and McCarty, M.; *Studies on the chemical structure of the streptococcal cell wall. II. The composition of group C cell walls and chemical basis for serologic specificity of the carbohydrate moiety.* J Exp Med, 1962. **115**: p. 49-62.
44. Caliot, E., et al.; *Role of the Group B antigen of Streptococcus agalactiae: a peptidoglycan-anchored polysaccharide involved in cell wall biogenesis.* PLoS Pathog, 2012. **8**(6): p. e1002756.
45. Le Breton, Y., et al.; *Essential Genes in the Core Genome of the Human Pathogen Streptococcus pyogenes.* Sci Rep, 2015. **5**: p. 9838.

46. Shields, R.C., et al.; *Genomewide Identification of Essential Genes and Fitness Determinants of Streptococcus mutans UA159.* mSphere, 2018. **3**(1).
47. Quivey, R.G., Jr., et al.; *Functional profiling in Streptococcus mutans: construction and examination of a genomic collection of gene deletion mutants.* Mol Oral Microbiol, 2015. **30**(6): p. 474-95.
48. Blanchard, A.M., et al.; *PIMMS (Pragmatic Insertional Mutation Mapping System) Laboratory Methodology a Readily Accessible Tool for Identification of Essential Genes in Streptococcus.* Front Microbiol, 2016. **7**: p. 1645.
49. Hooven, T.A., et al.; *The essential genome of Streptococcus agalactiae.* BMC Genomics, 2016. **17**: p. 406.
50. Charbonneau, A.R.L., et al.; *Defining the ABC of gene essentiality in streptococci.* BMC Genomics, 2017. **18**(1): p. 426.
51. Beaussart, A., et al.; *Molecular mapping of the cell wall polysaccharides of the human pathogen Streptococcus agalactiae.* Nanoscale, 2014. **6**(24): p. 14820-7.
52. De, A., et al.; *Deficiency of RgpG Causes Major Defects in Cell Division and Biofilm Formation, and Deficiency of LytR-CpsA-Psr Family Proteins Leads to Accumulation of Cell Wall Antigens in Culture Medium by Streptococcus mutans.* Appl Environ Microbiol, 2017. **83**(17).
53. Henningham, A., et al.; *Virulence Role of the GlcNAc Side Chain of the Lancefield Cell Wall Carbohydrate Antigen in Non-M1-Serotype Group A Streptococcus.* MBio, 2018. **9**(1).
54. Kovacs, C.J., Faustoferri, R.C., and Quivey, R.G., Jr.; *RgpF Is Required for Maintenance of Stress Tolerance and Virulence in Streptococcus mutans.* J Bacteriol, 2017. **199**(24).
55. Nakano, K., et al.; *Role of glucose side chains with serotype-specific polysaccharide in the cariogenicity of Streptococcus mutans.* Caries Res, 2005. **39**(4): p. 262-8.
56. Dewhirst, F.E., et al.; *The human oral microbiome.* J Bacteriol, 2010. **192**(19): p. 5002-17.
57. Koo, H., Falsetta, M.L., and Klein, M.I.; *The exopolysaccharide matrix: a virulence determinant of cariogenic biofilm.* J Dent Res, 2013. **92**(12): p. 1065-73.
58. Loesche, W.J.; *Role of Streptococcus mutans in human dental decay.* Microbiol Rev, 1986. **50**(4): p. 353-80.
59. Tanner, A.C.R., et al.; *The Caries Microbiome: Implications for Reversing Dysbiosis.* Adv Dent Res, 2018. **29**(1): p. 78-85.
60. Zhan, L.; *Rebalancing the Caries Microbiome Dysbiosis: Targeted Treatment and Sugar Alcohols.* Adv Dent Res, 2018. **29**(1): p. 110-116.
61. Nagata, E., et al.; *Serotype-specific polysaccharide of Streptococcus mutans contributes to infectivity in endocarditis.* Oral Microbiol Immunol, 2006. **21**(6): p. 420-3.
62. Chia, J.S., et al.; *Platelet aggregation induced by serotype polysaccharides from Streptococcus mutans.* Infection and Immunity, 2004. **72**(5): p. 2605-2617.
63. Ford, I. and Douglas, C.W.I.; *The role of platelets in infective endocarditis.* Platelets, 1997. **8**(5): p. 285-294.
64. Zhu, L., et al.; *Novel Genes Required for the Fitness of Streptococcus pyogenes in Human Saliva.* mSphere, 2017. **2**(6).
65. Le Breton, Y., et al.; *Genome-wide identification of genes required for fitness of group A Streptococcus in human blood.* Infect Immun, 2013. **81**(3): p. 862-75.
66. Wilson, A.T.; *Loss of Group Carbohydrate during Mouse Passages of a Group a Hemolytic Streptococcus.* J Exp Med, 1945. **81**(6): p. 593-6.
67. McCarty, M. and Lancefield, R.C.; *Variation in the group-specific carbohydrate of group A streptococci. I. Immunochemical studies on the carbohydrates of variant strains.* J Exp Med, 1955. **102**(1): p. 11-28.
68. McCarty, M.; *Variation in the group-specific carbohydrate of group A streptococci. II. Studies on the chemical basis for serological specificity of the carbohydrates.* J Exp Med, 1956. **104**(5): p. 629-43.
69. Lazarevic, V. and Karamata, D.; *The tagGH operon of Bacillus subtilis 168 encodes a two-component ABC transporter involved in the metabolism of two wall teichoic acids.* Mol Microbiol, 1995. **16**(2): p. 345-55.

70. Chan, Y.G., et al.; *Staphylococcus aureus mutants lacking the LytR-CpsA-Psr family of enzymes release cell wall teichoic acids into the extracellular medium.* J Bacteriol, 2013. **195**(20): p. 4650-9.
71. Kawai, Y., et al.; *A widespread family of bacterial cell wall assembly proteins.* EMBO J, 2011. **30**(24): p. 4931-41.
72. Swoboda, J.G., et al.; *Wall teichoic acid function, biosynthesis, and inhibition.* Chembiochem, 2010. **11**(1): p. 35-45.
73. Xia, G., Kohler, T., and Peschel, A.; *The wall teichoic acid and lipoteichoic acid polymers of Staphylococcus aureus.* International Journal of Medical Microbiology, 2010. **300**(2-3): p. 148-154.
74. Brown, S., Zhang, Y.H., and Walker, S.; *A revised pathway proposed for Staphylococcus aureus wall teichoic acid biosynthesis based on in vitro reconstitution of the intracellular steps.* Chem Biol, 2008. **15**(1): p. 12-21.
75. Brown, S., et al.; *Staphylococcus aureus and Bacillus subtilis W23 make polyribitol wall teichoic acids using different enzymatic pathways.* Chem Biol, 2010. **17**(10): p. 1101-10.
76. Xu, P., et al.; *Genome-wide essential gene identification in Streptococcus sanguinis.* Sci Rep, 2011. **1**: p. 125.
77. Kawada-Matsuo, M., et al.; *GlmS and NagB regulate amino sugar metabolism in opposing directions and affect Streptococcus mutans virulence.* PLoS One, 2012. **7**(3): p. e33382.
78. Zhang, Y., et al.; *Genetic and biochemical characterization of genes involved in hyaluronic acid synthesis in Streptococcus zooepidemicus.* Appl Microbiol Biotechnol, 2016. **100**(8): p. 3611-20.
79. Liu, X.D., Duan, J., and Guo, L.H.; *Role of phosphoglucosamine mutase on virulence properties of Streptococcus mutans.* Oral Microbiol Immunol, 2009. **24**(4): p. 272-7.
80. Zhang, W., et al.; *Expression, essentiality, and a microtiter plate assay for mycobacterial GlmU, the bifunctional glucosamine-1-phosphate acetyltransferase and N-acetylglucosamine-1-phosphate uridyltransferase.* Int J Biochem Cell Biol, 2008. **40**(11): p. 2560-71.
81. Mann, P.A., et al.; *Chemical Genetic Analysis and Functional Characterization of Staphylococcal Wall Teichoic Acid 2-Epimerases Reveals Unconventional Antibiotic Drug Targets.* PLoS Pathog, 2016. **12**(5): p. e1005585.
82. D'Elia, M.A., et al.; *Lesions in teichoic acid biosynthesis in Staphylococcus aureus lead to a lethal gain of function in the otherwise dispensable pathway.* J Bacteriol, 2006. **188**(12): p. 4183-9.
83. Tsukioka, Y., et al.; *Identification of a fourth gene involved in dTDP-rhamnose synthesis in Streptococcus mutans.* J Bacteriol, 1997. **179**(13): p. 4411-4.
84. Tsukioka, Y., et al.; *Biological function of the dTDP-rhamnose synthesis pathway in Streptococcus mutans.* J Bacteriol, 1997. **179**(4): p. 1126-34.
85. Yamashita, Y., et al.; *Genes involved in cell wall localization and side chain formation of rhamnose-glucose polysaccharide in Streptococcus mutans.* J Bacteriol, 1998. **180**(21): p. 5803-7.
86. Hartley, M.D. and Imperiali, B.; *At the membrane frontier: a prospectus on the remarkable evolutionary conservation of polyprenols and polyprenyl-phosphates.* Arch Biochem Biophys, 2012. **517**(2): p. 83-97.
87. Yamashita, Y., et al.; *A novel gene required for rhamnose-glucose polysaccharide synthesis in Streptococcus mutans.* J Bacteriol, 1999. **181**(20): p. 6556-9.
88. Rush, J.S., et al.; *The molecular mechanism of N-acetylglucosamine side-chain attachment to the Lancefield group A carbohydrate in Streptococcus pyogenes.* J Biol Chem, 2017. **292**(47): p. 19441-19457.
89. Shibata, Y., et al.; *Expression and characterization of streptococcal rgp genes required for rhamnan synthesis in Escherichia coli.* Infect Immun, 2002. **70**(6): p. 2891-8.
90. Schalch, W., et al.; *Distinct functions of monoclonal IgG antibody depend on antigen-site specificities.* J Exp Med, 1979. **149**(4): p. 923-37.
91. D'Elia, M.A., et al.; *The N-acetylmannosamine transferase catalyzes the first committed step of teichoic acid assembly in Bacillus subtilis and Staphylococcus aureus.* J Bacteriol, 2009. **191**(12): p. 4030-4.
92. Bhavsar, A.P., Truant, R., and Brown, E.D.; *The TagB protein in Bacillus subtilis 168 is an intracellular peripheral membrane protein that can incorporate glycerol phosphate onto a membrane-bound acceptor in vitro.* J Biol Chem, 2005. **280**(44): p. 36691-700.

93. Schirner, K., Stone, L.K., and Walker, S.; *ABC transporters required for export of wall teichoic acids do not discriminate between different main chain polymers.* ACS Chem Biol, 2011. **6**(5): p. 407-12.
94. Torii, M., Kabat, E.A., and Bezer, A.E.; *Separation of Teichoic Acid of Staphylococcus Aureus into Two Immunologically Distinct Specific Polysaccharides with Alpha- and Beta-N-Acetylglucosaminyl Linkages Respectively. Antigenicity of Theichoic Acids in Man.* J Exp Med, 1964. **120**: p. 13-29.
95. Koc, C., et al.; *Structural and enzymatic analysis of TarM glycosyltransferase from Staphylococcus aureus reveals an oligomeric protein specific for the glycosylation of wall teichoic acid.* J Biol Chem, 2015. **290**(15): p. 9874-85.
96. Brown, S., et al.; *Methicillin resistance in Staphylococcus aureus requires glycosylated wall teichoic acids.* Proc Natl Acad Sci U S A, 2012. **109**(46): p. 18909-14.
97. Xia, G., et al.; *Glycosylation of wall teichoic acid in Staphylococcus aureus by TarM.* J Biol Chem, 2010. **285**(18): p. 13405-15.
98. Perez, C., et al.; *Structure and mechanism of an active lipid-linked oligosaccharide flippase.* Nature, 2015.
99. Cuthbertson, L., Kos, V., and Whitfield, C.; *ABC transporters involved in export of cell surface glycoconjugates.* Microbiol Mol Biol Rev, 2010. **74**(3): p. 341-62.
100. Ozaki, K., et al.; *A novel mechanism for glucose side-chain formation in rhamnose-glucose polysaccharide synthesis.* FEBS Lett, 2002. **532**(1-2): p. 159-63.
101. Taron, D.J., Childs, W.C., 3rd, and Neuhaus, F.C.; *Biosynthesis of D-alanyl-lipoteichoic acid: role of diglyceride kinase in the synthesis of phosphatidylglycerol for chain elongation.* J Bacteriol, 1983. **154**(3): p. 1110-6.
102. Koch, H.U., Haas, R., and Fischer, W.; *The role of lipoteichoic acid biosynthesis in membrane lipid metabolism of growing Staphylococcus aureus.* Eur J Biochem, 1984. **138**(2): p. 357-63.
103. Heymann, H., Manniello, J.M., and Barkulis, S.S.; *Structure of streptococcal cell walls. V. Phosphate esters in the walls of group A Streptococcus pyogenes.* Biochem Biophys Res Commun, 1967. **26**(4): p. 486-91.
104. Munoz, E., Ghuysen, J.M., and Heymann, H.; *Cell walls of Streptococcus pyogenes, type 14. C polysaccharide-peptidoglycan and G polysaccharide-peptidoglycan complexes.* Biochemistry, 1967. **6**(12): p. 3659-70.
105. Emdur, L.I., et al.; *Glycerolphosphate-containing cell wall polysaccharides from Streptococcus sanguis.* J Bacteriol, 1974. **120**(2): p. 724-32.
106. Vinogradov, E., et al.; *Structural elucidation of the extracellular and cell-wall teichoic acids of Staphylococcus aureus MN8m, a biofilm forming strain.* Carbohydr Res, 2006. **341**(6): p. 738-43.
107. Sanderson, A.R., Strominger, J.L., and Nathenson, S.G.; *Chemical structure of teichoic acid from Staphylococcus aureus, strain Copenhagen.* J Biol Chem, 1962. **237**: p. 3603-13.
108. Perego, M., et al.; *Incorporation of D-alanine into lipoteichoic acid and wall teichoic acid in Bacillus subtilis. Identification of genes and regulation.* J Biol Chem, 1995. **270**(26): p. 15598-606.
109. Reichmann, N.T., Cassona, C.P., and Grundling, A.; *Revised mechanism of D-alanine incorporation into cell wall polymers in Gram-positive bacteria.* Microbiology, 2013. **159**(Pt 9): p. 1868-77.
110. Koch, H.U., Doker, R., and Fischer, W.; *Maintenance of D-Alanine Ester Substitution of Lipoteichoic Acid by Reesterification in Staphylococcus-Aureus.* Journal of Bacteriology, 1985. **164**(3): p. 1211-1217.
111. Haas, R., Koch, H.U., and Fischer, W.; *Alanyl Turnover from Lipoteichoic Acid to Teichoic-Acid in Staphylococcus-Aureus.* Fems Microbiology Letters, 1984. **21**(1): p. 27-31.
112. Hubscher, J., et al.; *Phylogenetic distribution and membrane topology of the LytR-CpsA-Psr protein family.* BMC genomics, 2008. **9**: p. 617.
113. Chan, Y.G., et al.; *The capsular polysaccharide of Staphylococcus aureus is attached to peptidoglycan by the LytR-CpsA-Psr (LCP) family of enzymes.* J Biol Chem, 2014. **289**(22): p. 15680-90.
114. Schaefer, K., et al.; *In vitro reconstitution demonstrates the cell wall ligase activity of LCP proteins.* Nat Chem Biol, 2017. **13**(4): p. 396-401.

115. Gale, R.T., et al.; *B. subtilis LytR-CpsA-Psr Enzymes Transfer Wall Teichoic Acids from Authentic Lipid-Linked Substrates to Mature Peptidoglycan In Vitro.* Cell Chem Biol, 2017. **24**(12): p. 1537-1546 e4.
116. Rowe, H.M., et al.; *Modification of the CpsA protein reveals a role in alteration of the Streptococcus agalactiae cell envelope.* Infect Immun, 2015. **83**(4): p. 1497-506.
117. Eberhardt, A., et al.; *Attachment of capsular polysaccharide to the cell wall in Streptococcus pneumoniae.* Microb Drug Resist, 2012. **18**(3): p. 240-55.
118. Baumgart, M., et al.; *Impact of LytR-CpsA-Psr Proteins on Cell Wall Biosynthesis in Corynebacterium glutamicum.* J Bacteriol, 2016. **198**(22): p. 3045-3059.
119. Marechal, M., et al.; *Enterococcus hirae LcpA (Psr), a new peptidoglycan-binding protein localized at the division site.* BMC Microbiol, 2016. **16**(1): p. 239.
120. Wen, Z.T., Baker, H.V., and Burne, R.A.; *Influence of BrpA on critical virulence attributes of Streptococcus mutans.* J Bacteriol, 2006. **188**(8): p. 2983-92.
121. Chatfield, C.H., Koo, H., and Quivey, R.G., Jr.; *The putative autolysin regulator LytR in Streptococcus mutans plays a role in cell division and is growth-phase regulated.* Microbiology, 2005. **151**(Pt 2): p. 625-31.
122. Bitoun, J.P., et al.; *BrpA is involved in regulation of cell envelope stress responses in Streptococcus mutans.* Appl Environ Microbiol, 2012. **78**(8): p. 2914-22.
123. Bitoun, J.P., et al.; *Psr is involved in regulation of glucan production, and double deficiency of BrpA and Psr is lethal in Streptococcus mutans.* Microbiology, 2013. **159**(Pt 3): p. 493-506.
124. Deng, L., et al.; *Characterization of the linkage between the type III capsular polysaccharide and the bacterial cell wall of group B Streptococcus.* J Biol Chem, 2000. **275**(11): p. 7497-504.
125. Han, Y.W. and Wang, X.; *Mobile microbiome: oral bacteria in extra-oral infections and inflammation.* J Dent Res, 2013. **92**(6): p. 485-91.
126. Kobayashi, M., et al.; *Group B Streptococcus vaccine development: present status and future considerations, with emphasis on perspectives for low and middle income countries.* F1000Res, 2016. **5**: p. 2355.
127. Mitchell, T.J.; *The pathogenesis of streptococcal infections: from tooth decay to meningitis.* Nat Rev Microbiol, 2003. **1**(3): p. 219-30.
128. Cunningham, M.W.; *Pathogenesis of group A streptococcal infections.* Clin Microbiol Rev, 2000. **13**(3): p. 470-511.
129. Ralph, A.P. and Carapetis, J.R.; *Group a streptococcal diseases and their global burden.* Curr Top Microbiol Immunol, 2013. **368**: p. 1-27.
130. Carapetis, J.R., et al.; *The global burden of group A streptococcal diseases.* Lancet Infect Dis, 2005. **5**(11): p. 685-94.
131. Lue, Y.A., Howit, I.P., and Ellner, P.D.; *Rapid grouping of beta-hemolytic streptococci by latex agglutination.* J Clin Microbiol, 1978. **8**(3): p. 326-8.
132. Ruch, F.E., Jr. and Smith, L.; *Monoclonal antibody to streptococcal group B carbohydrate: applications in latex agglutination and immunoprecipitin assays.* J Clin Microbiol, 1982. **16**(1): p. 145-52.
133. D'Elia, M.A., et al.; *Wall teichoic acid polymers are dispensable for cell viability in Bacillus subtilis.* J Bacteriol, 2006. **188**(23): p. 8313-6.
134. Weidenmaier, C., et al.; *Role of teichoic acids in Staphylococcus aureus nasal colonization, a major risk factor in nosocomial infections.* Nat Med, 2004. **10**(3): p. 243-5.
135. Campbell, J., et al.; *Synthetic lethal compound combinations reveal a fundamental connection between wall teichoic acid and peptidoglycan biosyntheses in Staphylococcus aureus.* ACS Chem Biol, 2011. **6**(1): p. 106-16.
136. Weidenmaier, C., et al.; *Lack of wall teichoic acids in Staphylococcus aureus leads to reduced interactions with endothelial cells and to attenuated virulence in a rabbit model of endocarditis.* J Infect Dis, 2005. **191**(10): p. 1771-7.

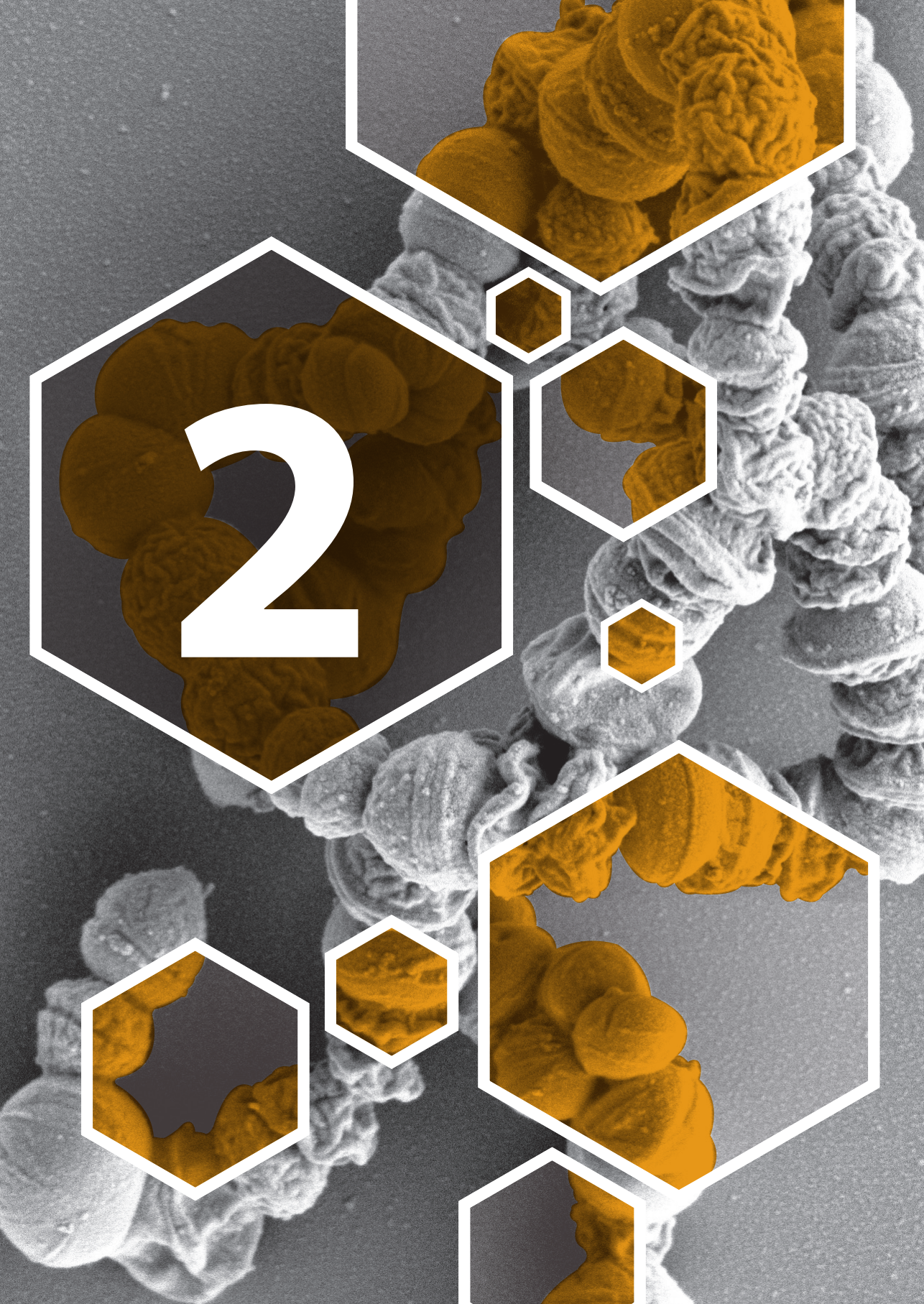
137. Winstel, V., et al.; *Wall Teichoic Acid Glycosylation Governs Staphylococcus aureus Nasal Colonization.* MBio, 2015. **6**(4).
138. Adibekian, A., et al.; *Comparative bioinformatics analysis of the mammalian and bacterial glycomes.* Chem. Sci., 2011. **2**(2): p. 337-344.
139. Xu, Y., et al.; *Analysis of a gene cluster of Enterococcus faecalis involved in polysaccharide biosynthesis.* Infect Immun, 2000. **68**(2): p. 815-23.
140. Rigottier-Gois, L., et al.; *The surface rhamnopolysaccharide epa of Enterococcus faecalis is a key determinant of intestinal colonization.* J Infect Dis, 2015. **211**(1): p. 62-71.
141. Li, W., et al.; *rmlB and rmlC genes are essential for growth of mycobacteria.* Biochem Biophys Res Commun, 2006. **342**(1): p. 170-8.
142. Ma, Y., Pan, F., and McNeil, M.; *Formation of dTDP-rhamnose is essential for growth of mycobacteria.* J Bacteriol, 2002. **184**(12): p. 3392-5.
143. Engels, W., et al.; *Role of lipopolysaccharide in opsonization and phagocytosis of Pseudomonas aeruginosa.* Infect Immun, 1985. **49**(1): p. 182-9.
144. Joiner, K.A.; *Complement evasion by bacteria and parasites.* Annu Rev Microbiol, 1988. **42**: p. 201-30.
145. Dong, C., et al.; *A structural perspective on the enzymes that convert dTDP-d-glucose into dTDP-l-rhamnose.* Biochem Soc Trans, 2003. **31**(Pt 3): p. 532-6.
146. Dong, C., et al.; *High-resolution structures of RmlC from Streptococcus suis in complex with substrate analogs locate the active site of this class of enzyme.* Structure, 2003. **11**(6): p. 715-23.
147. van der Beek, S.L., et al.; *GacA is Essential for Group A Streptococcus and Defines a New Class of Monomeric dTDP-4-dehydrorhamnose Reductases (RmlD).* Mol Microbiol, 2015.
148. Giraud, M.F., et al.; *RmlC, the third enzyme of dTDP-L-rhamnose pathway, is a new class of epimerase.* Nat Struct Biol, 2000. **7**(5): p. 398-402.
149. Beis, K., et al.; *The structure of NADH in the enzyme dTDP-d-glucose dehydratase (RmlB).* J Am Chem Soc, 2003. **125**(39): p. 11872-8.
150. Kantardjieff, K.A., et al.; *Mycobacterium tuberculosis RmlC epimerase (Rv3465): a promising drug-target structure in the rhamnose pathway.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 5): p. 895-902.
151. Allard, S.T., et al.; *The crystal structure of dTDP-D-Glucose 4,6-dehydratase (RmlB) from Salmonella enterica serovar Typhimurium, the second enzyme in the dTDP-l-rhamnose pathway.* J Mol Biol, 2001. **307**(1): p. 283-95.
152. Allard, S.T., et al.; *Toward a structural understanding of the dehydratase mechanism.* Structure, 2002. **10**(1): p. 81-92.
153. Dong, C., et al.; *RmlC, a C3' and C5' carbohydrate epimerase, appears to operate via an intermediate with an unusual twist boat conformation.* J Mol Biol, 2007. **365**(1): p. 146-59.
154. Blankenfeldt, W., et al.; *The structural basis of the catalytic mechanism and regulation of glucose-1-phosphate thymidylyltransferase (RmlA).* EMBO J, 2000. **19**(24): p. 6652-63.
155. Babaoglu, K., et al.; *Novel inhibitors of an emerging target in Mycobacterium tuberculosis; substituted thiazolidinones as inhibitors of dTDP-rhamnose synthesis.* Bioorg Med Chem Lett, 2003. **13**(19): p. 3227-30.
156. Ma, Y., et al.; *Drug targeting Mycobacterium tuberculosis cell wall synthesis: genetics of dTDP-rhamnose synthetic enzymes and development of a microtiter plate-based screen for inhibitors of conversion of dTDP-glucose to dTDP-rhamnose.* Antimicrob Agents Chemother, 2001. **45**(5): p. 1407-16.
157. Ren, J.X., et al.; *Virtual screening for the identification of novel inhibitors of Mycobacterium tuberculosis cell wall synthesis: inhibitors targeting RmlB and RmlC.* Comput Biol Med, 2015. **58**: p. 110-7.
158. Sivendran, S., et al.; *Identification of triazinoindol-benzimidazolones as nanomolar inhibitors of the Mycobacterium tuberculosis enzyme TDP-6-deoxy-d-xylo-4-hexopyranosid-4-ulose 3,5-epimerase (RmlC).* Bioorg Med Chem, 2010. **18**(2): p. 896-908.

159. Wang, Y., et al.; *Novel inhibitors of Mycobacterium tuberculosis dTDP-6-deoxy-L-lyxo-4-hexulose reductase (RmlD) identified by virtual screening.* Bioorg Med Chem Lett, 2011. **21**(23): p. 7064-7.
160. Wang, H., et al.; *Discovery of wall teichoic acid inhibitors as potential anti-MRSA beta-lactam combination agents.* Chem Biol, 2013. **20**(2): p. 272-84.
161. Ling, L.L., et al.; *A new antibiotic kills pathogens without detectable resistance.* Nature, 2015. **517**(7535): p. 455-9.
162. Marques, M.B., et al.; *Functional activity of antibodies to the group B polysaccharide of group B streptococci elicited by a polysaccharide-protein conjugate vaccine.* Infect Immun, 1994. **62**(5): p. 1593-9.
163. Sabharwal, H., et al.; *Group A streptococcus (GAS) carbohydrate as an immunogen for protection against GAS infection.* J Infect Dis, 2006. **193**(1): p. 129-35.
164. Pandey, M., Batzloff, M.R., and Good, M.F.; *Vaccination against rheumatic heart disease: a review of current research strategies and challenges.* Curr Infect Dis Rep, 2012. **14**(4): p. 381-90.
165. Steer, A.C., et al.; *Group A streptococcal vaccines: facts versus fantasy.* Curr Opin Infect Dis, 2009. **22**(6): p. 544-52.
166. Steer, A.C., Dale, J.B., and Carapetis, J.R.; *Progress toward a global group a streptococcal vaccine.* Pediatr Infect Dis J, 2013. **32**(2): p. 180-2.
167. Karakawa, W.W., Osterland, C.K., and Krause, R.; *Detection of Streptococcal Group-Specific Antibodies in Human Sera.* J Exp Med, 1965. **122**: p. 195-205.
168. Ayoub, E.M., Hawthorne, T., and Miller, J.; *Assay for antibodies to group C and G streptococcal carbohydrate by enzyme-linked immunosorbent assay.* J Lab Clin Med, 1986. **107**(3): p. 204-9.
169. Araj, G.F., et al.; *Immunoglobulin isotype response to the group-A streptococcal carbohydrate in humans.* APMIS Suppl, 1988. **3**: p. 8-12.
170. Rizkallah, M.F., Hoffler, E., and Ayoub, E.M.; *Serological confirmation of group C streptococcal pneumonia.* J Infect Dis, 1988. **158**(5): p. 1092-4.
171. Zimmerman, R.A., Auernheimer, A.H., and Taranta, A.; *Precipitating antibody to group A streptococcal polysaccharide in humans.* J Immunol, 1971. **107**(3): p. 832-41.
172. Schmidt, W.C. and Moore, D.J.; *The Determination of Antibody to Group a Streptococcal Polysaccharide in Human Sera by Hemagglutination.* J Exp Med, 1965. **121**: p. 793-806.
173. Riesen, W.F., Skvaril, F., and Braun, D.G.; *Natural infection of man with group A streptococci. Levels; restriction in class, subclass, and type; and clonal appearance of polysaccharide-group-specific antibodies.* Scandinavian journal of immunology, 1976. **5**(4): p. 383-90.
174. Salvadori, L.G., et al.; *Group A streptococcus-liposome ELISA antibody titers to group A polysaccharide and opsonophagocytic capabilities of the antibodies.* J Infect Dis, 1995. **171**(3): p. 593-600.
175. Zabriskie, J.B., et al.; *Phagocytic, serological, and protective properties of streptococcal group A carbohydrate antibodies.* Adv Exp Med Biol, 1997. **418**: p. 917-9.
176. Ayoub, E.M. and Dudding, B.A.; *Streptococcal group A carbohydrate antibody in rheumatic and nonrheumatic bacterial endocarditis.* J Lab Clin Med, 1970. **76**(2): p. 322-32.
177. Appleton, R.S., et al.; *Specificity of persistence of antibody to the streptococcal group A carbohydrate in rheumatic valvular heart disease.* J Lab Clin Med, 1985. **105**(1): p. 114-9.
178. Cunningham, M.W.; *Streptococcus and rheumatic fever.* Curr Opin Rheumatol, 2012. **24**(4): p. 408-16.
179. Dudding, B.A. and Ayoub, E.M.; *Persistence of Group a Streptococcal Carbohydrate Antibodies in Patients with Chronic Inactive Rheumatic Heart Disease.* Journal of Clinical Investigation, 1968. **47**(6): p. A26-&.
180. Cunningham, M.W., et al.; *Human monoclonal antibodies reactive with antigens of the group A Streptococcus and human heart.* J Immunol, 1988. **141**(8): p. 2760-6.
181. Malkiel, S., et al.; *T-Cell-dependent antibody response to the dominant epitope of streptococcal polysaccharide, N-acetyl-glucosamine, is cross-reactive with cardiac myosin.* Infect Immun, 2000. **68**(10): p. 5803-8.

182. Lancefield, R.C., McCarty, M., and Everly, W.N.; *Multiple mouse-protective antibodies directed against group B streptococci. Special reference to antibodies effective against protein antigens.* J Exp Med, 1975. **142**(1): p. 165-79.
183. Shigeoka, A.O., et al.; *Protective efficacy of hybridoma type-specific antibody against experimental infection with group-B Streptococcus.* J Infect Dis, 1984. **149**(3): p. 363-72.
184. Gotoff, S.P., et al.; *Human IgG antibody to group b Streptococcus type III: comparison of protective levels in a murine model with levels in infected human neonates.* J Infect Dis, 1986. **153**(3): p. 511-9.
185. Anthony, B.F., Concepcion, N.F., and Concepcion, K.F.; *Human antibody to the group-specific polysaccharide of group B Streptococcus.* J Infect Dis, 1985. **151**(2): p. 221-6.
186. Yoder, M.C. and Polin, R.A.; *Immunotherapy of neonatal septicemia.* Pediatr Clin North Am, 1986. **33**(3): p. 481-501.
187. Raff, H.V., et al.; *Human monoclonal antibodies to group B streptococcus. Reactivity and in vivo protection against multiple serotypes.* J Exp Med, 1988. **168**(3): p. 905-17.
188. Russell, M.W., et al.; *A Caries Vaccine? The state of the science of immunization against dental caries.* Caries Res, 2004. **38**(3): p. 230-5.
189. Smith, G.E.; *The control of dental caries: can vaccines prevent cavities?* Sci Total Environ, 1987. **63**: p. 1-11.
190. Talbman, M.A. and Smith, D.J.; *Effects of local immunization with Streptococcus mutans on induction of salivary immunoglobulin A antibody and experimental dental caries in rats.* Infect Immun, 1974. **9**(6): p. 1079-91.
191. Michalek, S.M. and McGhee, J.R.; *Effective immunity to dental caries: passive transfer to rats to antibodies to Streptococcus mutans elicits protection.* Infect Immun, 1977. **17**(3): p. 644-50.
192. McGhee, J.R., et al.; *Effective immunity to dental caries: studies of active and passive immunity to Streptococcus mutans in malnourished rats.* J Dent Res, 1976. **55 Spec No**: p. C206-14.
193. Morisaki, I., et al.; *Effective immunity to dental caries: enhancement of salivary anti-Streptococcus mutans antibody responses with oral adjuvants.* Infect Immun, 1983. **40**(2): p. 577-91.
194. Van de Rijn, I., Bleiweis, A.S., and Zabriskie, J.B.; *Antigens in Streptococcus mutans cross reactive with human heart muscle.* J Dent Res, 1976. **55 Spec No**: p. C59-64.
195. Fukuizumi, T., et al.; *Tonsillar application of killed Streptococcus mutans induces specific antibodies in rabbit saliva and blood plasma without inducing a cross-reacting antibody to human cardiac muscle.* Infection and Immunity, 1997. **65**(11): p. 4558-4563.
196. Fukuizumi, T., et al.; *Streptococcus sobrinus antigens that react to salivary antibodies induced by tonsillar application of formalin-killed S-sobrinus in rabbits.* Infection and Immunity, 2000. **68**(2): p. 725-731.
197. Kato, H., et al.; *Monoclonal-Antibody to Streptococcus-Mutans Type-E Cell-Wall Polysaccharide Antigen.* Infection and Immunity, 1986. **52**(2): p. 628-630.
198. Bammann, L.L. and Gibbons, R.J.; *Immunoglobulin-a Antibodies Reactive with Streptococcus-Mutans in Saliva of Adults, Children, and Predentate Infants.* Journal of Clinical Microbiology, 1979. **10**(4): p. 538-543.
199. Russell, M.W., et al.; *Serum Antibody-Responses to Streptococcus Mutans Antigens in Humans Systemically Infected with Oral Streptococci.* Oral Microbiology and Immunology, 1992. **7**(6): p. 321-325.
200. Wachsmann, D., et al.; *Serum and Salivary Antibody-Responses in Rats Orally Immunized with Streptococcus-Mutans Carbohydrate Protein Conjugate Associated with Liposomes.* Infection and Immunity, 1986. **52**(2): p. 408-413.
201. Bruyere, T., et al.; *Local response in rat to liposome-associated Streptococcus mutans polysaccharide-protein conjugate.* Vaccine, 1987. **5**(1): p. 39-42.
202. Lett, E., et al.; *Immunogenicity of polysaccharides conjugated to peptides containing T- and B-cell epitopes.* Infect Immun, 1994. **62**(3): p. 785-92.

203. Lett, E., et al.; *Mucosal Immunogenicity of Polysaccharides Conjugated to a Peptide or Multiple-Antigen Peptide-Containing T-Cell and B-Cell Epitopes.* Infection and Immunity, 1995. **63**(7): p. 2645-2651.
204. Koga, T., et al.; *Immunization against dental caries.* Vaccine, 2002. **20**(16): p. 2027-44.
205. Gorski, A., et al.; *Phage Therapy: What Have We Learned?* Viruses, 2018. **10**(6).
206. Oechslin, F.; *Resistance Development to Bacteriophages Occurring during Bacteriophage Therapy.* Viruses, 2018. **10**(7).
207. Shibata, Y., Yamashita, Y., and van der Ploeg, J.R.; *The serotype-specific glucose side chain of rhamnose-glucose polysaccharides is essential for adsorption of bacteriophage M102 to Streptococcus mutans.* FEMS Microbiol Lett, 2009. **294**(1): p. 68-73.
208. Friend, P.L. and Slade, M.D.; *Characteristics of group A streptococcal bacteriophages.* J Bacteriol, 1966. **92**(1): p. 148-54.
209. Krause, R.M.; *Studies on bacteriophages of hemolytic streptococci. I. Factors influencing the interaction of phage and susceptible host cell.* J Exp Med, 1957. **106**(3): p. 365-84.
210. Fischetti, V.A. and Zabriskie, J.B.; *Studies on streptococcal bacteriophages. II. Adsorption studies on group A and group C streptococcal bacteriophages.* J Exp Med, 1968. **127**(3): p. 489-505.
211. Delisle, A.L., et al.; *Biology and genome sequence of Streptococcus mutans phage M102AD.* Appl Environ Microbiol, 2012. **78**(7): p. 2264-71.
212. Araujo, P. and Krause, R.M.; *Group-Specific Carbohydrate of Group C-Variant Hemolytic Streptococci.* J Exp Med, 1963. **118**: p. 1059-67.


1


2

2

GacA is Essential for Group A *Streptococcus* and Defines a New Class of Monomeric dTDP-4-dehydrorhamnose Reductases (RmlD)

Samantha L. van der Beek[1], Yoann Le Breton[2], Andrew T. Ferenbach[3], Robert N. Chapman[4], Daan M.F. van Aalten[3], Iva Navratilova[5], Geert-Jan Boons[4], Kevin McIver[2], Nina M. van Sorge[1] and Helge C. Dorfmueller[3,6]

[1] University Medical Center Utrecht, Medical Microbiology, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

[2] Department of Cell Biology and Molecular Genetics, Maryland Pathogen Research Institute, University of Maryland, 3124 Biosciences Research Building, College Park, MD 20742, United States of America

[3] Division of Molecular Microbiology, University of Dundee, College of Life Sciences, Dow Street, DD1 5EH, Dundee, United Kingdom

[4] Complex Carbohydrate Research Center, Department of Chemistry, The University of Georgia, 315 Riverbend Road, Athens, USA

[5] Division of Biological Chemistry and Drug Discovery, University of Dundee, College of Life Sciences, Dow Street, DD1 5EH, Dundee, United Kingdom

[6] Rutherford Appleton Laboratory, Research Complex at Harwell, Didcot OX11 0FA, United Kingdom

Published in: Mol. Microbiol. (2015) DOI: 10.1111/mmi.13169

ABSTRACT

The sugar nucleotide dTDP-L-rhamnose is critical for the biosynthesis of the Group A Carbohydrate, the molecular signature and virulence determinant of the human pathogen Group A *Streptococcus* (GAS). The final step of the four-step dTDP-L-rhamnose biosynthesis pathway is catalyzed by dTDP-4-dehydrorhamnose reductases (RmlD). RmlD from the Gram-negative bacterium *Salmonella* is the only structurally characterized family member and requires metal-dependent homo-dimerization for enzymatic activity. Using a biochemical and structural biology approach, we demonstrate that the only RmlD homologue from GAS, previously renamed GacA, functions in a novel monomeric manner. Sequence analysis of 213 Gram-negative and Gram-positive RmlD homologues predicts that enzymes from all Gram-positive species lack a dimerization motif and function as monomers. The enzymatic function of GacA was confirmed through heterologous expression of *gacA* in a *S. mutans rmlD* knockout, which restored attenuated growth and aberrant cell division. Finally, analysis of a saturated mutant GAS library using Tn-sequencing and generation of a conditional-expression mutant identified gacA as an essential gene for GAS. In conclusion, GacA is an essential monomeric enzyme in GAS and representative of monomeric RmlD enzymes in Gram-positive bacteria and a subset of Gram-negative bacteria. These results will help future screens for novel inhibitors of dTDP-L-rhamnose biosynthesis.

INTRODUCTION

The cell wall of Gram-positive bacteria is an intricate network of peptidoglycan, proteins, and secondary cell wall polymers (SCWP) that are covalently linked to peptidoglycan. Teichoic or teichuronic acids are typical and well-studied SCWP in Gram-positive bacteria and play an important role in normal cell function and infection [1]. Many β-hemolytic streptococcal species appear to lack expression of typical teichoic or teichuronic acid structures [2, 3] and instead express a rhamnose-rich polymer, which comprises approximately half of the cell wall mass [4]. Historically, expression of these evolutionary conserved glycans underlies classification of β-hemolytic streptococci in Lancefield groups (A, B, C, G, ...) [5], a feature that is still applied in contemporary rapid test kits to diagnose streptococcal infections.

Streptococcus pyogenes, also referred to as Group A *Streptococcus* (GAS), is a β-hemolytic human-restricted pathogen and ranks in the top ten of infection-related causes of mortality worldwide [6]. GAS is the causative agent of a wide spectrum of clinical disease, including common localized infections (~ 700 million cases per year worldwide) and approximately 1.8 million cases of severe disease [6], including necrotizing fasciitis, streptococcal toxic shock syndrome, and post-infectious streptococcal sequelae, i.e. acute rheumatic fever and rheumatic heart disease. A better understanding of GAS pathogenesis and development of new drugs and protective vaccines is crucial. GAS expresses a characteristic SCWP known as Lancefield Group A Antigen or Group A Carbohydrate (GAC) [5]. The GAC structure consists of a polyrhamnose core decorated with alternating immunodominant *N*-acetylglucosamine (GlcNAc) side chains [7]. Recently, van Sorge *et al.* identified the gene cluster responsible for GAC biosynthesis and demonstrated that the GAC GlcNAc side chain contributes to GAS virulence [8].

In contrast to detailed insight into the biosynthesis of classical SCWP like teichoic acids, information regarding the biosynthesis of rhamnose-rich polysaccharides like GAC is limited. The production of dTDP-L-rhamnose is critical for GAC biosynthesis but also more broadly for the viability or virulence of other medically-important bacteria including *Mycobacterium spp*. [9], *Pseudomonas spp*. [10], and *Enterococcus faecalis* [11]. L-rhamnose is synthesized from α-glucose-1-phosphate through a four-step enzymatic process catalyzed by the enzymes RmlA-D [12, 13]. Structural analysis of RmlA-D from *Salmonella enterica* (*Se*) has provided valuable insight into the enzymatic mechanism of action for these enzymes, including requirement for Mg^{2+}-dependent dimerization in the case of RmlD [14]. In general, RmlD enzymes are members of the large short-chain dehydro-genases/reductases (SDR) superfamily,

which act on a wide family of substrates and commonly form homo-dimeric or multimeric complexes [15].

In GAS, homologues of RmlA, B and C that catalyze the first steps of the dTDP-rhamnose biosynthesis pathway can be identified through bioinformatics and are clustered on the genome. The GAS RmlD homologue appears to be encoded by the *gacA* gene, which is annotated as a dTDP-4-dehydrorhamnose reductase, but experimental data supporting this function is currently lacking. The goal of this study was to identify the function and structure of the *gacA* gene product through biochemistry, structural biology and bacterial genetics. We show that *gacA* is an essential gene of GAS that encodes a metal-independent dTDP-4-dehydrorhamnose reductase representative of a new class of monomeric RmlD enzymes.

RESULTS AND DISCUSSION

GacA encodes a functional metal-independent dTDP-4-dehydrorhamnose reductase (RmlD)

Bioinformatics analysis suggests that *gacA* encodes a dTDP-4-dehydrorhamnose reductase, an enzyme that catalyzes the final step in the production of dTDP-L-rhamnose [16]. In contrast to the *rmlD* genes in other species like *Shigella flexneri* [17] and *Streptococcus pneumoniae* serotype 19F [18], the only GAS *rmlD* homologue *gacA* is not part of an *rmlABCD* rhamnose biosynthesis operon. Instead, *gacA* is located at the beginning of the recently identified GAC gene cluster and hence named *gacA* [8]. A similar split genomic architecture of the rhamnose biosynthesis genes *rmlA-C* and *rmlD* was previously observed in *Streptococcus mutans* (*S. mutans*), a cariogenic Gram-positive bacterium [19].

We set out to investigate the potential function of GacA as a dTDP-4-dehydrorhamnose reductase. We cloned and expressed full-length GacA fused to a cleavable GST-His_6 tag. The GAS GacA protein sequence is 36% identical to the *Salmonella enterica* serovar Typhimurium RmlD protein (*Se*RmlD) (Fig. 1), the only reported RmlD structure [14]. To confirm the enzymatic activity of GacA *in vitro*, we also cloned, expressed and purified the putative RmlB and RmlC GAS homologues, since the RmlD substrate dTDP-4-dehydrorhamnose is not commercially available. In this biochemical assay, oxidation of NADPH is a read-out for GacA activity. No activity was observed with any combination of two enzymes (Fig. 2A). However, we observed significant oxidation of NADPH when all three enzymes were present, suggesting that GacA is indeed acting as a dTDP-4-dehydrorhamnose reductase

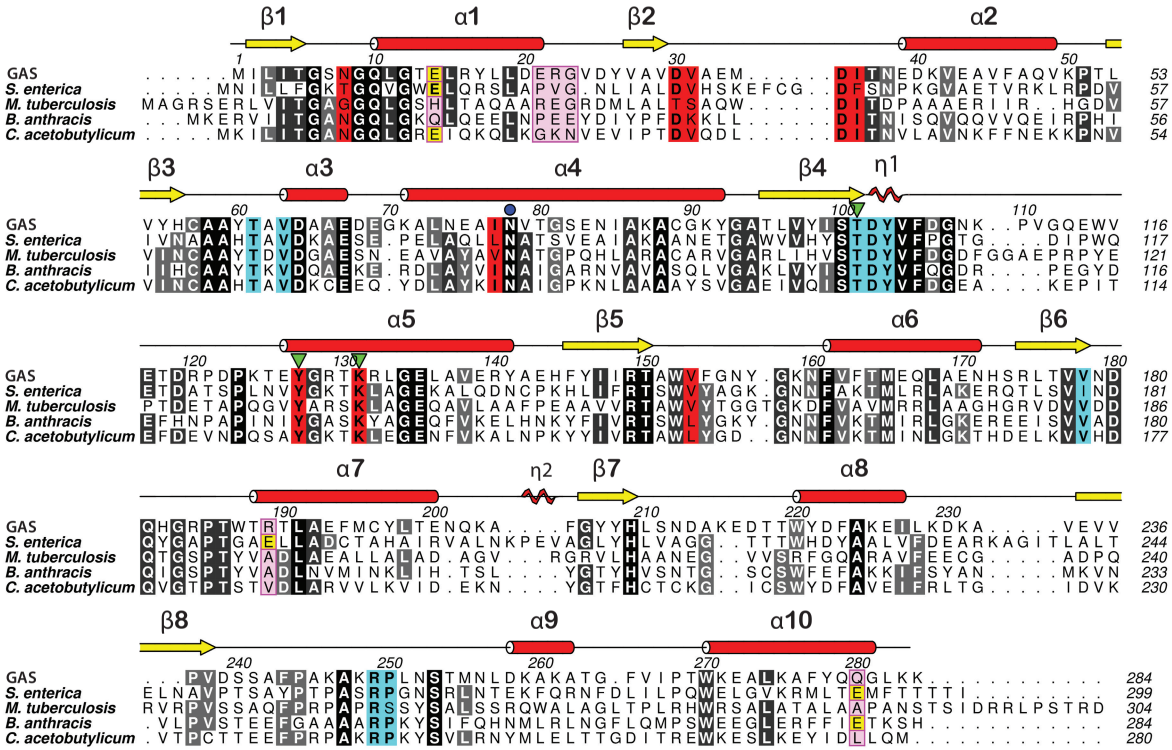


Figure 1. Sequence alignment for GacA and RmlD homologues. Sequence alignment of GAS GacA with RmlD from *Salmonella enterica* serovar Typhimurium [14], *Mycobacterium tuberculosis* (accession number WP_009938025), the hypothetical RmlD homologues from *Bacillus anthracis* str. Ames (accession number NP_843703) and *Clostridium acetobutylicum* (accession number WP_010965612). Conserved residues are colored in black (>80%) and grey (60-80%). Substrate binding site residues are colored in turquois and co-factor binding site residues in red. Secondary structure elements from GacA are indicated and labeled with red α-helices and yellow β-strands. Mg^{2+}-binding site residues from *Se*RmlD are colored in yellow (E1-E3) and the corresponding non-conserved residues in the other species are colored in magenta (R1-R3). The terminal residues of the α1-helix motif are highlighted with magenta boxes. N78 is indicated with a blue dot and the catalytic triad is indicated with green triangles.

(Fig. 2A). As mentioned above, the enzyme activity of *Se*RmlD was shown to be metal dependent [14]. Similarly, RmlD from *Mycobacterium tuberculosis* (*Mtb*RmlD) was assayed in presence of 10 mM $MgCl_2$, despite the lack of biochemical evidence that the enzyme requires a divalent cation for enzymatic activity [20]. To investigate the metal-dependent activity of GacA, we performed the assay in presence and absence of 10 mM $MgCl_2$ and / or 10 mM EDTA. No significant changes in enzymatic activity were observed for GacA (Fig. 2A), demonstrating that unlike *Se*RmlD [14], the proper positioning of the co-factor is not dependent on a metal-ion [21]. The GacA enzyme activity assay was subsequently performed in absence of $MgCl_2$.

Next, we studied Michaelis-Menten kinetics of GacA. The K_m^{app} value for dTDP-α-glucose was determined to be 370 μM (Fig. 2B), in agreement to the 110 μM K_m value determined for the *Se*RmlD homologue [14]. Using surface plasmon resonance, we investigated the binding affinity of NADPH and NADH to GacA, which were previously shown to be functional co-factors for RmlD enzymes [14]. Both substrates bind to GacA with binding affinities of K_D = 390 ± 2 μM (NADPH) and K_D = 110 μM ± 1 μM (NADH) (Fig. 2C).

GacA is active as a monomer

Many SDR family members, including *Se*RmlD, require dimerization or oligomerization to be functional [14, 15]. Results from our functional enzymatic assay demonstrated that GacA did not require metal for its activity, suggesting that it might be functional as a monomer. GacA was analyzed by size-exclusion chromatography (SEC) and SEC-MALLS and a molecular mass of the elution peak was calculated. The calculated mass of GacA is 27.5 kDa (+/- 2.4 kDa) (Fig. 2D). This is in good agreement with the theoretical calculated monomeric mass of 32 kDa of the GacA polypeptide sequence and represents the first monomeric RmlD enzyme. We subsequently investigated the effect of Mg^{2+}-ions on protein size. The purified protein was incubated with and without 10 mM $MgCl_2$ overnight at 4 ˚C and analyzed via SEC in the corresponding buffers (Fig. 2D). Both protein samples show identical elution volumes at 15.8 ml, with an average calculated molecular mass of 27.5 kDa, indicating that no mass change / dimerization occurs in presence of Mg^{2+} (Fig. 2D).

Mycobacterium tuberculosis RmlD inhibitors inhibit recombinant GacA

Wang *et al.* have identified a series of *Mtb*RmlD inhibitors by virtual screening using the crystal structure of *Se*RmlD [20]. The sequence identity between *Mtb*RmlD and *Se*RmlD is 31%, whilst the sequence identity between GacA and *Mtb*RmlD is 36% (Fig. 1). We assessed the inhibitory effect of two of the four previously described

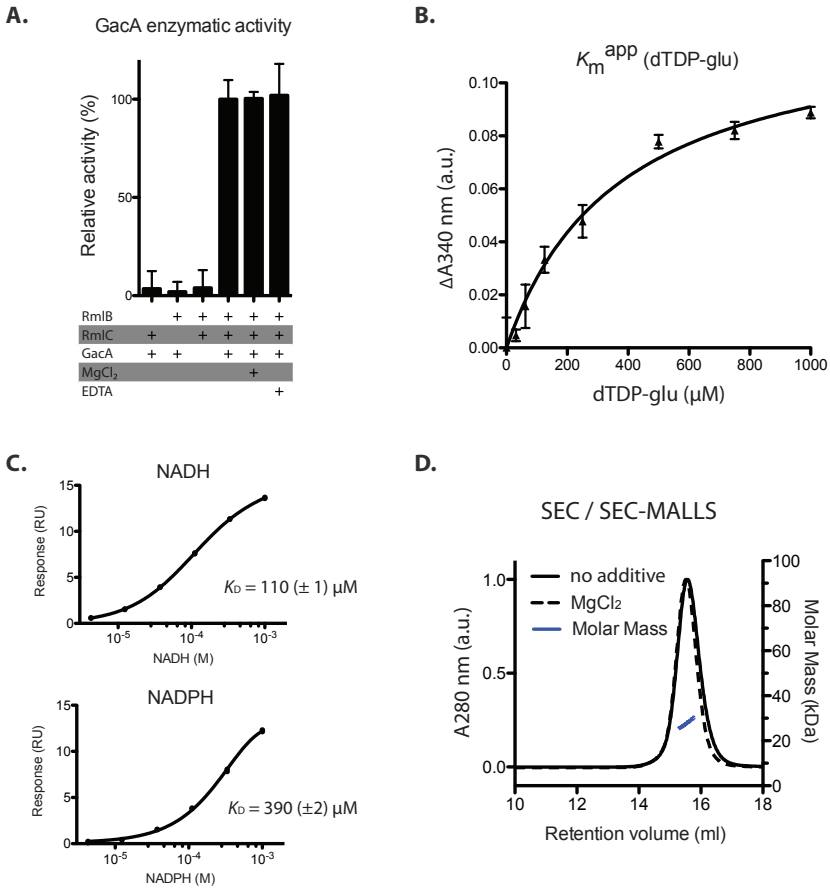


Figure 2. GacA is a functional metal-independent monomeric dTDP-4-dehydrorhamnose reductase. (A) Biochemical confirmation that GacA is a functional and metal-independent dTDP-4-dehydrorhamnose reductase. Read-out of GacA enzyme activity is change in absorbance at 340 nm, indicating oxidation of the GacA co-factor NADPH to $NADP^+$, was observed. GacA is only active in the presence of coupled enzymes *Sp*RmlB and *Sp*RmlC, which provide the substrate for GacA. GacA metal dependence was investigated in presence and absence of 10 mM $MgCl_2$ and 10 mM EDTA. **(B)** GacA Michaelis-Menten kinetics were determined for dTDP-glucose (K_m value of 370 μM). The second substrate NADPH was present in excess. **(C)** Surface plasmon resonance sensorgrams (insets) and equilibrium fit for the binding of NADPH and NADH to GacA. NADPH and NADH were injected over a concentration range of 4.1 μM to 1000 μM. Equilibrium affinity fits are shown in the bottom panel of the figure. **(D)** Size-exclusion chromatography (SEC) and SEC-MALLS of recombinant purified GacA in presence (dashed line) and absence (solid line) of 10 mM $MgCl_2$. All samples reveal the same retention volume, corresponding to a calculated average molecular mass of 27.5 kDa (blue line) demonstrating GacA is a monomer.

inhibitors using enzyme kinetics. Our experiments show that both compounds inhibit the activity of recombinant GacA with an IC_{50} of approx. 2 µM for compound 3 and approx. 10 µM for compound 2. This data is in good agreement to the *Mtb*RmlD inhibition with corresponding IC_{50} values of 0.9 µM and 15 µM, respectively [20].

GacA crystal structure reveals novel monomeric RmlD form with a conserved catalytic triad

To unravel why GacA functions as a monomer, we employed X-ray crystallography. Crystals diffracted routinely below 1.2 Å and were subjected to synchrotron data collection. Results were compared to the four GacA (RmlD) homologues that have been deposited in the Protein Data Bank (PDB) [22] (Supp. Fig. 1): The Gram-negative *Se*RmlD (pdb entry 1kbz, 1n2s, 1kc1, 1kc3) [14], *Clostridium acetobutylicum* RmlD (*Ca*RmlD; pdb entry 1vl0), the Gram-positive *Bacillus anthracis* RmlD (*Ba*RmlD; pdb entry 3sc6) and the archaea *Sulfolobus tokodaii* RmlD (*St*RmlD; pdb entry 2ggs).

Consistent with our SEC and activity studies, GacA crystallized as a monomer. The overall structure of the GacA monomer aligns well with all four RmlD structures in the database, with Cα rmsd of 1.4 Å (apo, *Se*RmlD, 1kbz, Fig. 3A), 1.1 Å (NADPH complex, *Ba*RmlD, 3sc6) and 1.0 Å (NADH complex, *Ca*RmlD, 1vl0) and 1.4 Å (NADPH complex, *St*RmlD, 2ggs, Supp. Fig. 1). GacA contains the typical α-helical/β-sheet arrangement known as Rossmann fold present in SDR domains (Fig 3A). In agreement with the well-characterized *Se*RmlD [14], the GacA β-sheet is formed of six β-strands, in the order 213457 (Fig. 3A). This β-sheet is flanked by three and four α-helices on either side, respectively. The previously described prominent kink in α-helix 4, a characteristic structural feature in SDR enzymes, is also present in GacA and is caused by the conserved Asn78 (Asn81 in *Se*RmlD) (Fig. 1, 3B). The second 3_{10}-helix η2 observed in *Se*RmlD [14] is missing in GacA due to a shorter loop connecting α-helix α7 and β-strand β7 (Figs. 1, 3A). The α-helices α1-5, α7 and α9 form the co-factor binding site, while α-helices α6, α8 and α10 form the substrate-binding domain (Fig. 3A, B). In comparison to *Se*RmlD, α8 from GacA lacks two α-helical turns, due to a deletion of eight amino acids in GacA (Fig. 1, 3A). Furthermore, we calculated the hydrodynamic radius (Stokes radius) of the soluble GacA and the crystallized monomer. The soluble GacA has a Rh = 26 Å (+/- 0.3%), in good agreement with the Rh calculated from the obtained crystal structure using HYDROPRO (Rh = 26 Å) [23]. No symmetry related GacA molecule forms protein-protein interactions via α1 and α10 as observed in *Se*RmlD (Fig. 3A) [14]. The GacA crystal structure confirms that GacA defines a new class of monomeric RmlD enzymes that do not require metal for enzymatic activity (Fig. 3A).

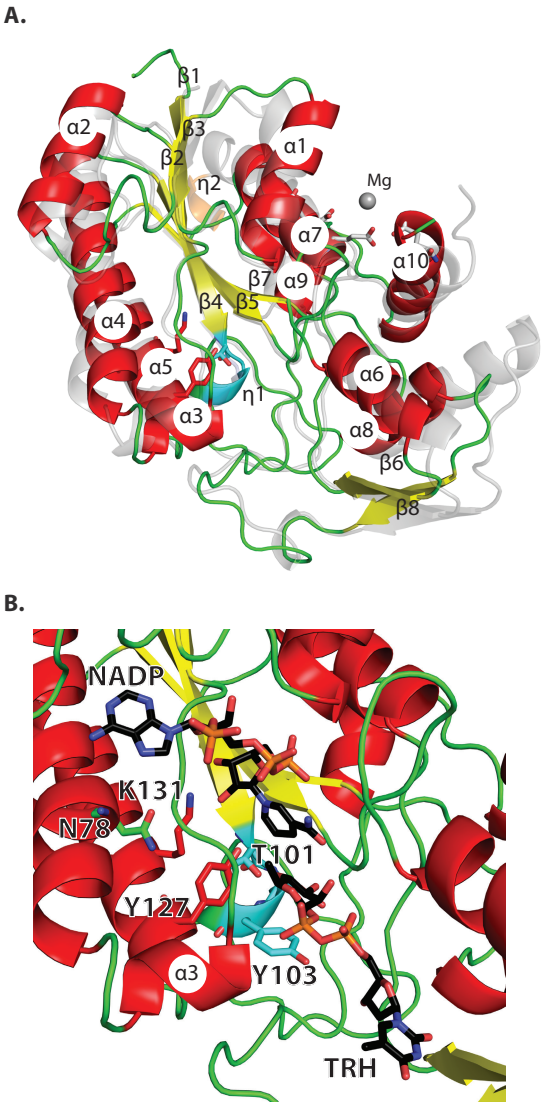


Figure 3. Structural insights into a monomeric Gram-positive RmlD enzyme. (A) Comparison of the GacA secondary structure elements with *Se*RmlD (PDB entry 1KBZ, [14]), a dTDP-4-dehydrorhamnose reductase. The structures are shown in a cartoon representation. GacA is colored with red helices and yellow strands; *Se*RmlD is colored in transparent grey. Secondary structure elements are labeled according to Figure 1. *Se*RmlD Mg^{2+}-binding site is shown with the three glutamic acids (grey sticks) coordinating the Mg^{2+}-ion (grey sphere). **(B)**. Active site view with cartoon and stick representation of GacA in complex with superimposed ligands NADPH and dTDP-L-rhamnose (TRH) from the ternary *Se*RmlD complex (PDB entry 1KC3, [14]). The conserved catalytic residues T101, Y127 and K131 (GacA) and N78 are shown as sticks, color-coded according to Figure 1.

We superimposed the ternary *Se*RmlD complex [14] onto the GacA crystal structure to investigate active site sequence conservation (Fig. 1, 3B). The GacA crystal structure is in the 'open' conformation, allowing binding of the co-factor NADPH and the acceptor substrate dTDP-4-dehydrorhamnose. All active site residues are conserved with the *Se*RmlD (Fig. 1, green triangles). The catalytic triad identified for *Se*RmlD [14], consisting of T104, Y128 and K132, occupies identical conformations in GacA (T101, Y127 and K131, Fig. 1, 3B), suggesting that GacA is a functional monomeric RmlD homologue using a conserved catalytic mechanism.

RmlD enzymes from Gram-positive bacteria are monomers due to lack of a conserved RmlD dimerization motif

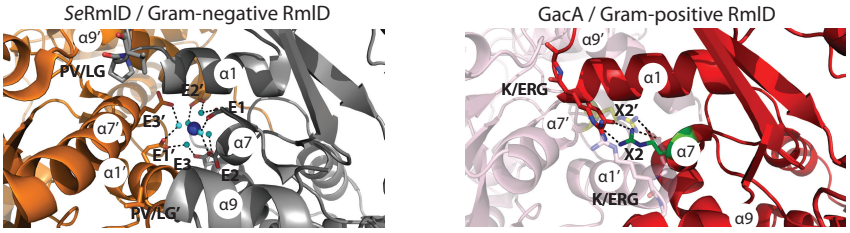
To investigate the molecular basis for the novel monomeric RmlD form, we focused on the amino acids at the *Se*RmlD dimerization interface. From the published *Se*RmlD structure [14], there appear to be two critical parameters for *Se*RmlD dimerization: 1) the three glutamate residues E15, E190, and E292 at the *Se*RmlD dimerization interface that help coordinate the Mg^{2+}-ion (Fig. 1, Fig. 4A), and 2) a shortened α1-helix caused by a proline residue (P22, *Se*RmlD) that allows binding of the second *Se*RmlD monomer via tight protein-protein interactions (Fig. 4A). To experimentally confirm the contribution of these two parameters to *Se*RmlD dimerization, we cloned, expressed and purified the wild-type RmlD enzyme from *Salmonella enterica* and designed a triple-mutant (3M), in which the putative key residues for *Se*RmlD dimerization (PVG, E, E) are replaced with the corresponding residues in the monomeric GacA (ERGV, R, Q). SEC analysis revealed that the *Se*RmlD triple mutant is a monomeric protein, identical to GacA (Supp. Fig 2A), whereas wild-type *Se*RmlD runs at a shorter retention time, in agreement to its higher dimeric mass [14]. Furthermore, EDTA treated *Se*RmlD elutes at the same retention time (Supp. Fig 2A), suggesting that the removal of the Mg^{2+} ion does not disrupt dimerisation. The peak fractions were analysed using fingerprint mass-spectrometry and contained the correct protein.

GacA lacks both of the confirmed critical dimerization parameters since it does not contain all conserved negatively-charged E residues and has an extended α1-helix due to the ERG motif that occupies the putative Mg^{2+}-binding pocket (Fig. 4A). Interestingly, the GacA structure reveals that a salt bridge is introduced (E21 and R189), stabilizing the tertiary structure of the α1- and α7-helices, which could contribute to the fact that GacA functions as a monomer (Fig. 4A). We have modeled an artificial GacA dimer (Fig. 4A), which further reveals structural features that support why GacA functions as a monomer. The large, basic 'R22' at the end of the α1-helix points with its side-chain into the modeled second monomer. Furthermore,

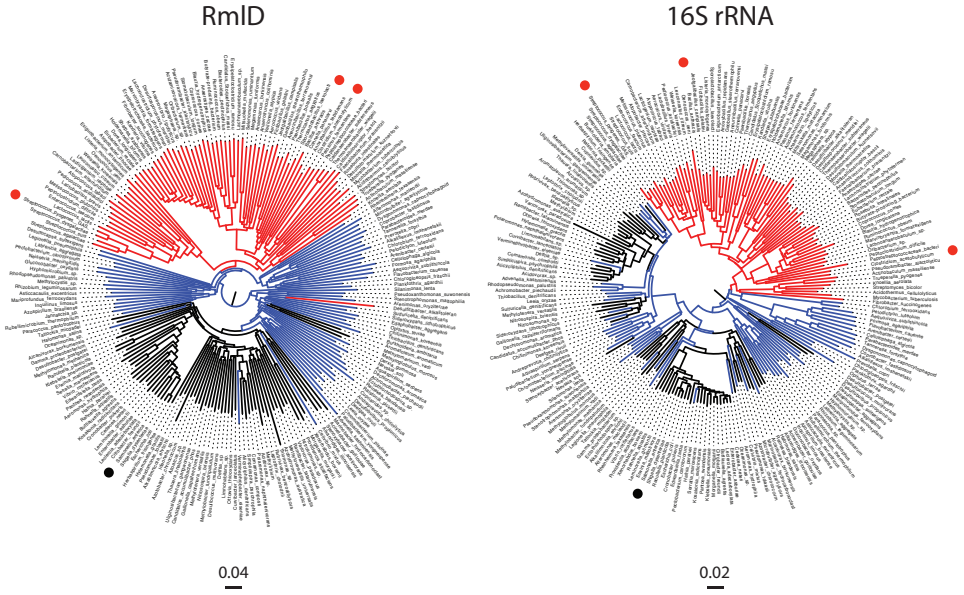
the extended α1-helix would clash with the α7-helix of the second GacA molecule (Fig. 4A). These features are incompatible with homo-dimerization, but most likely stabilize the monomeric enzyme.

We analyzed whether the lack of structural dimerization features is unique to GacA or are present in RmlD enzymes from other species. Using two independent psi-BLAST runs, one starting with *Se*RmlD as a representative of dimeric RmlD enzymes and the other starting with GacA as a representative of monomeric RmlD enzymes, we identified 213 bacterial (putative) RmlD homologues, including 78 from Gram-positive and 135 from Gram-negative species (Suppl. Tables 1-3). All full-length RmlD protein sequences as well as the corresponding 16S rRNA DNA sequences of the same 213 species were aligned and used to build bootstrapped Neighbor-Joining trees (Fig. 4B). Interestingly, RmlD sequences from all Gram-positive bacteria, except *Desulfitibacter alkalitolerans*, clustered separately from the RmlD sequences of Gram-negative bacteria. We aimed to determine whether the observed phylogenetic split was associated with sequence differences at the dimerization interface. Therefore, we aligned the 135 Gram-negative and 78 Gram-positive RmlD sequences corresponding to the critical negatively charged amino acids and the PLG sequence identified at the *Se*RmlD dimerization interface (Supp. Table 1-3). Interestingly, the RmlD sequences from Gram-negative species could be divided in two groups. The group named 'Gram-negative 1' contained the fully conserved dimerization site represented as an E1-PLG-E/D2-E3 motif (Fig. 4C 'Gram-negative 1') and similar to the previously described metal-dependent dimeric *Se*RmlD (Supp. Table 1). This strongly suggests that RmlD proteins from this subset of Gram-negative species form dimers. The remaining ~45% of Gram-negative RmlD enzymes lacked one or both of the critical dimerization parameters ('Gram-negative 2'; Fig. 4C, Supp. Table 2) and likely function as monomers similar to GacA. Superimposing the presence/absence of the dimerization motif on the 16S rRNA tree suggests that Gram-negative bacteria have gained or lost the RmlD dimerization motif multiple times during evolution (Fig. 4B). Strikingly, RmlD sequences from all Gram-positive species also lacked a conserved 'Gram-negative 1' dimerization motif (Fig. 4C 'Gram-negative 1' and 'Gram-positive'). For example, most Gram-positive RmlD enzymes replaced the negatively charged glutamate at 'E2' with a large positively charged arginine or aromatic tyrosine substitute ('X2' in Fig. 4C) and the position at 'E3' is also substituted in Gram-positive bacteria compared to 'Group 1' Gram-negative species ('X3' in Fig. 4C). In addition, RmlD in Gram-positive bacteria have an extended α1-helix by half a turn due to lack of the 'PLG' motif. This is illustrated in the artificial GacA dimer that we modeled based on the *Se*RmlD homo-dimer (Fig. 4A) as well as in the deposited RmlD crystal structures from the Gram-positive species *Clostridium acetobutylicum Ca*RmlD (1VL0.pdb, Suppl. Fig. 1) and *Bacillus*

A.


*Se*RmlD / Gram-negative RmlD
α9′
PV/LG
α1
E2
E1
E3′
α7′
α7
E1′
E3
α1′
E2′
PV/LG′
α9
GacA / Gram-positive RmlD
α9′
α1
K/ERG
X2′
α7′
X2
α7
α1′
K/ERG
α9

B.


RmlD
16S rRNA
0.04
0.02

C.

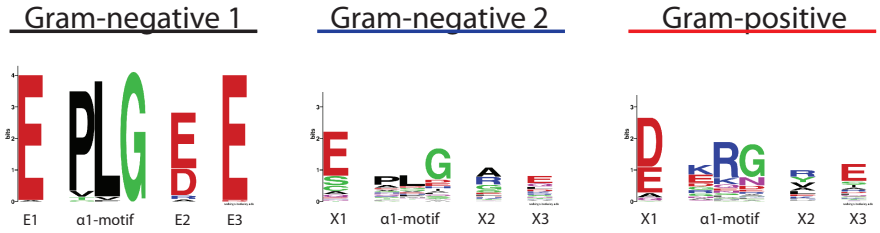

Gram-negative 1
Gram-negative 2
Gram-positive
E1
α1-motif
E2
E3
X1
α1-motif
X2
X3
X1
α1-motif
X2
X3

Figure 4. GacA represents a new class of monomeric RmlD enzymes. (A) Structural insight in dimerization interface of RmlD reveals steric hindrances that would prevent homo-dimerization in Gram-positive RmlD enzymes. Left: View of the *Se*RmlD homo-dimer Mg^{2+}-binding site, representing Gram-negative RmlD enzymes [14]. The E1-E/D2-E3 motif coordinates the Mg^{2+}-ion. The α1-helix contains the 'PLG' motif, which allows dimerization, since the 'P' introduces a turn into the α1-helix. Right: Hypothetical model of a GacA homo-dimer based on the *Se*RmlD homo-dimer (red and transparent cartoon representation). The 'K/ERG' motif prevents dimerization, since it extends the α1-helix by half a turn, which would clash with the α7'-helix of the second molecule. The X2 side-chain (R189) points into the Mg^{2+}-binding pocket and prevents dimerization (salmon and red arginine side chains). Large, basic or hydrophobic residues, mainly 'R' or 'Y', replace the 'E2' motif in Gram-positive RmlD homologues. A salt bridge between GacA E21 from the ERG motif and X2 (R189) stabilizes the tertiary structure of the α1- and α7-helices, which is located at the corresponding position of the Gram-negative Mg^{2+}-binding site (green and yellow side-chains). **(B)** Neighbor-Joining trees of 213 RmlD orthologous sequences (left) and their corresponding 16S rRNA sequences (right). Gram-positive bacteria are colored in red, Gram-negative that contain the conserved dimerization sequence logo E1-PLG-E/D2-E3 are colored in black ('Gram-negative 1'), and Gram-negative bacteria that contain a lack one or both of the dimerization criteria are colored in blue ('Gram-negative 2'). *Streptococcus pyogenes* (GAS), *Clostridium acetolyticum* and *Bacillus anthracis* are marked with red dots, *Salmonella enterica* with a black dot. **(C)** Left: Sequence logo of dimerization interface for Gram-negative RmlD enzymes. 135 (putative) orthologous RmlD sequences from Gram-negative bacteria (Supp. Table 1) were analyzed for their E1, E2, E3 and α1-helix motifs, which is critical for Mg^{2+}-binding and dimerization as described by Blankenfeldt *et al.* [14]. Seventy RmlD homologues contain the 'Gram-negative 1' motif 'E-PLG-E/D-E' (Supp. Table 1). Sixty-five RmlD homologues are lacking these conserved residues ('Gram-negative 2') and are therefore expected to function as monomers (Supp. Table 2). Right: Seventy-eight RmlD sequences from (putative) Gram-positive bacteria (Supp. Table 3) were analyzed for the same motifs. The Gram-positive RmlD homologues lack a distinctive motif and therefore lack the ability to coordinate a divalent metal ion. The 'Gram-negative 1' α1-helix 'PLG'-motif is replaced in Gram-positive RmlD sequences by a 'K/ERG'-motif.

*anthracis Ba*RmlD (3SC6.pdb, Suppl. Fig. 1). These enzymes have not been biochemically characterized, however, both proteins did not crystallize as dimers. Furthermore, these two enzymes are similar to the GacA structure and primary sequence (Fig. 1) containing large residues at the end of the α1-helix and missing the three conserved negatively charged E1-E/D2-E3 residues to accommodate a metal ion, hindering the *Se*RmlD typical dimerization (Supp. Fig. 1).

Overall, we confirmed that RmlD dimerization requires specific structural features that are absent in GacA. Additionally, comprehensive sequence analysis suggests that GacA is representative of a new class of monomeric RmlD enzymes that is present in all Gram-positive species and a subset of Gram-negative species.

GacA is an essential gene for GAS during growth in rich medium

To confirm the function of GacA in GAS, we attempted to generate mutants in GAS by plasmid insertion. However, we were unable to obtain mutants on multiple attempts, suggesting that *gacA* is essential for GAS [8]. Additional proof for potential essentiality of *gacA* was investigated as part of a larger screen for essential genes in GAS using the mariner transposon Krmit [24]. Saturated mutant libraries were produced and analyzed by Tn-seq to identify the insertion sites within each mutant pool that survive growth in THY rich medium at 37 ˚C. Chromosomal position and abundance of Tn-seq reads were mapped to the *gacA* genome sequence and a Bayesian statistical analysis was performed to identify regions with limited *Krmit* insertions compared to surrounding sequences indicative of gene essentiality (Table 1, 2). For known essential genes *dnaG* and *rpoD* < 100 insertions per kB are observed, whereas 10-50-fold more insertions are observed for non-essential control genes M5005_Spy_0601 and *emm1* (Fig. 5A, Tables 1, 2). Insertions for *gacA* demonstrate that *gacA* is indeed essential in the GAS strains 5448 (M1T1) and NZ131 (M49) when growing in rich media. This data is in agreement with a previous study conducted on *Mycobacterium smegmatis* [9], where *rmlD* was shown to be essential for mycobacterial growth. To validate *gacA* essentiality in an independent manner, we employed a previously published conditionally lethal approach that takes advantage of a theophylline-sensitive synthetic riboswitch functional in GAS [24]. In the presence of theophylline, which results in expression of *gacA*, the GAS gacAi strain was normally viable, whereas lack of theophylline significantly compromised growth of the bacteria (Fig. 5B). Visual inspection of the inducible *gacA* mutant bacteria (without theophylline) using scanning electron microscopy (SEM) indeed shows aberrant cell morphology, which defects in cell separation resulting in long chains and aberrant septum placement resulting in irregularly shaped cocci (Fig. 5C). These data underpin the critical role of rhamnose production in GAS physiology.

GAS GacA can functionally replace S. mutans RmlD – To confirm the role of GacA in dTDP-L-rhamnose production in live bacteria we made use of heterologous expression. It was previously shown that classical targeted disruption of *rmlD* in *S. mutans* is feasible [19, 25]. Rhamnose is incorporated in the serotyping cell wall-anchored carbohydrate composed of rhamnose decorated with a glucose sidechain [25]. Consequently, disruption of *rmlD* results in complete loss of rhamnose and glucose) from the cell wall but also significantly attenuates growth [19], ([25]. *Sm*RmlD and GacA are 82% identical (234 out of 284 residues), suggesting that they catalyze the same enzymatic reaction. We constructed a *S. mutans rmlD* mutant strain (SMU Δ*rmlD*) by replacing *rmlD* in frame with an Erm resistance cassette. Similar to disruption of *gacA* in GAS, loss of RmlD significantly affected cell morphology (Fig. 6A), bacterial growth (Fig. 6B) and resulted in complete loss of rhamnose (Fig.

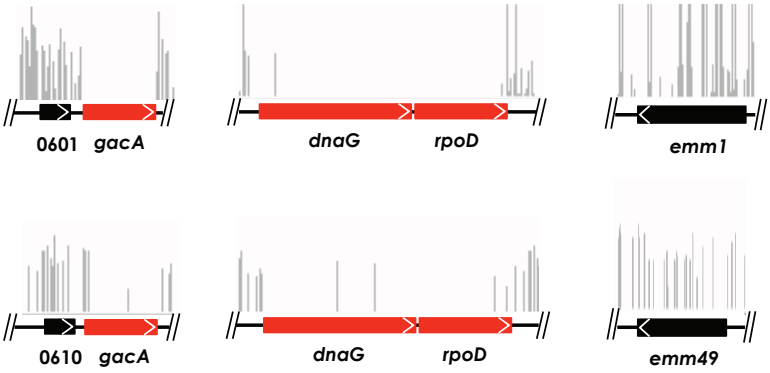
6C) as previously published [19]. We next complemented SMU Δ*rmlD* with GAS *gacA* on a complementation plasmid (SMU Δ*rmlD* + p*GacA*). Expression of GacA restored *S. mutans* growth (Fig. 6B), rhamnose production (Fig. 6C) and almost completely restored cell appearance (Fig. 6A). These results demonstrate that *gacA* can functionally replace *rmlD* in *S. mutans*, supporting the role of GacA as a dTDP-4-dehydrorhamnose reductase.

CONCLUSIONS

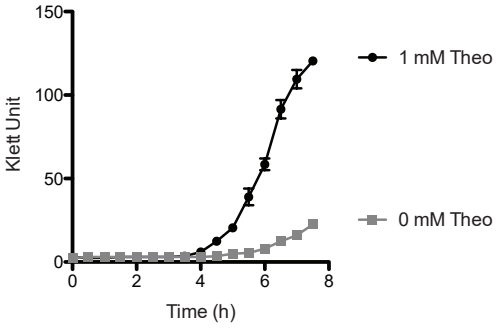
We have demonstrated through genetics, biochemistry, functional and structural analysis that GacA is the dTDP-4-dehydrorhamnose reductase RmlD homologue in GAS and is critical for normal growth. GacA represents the first structurally and biochemically characterized Gram-positive RmlD homologue. In contrast to the published RmlD structure from Gram-negative *Salmonella enterica* [14], GacA is representative of a new class of RmlD enzymes that are functional as a monomer. Bioinformatics analysis of of 213 RmlD sequences complemented with experimental evidence from wild-type and mutated *Se*RmlD, we have identified two sequence features that characterize the (putative) metal-dependent dimeric RmlDs that are not conserved in monomeric RmlDs. Interestingly, dimeric RmlD enzymes cluster within Gram-negative species, whereas monomeric RmlD enzymes are predominantly present in Gram-positive bacteria. *Se*RmlD was described as a metal-dependent enzyme since addition of EDTA reduced activity by about 70% [14, 21]. However, it remained unknown whether absence of Mg^{2+} result in loss of dimerization. It was speculated that the presence of Mg^{2+} stabilizes cofactor binding [14]. Our data shows that the EDTA treated *Se*RmlD wild-type enzyme remains dimeric (Supp. Fig. 2A), suggesting that the removal of Mg^{2+}-ion disturbs the proper tertiary structure of each monomer therefore inactivates the enzyme.

From targeted mutagenesis attempts and Tn-Seq studies we conclude that *gacA* is essential for growth of GAS in rich medium, in agreement to a previous study that suggested that *gacA* might be essential [8]. We validated that *gacA* is essential by a riboswitch-inducible expression system resulting in attenuated growth and severe cellular abnormalities in the absence of *gacA*. Additional experiments in *S. mutans* confirm the critical role of GacA in rhamnose biosynthesis since *gacA* could functionally replace *rmlD* in *S. mutans* resulting in restored rhamnose in cell wall, growth and morphology.

A.


0601 gacA
dnaG
rpoD
emm1
0610 gacA
dnaG
rpoD
emm49

B.


150
100
50
0
Klett Unit
0
2
4
6
8
Time (h)
1 mM Theo
0 mM Theo

C.

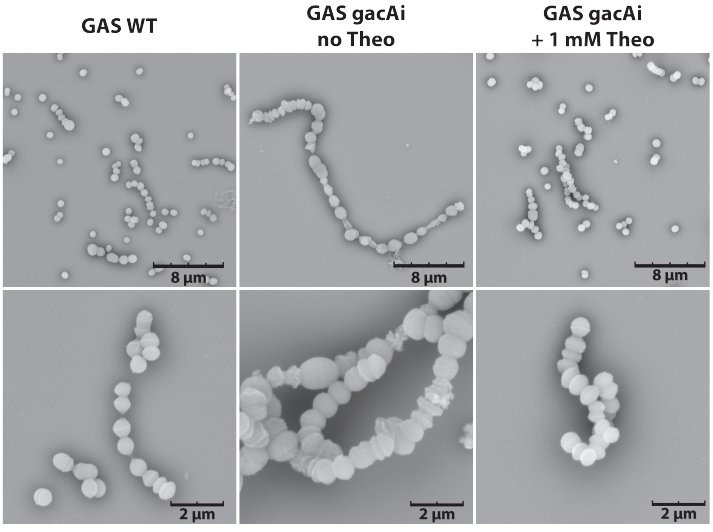

GAS WT
GAS gacAi
no Theo
GAS gacAi
+ 1 mM Theo
8 µm
8 µm
8 µm
2 µm
2 µm
2 µm

FIGURE 5. GacA is essential for GAS growth *in vitro.* (A) Essentiality of *gacA* in GAS 5448 and NZ131 as determined by Tn-seq. Complex *Krmit* transposon libraries were generated in the GAS strains 5448 and NZ131 and analyzed during growth in rich medium (THY) at 37 ˚C. Tn-seq analyses mapped the location and relative abundance of *Krmit* insertions (vertical lines) to the relevant GAS genome sequences. Using a Bayesian statistical analysis, *gacA* was determined to be essential (red arrows) based on limited transposon insertions (vertical lines) similar to the known essential genes *dnaG* and *rpoD*. Insertions for non-essential genes Spy_0601/0610 and *emm* from each strain are shown for comparison. **(B)** Conditional interference with GacA expression results in severe growth attenuation of GAS. Growth parameters as followed by Klett measurements of the GAS 5448 gacAi mutant in THY rich medium in the presence (1 mM, black circles, GacA expression) or absence (grey squares, no GacA expression) of theophylline. **(C)** Representative SEM images of GAS wild-type strain and the generated riboswitch strain gacAi in absence and presence of 1 mM of theophylline (theo).

Table 1. Bayesian analysis of Tn-seq data from GAS 5448 and NZ131.

	GAS 5448			GAS NZ131		
Gene[a]	Locus[b]	Zbar[c]	Score[d]	Locus[e]	Zbar[c]	Score[d]
	Spy0601	0	NE	Spy0610	0	NE
gacA	Spy0602	1	E	Spy0611	1	E
dnaG	Spy0599	1	E	Spy0608	1	E
rpoD	Spy0600	1	E	Spy0609	1	E
emm	Spy1719	0	NE	Spy1671	0.57	NE

[a]When available, gene name is provided.

[b]Spy numbers from the MGAS5005 genome.

[c]Essentiality values as determined by the Bayesian analysis. Genes with a Z value over 0.992 is classified as essential, genes with Z value under 0.03 are non-essential.

[d]Essentiality score. E, essential; NE, non-essential.

[e]Spy numbers for the M49 NZ131 GAS genome sequence.

Table 2. Insertion analysis of Tn-seq data from GAS 5448.

Locus	Gene	Total number of insertions per gene	Gene size (kb)	Total insertions per gene per kb
Spy0599	*dnaG*	86	1.812	**47.46**
Spy0600	*rpoD*	86	1.107	**77.69**
Spy0601		409	0.336	**1217.26**
Spy0602	*gacA*	85	0.852	**99.77**
Spy1719	*emm1*	7877	1.452	**5424.93**
Average		5297	0.851	**6224.44**

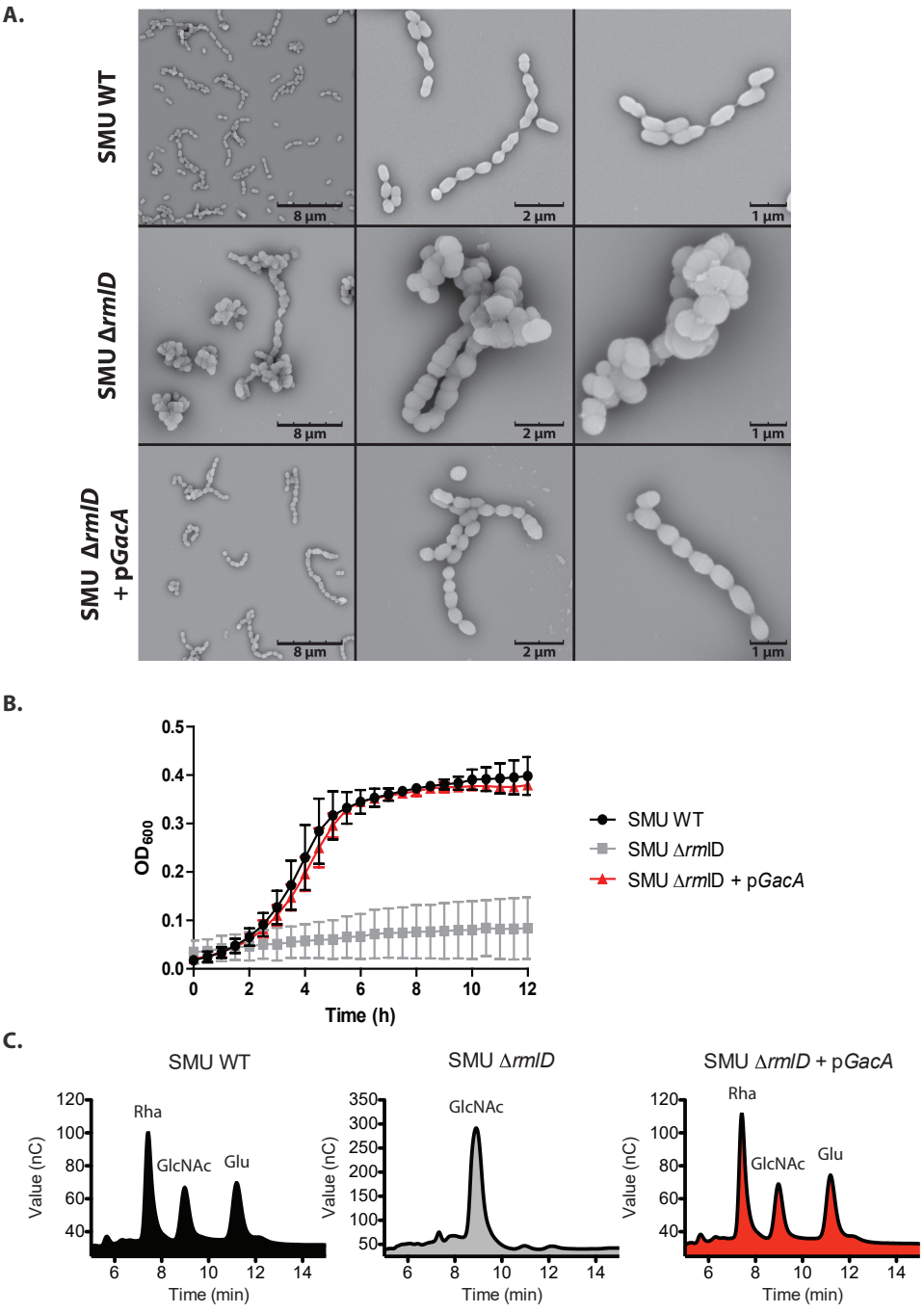

A.
SMU WT
SMU ΔrmlD
SMU ΔrmlD + pGacA
8 µm
2 µm
1 µm
B.
OD600
Time (h)
SMU WT
SMU ΔrmlD
SMU ΔrmlD + pGacA
C.
SMU WT
SMU ΔrmlD
SMU ΔrmlD + pGacA
Rha
GlcNAc
Glu
Value (nC)
Time (min)

FIGURE 6. GacA functionally replaces *S. mutans* RmlD *in vivo*. (A) Representative SEM images of *S. mutans* Xc (SMU) wild-type (WT), SMUΔ*rmlD* and SMUΔ*rmlD* + p*GacA*. Cells of the wild-type strain appear as short chains with division in a single plane, whereas the Δ*rmlD* mutant forms clumps with long chains of swollen cocci and aberrant multi-directional cell division. This aberrant morphology is almost completely restored by introduction of GAS *gacA* on an expression plasmid. **(B)** Representative growth curves (mean +/- SD, n = 3) of SMU wild-type, SMU Δ*rmlD*, and SMU Δ*rmlD* + p*GacA* complemented strains cultured at 37 °C without 5% CO_2 for 12 h. **(C)** Analysis of monosaccharide composition of SMU WT, SMUΔ*rmlD* and SMUΔ*rmlD* + p*GacA* by chromatography. Rha, rhamnose; NAG, GlcNAc; Glu, glucose.

In this study, we have identified GacA as an attractive drug target for the development of novel antimicrobial compounds against GAS. More importantly, these inhibitors could serve as lead compounds to inhibit L-rhamnose biosynthesis in other bacteria. The dTDP-L-rhamnose biosynthesis is an interesting target for the development of new drugs since 1) the pathway affects either the viability or virulence of many bacteria, including *Mycobacterium spp.* [9], *Pseudomonas spp.* [10], and *Enterococcus faecalis* [26] and 2) the pathway does not exist in humans, reducing the risk of side effects by off-target effects. Known *Mtb*RmlD inhibitors inhibit recombinant GAS GacA in the mid- to low micromolar range in kinetic assays. However, we were unable to demonstrate a reliable effect of these compounds on GAS growth since the compounds are highly-water insoluble. This implies that these compounds are not suitable as a starting point for structure-based drug design on the basis of their chemical properties. The genetic, biochemical and structural data presented here on GAS GacA forms the framework for future screenings to identify novel inhibitors that target GAC biosynthesis in GAS and dTDP-L-rhamnose biosynthesis through related RmlD enzymes in other human pathogens.

MATERIALS AND METHODS

GacA cloning, expression and purification – Full-length *gacA* (accession number AAZ51220.1; M5005_Spy_0602) was PCR amplified from GAS genomic DNA (M1T1 5448) and cloned by restriction-free cloning into a modified pGEX vector, with an N-terminal hexa-histidine-GST-tag followed by a PreScission Protease cleavage site (His_6GST-GacA). His_6GST-GacA was transformed in *E. coli* BL21(DE3) cells and recombinant protein expression was induced with 0.5 mM isopropyl-1-thio-β-D-galactopyranoside for 18 h. Cells were harvested by centrifugation and resuspended in buffer A (50 mM Tris-base, pH 8.0, 250 mM NaCl), supplemented with 10% glycerol, 0.5 complete Protease Inhibitor Cocktail Tablets (Roche) and 2 mM Tris (2-Carboxyethyl)-phosphine hydrochloride (TCEP). All purification steps were

carried out at 4 ˚C. Cells were disrupted and the supernatant from two subsequent centrifugation runs at 15,000 x *g* for 20 min and 100,000 x *g* for 1.5 h was adjusted to 10 mM imidazole. The sample was passed over a 5 ml His-trap column HP charged with Co^{2+}, washed with 10 column volumes (CV) of 25 mM imidazole in buffer A and the protein was eluted with a gradient over 10 CV of 500 mM imidazole in buffer A. The eluted protein was concentrated using a 50k Da MW cutoff concentrator and passed over a desalting column, equilibrated in buffer A. The His_6GST-tag was cleaved using PreScission Protease overnight, passed over a 5 ml His-trap column equilibrated in buffer A supplemented with 20 mM imidazole, the flow through was collected and concentrated using a 30k Da MW cutoff concentrator and injected into a Superdex 75 26/60 column equilibrated in TBS-buffer, supplemented with 0.2 mM TCEP. The fractions containing GacA were collected and concentrated to 25 mg/ml. The purified protein was confirmed by tryptic fingerprint mass spec (University of St. Andrews).

Enzymatic Activity of GacA – To analyze GacA enzyme kinetics, we cloned and expressed GAS homologues of *rmlB* (*Sp*RmlB; accession number AAZ51354.1; M5005_Spy_0736) and *rmlC* (*Sp*RmlC; accession number AAZ51353.1; M5005_Spy_0735) using a modified pET vector, harboring an N-terminal octa-histidine tag. The RmlB and RmlC-fusion proteins were expressed and purified as described for GacA. The assay was performed following the protocol from Sivendran *et al* [27] with the following changes: the assay buffer system contained 25 mM Tris-base (pH 7.5), 150 mM NaCl, 0.1 mM NADPH and 2 pM GacA. *Sp*RmlB and *Sp*RmlC were added to the assay in 25-fold molar excess relative to GacA. The assay was started with the addition of 400 µM dTDP-D-glucose. Michaelis-Menten kinetics for GacA for dTDP-glucose was calculated using a concentration range from 0.025 - 1 mM of dTDP-glucose. The oxidation of NADPH to $NADP^+$ was measured by the change in intrinsic absorbance at 340 nm using a SpectraMax M2 plate reader. Background absorbance was subtracted and data interpreted using the Michaelis-Menten model in GraphPad Prism. RmlD inhibitors 2 and 3 (Sigma) described for *Mycobacterium tuberculosis* [20] were dissolved to 10 mM stock concentration in 100% DMSO, diluted in enzyme assay buffer and added at the same time-point as the GacA enzyme. All reactions contained a final concentration of 2% DMSO.

SeRmlD cloning, expression and purification – Full-length *rmlD* (GI 16420628) was PCR amplified from *Salmonella enterica* genomic DNA (LT2) using the procedure described for *gacA*. Point mutations were inserted using standard mutagenesis procedures. Wild type and mutant enzymes were expressed and purified as described for GacA. EDTA treatment was performed by incubation of *Se*RmlD (0.1 mg/ml) with 10 mM

EDTA for 30 min at room temperature, followed by a concentration-step and size-exclusion chromatography in buffer A containing 10 mM EDTA.

Surface plasmon resonance (SPR) experiments - Recombinant, purified GacA was chemically biotinylated and captured on a streptavidin surface of a Biacore T200 instrument (GE-Healthcare) at densities of 3k–4k RU. To stabilize captured protein over time, all experiments were run at 10 ˚C. Ligands were injected over captured protein at flow rate 30 µl min^{-1} in running buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 0.05%Tween, 3% DMSO), with each compound injected in duplicates in concentration series 4-1,000 µM. Association was measured for 60 s and dissociation for 120 s. All data were double referenced for blank injections of buffer and biotin-blocked Streptavidin surface. Scrubber 2 (BioLogic Software) was used to process and analyse the data.

Crystallization and structure determination - Vapor diffusion sitting-drop crystallization was carried out at 20 ˚C. Crystals appeared after mixing equal amounts of protein and crystallization buffer containing 12.5% w/v PEG 1000, 12.5% w/v PEG 3350, 12.5% v/v MPD, 0.02 M of each carboxylic acid (0.2 M sodium formate, 0.2 M ammonium acetate, 0.2 M trisodium citrate, 0.2 M sodium potassium L-tartrate, 0.2 M sodium oxamate) and 0.1 M MES/imidazole pH 6.5 [28] after one day. Crystals were flash frozen into liquid nitrogen prior to data collection. Data were collected at beamline ID23-1 at ERSF and processed with iMOSFLM [29]. Data collection statistics are summarized in Table 3. On the basis of the high-resolution dataset, the structure was solved *ab initio* using the program 'Acorn' in the CCP4 suite [30]. The initial model was used for autobuilding in Phenix [31]. With the exception of the first and last residue, 288 residues out of 290 residues were built and the structural model was refined in iterative cycles using Coot [32] and Refmac [33] to the statistics shown in Table 3. The final model was refined to 1.1 Å resolution with an R_{factor} of 12.9% and R_{free} of 15.5%. The Ramachandran plot revealed that 98% of all residues are in favored regions, with no outliers as calculated by MolProbity [34]. The coordinates and structure factors have been deposited with the RCSB Protein Data Bank with PDB ID code 4WPG.

Neighbor-joining tree construction - To construct midpoint-rooted Neighor-Joining trees of RmlD sequences, we used two-independent psi-BLAST runs using *Se*RmlD and GacA as the query sequences. We obtained 135 Gram-negative RmlD homologues of *Se*RmlD using an E-value cutoff of > $7e^{-58}$, and we obtained 78 Gram-positive homologues of GAS GacA using a cut-off of > $1e^{-48}$. All hypothetical sequences and all RmlD sequences for which no corresponding 16S rRNA sequences were available were rejected. We used only one representative for each bacterium,

avoiding the use of multiple strain variants. Multiple sequence alignments of RmlD protein sequences and 16S rRNA sequences were constructed using ClustalW2 [35]. Neighbor-Joining trees were built from the RmlD and 16S rRNA alignments using ClustalX (2000 bootstraps were run)[36]. Neighbor-Joining trees were displayed using FigTree (http://tree.bio.ed.ac.uk/software/figtree/). All bacteria analyzed were assigned to the corresponding monomer/dimerization motif groups and were marked with related colors.

Table 3. Crystallographic data collection and refinement statistics.

Spacegroup	P1
Unit cell dimension (Å)	36.48 45.98 48.87
Resolution range (last shell) (Å)	43.8 - 1.1 (1.139 - 1.1)
Unit cell angles (˚)	66.3 81.21 95.27
Unique reflections	108172 (9823)
Completeness (%)	94.0 (85.5)
I/σI	22.5 (5.7)
Wilson B-factor	12.20
R_{factor}	0.129 (0.175)
R_{free}	0.155 (0.205)
Number of atoms	5062
Macromolecules	2314
Ligands	21
Water	468
Protein residues	287
Rms bond length (Å)	0.023
Rms bond angles (˚)	2.08
Ramachandran favored (%)	98
Ramachandran outliers (%)	0
Average B-factor	18.10
Macromolecules	16.10
Solvent	27.90

I/σI, intensity divided by standard deviation of intensity, averaged over all measurement.
Rms is root-mean-square deviation from ideal value.

Tn-seq to identify genes essential for GAS growth in THY - The pKRMIT plasmid contains a *mariner* mini-transposon named *Krmit* (Kanamycin-resistant element for massive identification of transposants) modified for Tn-seq [24] and was used to perform saturating transposition for random mutagenesis in GAS 5448 and NZ131 as previously described for the *mariner* transposon *Oskar* [37, 38]. Tn-seq [39] was performed

as described recently (Le Breton *et al*., 2015) with the following primers oKrmit-Tnseq2 (5'-CAAGCAGAAGACGGCATACGAAGCGCCTACG-AGGAATTTGTATCG-3') and oAdapterPCR (5'-ACACTCTTTCCCTACACGACGCTCTT-CCGATCT-3'), resulting in the production of 176 bp *Krmit* insertion tags. Quality and yield of the resulting tags was assessed using a NanoDrop spectrophotometer (Thermo Scientific) and an Agilent Bioanalyzer. *Krmit* insertion tags were analyzed by Illumina sequencing (50 nt single end reads) on a HiSeq 1500 platform in the Institute for Bioscience and Biotechnology Research (IBBR) Sequencing Facility located at the University of Maryland, College Park. The quality of read datasets (Sanger FastQ format) was determined using FastQC [40], data was filtered and trimmed using Biopieces (biopieces.org) to select for reads containing the Tn-seq barcodes and *Krmit* ITR ends. Reads were then de-multiplexed and count tables generated using SamTools [41] and HTseq [42]. Reads were mapped to the GAS 5448 or NZ131 genome using Bowtie [43] and data relevant to the *gacA-L* locus visualized using the Integrative Genomics Viewer (IGV) browser (broadinstitute.org/igv/home). Gene essentiality was determined using a Bayesian statistical model based on the Metropolis-Hastings algorithm using the Python script (saclab.tamu.edu/essentiality) developed by DeJesus *et al*. [44].

Construction of a conditionally lethal gacA mutant – A GAS 5448 merodiploid mutant, gacAi, was constructed using the pSinS/pHplK system for gacA gene expression to be under the control of a theophylline-dependent riboswitch [45] as recently described [24]. Succinctly, a ca. 600-nu fragment of the 5'-end of the gacA gene was amplified using primers oGacA-1 (5'-CCCTGCTAAGGAGGTAACAACAAGATGATTTTAATTACAGGAAGCAATGG-3') and oGacA-2 (5'-cccGGATCCGTCAAATAACACATGAATTCTGC-3') and subsequently fused by SOE-PCR to the Psag promoter along with the synthetic riboswitch E as previously described (Le Breton et al., 2015). The resulting PCR product was then cloned into the BamHI site of the pSinS plasmid producing pGacA and mutation carried out as described [24], creating the GAS 5448 gacAi mutant; and the junction between the Psag promoter, the synthetic riboswitch E and the gacA gene in the GAS 5448 gacAi mutant was verified by DNA sequencing (data not shown).

Bacterial strains and growth conditions - S. mutans Xc is a serotype C wild-type strain [46] and was kindly provided by Dr. Y Yamashita (Kyushu University, Japan). *S. mutans* was routinely cultured in Todd-Hewitt Broth (THB; Oxoid) or on THB agar at 37 ˚C with 5% CO_2. GAS strain 5448 is a representative of the serotype M1T1 clone [47] and GAS NZ131 [48] is an invasive strain of the M49 serotype. GAS was grown in THB (Becton Dickinson) supplemented with 1% yeast extract (THY) or on THY agar at 37 ˚C. When required, growth medium was supplemented with 10 μg/ml erythromycin

(Erm) or 3 μg/ml chloramphenicol (Cm) for *S. mutans* or with 300 μg/ml kanamycin (Km) or 100 μg/ml spectinomycin (Spec) for GAS. For cloning purposes, *E. coli* strain MC1061 was grown in Luria-Burtani (LB; Oxoid) or on LB agar with ampicillin (100 μg/ml, Amp), Erm (500 μg/ml) or Cm 10 μg/ml.

Genetic manipulation of S. mutans - To confirm the function of GacA in the production of dTDP-L-rhamnose in bacteria, we heterologously expressed GAS *gacA* (M5005_Spy_0602; accession number AAZ51220.1) in a *S. mutans rmlD* deletion mutant (SMU Δ*rmlD*), which is unable to produce dTDP-L-rhamnose [19]. For complementation, full length *gacA* was amplified from the GAS 5448 chromosome using primers XbaI_*gacA*F 5′-GCTCTAGAATGATTTTAATTACAGGAAGCAATGGTC-3′ and BamHI_*gacA*R 5′-CGCGGATCCTACTTACTTTTTCAGTCCTTGTTGGT-3′ and cloned into expression vector pDC123 using XbaI and BamHI restriction sites, yielding p*GacA*. p*GacA* was transformed into *S. mutans* wild-type and selected for Cm resistance. The presence of p*GacA* was confirmed by PCR analysis. Subsequently, *rmlD* was knocked out by precise in-frame allelic replacement of *rmlD* with an erythromycin (erm) resistance gene in *S. mutans* + p*GacA*. Briefly, 700 bp immediately upstream of *rmlD* was amplified with the primers *rmlD*upF, 5′-CGCAGCAAGCAGTTACGTGATTTTGTTGAAG-3′, and *rmlD*upR+erm 5′-GTTTTGAGAATATTTTATATTTTTGTTCATTATTTTTTCTCCTT-TAAAAAGCTTTATTACTATTACC-3′, and 674 bp immediately downstream of *rmlD* was amplified with the primers *rmlD*downF+erm 5′-AGTTATCTATTATTTAACGGGAGGAAATAATATTTTAGCAAA-GAAGGACAGGTTTAAACC-3′, and *rmlD*downR, 5′-CTGAAGGTGATAA-ATCCGTGCCATA-3′. The *rmlD*upR+erm and *rmlD*downF+erm primers were constructed with 30-bp 5′ extensions (underlined) corresponding to the 5′ and 3′ ends of the *erm* gene, respectively. The upstream and downstream PCR fragments were combined with the 738 bp amplicon of the *erm* gene (amplified off the pDCerm plasmid [49]) as templates in a second round of PCR using primers *rmlD*upF and *rmlD*downR. The resultant PCR amplicon, containing an in-frame substitution of *rmlD* with *erm*, was transformed into *S. mutans* + p*GacA* and selected for Erm and Cm resistance as previously described [50], to yield *S. mutans* Δ*rmlD* + p*GacA* (SMU Δ*rmlD* + p*GacA*). PCR analysis was used to confirm the deletion of *rmlD*.

Scanning electron microscopy - Overnight cultures of *S. mutans* strains were diluted and grown to mid-log phase (*S. mutans* wild-type and *S. mutans* Δ*rmlD* + p*GacA* OD_{600nm} of 0.3, *S. mutans* Δ*rmlD* OD_{600nm} of 0.15). GAS WT strains were grown in THY medium overnight. GAS 5448 gacAi were selected from a plate containing Spec and 2 mM theophyllin and grown overnight cultures in the presence of different concentrations of theophylline (0, 0.5, 1, 1.5 and 2 mM) in THY. All cultures reached an OD_{600nm} of 0.5/0.6 except for strains without theophyllin, which reached OD_{600nm}

of 0.35. Cultures were diluted in THY with or without theophylline to an OD_{600nm} of 0.06 and grown to midlog phase. Samples were washed, fixed, and dehydrated as described previously [51], mounted onto 12.5 mm specimen stubs (Agar scientific, Stansted, Essex, UK) and coated with gold to 1 nm using a Quorum Q150R S sputter coater at 20 mA. Visual examination was performed with a Phenom PRO desktop scanning electron microscope (SEM; Phenom-World BV). The SEM was operated with an acceleration voltage of 10 kV.

S. mutans growth curves - Growth curves of *S. mutans* Xc wild-type, Δ*rmlD* and Δ*rmlD* + p*GacA* were obtained after dilution of an overnight culture to OD_{600nm} to 0.025 in THB. Optical density was recorded every 30 min over 12 h at 37 °C without 5% CO_2 in a 100 Honeycomb plate using a Bioscreen C MBR machine.

Carbohydrate analysis of S. mutans strains – For the isolation of cell wall carbohydrates in *S. mutans*, 2 L bacterial cultures were centrifuged (8,000 x g, 30 min, 4 °C) and washed with ice cold water. Five gram of wet bacterial cells was resuspended in 0.1 M citrate buffer (pH 4.6, 4˚C) and disrupted with a bead-beater (Biospec). The bacterial lysate was centrifuged (1,000 x g, 5 min, 4˚C), the white suspension was isolated and centrifuged (35,000 x g, 30 min, 4˚C) to collect the cell walls. The white pellet was washed with citrate buffer, resuspended in 0.1 M sodium acetate (pH 4.6) containing 4% sodium dodecyl sulfate (SDS) and boiled for 1 h at 100°C with agitation. The suspension was centrifuged (35,000 x g, 30 min, 21 °C) and washed with SuperQ to remove SDS. The pellet was resuspended in 10 ml MilliQ, and treated with RNase (10 µg/ml) and DNase (10 µg/ml, 37 °C, 2 h) and afterwards with pronase E (100 µg/ml, 37 °C for 24 h). Cell wall carbohydrates were extensively washed with SuperQ (35,000 x g for 30 min), treated with trypsin (100 µg/ml, 37 °C, 2 h), washed with SuperQ and lyophilized before the sample was subjected to hydrolysis with TFA using according to published procedures [52]. Carbohydrate analysis was performed on a Dionex ICS-3000 Ion Chromatography System using a CarboPac PA20 3 x 150mm column, equipped with a ICS-3000 Electrochemical Detector. The "Carbohydrates (Standard Quad)" waveform was used for detection. Eluent was run at 92% H_2O and 8% 0.2M NaOH(aq) for 25 min and increased to 100% 0.2 M NaOH over a 2 min period and held at this concentration for 10 minutes. The concentration was then returned to 92% H_2O and 8% 0.2 M NaOH(aq) over a 2 min period and held at this concentration for 11 min before a new sample was injected. The flow rate was 0.5 mL/min.

ACKNOWLEDGEMENTS

This work was supported by a Sir Henry Wellcome Postdoctoral Fellowship (092193) to HCD, a Wellcome Trust Senior Fellowship (WT087590MA) and MRC Programme Grant (M004139) to DvA, an NWO-VIDI grant (91713303) to NvS, and a NIH/NIAID Grant (AI047928) to KSM.

The authors thank Drs. Ashton T. Belew and Najib El-Sayed for support in bioinformatics design and implementation for the Tn-seq analysis. The authors acknowledge the ESRF ID23-1 for synchrotron beam time and Dr. Stephen Carr for assistance with SEC-MALLS. The authors acknowledge Dr. Mark de Been for assistance to compose the Neighbor-Joining trees and Miss Tonia Aristotelous for assistance in performing the SPR experiments. The authors thank Emrul Islam for his assistance in generating the gacAi plasmids and strains and Dr. Konstantinos Beis and Prof. So Iwata for suggestions and support.

PDB accession code **-** Coordinates and structure factors have been deposited with the Protein Data Bank (PDB entry 4WPG).

REFERENCES

1. Weidenmaier, C. and Peschel, A.; *Teichoic acids and related cell-wall glycopolymers in Gram-positive physiology and host interactions.* Nat Rev Microbiol, 2008. **6**(4): p. 276-87.
2. Caliot, E., et al.; *Role of the Group B antigen of Streptococcus agalactiae: a peptidoglycan-anchored polysaccharide involved in cell wall biogenesis.* PLoS Pathog, 2012. **8**(6): p. e1002756.
3. Sutcliffe, I.C., Black, G.W., and Harrington, D.J.; *Bioinformatic insights into the biosynthesis of the Group B carbohydrate in Streptococcus agalactiae.* Microbiology, 2008. **154**(5): p. 1354-1363.
4. McCarty, M.; *The lysis of group A hemolytic streptococci by extracellular enzymes of Streptomyces albus. II. Nature of the cellular substrate attacked by the lytic enzymes.* J Exp Med, 1952. **96**(6): p. 569-80.
5. Lancefield, R.C.; *A Serological Differentiation of Human and Other Groups of Hemolytic Streptococci.* J Exp Med, 1933. **57**(4): p. 571-95.
6. Carapetis, J.R., et al.; *The global burden of group A streptococcal diseases.* Lancet Infect Dis, 2005. **5**(11): p. 685-94.
7. Coligan, J.E., Kindt, T.J., and Krause, R.M.; *Structure of the streptococcal groups A, A-variant and C carbohydrates.* Immunochemistry, 1978. **15**(10-11): p. 755-60.
8. van Sorge, N.M., et al.; *The classical lancefield antigen of group a streptococcus is a virulence determinant with implications for vaccine design.* Cell Host Microbe, 2014. **15**(6): p. 729-40.
9. Ma, Y., Pan, F., and McNeil, M.; *Formation of dTDP-rhamnose is essential for growth of mycobacteria.* J Bacteriol, 2002. **184**(12): p. 3392-5.
10. Engels, W., et al.; *Role of lipopolysaccharide in opsonization and phagocytosis of Pseudomonas aeruginosa.* Infect Immun, 1985. **49**(1): p. 182-9.
11. Teng, C.H., et al.; *Escherichia coli K1 RS218 interacts with human brain microvascular endothelial cells via type 1 fimbria bacteria in the fimbriated state.* Infect Immun, 2005. **73**(5): p. 2923-31.
12. Kornfeld, S., and Glaser, L.; *The enzymatic synthesis of thymidine-linked sugars.* J. Biol. Chem., 1961. **236**: p. 1791-1794.
13. Pazur, J.H., and Shuey, E.W.; *The enzymatic synthesis of thymidine diphosphate glucose and its conversion to thymidine diphosphate rhamnose.* J. Biol. Chem., 1961. **236**(1780-1785).
14. Blankenfeldt, W., et al.; *Variation on a theme of SDR. dTDP-6-deoxy-L- lyxo-4-hexulose reductase (RmlD) shows a new Mg2+-dependent dimerization mode.* Structure, 2002. **10**(6): p. 773-86.
15. Kavanagh, K.L., et al.; *Medium- and short-chain dehydrogenase/reductase gene and protein families : the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes.* Cell Mol Life Sci, 2008. **65**(24): p. 3895-906.
16. Giraud, M.F. and Naismith, J.H.; *The rhamnose pathway.* Curr Opin Struct Biol, 2000. **10**(6): p. 687-96.
17. Macpherson, D.F., Manning, P.A., and Morona, R.; *Characterization of the dTDP-rhamnose biosynthetic genes encoded in the rfb locus of Shigella flexneri.* Mol Microbiol, 1994. **11**(2): p. 281-92.
18. Morona, J.K., Morona, R., and Paton, J.C.; *Characterization of the locus encoding the Streptococcus pneumoniae type 19F capsular polysaccharide biosynthetic pathway.* Mol Microbiol, 1997. **23**(4): p. 751-63.
19. Tsukioka, Y., et al.; *Identification of a fourth gene involved in dTDP-rhamnose synthesis in Streptococcus mutans.* J Bacteriol, 1997. **179**(13): p. 4411-4.
20. Wang, Y., et al.; *Novel inhibitors of Mycobacterium tuberculosis dTDP-6-deoxy-L-lyxo-4-hexulose reductase (RmlD) identified by virtual screening.* Bioorg Med Chem Lett, 2011. **21**(23): p. 7064-7.
21. Graninger, M., et al.; *Characterization of dTDP-4-dehydrorhamnose 3,5-epimerase and dTDP-4-dehydrorhamnose reductase, required for dTDP-L-rhamnose biosynthesis in Salmonella enterica serovar Typhimurium LT2.* J Biol Chem, 1999. **274**(35): p. 25069-77.
22. Bernstein, F.C., et al.; *The Protein Data Bank: a computer-based archival file for macromolecular structures.* J Mol Biol, 1977. **112**(3): p. 535-42.

23. Ortega, A., Amoros, D., and Garcia de la Torre, J.; *Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models.* Biophys J, 2011. **101**(4): p. 892-8.
24. Le Breton, Y., et al.; *Essential Genes in the Core Genome of the Human Pathogen Streptococcus pyogenes.* Sci Rep, 2015. **5**: p. 9838.
25. Nakano, K. and Ooshima, T.; *Serotype classification of Streptococcus mutans and its detection outside the oral cavity.* Future Microbiol, 2009. **4**(7): p. 891-902.
26. Teng, F., et al.; *Further characterization of the epa gene cluster and Epa polysaccharides of Enterococcus faecalis.* Infect Immun, 2009. **77**(9): p. 3759-67.
27. Sivendran S, J.V., Sun D, Wang Y, Grzegorzewicz AE, Scherman MS, Napper AD, McCammon JA, Lee RE, Diamond SL, McNeil M; *Identification of triazinoindol-benzimidazolones as nanomolar inhibitors of the Mycobacterium tuberculosis enzyme TDP-6-deoxy-d-xylo-4-hexopyranosid-4-ulose 3,5-epimerase (RmlC).* Bioorg Med Chem, 2010. **18**((2)): p. 896-908.
28. Gorrec, F.; *The MORPHEUS protein crystallization screen.* J Appl Crystallogr, 2009. **42**(Pt 6): p. 1035-1042.
29. Battye TG, K.L., Johnson O, Powell HR, Leslie AG; *iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM.* Acta Crystallogr D Biol Crystallogr, 2011. **67**((Pt4)): p. 271-281.
30. McCoy, A.J., et al.; *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
31. Adams PD, A.P., Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH.; *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**((Pt2)): p. 213-221.
32. Emsley P1, L.B., Scott WG, Cowtan K; *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**((Pt4)): p. 486-501.
33. Murshudov GN, V.A., Dodson EJ; *Refinement of Macromolecular Structures by the Maximum-Likelihood Method.* Acta Crystallogr D Biol Crystallogr, 1997. **53**((Pt3)): p. 240-255.
34. Lovell SC, D.I., Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC; *Structure validation by Calpha geometry: phi,psi and Cbeta deviation.* Proteins, 2003. **15**((50)): p. 437-50.
35. McWilliam, H., et al.; *Analysis Tool Web Services from the EMBL-EBI.* Nucleic Acids Res, 2013. **41**(Web Server issue): p. W597-600.
36. Larkin, M.A., et al.; *Clustal W and Clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-8.
37. Le Breton, Y. and McIver, K.S.; *Genetic manipulation of Streptococcus pyogenes (the Group A Streptococcus, GAS).* Curr Protoc Microbiol, 2013. **30**: p. Unit 9D 3.
38. Le Breton, Y., et al.; *Genome-wide identification of genes required for fitness of group A Streptococcus in human blood.* Infect Immun, 2013. **81**(3): p. 862-75.
39. van Opijnen, T. and Camilli, A.; *Genome-wide fitness and genetic interactions determined by Tn-seq, a high-throughput massively parallel sequencing method for microorganisms.* Curr Protoc Microbiol, 2010. **Chapter 1**: p. Unit1E 3.
40. Ramirez-Gonzalez RH, L.R., Waite D, Thanki A, Drou N, Caccamo M, Davey R; *StatsDB: platform-agnostic storage and understanding of next generation sequencing run metrics.* F1000Res, 2013. **2**(248).
41. Li, H., et al.; *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
42. Chandramohan, R., et al.; *Benchmarking RNA-Seq quantification tools.* Conf Proc IEEE Eng Med Biol Soc, 2013. **2013**: p. 647-50.
43. Langmead, B., et al.; *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.
44. DeJesus, M.A., et al.; *Bayesian analysis of gene essentiality based on sequencing of transposon insertion libraries.* Bioinformatics, 2013. **29**(6): p. 695-703.
45. Topp, S., et al.; *Synthetic riboswitches that induce gene expression in diverse bacterial species.* Appl Environ Microbiol, 2010. **76**(23): p. 7881-4.

46. Koga, T., et al.; *Effect of subculturing on expression of a cell-surface protein antigen by Streptococcus mutans.* J Gen Microbiol, 1989. **135**(12): p. 3199-207.
47. Kansal, R.G., et al.; *Inverse relation between disease severity and expression of the streptococcal cysteine protease, SpeB, among clonal M1T1 isolates recovered from invasive group A streptococcal infection cases.* Infect Immun, 2000. **68**(11): p. 6362-9.
48. Simon, D. and Ferretti, J.J.; *Electrotransformation of Streptococcus pyogenes with plasmid and linear DNA.* FEMS Microbiol Lett, 1991. **66**(2): p. 219-24.
49. Jeng, A., et al.; *Molecular genetic analysis of a group A Streptococcus operon encoding serum opacity factor and a novel fibronectin-binding protein, SfbX.* J Bacteriol, 2003. **185**(4): p. 1208-17.
50. Perry, D., Wondrack, L.M., and Kuramitsu, H.K.; *Genetic transformation of putative cariogenic properties in Streptococcus mutans.* Infect Immun, 1983. **41**(2): p. 722-7.
51. Garufi, G., et al.; *Synthesis of lipoteichoic acids in Bacillus anthracis.* J Bacteriol, 2012. **194**(16): p. 4312-21.
52. Fan, J.Q., et al.; *Comparison of acid hydrolytic conditions for Asn-linked oligosaccharides.* Anal Biochem, 1994. **219**(2): p. 375-8.

SUPPLEMENTAL MATERIALS

Supplemental tables and figures can be found online:
https://onlinelibrary.wiley.com/doi/abs/10.1111/mmi.13169


3

Functional characterization and lead compound identification of streptococcal dTDP-L-rhamnose biosynthesis enzymes

3

Samantha L. van der Beek[1], Azul Zorzoli[2], Ebru Çanak[1], Robert N. Chapman[3], Benjamin H. Meyer[2], Geert-Jan Boons[3,4], Helge C. Dorfmueller[2] and Nina M. van Sorge[1]

[1] University Medical Center Utrecht, Utrecht University, Medical Microbiology, Heidelberglaan 100, 3584 CX, Utrecht, The Netherlands.

[2] Division of Molecular Microbiology, University of Dundee, School of Life Sciences, Dow Street, DD1 5EH, Dundee, UK.

[3] Complex Carbohydrate Research Center, Department of Chemistry, The University of Georgia, 315 Riverbend Road, Athens, USA

[4] University Utrecht, Utrecht Institute Pharmaceutical Science, Department of Medical Chemistry and Chemical Biology, 3508 TB, Utrecht, The Netherlands

Manuscript under revision in: Molecular Microbiology

ABSTRACT

Biosynthesis of the nucleotide sugar precursor dTDP-L-rhamnose is critical for the viability and virulence of many human pathogenic bacteria, including *Streptococcus pyogenes* (Group A *Streptococcus*; GAS) and *Streptococcus mutans*. Both pathogens require dTDP-L-rhamnose for the production of structurally similar rhamnose-containing polysaccharides in their cell wall. Via heterologous expression in *S. mutans*, we confirm that GAS RmlB and RmlC are critical for dTDP-L-rhamnose biosynthesis through their action as dTDP-glucose-4,6-dehydratase and dTDP-4-keto-6-deoxyglucose-3,5-epimerase enzymes, respectively. Complementation with GAS RmlB and RmlC containing specific point mutations corroborated the conservation of previous identified catalytic residues in these enzymes. Bio-layer interferometry was used to identify and confirm inhibitory lead compounds that bind to GAS dTDP-L-rhamnose biosynthesis enzymes RmlB, RmlC and GacA. One of the identified compounds, Ri03, inhibited growth of GAS as well as several other rhamnose-dependent streptococcal pathogens with an IC_{50} of 120-410 μM. Furthermore, Ri03 displayed no cytotoxity in U937 monocytic cells up to a concentration of 15 mM. We therefore conclude that inhibitors of dTDP-L-rhamnose biosynthesis such as Ri03 affect streptococcal viability and can serve as a lead compound for the development of a new class of antibiotics that targets dTDP-L-rhamnose biosynthesis in pathogenic bacteria.

INTRODUCTION

The typical cell wall architecture of a Gram-positive bacterium consists of a thick peptidoglycan layer that is decorated with wall teichoic acids, proteins and capsular polysaccharides. However, certain lactic acid bacteria, particularly streptococcal species, lack the expression of wall teichoic acids and instead express rhamnose cell wall polysaccharides, which are covalently anchored to peptidoglycan [1]. Rhamnose cell wall polysaccharides encompass about 40-60% of the cell wall mass and are considered to be functional homologs of wall teichoic acids [2, 3]. Examples of rhamnose cell wall polysaccharides are the group A carbohydrate (GAC) of Group A *Streptococcus* (GAS; *Streptococcus pyogenes*) and the serotype-determining polysaccharides, referred to as rhamnose-glucose polysaccharides (RGP), of *Streptococcus mutans*. The GAC and RGP share structural similarities; both consist of an α-1,2-/α-1,3-linked polyrhamnose backbone with alternating *N*-acetylglucosamine side chains for GAS and glucose or galactose side chains for *S. mutans* [4-8] (Fig. 1). For *S. mutans*, the type of side chain as well as their linkage to the rhamnan backbone determines the *S. mutans* serotype [6, 9] (Fig. 1). In contrast, all GAS serotypes express a structurally invariant GAC [10]. Similar to classical wall teichoic acids, rhamnose cell wall polysaccharides are critical for maintaining cell wall shape, bacterial physiology and virulence, but in-depth knowledge of their biosynthesis or host interactions at a molecular level is limited [11-14]. A better understanding of these mechanisms could aid the development of new classes of antibiotics, antibiotic adjuvants or vaccines [6, 12, 15].

L-Rhamnose is the main building block for both the GAC and RGP. The biosynthesis pathway of the nucleotide precursor, dTDP-L-rhamnose, is highly conserved among both Gram-positive and Gram-negative bacteria [1, 16]. The dTDP-L-rhamnose biosynthesis pathway is critical or even essential for viability or virulence of a wide range of human pathogens including GAS [17, 18], Group B *Streptococcus* (GBS) [2, 19], some serotypes of *Streptococcus pneumoniae* [20, 21], *S. mutans* [13, 22-24], *Enterococcus faecalis* [25, 26], *Mycobacterium spp.* [27, 28], *Pseudomonas spp.* [29] and *Salmonella enterica* serovar *Typhimurium* [30]. Consequently, this pathway is considered to be an interesting drug target, especially since dTDP-L-rhamnose is not produced or used by humans [31]. dTDP-L-rhamnose is produced through a four-step enzymatic pathway catalyzed by the enzymes RmlABCD. In the first step of the pathway, RmlA, a glucose-1-phosphate thymidyltransferase, converts glucose-1-phosphate into dTDP-glucose [32], which is subsequently oxidized and dehydrated to form dTDP-4-keto-6-deoxy-D-glucose by the dTDP-D-glucose 4,6-dehydratase RmlB [33]. RmlC catalyzes an unusual double epimerization reaction [34-36], the product of which is finally reduced by RmlD, a dTDP-4-dehydrorhamnosereductase, to form dTDP-L-

rhamnose [37, 38]. The mechanisms of action as well as structural characteristics of the RmlABCD enzymes have been studied extensively [11, 32-36, 39-42]. Indeed, crystal structures are available for all four Rml enzymes from different Gram-positive and Gram-negative bacterial species and show high structural conservation. This structural and functional information has enabled the development of different screening methods to discover inhibitors against RmlBCD, yielding compounds that can inhibit dTDP-L-rhamnose biosynthesis in the low micromolar range in a biochemical assay [43-47]. Given the importance of L-rhamnose to virulence and viability of many human pathogens, potent inhibitors of these pathways are of therapeutic interest, especially with the alarming development of antibiotic resistant bacteria.

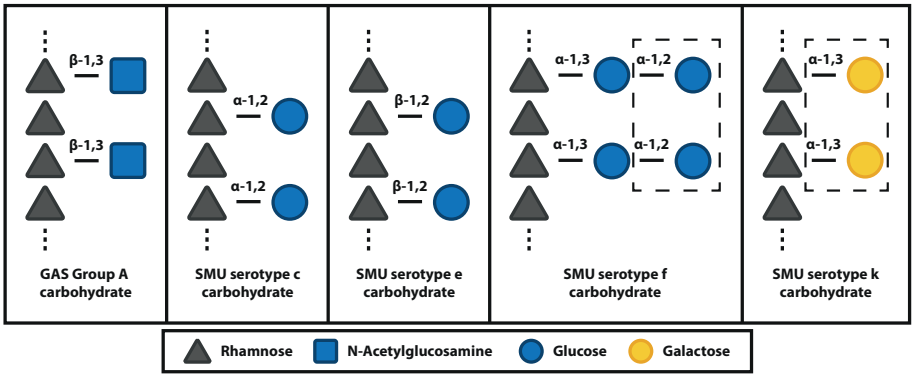


Figure 1. Schematic overview of streptococcal rhamnose polysaccharide structures. Schematic representation of the chemical composition of the group A carbohydrate from GAS and the serotype-specific (c, e, f and k) carbohydrates from *S. mutans* (SMU). All carbohydrates share an α-1,2/α-1,3 linked polyrhamnose backbone. Sugar residues in dashed boxes were recently identified by St Michael *et al*. [6].

Recently, we characterized GAS GacA as an RmlD homolog catalyzing the last step in dTDP-L-rhamnose biosynthesis [11]. Surprisingly, this GAS RmlD enzyme was confirmed to be a monomer as opposed to previously characterized RmlD dimers. Additional bioinformatics analysis of 213 RmlD homologs revealed that the majority of RmlD enzymes are predicted to be monomeric, while only a subclass of RmlD enzymes in Gram-negatives form metal-dependent homodimers [37]. In this study, we extended our work to study GAS RmlB and RmlC on a functional level through heterologous expression in *S. mutans* and subsequent analysis of growth, morphology and cell wall composition. In addition, we report the identification of small chemical fragments that bind these enzymes and inhibit GAS growth with IC_{50} values ranging from 110 μM to 6.2 mM. Furthermore, the most potent fragment,

Ri03, could inhibit the growth of *S. mutans* and Group C *Streptococcus* (GCS; *S. equi* subsp. *zooepidemicus*) with similar efficacy. Moreover, Ri03 shows minimal off-target effects in bacterial species that do not express dTDP-L-rhamnose or for which dTDP-L-rhamnose expression is not essential and displays low toxicity towards human cells. These results demonstrate that rhamnose biosynthesis inhibitors can directly interfere with bacterial viability and could form a new class of antibiotics targeting nucleotide sugar production.

RESULTS

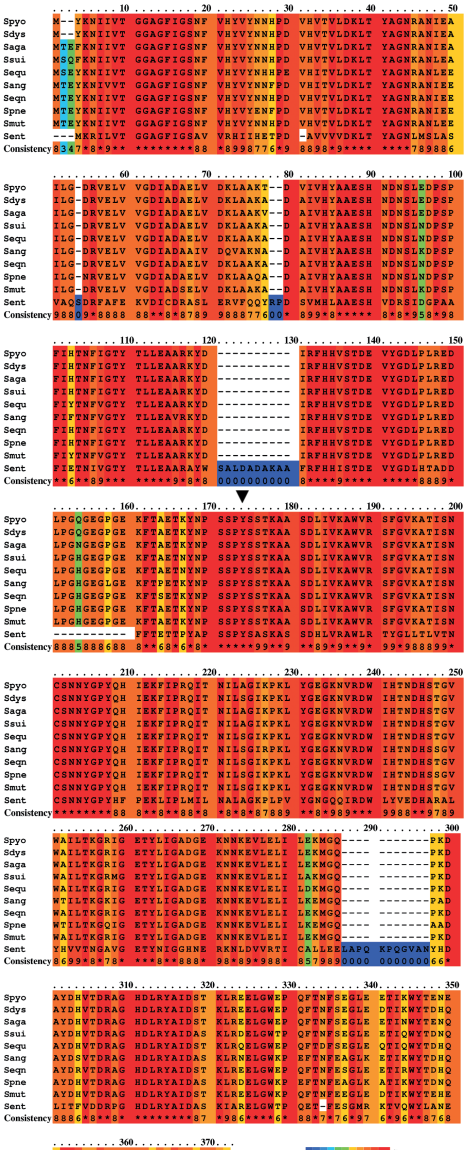
Protein sequence analysis of GAS RmlB and RmlC

As an extension of our previous work on GAS GacA [11], we sought to characterize GAS RmlB (GAS5448_RS05645) and RmlC (GAS5448_RS05650), the putative dTDP-glucose-4,6-dehydratase and dTDP-4-keto-6-deoxyglucose-3,5-epimerase, respectively. GAS *rmlB* and *rmlC* are clustered in an operon together with *rmlA* in the order *rmlACB*. Structural and biochemical analysis of RmlB and RmlC from *S. suis*, *S. enterica* and *M. tuberculosis* revealed that both enzymes are functional homodimers in these organisms [33, 35, 36, 40, 41]. Protein sequence alignment of RmlB and RmlC homologs from several streptococcal species and *S. enterica* displayed high homology (Fig. 2 and Table S1). Importantly, all catalytic residues in RmlB and RmlC are conserved (Fig. 2A, B and Table S1).

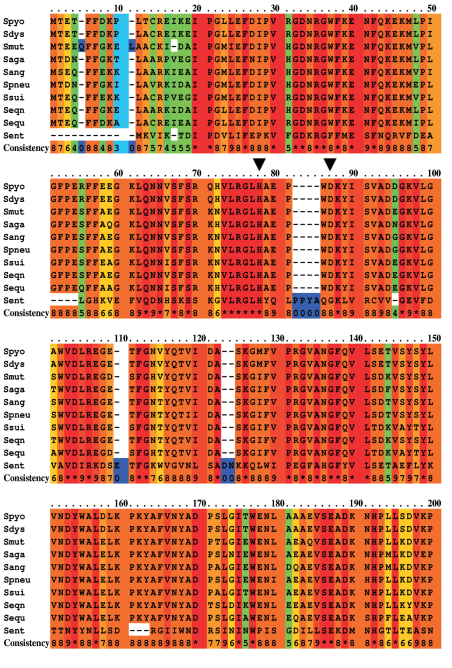
GAS RmlB and RmlC functionally replace S. mutans homologs

Most genes directly or indirectly involved in the GAC biosynthesis pathway are essential for GAS viability, including all four dTDP-L-rhamnose biosynthesis genes *rmlABC* and *gacA* [11, 14, 17, 18]. In *S. mutans*, the dTDP-L-rhamnose biosynthesis genes are essential in a transposon library [23]. Under non-competitive conditions however, it is possible to inactivate these genes in *S. mutans* Xc although this has a dramatic impact on growth and morphology [11, 48]. GAS RmlB and RmlC proteins share 94% and 83% protein sequence identity with the *S. mutans* homologs, respectively. In addition, the organization of these genes is identical in both species with the exception of a gene (174 nucleotides, 57 amino acids) between *rmlC* and *rmlB* in *S. mutans* encoding a hypothetical protein.

A. **RmlB alignment**



B. **RmlC alignment**



C.

Species	% Identity with *Streptococcus pyogenes* RmlB	% Identity with *Streptococcus pyogenes* RmlC
Salmonella enterica	55	34
Streptococcus agalactiae	98	83
Streptococcus anginosus	87	85
Streptococcus dysgalactiae	96	99
Streptococcus equi	92	83
Streptococcus equinus	95	85
Streptococcus mutans	94	83
Streptococcus pneumoniae	91	84
Streptococcus suis	94	84

Figure 2. Protein sequence alignment and identity matrix of RmlB and RmlC homologs. Color-coded representation of amino acid conservation for **(A)** GAS RmlB and **(B)** GAS RmlC to S. *enterica* and different streptococcal species. The amino acid conservation is scored from 0 to 10, with 0 (color blue) assigned to the least conserved residue and 10 (color red) to the most conserved residue. Critical enzymatic residues for RmlB (Y159) and RmlC (H76 and K82) are indicated with an inverted triangle. *S. pyogenes* (Spyo); *Salmonella enterica*; (Sent); *S. agalactiae* (Saga); *Streptococcus anginosus* (Sang); *Streptococcus dysgalactiae* (Sdys); *Streptococcus equi* (Sequ); *Streptococcus equinus* (Seqn); *S. mutans* (Smut); *S. pneumoniae* (Spneu); *Streptococcus suis* (Ssui). Protein accession numbers are described in the supplementary data (Table S1). **(C)** Percentage identity matrix of RmlB and RmlC homologs.

To confirm the function of GAS RmlB and RmlC in dTDP-L-rhamnose biosynthesis, we heterologously expressed GAS RmlB and GAS RmlC-encoding genes in *S. mutans* strains lacking *rmlB* or *rmlC*, respectively. Deletion of *S. mutans rmlB* (SMU Δ*rmlB*) or *rmlC* (SMU Δ*rmlC*) by replacement with an erythromycin resistance cassette severely attenuated bacterial growth compared to the wild-type strain (WT) (Fig. 3A, B), which is characteristic for a rhamnose-deficient *S. mutans* strain and in line with our previously constructed *S. mutans rmlD* deletion strain [11]. Morphological analysis of *rmlB* and *rmlC* mutant bacteria by scanning electron microscopy revealed swelling and clumping of bacteria as a result of misplaced septa resulting in division errors and multidirectional growth (Fig. 3C). Subsequent analysis of the cell wall carbohydrate composition by HPLC/MS confirmed that SMU Δ*rmlB* and SMU Δ*rmlC* lacked rhamnose in their cell walls, which concordantly resulted in the loss of the glucose side chains (Fig. 3D). Introduction of either homologous *S. mutans rmlB* or heterologous GAS *rmlB* on an expression plasmid in the corresponding SMU Δ*rmlB* mutant restored rhamnose incorporation in the cell wall (Fig. 3D) as well as the defective morphological phenotype and growth (Fig. 3A, C). Initially, we were unable to complement SMU Δ*rmlC* with *rmlC* from *S. mutans*, whereas heterologous GAS *rmlC* could restore growth, morphology and rhamnose production of the SMU Δ*rmlC* mutant (Fig. 3B, C, D). Upon reexamination of the UA159 genome, a *S. mutans* serotype c strain that we used as a reference genome for *S. mutans* Xc, we located an alternative ATG-start site 135 bp upstream of the annotated *rmlC* gene, which is in line with the annotation of *rmlC* in GAS. Importantly, available structural information indicates that the first 45 amino acids are part of the RmlC dimerization interface, in particular forming the extension of the beta-sheet with two additional beta-strands, which are required for nucleotide binding [35, 49]. In agreement with structural and genetic predictions, complementation of SMU Δ*rmlC* with the extended PCR product complemented all observed defects (Fig. 3B, C, D), indicating that the extended genomic PCR product encodes a functional RmlC enzyme (198 amino acids), which is similar to the size of GAS and *S. suis* RmlC (197 amino acids).

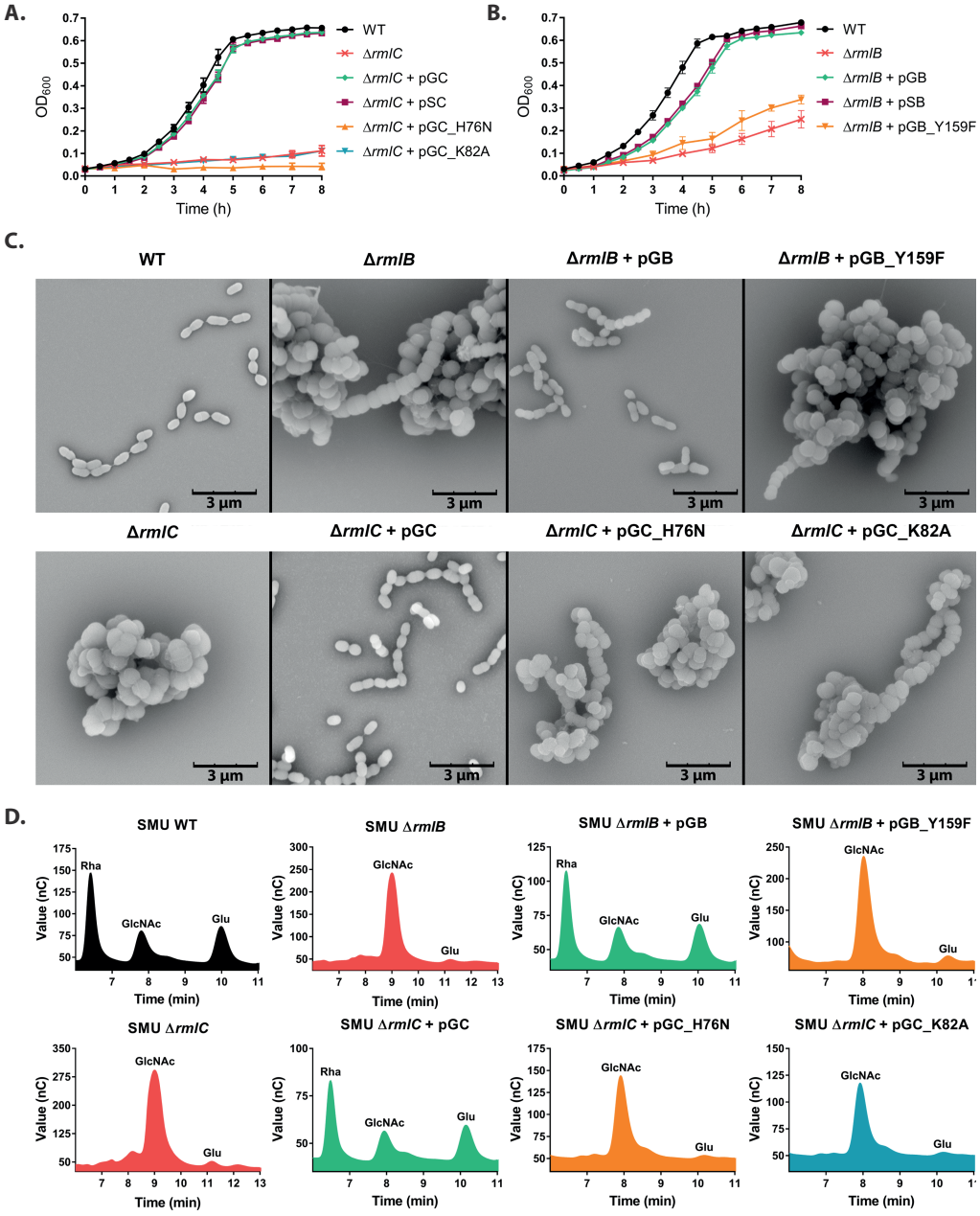

A.
OD600
Time (h)
WT
ΔrmlC
ΔrmlC + pGC
ΔrmlC + pSC
ΔrmlC + pGC_H76N
ΔrmlC + pGC_K82A
B.
OD600
Time (h)
WT
ΔrmlB
ΔrmlB + pGB
ΔrmlB + pSB
ΔrmlB + pGB_Y159F
C.
WT
ΔrmlB
ΔrmlB + pGB
ΔrmlB + pGB_Y159F
ΔrmlC
ΔrmlC + pGC
ΔrmlC + pGC_H76N
ΔrmlC + pGC_K82A
3 µm
D.
SMU WT
SMU ΔrmlB
SMU ΔrmlB + pGB
SMU ΔrmlB + pGB_Y159F
SMU ΔrmlC
SMU ΔrmlC + pGC
SMU ΔrmlC + pGC_H76N
SMU ΔrmlC + pGC_K82A
Rha
GlcNAc
Glu
Value (nC)
Time (min)

Figure 3. Heterologous expression of GAS RmlB and RmlC and catalytically-inactive enzymes in *S. mutans*. (A, B) Growth curves, **(C)** representative scanning electron microscopy images and **(D)** cell wall carbohydrate composition analysis of *S. mutans* wild-type (SMU WT), SMU Δ*rmlB*, SMU Δ*rmlC*, SMU Δ*rmlB* + pSB, SMU Δ*rmlB* + pGB(_Y159F), SMU Δ*rmlC* + pSC and SMU Δ*rmlC* + pGC(_H76N/K82A). Rha, rhamnose; GlcNAc, *N*-acetylglucosamine; Glu, glucose. Growth curves represent mean ± standard error of mean (SEM) of at least three biological repeats.

Catalytic residues of GAS RmlB and RmlC are conserved

Based on biochemical conformation of catalytic residues in *S. suis* RmlB and RmlC [36, 41] and protein sequence alignment with the GAS RmlB and RmlC homologs, we set out to functionally validate predicted catalytic residues of GAS RmlB (Y159) and RmlC (H76 and K82). Point mutations were individually introduced in GAS RmlB (Y159F) and RmlC (H76N and K82A) and overexpression vectors carrying these mutant genes were expressed in SMU Δ*rmlB* and SMU Δ*rmlC*, respectively. All mutated residues are involved in catalysis and mutated genes were unable to complement the characteristic rhamnose-depleted phenotype consisting of growth retardation and an aberrant morphology (Fig. 3A-C). Complementary, our cell wall composition analysis revealed that the *S. mutans* strains carrying these three mutated constructs completely lacked incorporation of rhamnose in their cell wall (Fig. 3D).

Inhibitor screen against GAS Rml proteins and hit confirmation

Given the importance of the dTDP-L-rhamnose biosynthesis pathway for viability or virulence of many human pathogens, we aimed to identify chemical scaffolds that could act as starting points for future optimization and drug development pipelines. Therefore, we conducted a bio-layer interferometry (BLI) inhibitor screen against the commercially available Maybridge Library using the three recombinant GAS dTDP-rhamnose biosynthesis enzymes RmlB, RmlC and GacA (Fig. S1A). The advantage of this approach is that it precludes the use of expensive or commercially unavailable enzyme substrates. Using a commercially available library of ~1,000 chemical fragments, our initial screen identified 12 hits of which seven fragments were found to specifically interact with GAS RmlB/RmlC/GacA as confirmed by BLI (Fig. S1B-D). In addition, we confirmed the identified hits in dilution series for their binding affinity to these three enzymes (Fig. S2 and Table S2). Importantly, this method allowed us to estimate enzyme specificity by comparing the data obtained from one enzyme to the other two. Using this strategy, we identified chemical scaffolds with specificity for one enzyme, but also several fragments that bound to more than one enzyme of the dTDP-rhamnose biosynthesis pathway (Table 1, S2, Fig. 4 and S2).

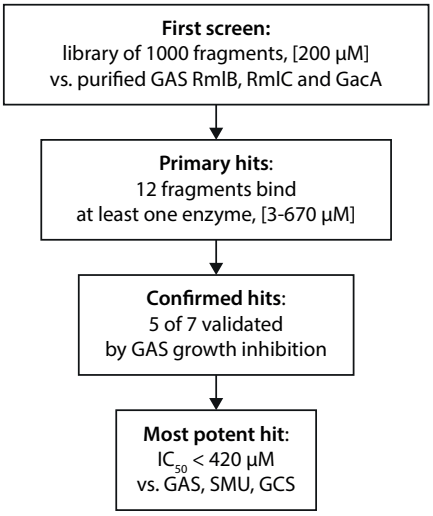


Figure 4. Screening and fragment validation flowchart.

Table 1. Overview of identified compounds and IC_{50} for GAS and *S. aureus* growth inhibition

Inhibitor	Compound	Binding target	IC_{50} (mM) GAS	IC_{50} (mM) *S. aureus*
Ri01	N-(1-Naphthyl)ethylenediamine dihydrochloride	RmlB + RmlC	0.27	0.68
Ri02	2,5-Difluorophenylhydrazine	RmlB + GacA	0.27	0.14
Ri03	5-(4-chlorophenyl)-2-furoic acid	RmlB + GacA	0.12	2.48
Ri06	4,4'-Thiodiphenol	RmlB + GacA	0.11	0.44
Ri08	2-(2,5-Dimethyl-1H-pyrrol-1-yl)benzoic acid	RmlB + RmlC	2.67	8.92

Bactericidal activity of identified fragments

Since dTDP-rhamnose is an essential nucleotide sugar in GAS [17, 50], we assessed the functional capacity of these scaffolds to impact GAS growth. All fragments identified in this study were water insoluble and were therefore dissolved in DMSO, except Ri07, which was already liquid at room temperature. GAS is tolerant to DMSO concentrations of 2%, therefore this was the maximum concentration used in bacterial assays. Even in DMSO, Ri04 and Ri07 were highly insoluble in bacterial culture medium and could therefore not be tested for inhibition of bacterial growth. The remaining five compounds were able to inhibit growth of GAS with IC_{50} values ranging from 110 µM to 6.2 mM (Fig. 5A, Table 1). In accordance with growth inhibition, GAS morphology was also severely affected upon exposure to either 200 µM Ri03 or 100 µM Ri06 (Fig. 5B). Especially for Ri03, streptococci were swollen and chains were longer, a phenotype reminiscent of an inducible *gacA* GAS knock-out strain [11].

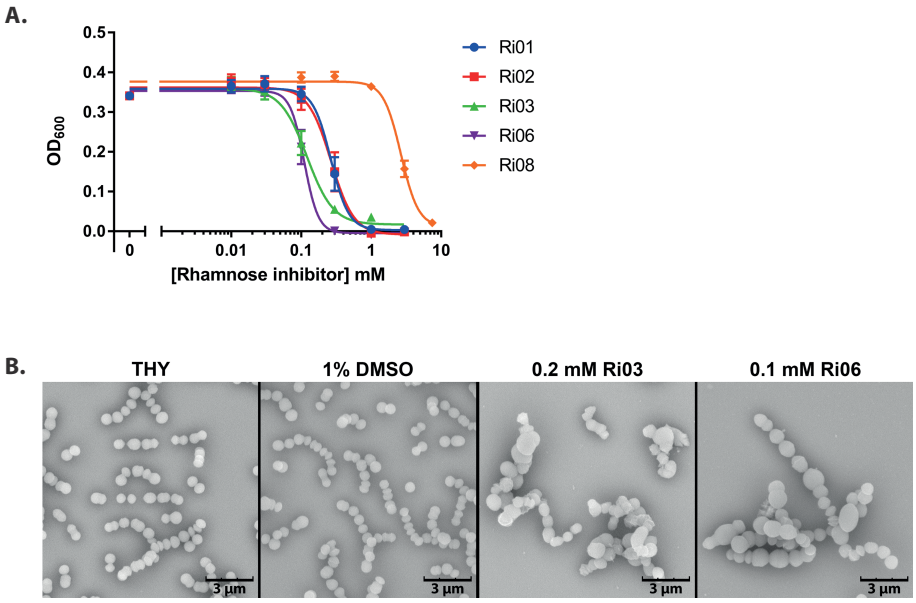


Figure 5. Identified fragments inhibit growth and affect morphology of GAS. (A) Dose-response curves for growth inhibition of GAS with various concentrations of Ri01, Ri02, Ri03, Ri06 or Ri08. IC_{50} values range from 110 µM to 2.66 mM (Table 1). **(B)** Representative scanning electron microscopy images of GAS after 16 h incubation in growth medium (THY), or THY with the addition of 1% DMSO, 0.2 mM Ri03 or 0.1 mM Ri06.

Target specificity of potential inhibitors

We aimed to gain more insight into target specificity of these five compounds. Therefore, we assessed whether Ri01, Ri02, Ri03, Ri06 and Ri08 inhibited growth of *Staphylococcus aureus* USA300 NRS384, a pathogenic methicillin-resistant bacterium that lacks expression of rhamnose (Fig. 6A, Table 1). Four of the five compounds only showed 2- or 4-fold differences in IC_{50} compared to GAS. However, compound Ri03 inhibited growth of GAS ~20-fold more effective compared to *S. aureus* (120 µM vs. 2.48 mM). We also assessed growth inhibition on an *E. coli* strain that produces non-essential rhamnose [51] as confirmed by deletion of *rfbA*, an *rmlA* homolog, which did not affect the growth rate. Similar to *S. aureus*, Ri03 was approximately 30-fold less effective on this *E. coli* compared to GAS with an IC_{50} of 3.61 mM (Fig 6B, Table S3) and was not affected by mutation of *rfbA* (Fig 6B, Table S3), suggesting that the observed growth inhibition, both in *E. coli* and *S. aureus*, is due to off-target effects of Ri03.

Ri03 prevents growth of other pathogenic streptococci

The observed window in Ri03 efficacy between GAS and *E. coli*/*S. aureus* suggested that Ri03-mediated growth inhibition is partially dependent on inhibition of dTDP-L-rhamnose biosynthesis. We extended our experiments to include other human and animal streptococcal species for which dTDP-rhamnose is an essential nucleotide sugar. GCS contains a related surface carbohydrate, the characteristic group C carbohydrate, composed of a polyrhamnose backbone decorated with a di-GalNAc side chain [52, 53]. Ri03 inhibited growth of GCS with an IC_{50} of 0.42 mM (Table S3, Fig. 7) and of *S. mutans* with an IC_{50} of 0.41 mM (Table S3, Fig. 7).

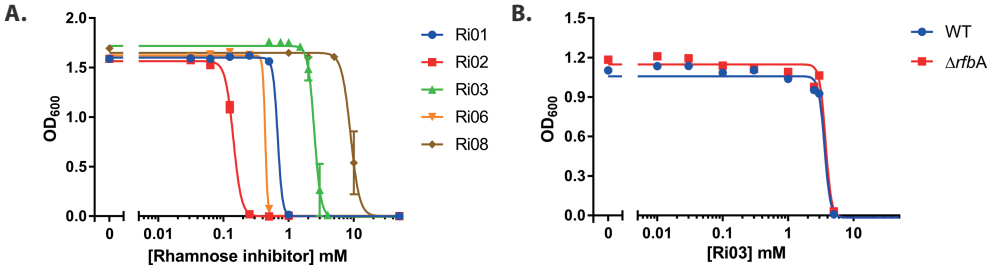


Figure 6. Growth inhibition of rhamnose-independent bacteria with rhamnose inhibitors. (A) Dose-response curves for growth inhibition of *S. aureus* NRS384, a strain that does not express dTDP-L-rhamnose, with various concentrations of Ri01, Ri02, Ri03, Ri06 or Ri08 to determine the IC_{50} (Table 1). **(B)** Dose-response curves for growth inhibition of *E. coli* BW25113 WT and Δ*rfbA*, a rhamnose-deficient mutant, with various concentrations of Ri03 to determine the IC_{50}. Dose-response curves represent mean ± SEM of at least three biological experiments performed in duplicate.

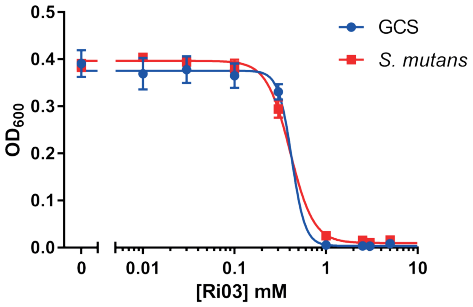


Figure 7. Identified fragments inhibit growth of streptococcal species. Dose-response curves for growth inhibition of GCS MGCS10565 and *S. mutans* Xc with various concentrations of Ri03. IC_{50} values are 420 µM and 410 µM, respectively. Dose-response curves represent mean ± SEM of at least three biological experiments performed in duplicate.

Cytotoxicity of Ri03 in human U937 cells

Human cells do neither produce L-rhamnose nucleotide sugars, nor is there any evidence for them incorporating L-rhamnose into their carbohydrates [31]. Nonetheless, blasting of bacterial rhamnose biosynthesis enzyme sequences against

the human genome revealed a possible human homolog of RmlB. The human gene *TGDS* is annotated to encode a dTDP-glucose-4,6-dehydratase, which shares 33-36% protein sequence identity with Gram-positive and Gram-negative RmlB enzymes (Fig. S3A, B). We therefore investigated whether TGDS could act as a functional RmlB homolog by complementing SMU Δ*rmlB* with codon-optimized *TGDS*, which should assist the production of TGDS in bacterial cells, on an expression plasmid. However, TGDS was not able to complement the attenuated growth rate or aberrant morphology of SMU Δ*rmlB* (Fig. S3C).

Given potential presence of structural homologs in eukaryotic cells, Ri03 was tested for cytotoxicity on U937 cells, a human monocytic cell line, by measuring the release of lactose dehydrogenase (LDH) into the supernatant. Ri03 induced 5% LDH release with a concentration of 20 mM Ri03 (Fig. 8). LDH release was comparable to DMSO controls for concentrations up to 15 mM. Compared to the IC_{50} of Ri03 on GAS (120 µM), there is a therapeutic window of more than 150-fold.

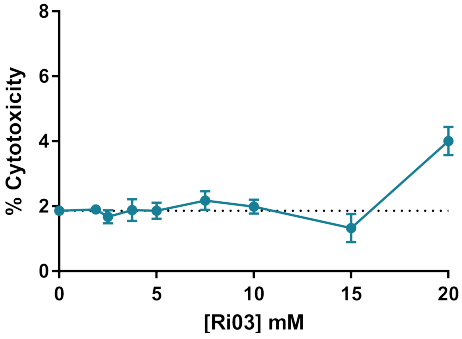


Figure 8. Cytotoxicity of Ri03 on human monocytic cells. Release of lactose dehydrogenase by U937 monocytic cells after Ri03 exposure as a measurement for cytotoxicity. Cytotoxicity curves represent mean ± SEM of three biological experiments performed in duplicate.

DISCUSSION

In this study we show that GAS RmlB and RmlC are the dTDP-glucose-4,6-dehydratase and dTDP-4-keto-6-deoxyglucose-3,5-epimerase enzymes, respectively. Both GAS RmlB and RmlC functionally replaced *S. mutans* homologs in a heterologous expression system. The rescue of growth and morphological phenotypes of the *rmlB* and *rmlC* mutans by plasmid overexpression excludes the occurrence of polar effects within the *rmlACB* operon. In addition to proving the function of both GAS RmlB and RmlC as rhamnose biosynthesis enzymes, we also identified Y159 in RmlB and H76 and K82 in RmlC as critical catalytic residues. Our complementation studies indicate that the mechanism of catalysis is conserved among streptococcal species.

Deletion of *rmlB* and *rmlC* in *S. mutans* resulted in a phenotype similar as previously observed for an *rmlD* deletion mutant [11], underscoring that inhibition of the dTDP-rhamnose biosynthesis pathway severely impacts bacterial viability. Indeed, the *rmlABCD* genes are essential for *S. mutans* in the competitive environment of a mutant transposon library [23], similar to GAS and GCS [17, 50, 54]. We are however able to construct *rml* deletion mutants in *S. mutans* in isolation in contrast to similar attempts in GAS [12]. The reason for this discrepancy is currently unclear but may be caused by differential compositions of the cell walls or possibly these deletions are partially rescued through interacting pathways.

Screening a chemical compound library to identify inhibitors of the dTDP-L-rhamnose biosynthesis pathway is methodically challenging. Previously, indirect methods were used to determine the production of the end product dTDP-L-rhamnose by monitoring the production of co-factor NAD(P)H [11, 46, 55]. A superior assay would involve a HPLC or mass-spectrometry approach where every single step of product formation and substrate consumption could be monitored. However, it is very challenging to develop this method suitable for medium or high throughput screening. We therefore investigated an approach using recombinant enzymes in their natural state, i.e. with bound co-factors in their active site. RmlB enzymes require the co-factors NAD(P), which appears to be tightly bound in the active site as it was co-purified during RmlB crystallizations [33, 42]. This is in agreement with the regeneration of NAD during dTDP-4-keto-6-deoxy-D-glucose synthesis [42]. GacA/RmlD enzymes also require NADPH as a co-factor, which appears to be less tightly bound as it was not observed in the GAS GacA and *S. suis* RmlD crystal structures and therefore should be regenerated by a different enzyme [11, 37]. Using the binding approach, we identified seven potential inhibitors that interacted with at least one enzyme of the dTDP-L-rhamnose biosynthesis pathway with low millimolar to high micromolar affinities. One of the most potent chemical fragments Ri03, 5-(4-chlorophenyl)-2-furoic acid, is a small molecule, which binds RmlB and GacA in the high μM range, and RmlC in the low millimolar range. Despite the fact that these are only fragments, which are generally known to have low affinity and limited specificity, we demonstrated that Ri03 prevented growth of several streptococcal strains, which express rhamnose as an essential component of their cell wall. The high sequence identity of RmlB and GacA/RmlD amongst GAS, GCS and *S. mutans* (92-94% and 82-87% protein sequence identity, respectively; Fig. 2C) supports the observation that Ri03 shows similar IC_{50} values against all three strains. Importantly, Ri03 only inhibits growth in the millimolar range of species that do not produce rhamnose (*S. aureus*) or for which rhamnose is not essential (*E. coli*). This data suggests that compound Ri03 has, in addition to on-target effects, also minor off-target effects for which the targets remain unknown.

To our surprise, bioinformatics revealed that the human protein TGDS shares 33-36% protein sequence identity with bacterial RmlB enzymes and has previously been annotated as a dTDP-glucose-4,6 dehydratase [56]. It seems unlikely that this annotation for TGDS is correct, considering that both dTDP-L-rhamnose and dTDP-D-glucose, the substrate of RmlB, have not been detected in human cells [31, 57]. In support of this, we were unable to complement SMU Δ*rmlB* with codon-optimized *TGDS*. However, it remains unknown whether TGDS was produced as a functional enzyme and whether it is a promiscuous enzyme that is able to convert dTDP-glucose into dTDP-4-keto-6-deoxy-D-glucose. We therefore advise to test potential inhibitors, such as Ri03 and future derivatives, against RmlB, such as Ri03, for binding to TGDS to exclude off target effects as a precautionary measure, given that mutations in TGDS cause the rare Catel-Manzke syndrome (Fig. S3A) [56, 58, 59].

Finally, encouraged by the apparent specificity of Ri03 to inhibit bacterial growth in a rhamnose-dependent manner, we tested the cytotoxicity of Ri03 on human cells and showed that no cytotoxicity of Ri03 occurred in U937 human monocytic cells with concentrations up to 15 mM. Certainly, Ri03 needs further optimization before it can be used as a therapeutic agent and more studies are required to test the cytotoxic effects of Ri03 (and subsequent derivatives) on other human cells. However, these first results suggest a reasonable therapeutic window, which may be increased after additional optimization of this chemical scaffold. Future experiments aim to investigate the binding mode of Ri03 on a structural level to the proteins RmlB, RmlC and GacA to elucidate whether Ri03 inhibits RmlB, RmlC and GacA by competing with the nucleotide-sugar and if a common binding mode is observed amongst the different enzymes. With more functional and structural knowledge, Ri03 can be further optimized to identify more potent and specific derivatives with greater potency targeting the GAS and related homologous Rml enzymes in other pathogenic species.

MATERIALS AND METHODS

Bacterial strains and growth conditions

Bacterial strains used in this study are described in Table 2. *S. mutans* Xc [60], a serotype c wild-type (WT) strain, was a kind gift of Dr. Y. Yamashita (Kyushu University, Japan) and was routinely grown in Todd-Hewitt Broth (THB, Oxoid) or on THB agar at 37°C with 5% CO_2. When appropriate, *S. mutans* was cultured with 10 µg/ml erythromycin (Erm) or 3 µg/ml chloramphenicol (Cm). GAS 5448, a representative of the epidemic M1T1 clone, was cultured in Todd-Hewitt Broth (Becton Dickinson) supplemented

with 1% yeast extract (THY; Oxoid) or on THY agar at 37°C. *E. coli* MC1061 was used for cloning purposes and was grown in lysogeny broth (LB, Oxoid) or on LB agar containing 10 µg/ml CHL at 37 °C.

Table 2. Bacterial strains used in this study

Bacterial strains	Abbreviation	Resistance
S. mutans Xc wild type (serotype c)	SMU WT	-
S. mutans Xc Δ*rmlB*	SMU Δ*rmlB*	Erm
S. mutans Xc Δ*rmlB* + pDC123_SMU_*rmlB*	SMU Δ*rmlB* + pSB	Erm + Cm
S. mutans Xc Δ*rmlB* + pDC123_GAS_*rmlB*	SMU Δ*rmlB* + pGB	Erm + Cm
S. mutans Xc Δ*rmlB* + pDC123_GAS_*rmlB*_Y159F	SMU Δ*rmlB* + pGB_Y159F	Erm + Cm
S. mutans Xc Δ*rmlB* + pDC123_*TGDS*	SMU Δ*rmlB* + pTGDS	Erm + Cm
S. mutans Xc Δ*rmlC*	SMU Δ*rmlC*	Erm
S. mutans Xc Δ*rmlC* + pDC123_SMU_*rmlC*	SMU Δ*rmlC* + pSC	Erm + Cm
S. mutans Xc Δ*rmlC* + pDC123_GAS_*rmlC*	SMU Δ*rmlC* + pGC	Erm + Cm
S. mutans Xc Δ*rmlC* + pDC123_GAS_*rmlC*_H76N	SMU Δ*rmlC* + pGC_H76N	Erm + Cm
S. mutans Xc Δ*rmlC* + pDC123_GAS_*rmlC*_K82A	SMU Δ*rmlC* + pGC_K82A	Erm + Cm
S. pyogenes 5448 wild type (serotype M1T1)	GAS	-
E. coli MC1061	-	-
E. coli BL21 (DE3)	-	-
E. coli DH5alpha	-	-
E. coli K-12 BW25113	-	-
E. coli JW2023 – Δ*rfbA* in BW25113	Δ*rfbA*	Kan
S. aureus USA300 NRS384	-	-
S. equi subsp. *zooepidemicus* MGCS10565	GCS	-

Cloning, expression and purification of GAS enzymes

GAS RmlB, RmlC and GacA proteins were produced and purified as described previously [11].

Genetic manipulation of S. mutans

S. mutans is naturally competent and was transformed as described previously. Shortly, bacteria were grown in THB containing 5% heat-inactivated horse serum (BioRad) and supplied with 500 ng knockout construct or complementation plasmid [11]. Cultures were plated on THB agar plates containing the appropriate antibiotics. Single colonies were selected and verified for the deletion of *rmlB*/*rmlC* and/or the presence of complementation plasmid using colony PCR and sequencing.

Deletion mutants of *rmlB* (SMU Δ*rmlB*) and *rmlC* (SMU Δ*rmlC*) were obtained via the addition of a knockout construct to competent *S. mutans* WT. This construct consisted of an Erm resistance cassette with ~700 bp flanking regions homologous to the up- and downstream regions of *rmlB* or *rmlC*. A detailed cloning strategy and a primer list are available in the supplemental material (Method S1 and Table S4).

S. mutans complementation plasmids

S. mutans rmlB/rmlC and GAS *rmlB/rmlC* were amplified from gDNA using primers containing XbaI, XhoI or BamHI restriction site (Table S4). Digested PCR products were subsequently ligated into expression plasmid pDC123 and propagated in *E. coli* MC1061 to obtain large quantities of complementation plasmid. Point mutations were introduced by mutagenesis PCR, using overlapping primers with the corresponding mutation site integrated. The constructed plasmids with the single point mutations were all confirmed by DNA sequencing.

For complementation of SMU Δ*rmlB* with the human gene *TGDS* (gene ID: 23483), the exons of *TGDS* were first codon optimized for SMU UA159. A gBlock of this optimized sequence with XbaI-pDC123 and BglII-pDC123 overhangs attached to the 5'- and 3'-ends, respectively, was used to ligate TGDS into pDC123 (pTGDS), which was confirmed by DNA sequencing.

SMU Δ*rmlB* and SMU Δ*rmlC* mutants were not transformable likely resulting from the severe growth defects. Therefore, SMU Δ*rmlB* and SMU Δ*rmlC* strains complemented with *S. mutans rmlB/rmlC* or GAS *rmlB/rmlC* were constructed using a two-step approach. First, *S. mutans* WT was transformed with the complementation plasmid as described above. Next, these complemented strains were transformed with the respective *rmlB* or *rmlC* knockout constructs and selected for double antibiotic resistance.

Scanning Electron Microscopy

Scanning electron microscopy was performed as described previously by van der Beek et al [11]. In short, bacteria were grown to mid-exponential phase, except when incubated with rhamnose inhibitor. In this case, bacteria were cultured as described below in bacterial growth inhibition assays and grown overnight. Bacteria were subsequently washed, fixed, dehydrated, mounted onto 12.5 mm specimen stubs (Agar scientific, Stansted, Essex, UK) and coated with gold to 1 nm using a Quorum Q150R S sputter coater at 20 mA. Samples were visualized with a Phenom PRO desktop scanning electron microscope (Phenom-World BV) at an acceleration voltage of 10 kV.

S. mutans growth curves

Overnight cultures of *S. mutans* strains with an optical density at 600 nm (OD_{600}) higher than 0.35 were diluted 10 times and grown for 1.5 h to early exponential phase. Cultures were then diluted to OD_{600} 0.03 and OD_{600} was measured manually every 30 min. For cultures with an overnight OD_{600} below 0.35, the initial 10-fold dilution and growth step was omitted. Instead, such overnight cultures were directly diluted to OD_{600} 0.03 and OD_{600} was measured every hour. All cultures were incubated at 37 °C with 5% CO_2 in between measurements.

Cell wall carbohydrate composition analysis

Cell wall polysaccharides were isolated from *S. mutans,* hydrolyzed and analyzed by chromatography as described previously by van der Beek *et al*. [11]. In short, bacterial cells were harvested from 2-5 L cultures and disrupted in 0.1 M citrate buffer (pH 4.6) using a bead beater (Biospec). Cell walls were collected by centrifugation and boiled for 1 h at 100 °C in 0.1 M sodium acetate (pH 4.6) containing 4% sodium dodecyl sulfate. Samples were subsequently treated with RNase, DNase, pronase E and trypsin. Cell wall polysaccharides containing peptidoglycan and the serotype c carbohydrate were lyophilized before hydrolysis with TFA. Finally, carbohydrate analysis of monosaccharides was performed on a Dionex ICS-3000 Ion Chromatography System (Dionex / Thermo Scientific, 1228 Titan Way, P.O. Box 3603, Sunnyvale, CA, 94088-3603, United States) using a CarboPac PA20 (Dionex / Thermo Scientific, 1228 Titan Way, P.O. Box 3603, Sunnyvale, CA, 94088-3603, United States) 3 × 150 mm column, equipped with a ICS-3000 Electrochemical Detector (Dionex / Thermo Scientific, 1228 Titan Way, P.O. Box 3603, Sunnyvale, CA, 94088-3603, United States). Monosaccharides from all samples were eluted from the column with 12% 0.2 M NaOH, except for monosaccharides from SMU Δ*rmlB* and Δ*rmlC*, which were eluted with 8% 0.2 M NaOH. The 'Carbohydrates (Standard Quad)' waveform was used for detection.

BLI screen using recombinant RmlB, RmlC and GacA

The Maybridge library of fragment-like molecules (Ro3) was purchased from Maybridge (USA). All three GAS proteins, RmlB, RmlC and GacA, were screened against the library using identical protocols. Proteins were biotinylated on primary amine amino acid side chains using the Thermo Scientific™ EZ-Link™ NHS-PEG_4-Biotin reagent. The enzymes were incubated with NHS-PEG_4-Biotin in a 1:1 molar ratio at 150 µM for 45 min at room temperature. The reaction buffer contained 25 mM HEPES, pH 7.5, 150 mM NaCl and 0.02 mM TCEP. Excess NHS-PEG_4-biotin reagent was removed using 2 mL Zeba desalting spin columns (Thermo Scientific). Successful biotinylation on primary amines was investigated via Western blotting

and probing with ExtrAvidin®–Peroxidase antibody, 1:10,000 (Sigma E2886, Fig. S1A). Before compound screening, the proteins were loaded onto a parallel set of superstreptavidin biosensors (SSA) by incubation in buffer for 900 sec. Protein concentrations suitable for the experiments were determined in a series of dilutions. Final screening was conducted at concentrations of: 50 µg/mL RmlB, 25 µg/mL RmlC and 12.5 µg/mL GacA. All compounds were at 2% DMSO concentration. The sensor was blocked by immersion in biocytin (10 µg/mL) for 30 sec.

Initial library screening was performed at a compound concentration of 200 µM. Hits were identified by plotting the response rate of every single compound after background subtraction. In total, 12 compounds were validated in 6-point concentration series using 3-fold dilutions, ranging from 670 µM to 3 µM. Data were processed and kinetic parameters were calculate using the ForteBio software. Binding curves were manually inspected and approximate binding constants K_D are listed in Table S2.

Bactericidal activity

GAS 5448, *S. aureus* NRS384, *E. coli* BW25113, GCS MGCS10565 and *S. mutans* Xc strains (Table 2) were grown in appropriate bacterial culture broth overnight at 37 °C. Bacterial cultures were diluted 10 times and grown to mid-log phase (OD_{600} 0.4). Cultures were diluted 200 times in culture medium followed by a 2-fold dilution with various concentrations of rhamnose inhibitor or DMSO. OD_{600} was recorded every 15 min at 37°C with 5 sec medium shaking before measurement for GAS, GCS, *S. mutans* and *E. coli* or medium continuous shaking for *S. aureus* using either a Bioscreen C MBR machine (Growth Curves AB Ltd, Oy, Finnland) or a Synergy 2 plate reader (Biotek). OD_{600} values were plotted against the concentration of compound after 9 hours of growth using GraphPad Prism 7. IC_{50} values were calculated using a non-linear four-parameter dose-response curve with variable slope. Because high concentrations were not always feasible due to solubility issues, a high concentration of compound (50 mM) was artificially introduced where appropriate and set to zero to obtain more reliable dose-response curves.

Cytotoxicity

U937 human monocytic cells (American Type Culture Collection) were grown at 37 °C with 5% CO_2 in RPMI 1640 (Biowest) containing 100 U/ml penicillin (Gibco), 100 µg/ml streptomycin (Gibco) and 5% fetal calf serum (Biowest). Cells were then centrifuged and adjusted to 10^7 cells/ml in THB medium. In a 96-well round bottom plate, 10^5 U937 cells were added to 100 µl of THB containing various concentrations of Ri03, DMSO or lysis buffer. Cells were incubated for 30 min at 37 °C with 5% CO_2 with mild

shaking. U937 cells were centrifuged and 50 µl of supernatant was transferred to a new 96-well flat bottom plate. Lactose dehydrogenase release was measured using the CytoTox 96® Non-Radioactive Cytotoxicity Assay kit (Promega) by adding 50 µl of substrate reagent. OD_{490} was measured after 30 min using an iMark™ microplate absorbance reader (Bio-Rad). No stop buffer was added to avoid precipitation of Ri03. Percentage of cytotoxicity is calculated relative to lysis buffer control.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Dr Olawale Raimi (University of Dundee) for assistance with BLI. We are grateful for funding to HCD and his laboratory from Tenovus Scotland, Wellcome Trust and Royal Society (grant number 109357/Z/15/Z.), FEBS and the University of Dundee. This work was supported by a VIDI grant (91713303) from the Dutch Scientific Organization (NWO) to N.M.v.S and S.v.d.B. The authors declare no conflict of interest.

REFERENCES

1. Mistou, M.Y., Sutcliffe, I.C., and van Sorge, N.M.; *Bacterial glycobiology: rhamnose-containing cell wall polysaccharides in Gram-positive bacteria.* FEMS Microbiol Rev, 2016. **40**(4): p. 464-79.
2. Caliot, E., et al.; *Role of the Group B antigen of Streptococcus agalactiae: a peptidoglycan-anchored polysaccharide involved in cell wall biogenesis.* PLoS Pathog, 2012. **8**(6): p. e1002756.
3. Sutcliffe, I.C., Black, G.W., and Harrington, D.J.; *Bioinformatic insights into the biosynthesis of the Group B carbohydrate in Streptococcus agalactiae.* Microbiology, 2008. **154**(5): p. 1354-1363.
4. Heymann, H., et al.; *Structure of Streptococcal Cell Walls. Ii. Group a Biose and Group a Triose from C-Polysaccharide.* J Biol Chem, 1964. **239**: p. 1656-63.
5. Huang, D.H., Rama Krishna, N., and Pritchard, D.G.; *Characterization of the group A streptococcal polysaccharide by two-dimensional 1H-nuclear-magnetic-resonance spectroscopy.* Carbohydr Res, 1986. **155**: p. 193-9.
6. St Michael, F., et al.; *Investigating the candidacy of the serotype specific rhamnan polysaccharide based glycoconjugates to prevent disease caused by the dental pathogen Streptococcus mutans.* Glycoconj J, 2017.
7. Pritchard, D.G., et al.; *Characterization of the serotype e polysaccharide antigen of Streptococcus mutans.* Mol Immunol, 1986. **23**(2): p. 141-5.
8. Linzer, R., Reddy, M.S., and Levine, M.J.; *Structural studies of the serotype-f polysaccharide antigen from Streptococcus mutans OMZ175.* Infect Immun, 1987. **55**(12): p. 3006-10.
9. Nakano, K. and Ooshima, T.; *Serotype classification of Streptococcus mutans and its detection outside the oral cavity.* Future Microbiol, 2009. **4**(7): p. 891-902.
10. Lancefield, R.C.; *A Serological Differentiation of Human and Other Groups of Hemolytic Streptococci.* J Exp Med, 1933. **57**(4): p. 571-95.
11. van der Beek, S.L., et al.; *GacA is Essential for Group A Streptococcus and Defines a New Class of Monomeric dTDP-4-dehydrorhamnose Reductases (RmlD).* Mol Microbiol, 2015.
12. van Sorge, N.M., et al.; *The classical lancefield antigen of group a streptococcus is a virulence determinant with implications for vaccine design.* Cell Host Microbe, 2014. **15**(6): p. 729-40.
13. Kovacs, C.J., Faustoferri, R.C., and Quivey, R.G., Jr.; *RgpF Is Required for Maintenance of Stress Tolerance and Virulence in Streptococcus mutans.* J Bacteriol, 2017. **199**(24).
14. Zhu, L., et al.; *Novel Genes Required for the Fitness of Streptococcus pyogenes in Human Saliva.* mSphere, 2017. **2**(6).
15. Sabharwal, H., et al.; *Group A streptococcus (GAS) carbohydrate as an immunogen for protection against GAS infection.* J Infect Dis, 2006. **193**(1): p. 129-35.
16. Giraud, M.F. and Naismith, J.H.; *The rhamnose pathway.* Curr Opin Struct Biol, 2000. **10**(6): p. 687-96.
17. Le Breton, Y., et al.; *Essential Genes in the Core Genome of the Human Pathogen Streptococcus pyogenes.* Sci Rep, 2015. **5**: p. 9838.
18. Le Breton, Y., et al.; *Genome-wide identification of genes required for fitness of group A Streptococcus in human blood.* Infect Immun, 2013. **81**(3): p. 862-75.
19. Hooven, T.A., et al.; *The essential genome of Streptococcus agalactiae.* BMC Genomics, 2016. **17**: p. 406.
20. Kim, J.O., et al.; *Relationship between cell surface carbohydrates and intrastrain variation on opsonophagocytosis of Streptococcus pneumoniae.* Infect Immun, 1999. **67**(5): p. 2327-33.
21. Magee, A.D. and Yother, J.; *Requirement for capsule in colonization by Streptococcus pneumoniae.* Infect Immun, 2001. **69**(6): p. 3755-61.
22. Engels-Deutsch, M., et al.; *Insertional inactivation of pac and rmlB genes reduces the release of tumor necrosis factor alpha, interleukin-6, and interleukin-8 induced by Streptococcus mutans in monocytic, dental pulp, and periodontal ligament cells.* Infect Immun, 2003. **71**(9): p. 5169-77.

23. Shields, R.C., et al.; *Genomewide Identification of Essential Genes and Fitness Determinants of Streptococcus mutans UA159.* mSphere, 2018. **3**(1).
24. Tsuda, H., et al.; *Role of serotype-specific polysaccharide in the resistance of Streptococcus mutans to phagocytosis by human polymorphonuclear leukocytes.* Infect Immun, 2000. **68**(2): p. 644-50.
25. Xu, Y., et al.; *Analysis of a gene cluster of Enterococcus faecalis involved in polysaccharide biosynthesis.* Infect Immun, 2000. **68**(2): p. 815-23.
26. Rigottier-Gois, L., et al.; *The surface rhamnopolysaccharide epa of Enterococcus faecalis is a key determinant of intestinal colonization.* J Infect Dis, 2015. **211**(1): p. 62-71.
27. Li, W., et al.; *rmlB and rmlC genes are essential for growth of mycobacteria.* Biochem Biophys Res Commun, 2006. **342**(1): p. 170-8.
28. Ma, Y., Pan, F., and McNeil, M.; *Formation of dTDP-rhamnose is essential for growth of mycobacteria.* J Bacteriol, 2002. **184**(12): p. 3392-5.
29. Engels, W., et al.; *Role of lipopolysaccharide in opsonization and phagocytosis of Pseudomonas aeruginosa.* Infect Immun, 1985. **49**(1): p. 182-9.
30. Joiner, K.A.; *Complement evasion by bacteria and parasites.* Annu Rev Microbiol, 1988. **42**: p. 201-30.
31. Adibekian, A., et al.; *Comparative bioinformatics analysis of the mammalian and bacterial glycomes.* Chem. Sci., 2011. **2**(2): p. 337-344.
32. Blankenfeldt, W., et al.; *The structural basis of the catalytic mechanism and regulation of glucose-1-phosphate thymidylyltransferase (RmlA).* EMBO J, 2000. **19**(24): p. 6652-63.
33. Beis, K., et al.; *The structure of NADH in the enzyme dTDP-d-glucose dehydratase (RmlB).* J Am Chem Soc, 2003. **125**(39): p. 11872-8.
34. Dong, C., et al.; *RmlC, a C3′ and C5′ carbohydrate epimerase, appears to operate via an intermediate with an unusual twist boat conformation.* J Mol Biol, 2007. **365**(1): p. 146-59.
35. Giraud, M.F., et al.; *RmlC, the third enzyme of dTDP-L-rhamnose pathway, is a new class of epimerase.* Nat Struct Biol, 2000. **7**(5): p. 398-402.
36. Dong, C., et al.; *High-resolution structures of RmlC from Streptococcus suis in complex with substrate analogs locate the active site of this class of enzyme.* Structure, 2003. **11**(6): p. 715-23.
37. Blankenfeldt, W., et al.; *Variation on a theme of SDR. dTDP-6-deoxy-L- lyxo-4-hexulose reductase (RmlD) shows a new Mg2+-dependent dimerization mode.* Structure, 2002. **10**(6): p. 773-86.
38. van der Beek, S.L., et al.; *GacA is Essential for Group A Streptococcus and Defines a New Class of Monomeric dTDP-4-dehydrorhamnose Reductases (RmlD).* Mol Microbiol, 2015.
39. Dong, C., et al.; *A structural perspective on the enzymes that convert dTDP-d-glucose into dTDP-l-rhamnose.* Biochem Soc Trans, 2003. **31**(Pt 3): p. 532-6.
40. Kantardjieff, K.A., et al.; *Mycobacterium tuberculosis RmlC epimerase (Rv3465): a promising drug-target structure in the rhamnose pathway.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 5): p. 895-902.
41. Allard, S.T., et al.; *The crystal structure of dTDP-D-Glucose 4,6-dehydratase (RmlB) from Salmonella enterica serovar Typhimurium, the second enzyme in the dTDP-l-rhamnose pathway.* J Mol Biol, 2001. **307**(1): p. 283-95.
42. Allard, S.T., et al.; *Toward a structural understanding of the dehydratase mechanism.* Structure, 2002. **10**(1): p. 81-92.
43. Babaoglu, K., et al.; *Novel inhibitors of an emerging target in Mycobacterium tuberculosis; substituted thiazolidinones as inhibitors of dTDP-rhamnose synthesis.* Bioorg Med Chem Lett, 2003. **13**(19): p. 3227-30.
44. Ma, Y., et al.; *Drug targeting Mycobacterium tuberculosis cell wall synthesis: genetics of dTDP-rhamnose synthetic enzymes and development of a microtiter plate-based screen for inhibitors of conversion of dTDP-glucose to dTDP-rhamnose.* Antimicrob Agents Chemother, 2001. **45**(5): p. 1407-16.

45. Ren, J.X., et al.; *Virtual screening for the identification of novel inhibitors of Mycobacterium tuberculosis cell wall synthesis: inhibitors targeting RmlB and RmlC.* Comput Biol Med, 2015. **58**: p. 110-7.
46. Sivendran, S., et al.; *Identification of triazinoindol-benzimidazolones as nanomolar inhibitors of the Mycobacterium tuberculosis enzyme TDP-6-deoxy-d-xylo-4-hexopyranosid-4-ulose 3,5-epimerase (RmlC).* Bioorg Med Chem, 2010. **18**(2): p. 896-908.
47. Wang, Y., et al.; *Novel inhibitors of Mycobacterium tuberculosis dTDP-6-deoxy-L-lyxo-4-hexulose reductase (RmlD) identified by virtual screening.* Bioorg Med Chem Lett, 2011. **21**(23): p. 7064-7.
48. Tsukioka, Y., et al.; *Biological function of the dTDP-rhamnose synthesis pathway in Streptococcus mutans.* J Bacteriol, 1997. **179**(4): p. 1126-34.
49. Christendat, D., et al.; *Crystal structure of dTDP-4-keto-6-deoxy-D-hexulose 3,5-epimerase from Methanobacterium thermoautotrophicum complexed with dTDP.* J Biol Chem, 2000. **275**(32): p. 24608-12.
50. Le Breton, Y., et al.; *Genome-wide discovery of novel M1T1 group A streptococcal determinants important for fitness and virulence during soft-tissue infection.* PLoS Pathog, 2017. **13**(8): p. e1006584.
51. Stevenson, G., et al.; *Structure of the O antigen of Escherichia coli K-12 and the sequence of its rfb gene cluster.* J Bacteriol, 1994. **176**(13): p. 4144-56.
52. Coligan, J.E., Kindt, T.J., and Krause, R.M.; *Structure of the streptococcal groups A, A-variant and C carbohydrates.* Immunochemistry, 1978. **15**(10-11): p. 755-60.
53. Neiwert, O., Holst, O., and Duda, K.A.; *Structural investigation of rhamnose-rich polysaccharides from Streptococcus dysgalactiae bovine mastitis isolate.* Carbohydr Res, 2014. **389**: p. 192-5.
54. Charbonneau, A.R.L., et al.; *Defining the ABC of gene essentiality in streptococci.* BMC Genomics, 2017. **18**(1): p. 426.
55. Graninger, M., et al.; *Characterization of dTDP-4-dehydrorhamnose 3,5-epimerase and dTDP-4-dehydrorhamnose reductase, required for dTDP-L-rhamnose biosynthesis in Salmonella enterica serovar Typhimurium LT2.* J Biol Chem, 1999. **274**(35): p. 25069-77.
56. Ehmke, N., et al.; *Homozygous and compound-heterozygous mutations in TGDS cause Catel-Manzke syndrome.* Am J Hum Genet, 2014. **95**(6): p. 763-70.
57. Tonetti, M., et al.; *The metabolism of 6-deoxyhexoses in bacterial and animal cells.* Biochimie, 1998. **80**(11): p. 923-31.
58. Pferdehirt, R., et al.; *Catel-Manzke Syndrome: Further Delineation of the Phenotype Associated with Pathogenic Variants in TGDS.* Mol Genet Metab Rep, 2015. **4**: p. 89-91.
59. Schoner, K., et al.; *Mutations in TGDS associated with additional malformations of the middle fingers and halluces: Atypical Catel-Manzke syndrome in a fetus.* Am J Med Genet A, 2017. **173**(6): p. 1694-1697.
60. Koga, T., et al.; *Effect of subculturing on expression of a cell-surface protein antigen by Streptococcus mutans.* J Gen Microbiol, 1989. **135**(12): p. 3199-207.

SUPPLEMENTAL MATERIALS

Method S1. Cloning strategy of knock-out constructs and complementation plasmids for *S. mutans*.

Knock-out constructs for the deletion of *rmlB* and *rmlC* were made in several steps. First, upstream regions of *rmlB* and *rmlC* were amplified from *S. mutans* gDNA using a forward primer ~700 bp upstream of *rmlB* (P1) or *rmlC* (P13) and a reverse primer directly upstream of *rmlB* (P2) or *rmlC* (P14) with an Erm overhang of 30 bp at the 5'-end. Similarly, downstream regions were amplified using reverse primers ~700 bp downstream of *rmlB* (P4) or *rmlC* (P16) and forward primers directly downstream of *rmlB* (P3) or *rmlC* (P15) with an Erm overhang of 30 bp attached to the 5'-end. The Erm resistance cassette was amplified from plasmid pDCerm (P5 + P6). The upstream and downstream regions were subsequently ligated to the ERY resistance cassette using a single pcr step to create the Δ*rmlB* and Δ*rmlC* constructs used for transformation.

Table S1. Protein accession numbers for RmlB and RmlC homologs used for sequence alignments

	Protein accession number	
Species	**RmlB**	**RmlC**
Streptococcus pyogenes	WP_011285494.1	WP_012560640.1
Salmonella enterica	WP_023226833.1	WP_038395428.1
Streptococcus agalactiae	WP_016502954.1	WP_017769218.1
Streptococcus anginosus	PLA73052.1	WP_049516748.1
Streptococcus dysgalactiae	WP_065361279.1	WP_003057688.1
Streptococcus equi	WP_014623057	WP_041790386.1
Streptococcus equinus	ARC34160.1	ARC34159.1
Streptococcus mutans	WP_002273831.1	WP_002283091
Streptococcus pneumoniae	NP_357916.1	WP_000131446
Streptococcus suis	WP_044759933.1	BAM95164.1

Table S2. Binding affinity of compounds to GAS RmlB, RmlC and GacA in mM

	Ri01	Ri02	Ri03	Ri04	Ri06	Ri07	Ri08
RmlB	4.8	0.4	0.2	30	17	1.8	>0.1
RmlC	7.8	n.b.	3	80	n.b.	n.b.	0.1
GacA	2.5	0.2	0.3	n.b.	1.1	n.b.	n.b.

n.b.: no binding

Table S3. Bacterial growth inhibition by Ri03 in different bacterial species

Strain	MIC_{50} (mM)	Rhamnose
GAS 5448	0.12	Essential
S. mutans Xc	0.41	Critical
GCS MGCS10565	0.42	Essential
S. aureus NRS384	2.48	Absent
E. coli BW25113 WT	3.61	Non-essential
E. coli BW25113 Δ*rfbA*	3.74	Non-essential

Table S4. Primer used for cloning of *S. mutans* (SMU) *rmlB* and *rmlC* mutants

Primer	Origin	Sequence	Template
Upstream Δ*rmlB* construct + ERY overhang			
P1	*rmlB* up Fwd	AGCTGATAAAATCAATCGCAAAGAC	SMU Xc
P2	*rmlB* up Rev + **ERY**	**GTTTTGAGAATATTTTATATTTTTGTTCAT**ATATTAAGATACAAAGGGGCGATTCA	SMU Xc
Downstream Δ*rmlB* construct + ERY overhang			
P3	*rmlB* down Fwd + **ERY**	**AGTTATCTATTATTTAACGGGAGGAAATAA**AGATAAATAAGAGGCTGGGACAAAAGT	SMU Xc
P4	*rmlB* down Rev	CAATAAAAGCTCGCACCCGTT	SMU Xc
Erythromycin resistance cassette			
P5	ERY Fwd	ATGAACAAAAATATAAAATATTCTCAAAACTTTTTAACG	pDCerm
P6	ERY Rev	TTATTTCCTCCCGTTAAATAATAGATAACT	pDCerm
pDC123_SMU_*rmlB*			
P7	SMU_*rmlB*.XbaI Fwd	GCTCTAGAATGACAGAATATAAAAACATTATCGTTACCG	SMU Xc
P8	SMU_*rmlB*.BamHI Rev	CGCGGATCCTTAATTAAGTATTTTTTGTGTTTTGGCATAG	SMU Xc
pDC123_GAS_*rmlB*(_Y159F)			
P9	GAS_*rmlB*.XbaI Fwd	GCTCTAGAATGTATAAAAATATTATCGTAACTGGTGGAGC	GAS 5448
P10	GAS_*rmlB*.BamHI Rev	CGCGGATCCTTATTTAATCACTTCTTGAGTTTTAGCATACTTG	GAS 5448
pDC123_GAS_*rmlB*_Y159F			
P11	GAS_*rmlB*. Y159F Fwd	CATCATCACCTTTCTCATCAACTAAGG	GAS 5448
P12	GAS_*rmlB*. Y159F Rev	CCTTAGTTGATGAGAAAGGTGATGATG	GAS 5448
Upstream Δ*rmlC* construct + ERY overhang			
P13	*rmlC* up Fwd	AGCCAGAACACCCTAAATCACACTA	SMU Xc
P14	*rmlC* up Rev + **ERY**	**GTTTTGAGAATATTTTATATTTTTGTTCAT**TTTAAGCTTCTCCAATCAAACGGA	SMU Xc
Downstream Δ*rmlC* construct + ERY overhang			
P15	*rmlC* down Fwd + **ERY**	**AGTTATCTATTATTTAACGGGAGGAAATAA**TGTTTGAAATCTTCTTAGAGATTGCAA	SMU Xc
P16	*rmlC* down Rev	TCGCCTTATTTCTCTAGTACTTTTTTCAG	SMU Xc
pDC123_SMU_*rmlC*			
P17	SMU_*rmlC* .XbaI Fwd	GCTCTAGAATGAACAAAAATATAAAATATTCTCAAAACTTTTTAACG	SMU Xc
P18	SMU_*rrmlC* .BamHI Rev	CGCGGATCCTCACAAATCTTCTGCTTTTAATGGTTT	SMU Xc

Table S4. Continued

pDC123_GAS_rmlC			
P19	GAS_rmlC.XbaI Fwd	GCTCTAGAATGACAGAAACTTTTTTTGACAAACCATT	GAS 5448
P20	GAS_rmlC.BamHI Rev	CGCGGATCCCTATAGGTCTTTTGGTTTCAGTGGTTTG	GAS 5448
pDC123_GAS_rmlC_H76N			
P21	GAS_rmlC. H76N Fwd	CCAAGGTTCAGCGTTAAGTCCACGCAGCAC	GAS 5448
P22	GAS_rmlC. H76N Rev	GTGCTGCGTGGACTTAACGCTGAACCTTGG	GAS 5448
pDC123_GAS_rmlC_K82A			
P23	GAS_rmlC. K82A Fwd	CAACTGAGATGTAAGCATCCCAAGGTTCAG	GAS 5448
P24	GAS_rmlC. K82A Rev	CTGAACCTTGGGATGCTTACATCTCAGTTG	GAS 5448
pDC123_GAS_rmlC_H76N/K82A			
P25	GAS_rmlC.XhoI Fwd	CAAGCTCGAGGAATGTCTTCACAACGC	GAS 5448
P26	GAS_rmlC. XbaI Rev	CGCGGAGCTCTAGAATGACAGAAAC	GAS 5448

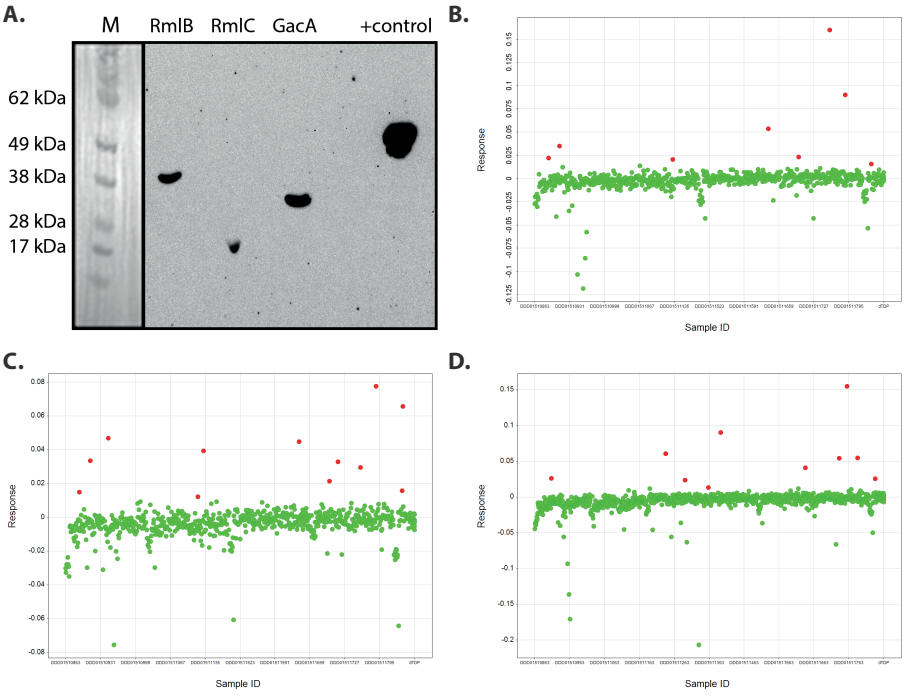


Figure S1. Maybridge library screening data. (A) Anti-biotin Western blot of purified RmlB, RmlC and GacA proteins (RmlB = 39 kDa, RmlC = 23 kDa, GacA = 32 kDa), with positive control (biotinylated Protein X; 55 kDa). **(B)** RmlB, **(C)** RmlC and **(D)** GacA hit cut-off value figures from BLI screen. The binding curve of all compounds were analyzed with robust calculations and plotted as compound (X-axis) vs. response. Potential hits above the hit cut-off are depicted in red. In total 17 compounds were selected and tested in dose-dependent BLI studies against all three target enzymes.

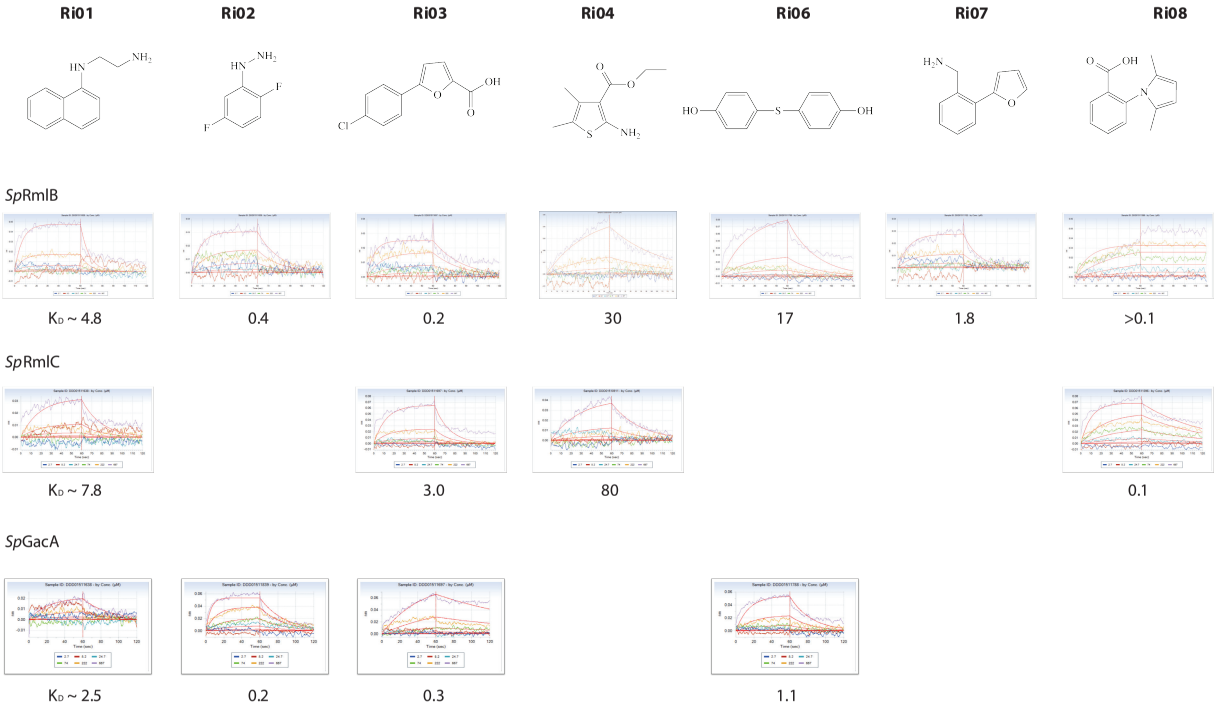


Figure S2. Binding studies of rhamnose inhibitors to RmlB, RmlC and GacA. Chemical structures of identified inhibitors, binding curves vs. purified recombinant proteins and calculated binding affinities (K_D) in mM.

A.

```
SMU_RmlB     -------------MTEYKNIIVTGGAGFIGSNFVHYVYNNHPDVHVTVLDKLTYAGNRAN 47
Human_TGDS   MSAACWEEPWGLPGGFAKRVLVTGGAGFIASHMIVSLVEDYPNYMIINLDKLDYCASLKN 60
                          *.::********.*:::  : :::*:  :  **** *...  *

                                          ▼▼       ▼ ▼
SMU_RmlB     LEEILGD-RVELVVGDIADSELVDKLA--AKADAIVHYAAESHNDNSLKDPSPFIYTNFV 104
Human_TGDS   LETISNKQNYKFIQGDICDSHFVKLLFETEKIDIVLHFAAQTHVDLSFVRAFEFTYVNVY 120
             ** * .. . ::: ***.**.:*. *    * * ::*:**::* * *:     * *.*.

                                  ★ ★                              ★
SMU_RmlB     GTYILLEAARKYD-IRFHHVSTDEVYGDLPLREDLPGHGEGPGEKFTAETKYNPSSPYSS 163
Human_TGDS   GTHVLVSAAHEARVEKFIYVSTDEVYGGSL------------DKEFDESSPKQPTNPYAS 168
             **::*:.**::    :* :********.              .::*  .:  :*:.**:*

             ★
SMU_RmlB     TKAASDLIVKAWVRSFGVKATISNCSNNYGPYQHIEKFIPRQITNILSGIKPKLYGEGKN 223
Human_TGDS   SKAAAECFVQSYWEQYKFPVVITRSSNVYGPHQYPEKVIPKFISLLQHNRKCCIHGSGLQ 228
             :***:: :*::: ..: . ..*:..** ***:*: **.**: *: :  . *  ::*.* :

                ▼
SMU_RmlB     VRDWIHTNDHSTGVWAILTKGRIGETYLIGADGEKNNK----EVLELILEKM--SQPKNA 277
Human_TGDS   TRNFLYATDVVEAFLTVLKKGKPGEIYNIGTNFEMSVVQLAKELIQLIKETNSESEMENW 288
             .*::::::.*   .. ::*.**: ** * **:: * .      *:::** *.   *: :*

                ▼▼
SMU_RmlB     YDHVTDRAGHDLRYAIDSTKLREELGWKPQFTNFEEGLEDTIKWYTEHEDWWKAEKEAVE 337
Human_TGDS   VDYVNDRPTNDMRYPMKSEKI-HGLGWRPKV-PWKEGIKKTIEWYRENFHNWKNVEKALE 346
              *:*.**  :*:** :.* *: . ***:*:.  ::**::.**:** *: . **  ::*:*

SMU_RmlB     ANYAKTQKILN 348
Human_TGDS   PFPV------- 350
                .
```

B.

RmlB homologs	% Identity with TGDS
S. pyogenes	33.75
S. mutans	34.47
S. suis	33.23
S. enterica	35.54
M. tuberculosis	33.44

C.

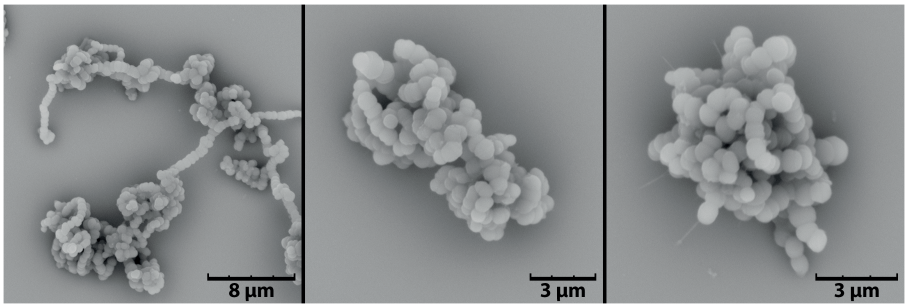


Figure S3. The human protein TGDS cannot replace RmlB in *S. mutans*. (A) Protein sequence alignment of human TGDS with *S. mutans* (SMU) RmlB. Catalytic residues as identified by Allard *et al.* [41] are indicated with a star; mutations in TGDS associated with the Catel-Manzke syndrome as previously identified [56, 58, 59] are indicated with an arrow. **(B)** Percentage protein sequence identity of several bacterial RmlB homologs with human TGDS. **(C)** Scanning electron microscopy images of SMU Δ*rmlB* + pTGDS.


3


4

Role of RgpC, RgpD and RgpE in streptococcal rhamnose polysaccharide transport

4

Samantha L. van der Beek, Hidde Nab, Abdurrahman Çitak, Nina M. van Sorge

University Medical Center Utrecht, Utrecht University, Medical Microbiology, Heidelberglaan 100, 3584 CX, Utrecht, The Netherlands.

ABSTRACT

Many streptococci, such as *Streptococcus mutans* and *Streptococcus pyogenes*, express a cell wall carbohydrate consisting of a polyrhamnose backbone with species- or serotype-specific glycan side chains. These streptococcal rhamnose polysaccharides (SRPs) are critical for viability and virulence, thereby gaining attention as therapeutic targets for new antibiotics and as vaccines antigens. However, knowledge regarding SRP biosynthesis is limited. Transport of these macromolecules across the membrane is a critical step. The possibility to generate viable SRP biosynthesis mutants in *S. mutans*, allowed us to functionally characterize SRP translocation. The *S. mutans* SRP transporter consists of a permease protein RgpC and ATPase protein RgpD. Deletion of RgpCD critically affected bacterial physiology, observed as slow growth and disturbed cell division, which could be restored by reintroduction of RgpCD. Interestingly, we could functionally interchange RgpCD with the *S. pyogenes* SRP transporter proteins GacD and GacE as a whole and as individual proteins in a chimeric complex, suggesting that substrate specificity of SRP ABC-transporters is side chain independent. Bioinformatics analyses indicated considerable sequence diversity in the C-terminal domain of RgpD, which currently has an unknown function. Deletion of the RgpD C-terminal domain did not affect bacterial growth or morphology. In contrast, additional deletion of the downstream gene *rgpE* greatly affected bacterial morphology, suggesting a genetic interaction between the *rgpD* C-terminal domain and *rgpE*. Possibly, RgpD and RgpE homologs present in many streptococcal strains are responsible for the addition of a yet unidentified capping residue to the SRPs.

INTRODUCTION

Streptococcus mutans is a common Gram-positive bacterium in the human oral cavity and considered to be the main etiological agent of dental caries by causing biofilm-mediated dysbiosis [1-5]. *S. mutans* can also cause systemic diseases outside the oral cavity, most prominently endocarditis [6-8]. Furthermore, *S. mutans* has been implicated in inflammatory bowel disease, colorectal cancer, and adverse pregnancy outcomes [9]. Targeted antimicrobial therapy and vaccination strategies against pathogenic bacteria in the oral microbiome, including *S. mutans* and *Streptococcus sobrinus*, could help to restore the microbial balance in the oral cavity, thereby preventing tooth decay and extra-oral diseases such as endocarditis [5, 10, 11].

S. mutans expresses a peptidoglycan-linked surface-exposed serotype-specific carbohydrate, known as the rhamnose-glucose polysaccharide, which is critical for survival and virulence of the bacterium [12-17]. The structural composition of this surface polysaccharide, specifically the identity and the linkage of the side chain to the conserved α-1,2/α-1,3-linked polyrhamnose backbone, determines the *S. mutans* serotype: c, e, f or k [18-22] (Fig 1A). Serotypes c, e and f all express side chains containing glucose, whereas a recent structure redetermination identified a galactose side chain in the serotype k polysaccharide [22] (Fig. 1A). This finding instigated a recent suggestion to refer to this family of surface polysaccharides as Streptococcal Rhamnose Polysaccharides (SRPs) instead of rhamnose-glucose polysaccharides [23]. SRPs do not only include the *S. mutans* serotype polysaccharides, but also the related cell wall polysaccharides from other streptococci containing rhamnose as main constituent such as the group A, C and G carbohydrates (Fig. 1B).

The structural conservation, surface localization, and importance in bacterial physiology and host-pathogen interaction makes the *S. mutans* SRPs attractive therapeutic targets. Indeed, St. Michaels et al. recently demonstrated that sera from mice and rabbits immunized with glycoconjugates of serotype f or k carbohydrates conjugated to bovine serum albumin, promoted antibody-induced killing of both homologous and heterologous *S. mutans* serotypes [22]. In addition, van der Beek et al. recently identified chemical fragments that target the production of dTDP-L-rhamnose, the activated sugar precursor required for biosynthesis of all SRPs (Chapter 3). One of these fragments inhibited the growth of *S. mutans* with an MIC_{50} of 410 μM and could be optimized and developed into a new antimicrobial therapy (Chapter 3).

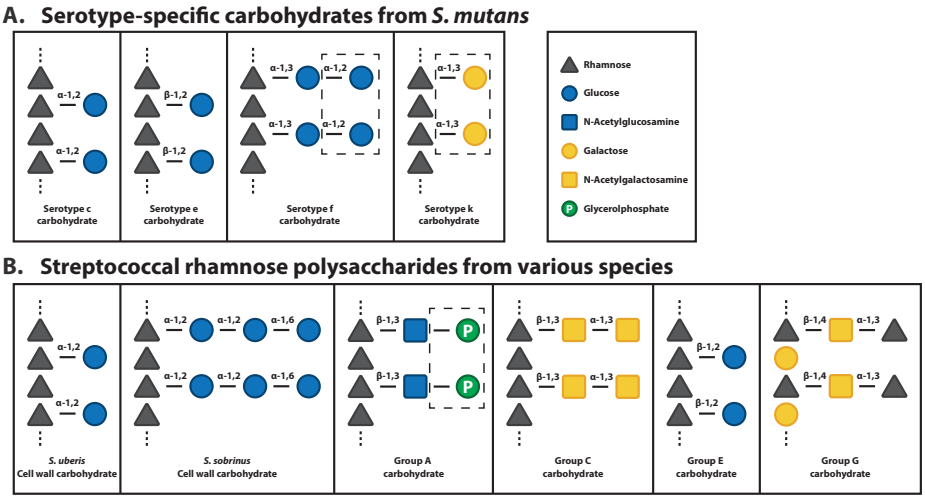


Figure 1. Schematic representation of structural composition of the **(A)** serotype-specific polysaccharides of *S. mutans* serotypes c, e, f and k, and **(B)** streptococcal rhamnose polysaccharides (SRPs) from a broad range of streptococcal species. Except for the group G carbohydrate, these polysaccharides share a common α-1,2/α-1,3 linked polyrhamnose backbone and are uniquely distinguished by the composition and linkage of their side chains. Structures were determined in *S. mutans* strains GS-5 [18] and 10449 (serotype c) [22], LM7 (serotype e) [20], OMZ175 (serotype f) [19, 21, 22], OM42X, OM88X and OM98X (serotype k) [22], *S. uberis* 233/ATCC 13387 [24], *S. sobrinus* B13 and 6715-T_2 [25], *Streptococcus pyogenes* J17A4/ATCC 12385 [26] and 5448 (group A) [23], *Streptococcus dysgalactiae* subsp. *equisimilis* C74/ATCC 12388 and *Streptococcus dysgalactiae* subsp. *dysgalactiae* 2023/ATCC 43078 (group C) [26, 27], *Streptococcus porcinus* K129/ATCC 12390 (group E) [28, 29], *Streptococcus dysgalactiae* subsp. *equisimilis* D166B/ATCC 12394 (Group G) [30]. Sugar residues indicated by a dashed box were recently identified during structure redetermination [22, 23].

In addition to targeting the nucleotide sugar production, other critical steps in the SRP biosynthesis pathways also represent interesting drug targets. Here, we aimed to characterize the transport of *S. mutans* SRPs across the membrane. Previous studies suggest that *rgpC* and *rgpD* encode a permease protein and ATPase protein, respectively, that together form the ABC-transporter RgpCD [31]. Indeed, deletion of these genes in *S. mutans* results in a characteristic rhamnose-deficient phenotype of impaired dividing and growing bacteria. Interestingly, these defects could be restored by reintroduction of the homologous *S. mutans* transporter but also by heterologous expression of the SRP transporter proteins GacD and GacE from *Streptococcus pyogenes*, providing insight into the substrate specificity of SRP ABC-transporters. Next, we focused on the extended C-terminal domain of RgpD, which currently has an unknown function and shows low sequence conservation among *S. mutans* strains. Using bioinformatics analysis and bacterial mutagenesis, we discovered a genetic interaction of the *rgpD* C-terminal domain with *rgpE*, a gene

downstream of *rgpD*, which may help to uncover the function of this extended domain.

MATERIALS AND METHODS

Bacterial strains and growth conditions

Bacterial strains used during this study are listed in Table 1. *Escherichia coli* MC1061 was routinely cultured in Lysogeny broth (Oxoid) at 37 °C with shaking or on Lysogeny broth agar at 37 °C with an addition of 10 µg/ml chloramphenicol (Cm) when harboring a plasmid. *Streptococcus mutans* Xc is a pathogenic serotype c strain [32] and was cultured in Todd-Hewitt broth (Oxoid) or on Todd-Hewitt broth agar at 37 °C with 5% CO_2. Antibiotic selection for *S. mutans* was 10 µg/ml erythromycin (Erm) or 3 µg/ml Cm.

Genetic manipulation of *S. mutans*

Genetic manipulation of *S. mutans* was performed as described previously by van der Beek et al [14].

Table 1. Bacterial strains used during this study.

Bacterial strains	Resistance
Escherichia coli	
MC1061	-
Streptococcus mutans Xc	
Wild-type (WT)	-
Δ*rgpCD*	Erm
Δ*rgpCD* + p*RgpCD*	Erm + Cm
Δ*rgpCD*+ p*GacDE*	Erm + Cm
Δ*rgpCD* + p*RgpC*/G*acE*	Erm + Cm
Δ*rgpCD* + p*GacD/RgpD*	Erm + Cm
$\Delta rgpCD_{775\text{-}1218}$	Erm
$\Delta rgpCD_{775\text{-}1218}E$	Erm
Δ*rgpE*	Erm
Δ*rgpE* + *pRgpE*	Erm + Cm
Δ*rgpCDE*	Erm
Δ*rgpCDE* + p*RgpCDE*	Erm + Cm

Erm: erythromycin; Cm: chloramphenicol

Growth curves of *S. mutans*

Growth curves of *S. mutans* strains were performed as described previously by van der Beek et al (Chapter 3). In short, overnight cultures of strains displaying normal growth were diluted and grown for 1.5 h to early-logarithmic phase (optical density at 600 nm (OD_{600}) of ~0.2-0.3). Cultures were subsequently diluted to an OD_{600} of 0.03-0.04. Bacterial mutants with an overnight OD_{600} below 0.35 were directly diluted to OD_{600} 0.03-0.04. Measurements were performed manually using a Thermo Spectronic Genesys 20 spectrometer at regular intervals. Samples were incubated at 37 °C with 5% CO_2 in between measurements. Growth curves represent mean ± standard error of mean (SEM) of at least three biological experiments.

Scanning Electron Microscopy

Sample preparation for scanning electron microscopy was performed as described previously by van der Beek et al, except samples were coated with 6-7 nm gold [14]. Images were made using a For scanning electron microscopy imaging, a Scios dual beam electron microscope (Thermo Scientific Inc., The Netherlands) was used, operated in Optiplan mode at high vacuum. Bacteria were imaged at an acceleration voltage of 2.0 kV and a current of 0.1 nA, at 7.0 mm working distance. High magnification images were collected at a 4.5 nm pixel size with a dwell time of 2 µs (scan area 3072×2048), using an Everhart-Thornley detector set to collect secondary electrons.

Bioinformatics analysis

Homologous proteins and protein sequence identities were identified by blast searches of the amino acid sequences of RgpC, RgpD and RgpE in the National Center for Biotechnology Information (NCBI; https://blast.ncbi.nlm.nih.gov/Blast.cgi) database (performed on 17-04-2018). Multiple protein sequence alignments were made using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).

RESULTS

Deletion of *rgpC* and *rgpD* results in a rhamnose-deficient phenotype

S. mutans rgpC and *rgpD* are located in the *S. mutans* SRP biosynthesis gene cluster and are assumed to form the ABC-transporter that transports the serotype-defining SRPs across the membrane [31]. Currently however, experimental evidence for this role is limited [33]. To gain more insight into the function of ABC-transporter RgpCD, we constructed an *rgpCD* deletion strain in *S. mutans* Xc (Δ*rgpCD*; Fig 2A) by genetic replacement of *rgpCD* with an Erm resistance cassette (*erm*). Compared to wild-type

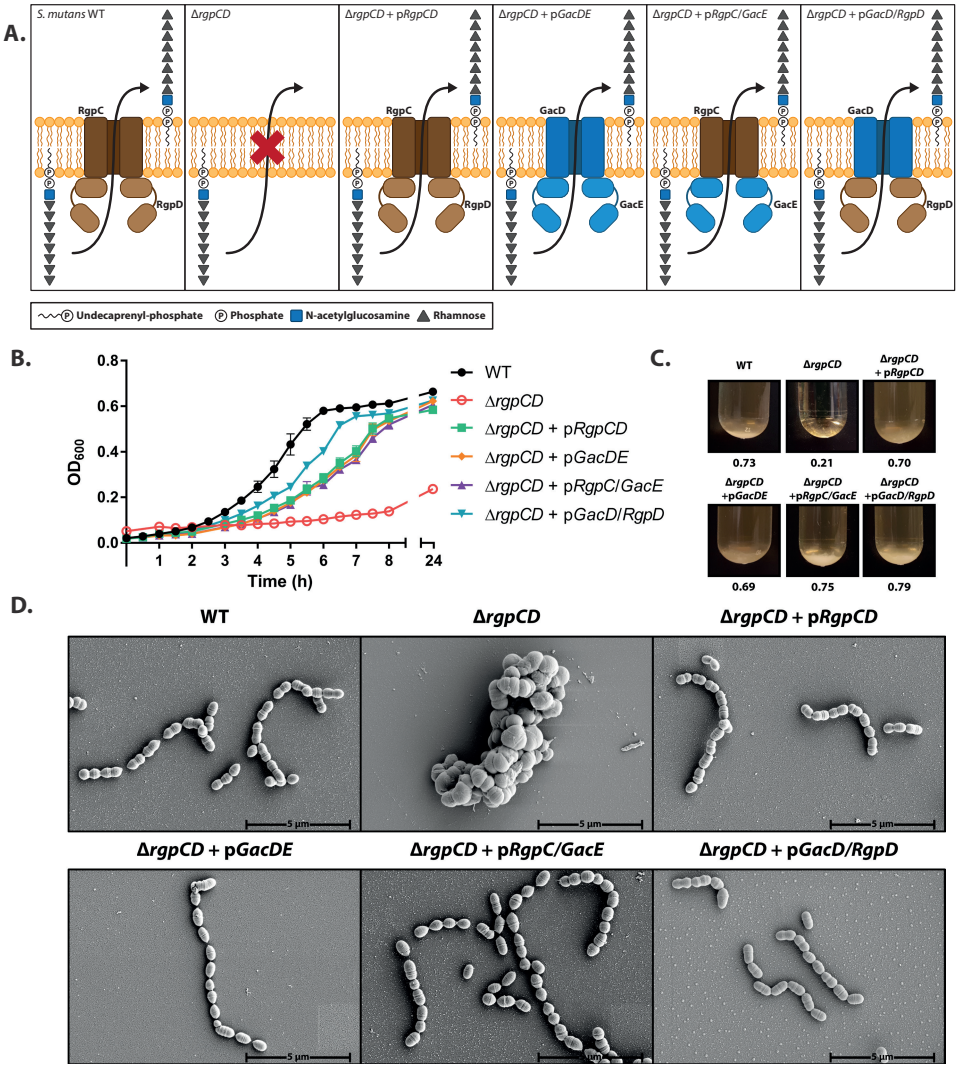


Figure 2. Deletion of *rgpCD* in *S. mutans*, homologous complementation with *S. mutans rgpCD* (brown) and heterologous complementation with *S. pyogenes gacDE* (blue). A) Schematic overview of SRP transporters in *S. mutans* WT and constructed mutants. Note: It is currently unknown, whether glucose side chain decoration occurs before or after translocation. B) Growth curves, C) buoyancy and overnight OD_{600} values, and D) representative scanning electron microscopy images of *rgpCD* deletion strains (Δ*rgpCD*) complemented with either p*RgpCD*, p*GacDE*, p*RgpC/GacE* or p*GacD/RgpD* overexpressed from plasmid pDC123. Growth curves show mean ± SEM from three biological replicates.

(WT), Δ*rgpCD* displayed severe growth defects and did not reach an OD_{600} above 0.35 (Fig 2B). Δ*rgpCD* also displayed altered buoyancy in overnight cultures, sinking to the bottom of culture tubes, while WT bacteria remained partially in suspension (Fig 2C). In addition, Δ*rgpCD* showed both cell division and cell separation defects visible as big clumps consisting of swollen cocci, which divided in multiple directions (Fig 2D). These results are in agreement with a rhamnose-deficient *rmlD* mutant, suggesting the absence of SRPs in the cell wall of Δ*rgpCD* [14]. These aberrant phenotypes could be restored by reintroduction of *rgpCD* on an expression plasmid (Δ*rgpCD* + p*RgpCD*; Fig 2A), excluding the deletion caused polar effects in the operon (Fig 2). The slightly suppressed growth rate compared to WT (Fig 2B) is thought to result from overexpression of the transporter.

RgpC and RgpD are highly conserved among streptococcal species

It is currently unknown how SRP transporters recognize their substrate. Sequence diversity in *rgpC* and *rgpD* between different *S. mutans* serotypes (c, e, f and k; Fig 1A) could provide an indication whether SRP transporters interact with structurally conserved or rather with unique features of SRPs. Protein blast search of *S. mutans* Xc RgpC against 70 different genome sequenced *S. mutans* strains representing all four serotypes showed that RgpC is highly conserved (99-100% protein sequence identity) (Table 2). This suggests that the substrate specificity of RgpC is likely independent of the serotype-specific SRP side chains. Remarkably, two RgpC homologs from serotype c strains SF1 and SM4 shared only 60% protein sequence identity (Table 2).

Several streptococcal species express SRPs that share structural features with the SRPs from *S. mutans*. Specifically, many of these species, including *S. pyogenes* [26], *Streptococcus uberis* [24], *Streptococcus sobrinus* (formerly known as *S. mutans* serotypes d and g) [25, 34, 35], *Streptococcus equi* subsp. *zooepidemicus* [26, 36, 37] and *Streptococcus dysgalactiae* subsp. *dysgalactiae* [27], express a cell wall polysaccharide containing an α-1,2/α-1,3-linked polyrhamnose backbone, identical to *S. mutans* (Fig. 1B). For *S. pyogenes*, it was recently demonstrated that only the rhamnan backbone is transported across the membrane, whereas the GlcNAc side chains of the group A carbohydrate are attached extracellularly by GacL [38]. A similar two-step biosynthesis process may occur for SRP transport in all these species, possibly putting fewer constraints on the specificity of the SRP transporter. Surprisingly however, RgpC shares only around 61-67% protein sequence identity to permease proteins from these other streptococcal species (Table 3). This roughly corresponds to the protein sequence identity (58%) of an RgpC homolog in a *S. dysgalactiae* subsp. *equisimilis* strain, which expresses the Lancefield group G carbohydrate that contains a $[\alpha\text{-}1,2\ \text{galactose}\ \alpha\text{-}1,3\ \text{rhamnose}]_n$ backbone [30]

(Fig. 1B). Protein sequence identity between RgpC homologs among species expressing the same SRP is much higher; 99% for species expressing the Lancefield group A carbohydrate, and 83-99% for species expressing the Lancefield group C carbohydrate. These findings suggest that substrate specificity of RgpC is not solely determined by the structure of the glycopolymer main chain or side chain modifications, but suggest RgpC recognizes additional yet unidentified unique features on these SRPs, such as capping or anchor residues.

Table 2. Protein sequence identity of RgpC, RgpD and RgpE among different *S. mutans* strains.

		% Identity compared to *S. mutans* Xc		
Strain	**Serotype**	**RgpC**	**RgpD**	**RgpE**
Representatives from group 1 (out of 61)				
UA159	c	100	100	99
1SM1	e	100	100	99
OMZ175	f	100	100	99
NLML4	e	99	99	100
Representatives from group 2 (out of 7)				
KK23	c	100	83	50
N29	c	100	82	50
LJ23	k	100	82	50
Group 3				
SF1	c	60	67	23
SM4	c	60	67	23

Table 3. Protein sequence identity of RgpC, RgpD and RgpE among different streptococci.

			% Identity compared to *S. mutans* Xc		
Species	**Strain**	**SRP**	**RgpC**	**RgpD**	**RgpE**
S. uberis	0140J/ATCC BAA-854	Serotype c[#]	67	66	29
S. uberis	C9359	Serotype c[#]	61	73	53
S. sobrinus	W1703	Serotype unknown	65	76	52
S. pyogenes	MGAS5005	Lancefield group A	64	74	51
S. equi subsp. *zooepidemicus*	MGCS10565	Lancefield group C	65	75	53
S. dysgalactiae subsp. *dysgalactiae*	NADC Z-8/ATCC 27957	Lancefield group C	65	75	50
S. dysgalactiae subsp. *equisimilis*	AC-2713/ATCC BAA-338	Lancefield group A	64	75	51
S. dysgalactiae subsp. *equisimilis*	167	Lancefield group C	64	76	50
S. dysgalactiae subsp. *equisimilis*	D166B/ATCC 12394*	Lancefield group G	58	54	NI

* Structure of rhamnose polysaccharide characterized in this strain

[#] Based on structural data from *S. uberis* 233/ATCC 13387

SRP; Streptococcal rhamnose polysaccharide, NI; Not identified

In contrast to RgpC, RgpD is slightly less conserved among the 70 *S. mutans* strains examined. Sixty-one strains contained a near identical (99-100%) homolog. In contrast, in strains SF1 and SM4 RgpD only shares 67% identity, which corresponds to the lower homology of their corresponding RgpC homolog. Interestingly, we also identified a third group consisting of seven strains classified as serotype c or k (24, KK23 - LJ23, 5SM3, N29, NMT4863, SA38) that displayed an intermediate level of RgpD protein sequence identity (82-83%; Table 2). Alignment of representative sequences from these three different groups shows that this sequence diversity localizes to the C-terminal region of RgpD (Fig. S1). RgpD homologs in aforementioned streptococcal species share a higher protein sequence identity (65-76%) as compared to RgpC (Table 3). Similar to *S. mutans*, sequence diversity is mainly confined to the C-terminal region of these RgpD homologs for these streptococcal species.

GacD and GacE can functionally replace RgpC and RgpD

To experimentally address potential differences in substrate specificity of SRP transporters, we heterologously expressed the GacDE transporter from *S. pyogenes* in the Δ*rgpCD S. mutans* background (Δ*rgpCD* + p*GacDE*; Fig 2A). To create this strain, we first introduced GacDE in the *S. mutans* WT strain and subsequently deleted *rgpCD*. Indeed, GacDE from *S. pyogenes* 5448 was able to restore the observed Δ*rgpCD* mutant phenotypes comparable to complementation with the homologous RgpCD transporter (Fig. 2B-D). Next, we investigated whether RgpC/GacE or RgpD/GacD could form functional chimeric complexes together (Δ*rgpCD* + p*RgpC*/*GacE* and Δ*rgpCD* + p*GacD*/*RgpD*; Fig 2A) when expressed in *S. mutans* Δ*rgpCD*. Again, both RgpC/GacE and GacD/RgpD complemented the Δ*rgpCD* phenotypes, with growth rates comparable to RgpCD complementation (Fig 2B-D).

The C-terminal domain of RgpD is not essential for *S. mutans* physiology

As indicated above, the RgpD N-terminal domain confers the ATPase activity and is highly conserved among the studied homologs of streptococcal strains. Indeed, most of the sequence heterogeneity is located in the C-terminal domain (~150 amino acids) of unknown function (Table 2, 3, Fig S1). RgpD and its homologs cluster phylogenetically with other ATPase proteins that that contain an extended C-terminal domain as previously described [39]. However, it is not possible to identify a mutual functional domain suggesting that they play diverse roles in glycopolymer biosynthesis. Indeed, blasting of this C-terminal protein domain did not provide any clues for a possible function.

To gain insight into the function of the RgpD C-terminal domain, we created a deletion mutant lacking the terminal 147 amino acids of RgpD (Δ*rgpD*$_{775\text{-}1218}$). Absence of this domain did not alter the growth rate, buoyancy or morphology of this mutant compared to WT (Fig 3).

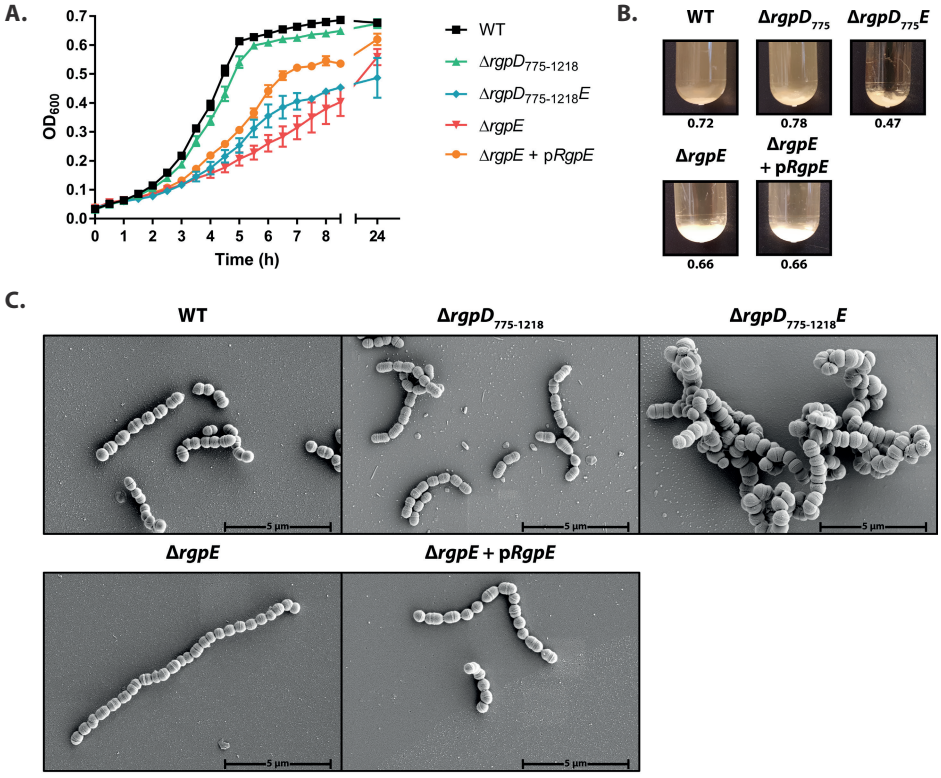


Figure 3. Phenotypic characterization of *S. mutans* strains carrying a deletion in either the C-terminus of *rgpD* (Δ*rgpD*$_{775\text{-}1218}$), *rgpE* (Δ*rgpE*) or both (Δ*rgpD*$_{775\text{-}1218}$*E*). **(A)** Growth curves **(B)** buoyancy and the OD_{600} values of overnight cultures, and **(C)** representative scanning electron microscopy images of *S. mutans* Xc WT, Δ*rgpD*$_{775\text{-}1218}$, Δ*rgpD*$_{775\text{-}1218}$E, Δ*rgpE* and Δ*rgpE* + p*RgpE*. Growth curves show mean ± SEM from four biological replicates.

Genetic interaction between RgpE and C-terminal domain of RgpD

The described functions of characterized C-terminal ATPase domains can be categorized in two groups: ATPase C-terminal domains that modify glycopolymers, i.e. through methylation, sulfation, or acetylation, and ATPase C-terminal domains that recognize a terminal modification on glycopolymers, which is added by the downstream gene product [39]. Experimental support for the latter function can be obtained by genetic studies targeting *rgpE*, which is downstream of *rgpD*,

and is annotated as a glycosyltransferase. It was previously suggested that RgpE participates in SRP glucose side chain formation, since insertional inactivation of *rgpE* causes a slight reduction in cell wall glucose levels of *S. mutans* Xc [33]. A blast search of RgpE from strain Xc in the NCBI database revealed that the majority of identified RgpE homologs in *S. mutans* (61 out of 70) is highly conserved (99-100%; Table 2). However, in seven *S. mutans* strains, RgpE shared only 50% protein sequence identity (Table 2). Interestingly, we observe a correlation between the sequence (dis) similarity of RgpE and the extended C-terminus of RgpD from these same strains (Table 2). For example, RgpE from strains SF1 and SM4 shared just 23% homology with Xc RgpE and also show the lowest homology for the C-terminal domain of RgpD (Table 2). Bioinformatics analysis of RgpE also resulted in the identification of possible RgpE homologs in other streptococci with 40-58% protein sequence identity (Table 3). Similar to *S. mutans*, low identity of RgpE homologs corresponded to low sequence identity of RgpD homologs, even among different *S. uberis* strains or species expressing the same Lancefield group C carbohydrate.

To investigate a possible interaction between the C-terminal domain of RgpD and RgpE, a panel of deletion strains was generated in *S. mutans* Xc. Deletion of RgpE alone (Δ*rgpE*) impaired growth (Fig 3A), increased streptococcal chain length (Fig 3C) and increased sedimentation of overnight cultures (Fig 3B) compared to WT and Δ$rgpD_{775\text{-}1218}$. This phenotype could only be partially restored by the introduction of RgpE on expression plasmid pDC123 (Δ*rgpE* + p*RgpE*) (Fig 3). Interestingly, we observed a genetic interaction between the C-terminal domain of RgpD and RgpE (Δ$rgpD_{775\text{-}1218}E$) as indicated by defects in cell division resulting in the formation of clumps of swollen cocci almost similar to Δ*rgpCD**slow (Fig. 3C). In addition, we observed a sinking phenotype similar to the Δ*rgpCD* strains. Surprisingly, Δ$rgpD_{775\text{-}1218}E$ appeared to grow faster than Δ*rgpE*, but displayed a lower overnight OD_{600} (Fig 3A, B).

Stoichiometry

A possible explanation for the incomplete complementation of Δ*rgpE* by *rgpE* is that the enzymes were overexpressed compared to WT bacteria, resulting in incorrect stoichiometry within the glycopolymer biosynthesis and transport machinery. Given the observed genetic interaction between RgpE with the RgpCD transporter, we attempted whether overexpression of all three genes (*rgpCDE*) in an Δ*rgpCDE* background (Δ*rgpCDE* + p*RgpCDE*) could overcome such potential imbalance. Deletion of *rgpCDE* (Δ*rgpCDE*) resulted in a phenotype of attenuated growth, low overnight OD_{600} and abnormal morphology similar to our observations with Δ*rgpCD* (Fig. 4). However, complementation with p*RgpCDE* only partially restored the observed phenotypes of the Δ*rgpCDE* strain (Fig. 4).

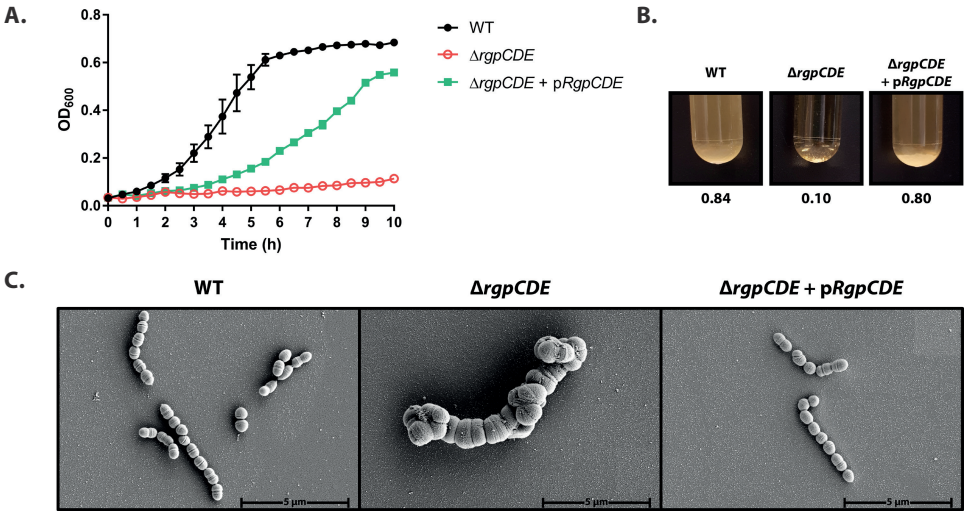


Figure 4. (A) Growth curves, **(B)** buoyancy and the OD_{600} values of overnight cultures, and **(C)** representative scanning electron microscopy images of *rgpCDE* deletion strain (Δ*rgpCDE*) and complemented with p*RgpCDE*. Growth curves show mean ± SEM from three biological replicates.

DISCUSSION

In this study, we investigated RgpCD as the ABC-transporter of the serotype-specific SRP of *S. mutans* using bioinformatics analysis, molecular genetics and phenotypic evaluation. Deletion of *rgpCD* resulted in a rhamnose-deficient phenotype corresponding to lack of the SRP in the cell wall [14]. The fact that we could generate this deletion mutant is in line with a previous report [33], but contradicts the observed essentiality reported from transposon mutant library screening of *S. mutans* strain UA159 [13]. The competitive environment of the UA159 mutant transposon library likely explains this discrepancy since mutants with a WT growth phenotype easily outgrow poor growing mutants such as Δ*rgpC* or Δ*rgpD*.

In an effort to gain insight into the substrate specificity of SRP ABC-transporters, we developed a heterologous expression model. The functional exchange of *S. mutans* RgpCD with *S. pyogenes* GacDE suggests that the transporter interacts with a conserved part, most likely the rhamnan backbone of the SRP (Fig. 1), or that the transporters are promiscuous regarding substrate interaction. In addition, RgpC and RgpD are individually interchangeable with their respective homologs from *S. pyogenes*, GacD and GacE, indicating that these chimeric transporters are also functional. A similar observation regarding transporter promiscuity was made by

Schirner et al. [40], who functionally replaced the wall teichoic acid ABC-transporter TagGH from *Bacillus subtilis* with TarGH from *S. aureus*. Apparently, these wall teichoic acid transporters do not distinguish between the main chain polymer structure, glycerolphosphate or ribitolphosphate, and substrate specificity may depend on recognition of the conserved diphospholipid-linked disaccharide portion of the teichoic acid precursor [40]. In this regard, our results suggest that substrate specificity of RgpCD and GacDE may be determined by the conserved α-1,2/α-1,3 linked polyrhamnose backbone. Interestingly, this idea is in line with recent data in *S. pyogenes* that demonstrated that the GlcNAc side-chain is linked to the rhamnan backbone after transport of the rhamnan backbone [38]. A similar two-step system where side chains are attached extracellularly likely also exists in *S. mutans*. This hypothesis is supported by the presence of unsubstituted polyrhamnose SRP in the cell walls of several *S. mutans* strains alongside the complete serotype-specific rhamnan polysaccharides [22]. Unmodified rhamnan backbones were also identified in cell walls of *S. pyogenes*, *S. dysgalactiae* subsp. *dysgalactiae* and *S. uberis*, suggesting that the two-step biosynthesis system is likely widespread among streptococcal species [23, 24, 27]. Alternatively, the SRP transporters recognize an as yet undetermined common lipid linked part of the SRP. Additional functional exchange experiments with homologs that transport SRPs with a different backbone, such as the group G carbohydrate (Fig. 1B), in addition to more in-depth structural analysis of these SRPs could shed some light on this.

Although RgpC and RgpD are highly conserved among *S. mutans* strains, there is striking sequence variation. Three distinct groups could be identified based on protein sequence identity: 1) strains that contain almost identical RgpC and RgpD homologs (99-100%), 2) strains expressing almost identical RgpC, but more distinct RgpD (82-83%) compared to group 1, and 3) two strains with quite dissimilar RgpC (60%) and RgpD (67%) compared to groups 1 and 2. Unexpectedly, this division is serotype independent as serotype c strains are represented in all three groups. The sequence diversity in RgpC and RgpD among these serotype c strains, may suggest there is an additional, yet unidentified, modification on the serotype-specific carbohydrates that results in microheterogeneity among SRPs. In line with these findings, *S. uberis* homologs of RgpCD share only 65-67% protein sequence identity, while *S. uberis* expresses a structurally identical polysaccharide as *S. mutans* serotype c enforcing our hypothesis of an additional modification [18, 24] (Fig. 1). Unfortunately, there is no genome sequence available of *S. uberis* strain 233, the strain used to determine the structure of its cell wall polysaccharide. Furthermore, it is currently unknown whether every *S. uberis* strain expresses an identical cell wall polysaccharide. A possible explanation for these suggested microheterogeneities may be found in

rgpE, the gene adjacent to *rgpD*. We identified a genetic interaction between the C-terminal domain of RgpD and RgpE. Low protein sequence identity between RgpD homologs was confined to the extended C-terminal domain, even among different streptococcal species, and correlates to sequential diversity in RgpE. Deletion of the C-terminus of RgpD did not alter bacterial physiology, while deletion of RgpE resulted in minor morphological changes and affected growth. However, deletion of both the C-terminus of RgpD and RgpE resulted in an enhanced aberrant phenotype. Interaction between the C-terminal domain of ABC-transporter ATPase proteins and a protein encoded by a downstream gene has been observed previously in *E. coli* as well as in other bacteria [39, 41]. According to these studies, ATPase proteins (Wzt), involved in the transport of lipopolysaccharide linked O-antigen polysaccharides, could only be exchanged between serotypes with co-exchange of the neighboring gene (*wbdD*) [41]. In this particular case, WbdD attaches a serotype-specific terminal cap to the O-antigen polymer, which is subsequently recognized by the C-terminus of Wzt to initiate transport. ATPase proteins that are phylogenetically related to Wzt by the presence of an extended C-terminal domain recognize a wide range of terminal modifications such as 2-*O*-methyl, 3-*O*-methyl, phosphorus, 3-deoxy-D-manno-octulosonic acid (Kdo), GlcNAc or MurNAc, which could explain the lack of a shared mutual domain in the C-terminal part of these ATPase proteins [39]. In a second phylogenetic group of ATPases with extended C-terminal domains, which includes RgpD, putative catalytic domains were identified in the extended domain conferring sulfo-, methyl- or acetyltransferase activity, suggesting a different role in polysaccharide transport. So far, no catalytic domain with known function was identified in either the C-terminus of RgpD or in RgpE, but insertional inactivation of *rgpE* did result in decreased glucose levels in the cell wall [33]. Compared to other glycopolymer components, terminal modifications are relatively less abundant in the cell wall and sometimes labile resulting in their removal by harsh chemical treatment during glycopolymer isolation. As a result, these modifications are frequently missed in structural analyses, as recently illustrated for the group A carbohydrate in *S. pyogenes* [23] or for the capsular polysaccharides of *Streptococcus agalactiae* [42]. Alternative purification methods could reveal structural insights into the composition of the serotype-specific SRP from *S. mutans*. Combined with targeted mutagenesis, we might be able to uncover the functions of RgpD, RgpE and their homologs.

Overall, the results of this study suggest that substrate specificity of SRP transporters is side chain independent. In addition, we conclude based on protein sequence conservation of predicted SRP transporter proteins among streptococcal strains and species, available SRP structural data and targeted mutagenesis, that the extended

C-terminal domain of SRP ABC-transporter ATPases and a gene directly downstream may be responsible for additional, yet unidentified, modifications on these SRPs, resulting in SRP microheterogeneity (Fig. 5). However, it is yet unclear how these modifications influence SRP translocation and how they are involved in transporter substrate recognition.

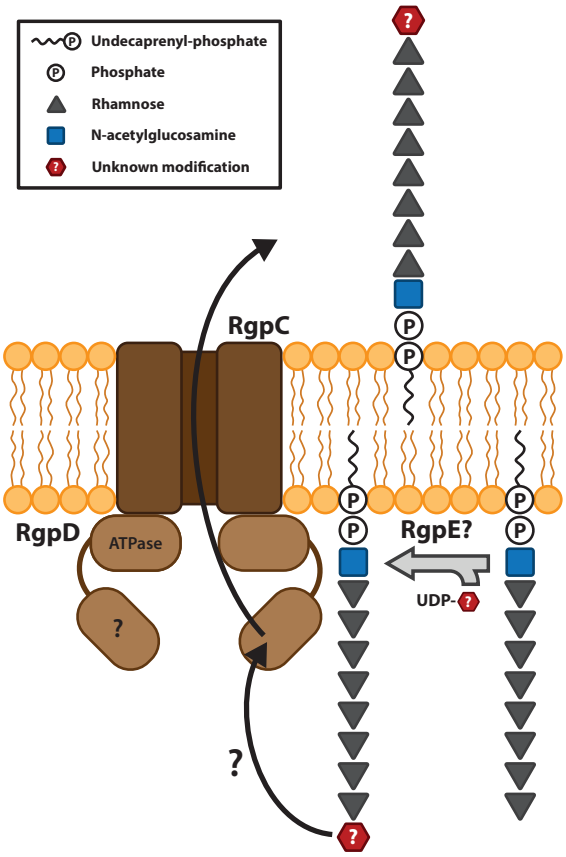


Figure 5. Proposed model for translocation of the SRP rhamnan backbone in *S. mutans*. RgpE attaches an unknown residue to the rhamnan backbone, possibly to terminate elongation. This residue is putatively recognized by the C-terminal domain of RgpD, which initiates transport across the membrane.

ACKNOWLEDGEMENT

The authors would like to thank J. Fermie from the Cell Microscopy Core, Section Cell Biology, Center for Molecular Medicine, University Medical Center Utrecht, The Netherlands for his assistance with the scanning electron microscope.

REFERENCES

1. Dewhirst, F.E., et al.; *The human oral microbiome.* J Bacteriol, 2010. **192**(19): p. 5002-17.
2. Koo, H., Falsetta, M.L., and Klein, M.I.; *The exopolysaccharide matrix: a virulence determinant of cariogenic biofilm.* J Dent Res, 2013. **92**(12): p. 1065-73.
3. Loesche, W.J.; *Role of Streptococcus mutans in human dental decay.* Microbiol Rev, 1986. **50**(4): p. 353-80.
4. Tanner, A.C.R., et al.; *The Caries Microbiome: Implications for Reversing Dysbiosis.* Adv Dent Res, 2018. **29**(1): p. 78-85.
5. Zhan, L.; *Rebalancing the Caries Microbiome Dysbiosis: Targeted Treatment and Sugar Alcohols.* Adv Dent Res, 2018. **29**(1): p. 110-116.
6. Han, Y.W. and Wang, X.; *Mobile microbiome: oral bacteria in extra-oral infections and inflammation.* J Dent Res, 2013. **92**(6): p. 485-91.
7. Nakano, K., et al.; *Detection of novel serotype k Streptococcus mutans in infective endocarditis patients.* J Med Microbiol, 2007. **56**(Pt 10): p. 1413-5.
8. Nakano, K. and Ooshima, T.; *Serotype classification of Streptococcus mutans and its detection outside the oral cavity.* Future Microbiol, 2009. **4**(7): p. 891-902.
9. Lucchese, A.; *Streptococcus mutans antigen I/II and autoimmunity in cardiovascular diseases.* Autoimmun Rev, 2017. **16**(5): p. 456-460.
10. Bowen, W.H.; *Dental caries - not just holes in teeth! A perspective.* Mol Oral Microbiol, 2016. **31**(3): p. 228-33.
11. Stone, V.N. and Xu, P.; *Targeted antimicrobial therapy in the microbiome era.* Mol Oral Microbiol, 2017.
12. Nagata, E., et al.; *Serotype-specific polysaccharide of Streptococcus mutans contributes to infectivity in endocarditis.* Oral Microbiol Immunol, 2006. **21**(6): p. 420-3.
13. Shields, R.C., et al.; *Genomewide Identification of Essential Genes and Fitness Determinants of Streptococcus mutans UA159.* mSphere, 2018. **3**(1).
14. van der Beek, S.L., et al.; *GacA is Essential for Group A Streptococcus and Defines a New Class of Monomeric dTDP-4-dehydrorhamnose Reductases (RmlD).* Mol Microbiol, 2015.
15. Engels-Deutsch, M., et al.; *Insertional inactivation of pac and rmlB genes reduces the release of tumor necrosis factor alpha, interleukin-6, and interleukin-8 induced by Streptococcus mutans in monocytic, dental pulp, and periodontal ligament cells.* Infect Immun, 2003. **71**(9): p. 5169-77.
16. Kovacs, C.J., Faustoferri, R.C., and Quivey, R.G., Jr.; *RgpF Is Required for Maintenance of Stress Tolerance and Virulence in Streptococcus mutans.* J Bacteriol, 2017. **199**(24).
17. Tsuda, H., et al.; *Role of serotype-specific polysaccharide in the resistance of Streptococcus mutans to phagocytosis by human polymorphonuclear leukocytes.* Infect Immun, 2000. **68**(2): p. 644-50.
18. Wetherell, J.R., Jr. and Bleiweis, A.S.; *Antigens of Streptococcus mutans: characterization of a polysaccharide antigen from walls of strain GS-5.* Infect Immun, 1975. **12**(6): p. 1341-8.
19. Linzer, R., Reddy, M.S., and Levine, M.J.; *Structural studies of the serotype-f polysaccharide antigen from Streptococcus mutans OMZ175.* Infect Immun, 1987. **55**(12): p. 3006-10.
20. Pritchard, D.G., et al.; *Characterization of the serotype e polysaccharide antigen of Streptococcus mutans.* Mol Immunol, 1986. **23**(2): p. 141-5.
21. Pritchard, D.G., et al.; *Structure of the serotype f polysaccharide antigen of Streptococcus mutans.* Carbohydr Res, 1987. **166**(1): p. 123-31.
22. St Michael, F., et al.; *Investigating the candidacy of the serotype specific rhamnan polysaccharide based glycoconjugates to prevent disease caused by the dental pathogen Streptococcus mutans.* Glycoconj J, 2017.
23. Edgar, R.J., et al., *Genetic insight into zinc and antimicrobial toxicity uncovers a glycerol phosphate modification on streptococcal rhamnose polysaccharides*. 2018: BioRxiv.

24. Czabanska, A., Holst, O., and Duda, K.A.; *Chemical structures of the secondary cell wall polymers (SCWPs) isolated from bovine mastitis Streptococcus uberis.* Carbohydr Res, 2013. **377**: p. 58-62.
25. Linzer, R., Reddy, M.S., and Levine, M.J.; *Structural studies of the rhamnose-glucose polysaccharide antigen from Streptococcus sobrinus B13 and 6715-T2.* Infect Immun, 1985. **50**(2): p. 583-5.
26. Coligan, J.E., Kindt, T.J., and Krause, R.M.; *Structure of the streptococcal groups A, A-variant and C carbohydrates.* Immunochemistry, 1978. **15**(10-11): p. 755-60.
27. Neiwert, O., Holst, O., and Duda, K.A.; *Structural investigation of rhamnose-rich polysaccharides from Streptococcus dysgalactiae bovine mastitis isolate.* Carbohydr Res, 2014. **389**: p. 192-5.
28. Soprey, P. and Slade, H.D.; *Chemical structure and immunological specificity of the streptococcal group e cell wall polysaccharide antigen.* Infect Immun, 1971. **3**(5): p. 653-8.
29. Pritchard, D.G. and Furner, R.L.; *Structure of the group-specific polysaccharide of group E Streptococcus.* Carbohydr Res, 1985. **144**(2): p. 289-96.
30. Pritchard, D.G., et al.; *Structure of the group G streptococcal polysaccharide.* Carbohydr Res, 1988. **173**(2): p. 255-62.
31. Shibata, Y., et al.; *Expression and characterization of streptococcal rgp genes required for rhamnan synthesis in Escherichia coli.* Infect Immun, 2002. **70**(6): p. 2891-8.
32. Koga, T., et al.; *Effect of subculturing on expression of a cell-surface protein antigen by Streptococcus mutans.* J Gen Microbiol, 1989. **135**(12): p. 3199-207.
33. Yamashita, Y., et al.; *Genes involved in cell wall localization and side chain formation of rhamnose-glucose polysaccharide in Streptococcus mutans.* J Bacteriol, 1998. **180**(21): p. 5803-7.
34. Prakobphol, A. and Linzer, R.; *Purification and immunological characterization of a rhamnose-glucose antigen from Streptococcus mutans 6517-T2 (serotype g).* Infect Immun, 1980. **30**(1): p. 140-6.
35. Prakobphol, A., Linzer, R., and Genco, R.J.; *Purification and characterization of a rhamnose-containing cell wall antigen of Streptococcus mutans B13 (serotype d).* Infect Immun, 1980. **27**(1): p. 150-7.
36. Krause, R.M. and McCarty, M.; *Variation in the group-specific carbohydrate of group C hemolytic Streptococci.* J Exp Med, 1962. **116**: p. 131-40.
37. Ma, Z., et al.; *Complete genome sequence of Streptococcus equi subsp. zooepidemicus strain ATCC 35246.* J Bacteriol, 2011. **193**(19): p. 5583-4.
38. Rush, J.S., et al.; *The molecular mechanism of N-acetylglucosamine side-chain attachment to the Lancefield group A carbohydrate in Streptococcus pyogenes.* J Biol Chem, 2017. **292**(47): p. 19441-19457.
39. Cuthbertson, L., Kos, V., and Whitfield, C.; *ABC transporters involved in export of cell surface glycoconjugates.* Microbiol Mol Biol Rev, 2010. **74**(3): p. 341-62.
40. Schirner, K., Stone, L.K., and Walker, S.; *ABC transporters required for export of wall teichoic acids do not discriminate between different main chain polymers.* ACS Chem Biol, 2011. **6**(5): p. 407-12.
41. Cuthbertson, L., Powers, J., and Whitfield, C.; *The C-terminal domain of the nucleotide-binding domain protein Wzt determines substrate specificity in the ATP-binding cassette transporter for the lipopolysaccharide O-antigens in Escherichia coli serotypes O8 and O9a.* J Biol Chem, 2005. **280**(34): p. 30310-9.
42. Lewis, A.L., Nizet, V., and Varki, A.; *Discovery and characterization of sialic acid O-acetylation in group B Streptococcus.* Proc Natl Acad Sci U S A, 2004. **101**(30): p. 11123-8.

SUPPLEMENTAL MATERIAL

Figure S1. Protein sequence alignment of RgpD from representative *S. mutans* strains. The grey box indicates the less conserved C-terminal domain deleted in SMU Δ*rgpD*$_{775\text{-}1218}$.

```
Xc          MTKNNIAVKVDHVSKYFKLPVESTQSLRTALVNRFKGIKGYKKQHVLRDIDFEVEKGDFF     60
NLML4       MTKNNIAVKVDHVSKYFKLPVESTQSLRTALVNRFKGIKGYKKQHVLRDIDFEVEKGDFF     60
KK23        MTKNNIAVKVDHVSKYFKLPVESTQSLRTALVNRFKGIKGYKKQHVLRDIDFEVEKGDFF     60
LJ23        MTKNNIAVKVDHVSKYFKLPVESTQSLRTALVNRFKGIKGYKKQHVLRDIDFEVEKGDFF     60
SF1         MSENQIAVKVDHVSKYFRLPTEATQSLRTALVNRFKGVKGYREQHVLKDISFDVQKGDFF     60
            *::*:************:**.*:**************:***::****:**.*:*:*****

Xc          GIVGRNGSGKSTLLKIISQIYVPEQGKVTVDGKLVSFIELGVGFNPELTGRENVYMNGAM     120
NLML4       GIVGRNGSGKSTLLKIISQIYVPEQGKVTVDGKLVSFIELGVGFNPELTGRENVYMNGAM     120
KK23        GIVGRNGSGKSTLLKIISQIYVPEQGKVTVDGKLVSFIELGVGFNPELTGRENVYMNGAM     120
LJ23        GIVGRNGSGKSTLLKIISQIYVPEQGKVTVDGKLVSFIELGVGFNPELTGRENVYMNGAM     120
SF1         GIVGRNGSGKSTLLKIISQIYVPEKGTVTIDGKLVSFIELGVGFNPELTGRENVYMNGAM     120
            ************************:*.**:******************************

Xc          LGFTTEEVDTMYQDIVDFAELQDFMNQKLKNYSSGMQVRLAFSVAIKAQGDVLILDEVLA     180
NLML4       LGFTTEEVDTMYQDIVDFAELQDFMNQKLKNYSSGMQVRLAFSVAIKAQGDVLILDEVLA     180
KK23        LGFTTEEVDTMYQDIVDFAELQDFMNQKLKNYSSGMQVRLAFSVAIKAQGDVLILDEVLA     180
LJ23        LGFTTEEVDTMYQDIVDFAELQDFMNQKLKNYSSGMQVRLAFSVAIKAQGDVLILDEVLA     180
SF1         LGFTTEEVDAMYDDIVDFAELHEFMNQKLKNYSSGMQVRLAFSVAIKAEGDVLILDEVLA     180
            *********:**:********::************************:***********

Xc          VGDEAFQRKCNDYFLERKNSGKTTILVTHDMAAVKKYCNKAVLIDDGLIKAIGEPFDVAN     240
NLML4       VGDEAFQRKCNDYFLERKNSGKTTILVTHDMAAVKKYCNKAVLIDDGFIKAIGEPFDVAN     240
KK23        VGDEAFQRKCNDYFLERKNSGKTTILVTHDMAAVKKYCNKAVLIDDGLIKAIGEPDDVAN     240
LJ23        VGDEAFQRKCNDYFLERKNSGKTTILVTHDMAAVKKYCNKAVLIDDGLIKAIGEPDDVAN     240
SF1         VGDEAFQRKCNDYFMERKDSGKTTILVTHDMAAVNKYCNKAVLIDEGVVKAYGEPFEVAN     240
            **************:***:***************:**********:*.:** *** :***

Xc          QYSLDNTDQIVEDKQEEEAAVQEEEQIVVDNLEVKLLSANRMTPRDSIRFEISYNVLAD-     299
NLML4       QYSLDNTDQIVEDKQEEEAAIQEEEQIVVDNLEVKLLSANRMTPRDSIRFEISYNVLAD-     299
KK23        QYSFDNAVPIDSAD-SEAAE-EKEEVVKVENLEVKLLSESSMTPKEDIVFEISYDVLDD-     297
LJ23        QYSFDNAVPIDSAD-SEAAEEEKEEVVKVENLEVKLLSESSMTPKEDIVFEISYDVLDD-     298
SF1         QYSLDNIKIEVP-KEEIPKEDDTEKIKFIDNFNAKLLTNNIVSPDEEIEIEFTYDVLKEN     299
            ***:**       . .     : *:   ::*::.***: . ::* :.* :*::*:** :

Xc          VGTYIALSLTDVDRNIWIYNDNSLDYLSSGSGKKRVFYECHLKSLNDIKLKLEVTVRDKQ     359
NLML4       VGTYIALSLTDVDRNIWIYNDNSLDYLSSGSGKKRVFYECHLKSLNDIKLKLEVTVRDKQ     359
KK23        TETHIVFSLTDIDRNIWIYNDNAWDIPTKGSGHKCYRYTCSIPNINDIKLKLQVTILGKS     357
LJ23        TETHIVFSLTDIDRNIWIYNDNAWDIPTKGSGHKSYRYTCSIPNINDIKLKLQVTILGKS     358
SF1         LKTYIAFSFVDISNGIGLYNDNSWNYRQESKGSKKLIYKCKLPYFNTVKLRVQAYIRDEK     359
              *:*.:*:.*:...* :****: :   ...* *   * * :  :* :**:::. : .:.

Xc          GQMLAFSSATNTPIISINRDDLEGDDKSAMDSASGLIQRNGQWQFS--                405
NLML4       GQMLAFSSATNTPIISINRDDLEGDDKSAMDSASGLIQRNGQWQFS--                405
KK23        KEMLAFANTANAPIIVINRDDIFLDDLSAVDSASGIIQRNGKWDILSE                405
LJ23        KEMLAFANTANAPIIVINRDDIFLDDLSAVDSASGIIQRNGKWDILSE                406
SF1         DHDLTVLSMSKPPVFTV-RREIDESDFSLKDSATGLLERQGDWTINED                406
             . *:. . :: *:: : * ::  .* *  ***:*:::*:*.* :
```


5

Selection of merodiploid mutants in *Streptococcus mutans* by deletion of the near essential serotype c carbohydrate transporter

Samantha L. van der Beek[1], Malbert R. C. Rogers[1], Anne-Stéphanie Rueff[2], Jan-Willem Veening[2], Nina M. van Sorge[1]

5

[1] University Medical Center Utrecht, Utrecht University, Medical Microbiology, Heidelberglaan 100, 3584 CX, Utrecht, The Netherlands.

[2] Department of Fundamental Microbiology, Faculty of Biology and Medicine, University of Lausanne, Biophore Building, CH-1015 Lausanne, Switzerland

ABSTRACT

The oral pathogen *Streptococcus mutans* Xc expresses the cell wall-anchored serotype c carbohydrate, which is composed of a polyrhamnose (rhamnan) backbone with alternating glucose side chains. Previously, we investigated the role of the putative ABC-transporter RgpCD in serotype c carbohydrate translocation in *S. mutans* by genetic deletion of the encoding *rgpCD* genes (chapter 4). Surprisingly, we obtained colonies with the correct knockout genotype but two distinct morphologies and growth characteristics; one knockout phenotype was characterized by severely attenuated growth, cell separation and cell division defects, which were all significantly repressed in the other phenotype, suggesting occurrence of suppressor mutations. These two phenotypes were consistently obtained in independent transformations. Here we aimed to identify the suppressor mechanism by genetics and genomics techniques. We did not observe single nucleotide polymorphisms in the suppressor strain that could explain the observed intermediate phenotype. Instead, the suppressor strain harbored a large chromosomal duplication representing about 30% of the genome (621.7 kb), which included the genes *rgpCD*. Only in one of the two duplicated regions, *rgpCD* was replaced by the antibiotic selection marker *erm* thereby generating a heterogenous merodiploid. We observed the same duplication event in suppressor colonies obtained from five independent transformations. The borders of the duplicated region consist of the repeat insertion sequence IS*199*, containing two open reading frames *orfB* and *orfA*. Most likely, a tandem duplication, where the duplicated regions lay adjacent to each other, occurred during transformation via unequal crossing-over of the repeat sequences during chromosomal replication. In summary, these results clearly show that merodiploid strains are readily formed in *S. mutans* and can be selected for under stress conditions, such as the inactivation of critical genes. Not only could our observation explain results obtained by others that describe successful deletion of an essential gene, it may even serve as a tool to study the physiological function of otherwise essential genes. Furthermore, gene duplication could potentially act as an antibiotic tolerance mechanism.

INTRODUCTION

Streptococcus mutans is a Gram-positive oral bacterium that is one of the major causative agents of dental caries [1, 2]. Outside the oral cavity, *S. mutans* has furthermore been associated with systemic diseases, such as endocarditis [3, 4]. The cell wall of this bacterium harbors a carbohydrate, which is covalently attached to peptidoglycan and consists of a rhamnan backbone with serotype-specific glucose or galactose side chains [5]. These serotype-specific carbohydrates play important roles in bacterial physiology, host colonization and pathogenesis of *S. mutans* [6-9]. Previously, we characterized RgpCD as the putative ABC-transporter responsible for transport of the serotype c carbohydrate across the membrane in *S. mutans* Xc (chapter 4). Deletion of *rgpCD* resulted in severe growth and morphological defects that are likely resulting from impaired cell separation and cell division. However, during the construction of the *rgpCD* mutant strains, we also obtained an *rgpCD* mutant that displayed an intermediate growth phenotype compared to the wild type (WT) and the other mutant (previously not described). To discriminate between *rgpCD* deletion strains with different phenotypes, we designated these mutants Δ*rgpCD**slow (knock-out strain) and Δ*rgpCD**fast (suppressor strain).

Considering that the Δ*rgpCD**slow growth and morphological phenotype resembles that of a rhamnose-deficient mutant [8], which completely lacks the serotype c carbohydrate in its cell wall, we expected this to be the true deletion mutant. We hypothesized that Δ*rgpCD**fast had acquired a suppressor mechanism to overcome the cell wall stress originating from *rgpCD* deletion. Various suppressor mechanisms have previously been described in prokaryotes that allow deletion of essential or critical genes. In general, these mechanisms include the inactivation or upregulation of particular genes. For example, deletion of *Staphylococcus aureus ltaS*, a gene involved in lipoteichoic acid biosynthesis, is only possible in the presence of osmoprotectants [10, 11]. An *ltaS* mutant displays both growth and cell division defects, but acquisition of null-mutations in *gdpP*, encoding a c-di-AMP-specific phosphodiesterase, allows survival of the *ltaS* deletion strain in the absence of osmoprotectants [10]. An alternative suppressor mechanism is gene duplication, resulting in merodiploid strains. One way to achieve this, is by alternative pairing of a repeat sequence in the chromosome by unequal crossing-over during replication leading to a tandem duplication. This type of duplication has been described in *Streptococcus pneumoniae* and often results in duplication of a large part of the chromosome. For example, deletion of *pbp2B*, encoding a peptidoglycan biosynthesis protein, which is essential for cell elongation in *S. pneumoniae* D39, resulted in tandem duplication of eight downstream genes (~8 kb) including *ftsA*

and *ftsZ*, encoding cell division proteins [12]. However, only overexpression of FtsA, but not FtsZ, allowed deletion of *pbp2B* and resulted in normal growing bacteria, albeit with altered morphology.

In this study, we aimed to unravel the underlying mechanism of the two different *rgpCD* deletion phenotypes. Using whole genome sequencing, we discovered that *S. mutans* Δ*rgpCD**fast strains did not harbor any relevant single nucleotide polymorphisms (SNP) or single nucleotide insertions or deletion (indels). Instead, natural transformation of *S. mutans* repeatedly generated merodiploid strains carrying a 621.7 kb chromosomal duplication. Interestingly, the duplicated region was bordered by a repeat sequence, suggesting a tandem duplication as a result of unequal crossover during chromosomal replication.

MATERIALS AND METHODS

Bacterial strains and growth conditions

Streptococcus mutans Xc is a pathogenic serotype c strain [13] that is cultured in Todd-Hewitt broth (THB, Oxoid) or on THB agar at 37 °C with 5% CO_2. For culture of *S. mutans* Δ*rgpCD* strains, THB or THB agar was supplemented with 10 µg/ml erythromycin (Erm). Bacterial strains used during this study are listed in Table 1. Mutant strains are numbered as follows: Δ*rgpCD*_[S(low)/F(ast)][transformation #].[colony #]. Merodiploidy is predicted based on sequencing or phenotypic analysis.

Genetic manipulation of *S. mutans*

S. mutans Δ*rgpCD* was generated as described previously by van der Beek et al. [8]. In short, a linear *rgpCD* knock-out construct of ~1.5 kb, consisting of an Erm resistance cassette gene (*erm*; 738 bp) with 329 bp up- and 422 bp downstream flanking regions of *rgpCD*, was generated by fusion PCR (Fig. 1). The construct was transformed in *S. mutans* Xc, cultured in heat-inactivated horse serum, to create an in-frame deletion mutant of *rgpCD*. Single colonies were re-streaked on fresh plates containing erm and *rgpCD* deletion was verified by colony PCR using various primer combinations (Table S1).

Table 1. List of bacterial strains used during this study

Strain	First appearance after transformation	Overnight OD (n=1)	Illumina seq	Nanopore seq	Genome prediction
Wild type (WT)	-	0.84	Yes	Yes	Haploid
Transformation 1 (chapter 4)					
Δ*rgpCD*_F1.1	not recorded	0.80	Yes	-	Merodiploid
Δ*rgpCD*_S1.1	not recorded	0.21	Yes	Yes	Merodiploid
Δ*rgpCD*_S1.2	not recorded	0.06	-	-	Haploid
Transformation 2 (this study)					
Δ*rgpCD*_F2.1	2 days	0.72	Yes	-	Merodiploid
Δ*rgpCD*_F2.2	3 days	0.73	-	-	Merodiploid
Δ*rgpCD*_S2.1	6 days	0.10	-	-	Haploid
Δ*rgpCD*_S2.2	6 days	0.16	Yes	-	Haploid
Δ*rgpCD*_S2.3	6 days	0.26	-	-	Haploid
Δ*rgpCD*_S2.4	6 days	0.11	-	-	Haploid
Δ*rgpCD*_S2.5	6 days	0.08	-	-	Haploid
Δ*rgpCD*_S2.6	6 days	0.10	-	-	Haploid
Transformation 3 (this study)					
Δ*rgpCD*_F3.1	< 3 days	0.72	Yes	Yes	Merodiploid
Δ*rgpCD*_F3.2	< 3 days	0.70	-	-	Merodiploid
Transformation 4 (this study)					
Δ*rgpCD*_F4.1	< 3 days	0.73	Yes	-	Merodiploid
Δ*rgpCD*_F4.2	< 3 days	0.76	-	-	Merodiploid
Transformation 5 (this study)					
Δ*rgpCD*_F5.1	< 3 days	0.74	Yes	-	Merodiploid
Δ*rgpCD*_F5.2	< 3 days	0.71	-	-	Merodiploid

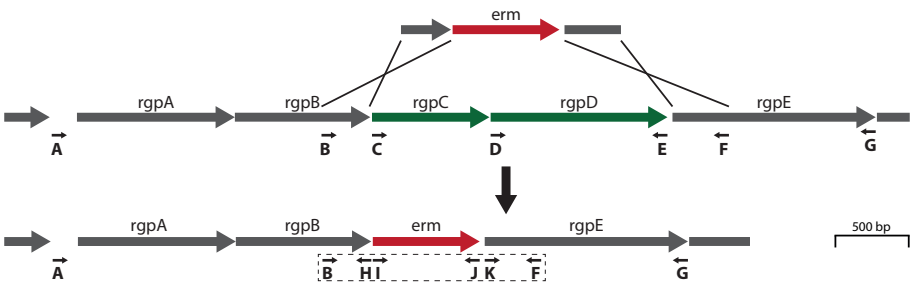


Figure 1. Schematic representation of the *rgpCD* locus in *Streptococcus mutans* and construction of an *rgpCD* deletion mutant. Small black arrows indicate the binding site of primers used during this study (Fig. 3, Table S1). Primers in a dashed box are used to create the *rgpCD* knock-out construct.

Growth curves of *S. mutans*

Growth curves of *S. mutans* strains were performed as described by van der Beek et al. in chapter 4 of this thesis.

Scanning Electron Microscopy

Scanning electron microscopy was performed as described previously by van der Beek et al. [8].

Genomic DNA isolation

Genomic DNA was isolated from *S. mutans* for Southern blot analysis and whole genome sequencing using the Wizard® Genomic DNA Purification Kit (Promega) according to manufacturer's protocol. Prior to lysis, bacterial pellets from 2 ml overnight cultures were resuspended in 125 µl protoplast buffer (20% sucrose, 20 mM Tris-HCl pH 7.0, 10 mM $MgCl_2$ and 0.05% Triton-X100 in MilliQ) and incubated for 1 hour at 37 °C with 5 µl (10kU/ml) mutanolysin.

Southern blot analysis

Chromosomal DNA of *S. mutans* WT, Δ*rgpCD**fast strains F1.1 and F3.1, and Δ*rgpCD**slow strain S1.1 was digested with HinDIII for 2 hours and run on a 1.2% agarose gel. Southern blot analysis was performed according to [14] using a non-radioactive DNA labelling and detection kit (DIG-High Prime DNA Labeling and Detection Starter Kit II from Sigma). In brief, probe 1 was generated by amplifying a fragment downstream of *rgpCD* (see Fig. 4A) using primers OVL1105 and OVL1106 and *S. mutans* Xc chromosomal DNA as template. Probe 2 (Fig. 4A) was generated using primers OVL1107 and OVL1108 using *S. mutans* Xc chromosomal DNA as template. Primer sequences are listed in Table S1. Phusion polymerase was used in all reactions.

Expected fragment sizes under complete HinDIII digestion for the WT with probe 1 are 3,551 and 518 bp, respectively. For a complete *rgpCD* replacement mutant with the *erm* cassette, expected fragment sizes with probe 1 are 3,551 and 2,229 (Δ*rgpCD**slow). A merodiploid mutant would yield three fragments of 3,551, 2,229 and 518 with hybridization to probe 1 (Δ*rgpCD**fast).

Expected fragment sizes under complete HinDIII digestion for the WT with probe 2 is a single band of 1,225 bp. For a complete *rgpCD* replacement mutant with the *erm* marker, no hybridization is to be expected (Δ*rgpCD**slow). A merodiploid mutant would yield the 1,225 bp band as well with hybridization to probe 2 (Δ*rgpCD**fast).

Illumina whole genome sequencing

Genomic DNA from *S. mutans* Xc WT, two Δ*rgpCD**slow (S1.1 and S2.2) and five Δ*rgpCD**fast (F1.1, F2.1, F3.1, F4.1 and F5.1) strains was used for Illumina sequencing. Library preparation was performed using the Illumina Nextera XT Library Prep Kit according to manufacturer's instructions. The obtained library was then sequenced on an Illumina NextSeq500 (2x150bp mid output) (Illumina inc.). The number of reads obtained for each strain is listed in Table 2.

Table 2. Number of sequencing reads per strain

	Number of reads (million)	
***S. mutans* strain**	**Illumina seq**	**Nanopore seq**
WT	3.20	0.19
Δ*rgpCD*_F1.1	1.78	-
Δ*rgpCD*_S1.1	2.20	0.12
Δ*rgpCD*_F2.1	1.34	-
Δ*rgpCD*_S2.2	2.50	-
Δ*rgpCD*_F3.1	1.10	0.34
Δ*rgpCD*_F4.1	0.45	-
Δ*rgpCD*_F5.1	0.59	-

Nanopore long read sequencing

Genomic DNA from *S. mutans* Xc WT, Δ*rgpCD*_F3.1 and Δ*rgpCD*_S1.1 was used for nanopore long read sequencing. Library preparation was performed using the Ligation Sequencing Kit 1D (SQK-LSK108) and the Native Barcoding Kit 1D PCR-free (EXP-NBD103) according to manufacturer's instructions. Sequencing was performed on a MinION Mk2 with an R9.4 flowcell (Oxford Nanopore Technologies). The number of reads obtained for each strain is listed in Table 2.

Analysis of sequencing data

To obtain a complete genome of *S. mutans* Xc, the Illumina and Nanopore reads were used together in a hybrid assembly using Unicycler [15] (version 0.4.1) in 'bold mode' with default options. This resulted in a single circular contig for *S. mutans* Xc WT, 6 contigs for Δ*rgpCD*_F3.1 and a single circular contig for Δ*rgpCD*_S1.1.

Reads from Δ*rgpCD**fast strains were compared with WT and Δ*rgpCD**slow strains for SNP and indel analysis, which was performed as described previously by van Mansfeld et al. [16, 17]. Only SNPs and indels that were unique for Δ*rgpCD**fast strains and were located in predicted protein-coding genes were selected.

Genomic reads were visualized with Integrative Genomics Viewer (IGV) version 2.3.34. Illumina reads of the wildtype and mutant strains were mapped against the obtained hybrid *S. mutans* Xc assembly using bowtie2 (version 2.3.4.1) with default parameters [18]. Circular whole genome coverage visualization was then obtained using Circleator [19]. Graphpad Prism 7.04 was used to visualize partial coverage graphs.

RESULTS

Deletion of *rgpCD* results in two different phenotypes

Previously, we characterized RgpC and RgpD as the putative ABC-transporter of the serotype c carbohydrate in *S. mutans* Xc (chapter 4) by constructing an *rgpCD* deletion strain (Δ*rgpCD*) via genetic in-frame replacement with an Erm resistance cassette (*erm*; Fig. 1). We recovered three colonies after transformation of the *S. mutans* Xc WT strain with the 1.5 kb *rgpCD* knock-out construct. In all three colonies, replacement of *rgpCD* by *erm* was confirmed by colony PCR using *erm* forward and downstream *rgpE* reverse primers. Remarkably however, one of the three mutants (Δ*rgpCD*_F1.1) displayed a different phenotype compared to the other two clones (Δ*rgpCD*_S1.1 and Δ*rgpCD*_S1.2) with respect to growth and morphology. The two clones that behaved similarly had maximum OD_{600} of 0.35, while Δ*rgpCD*_F1.1 reached an OD_{600} comparable to WT of ~0.7 (Table 1 and Fig. 2A). Furthermore, growth of Δ*rgpCD*_F1.1 in time was intermediate to WT and Δ*rgpCD*_S1.1 (Fig. 2A). Based on their growth characteristics, we designated Δ*rgpCD*_S1.1 and S1.2 as Δ*rgpCD**slow and Δ*rgpCD*_F1.1 as Δ*rgpCD**fast. In addition, the morphology of these clones also differed considerably; Δ*rgpCD**slow bacteria formed big clumps consisting of swollen cocci, which divided in multiple directions (Fig. 2B). In contrast, Δ*rgpCD**fast bacteria were only slightly swollen but formed normal streptococcal chains albeit of increased chain length indicating cell separation, but not cell division, errors (Fig. 2B). All three mutants displayed altered buoyancy in overnight cultures; Δ*rgpCD* bacteria sank to the bottom of a culture tube, WT bacteria remained partially in suspension (data now shown).

To investigate whether the two different Δ*rgpCD* phenotypes were reproducible, *S. mutans* Xc WT was additionally transformed four independent times with the 1.5 kb *rgpCD* knock-out construct. Indeed, both phenotypes were recovered in a second transformation, while three additional independent transformations yielded exclusively Δ*rgpCD**fast strains (Table 1). Colonies of Δ*rgpCD**fast strains generally appeared within 3 days after transformation, while Δ*rgpCD**slow colonies appeared

only after 6 days of culture. Further phenotypic characteristics of newly obtained Δ*rgpCD* strains, as assessed by scanning electron microscopy and recordings of the optical density and buoyancy of overnight cultures, corresponded with the phenotypes of the original Δ*rgpCD**slow and Δ*rgpCD**fast strains (Table 1, Fig. 2B and S1).

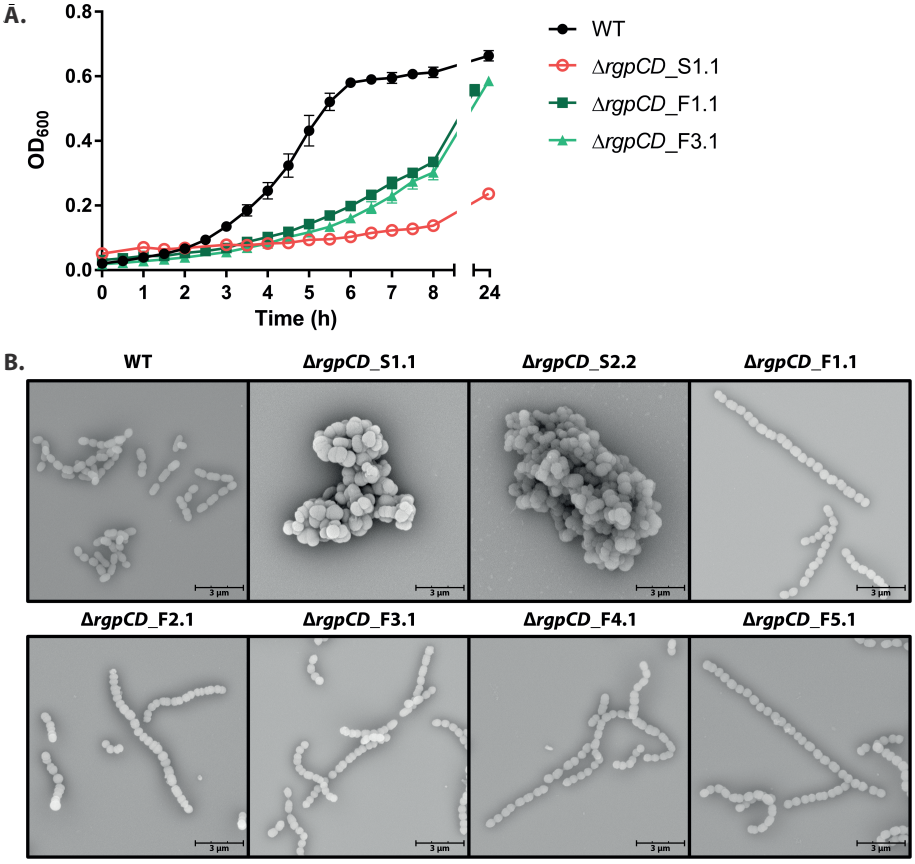


Figure 2. Deletion of *rgpCD* in *S. mutans* results in two phenotypes: Δ*rgpCDslow and Δ*rgpCD**fast. (A)** Growth curves of *S. mutans* WT, Δ*rgpCD*_S1.1 and Δ*rgpCD*_F1.1. Mutants were collected from a single transformation. **(B)** Scanning electron microscopy images of WT, Δ*rgpCD**slow (S1.1 and S2.2) and Δ*rgpCD**fast (F1.1, F2.1, F3.1, F4.1 and F5.1) strains collected from five independent transformations. Growth curves show mean ± SEM from three biological replicates.

Δ*rgpCD**fast still contains a copy of *rgpCD*

Δ*rgpCD**slow bacteria experience increased cell wall stress likely due to the lack of serotype-specific carbohydrates in their cell wall. Instead, these carbohydrates accumulate on the intracellular side of the membrane and prevent recycling of the membrane-anchor undecaprenyl-phosphate, which is also required for peptidoglycan biosynthesis. We hypothesized that Δ*rgpCD**fast bacteria had acquired suppressor mutations elsewhere on the genome to compensate for the toxicity due to loss of serotype-specific carbohydrate transport. We therefore performed Illumina whole genome sequencing on the *S. mutans* Xc WT, two Δ*rgpCD**slow (S1.1 and S2.2) and five Δ*rgpCD**fast (F1.1, F2.1, F3.1, F4.1 and F5.1) strains, obtained from independent transformations. SNP and indel analysis identified only a single point mutation in *pycB* (G1073C), annotated as an alpha subunit pyruvate carboxylase/ oxaloacetate decarboxylase, in one of the Δ*rgpCD**fast strains (Δ*rgpCD*_F4.1). However, considering this was the only mutation identified in just a single strain, it is unlikely that this is the suppressor mutation underlying the Δ*rgpCD**fast phenotype. Subsequent mapping of sequencing reads from Δ*rgpCD**slow strains against the predicted genome of *S. mutans* Xc WT confirmed the absence of *rgpCD* in both Δ*rgpCD**slow strains (Fig. 3A, S2). Surprisingly however, Δ*rgpCD**fast strains contained reads for *rgpCD* in all five strains, albeit the number of reads was about half compared to reads of the surrounding genes (Fig. 3A, S2).

To rule out the possibility that the recovered Δ*rgpCD**fast strains are actually a heterologous population of WT and *rgpCD* deletion mutants, single colonies from strain Δ*rgpCD*_F3.1 were re-streaked several times with an intermittent sonication step to create individual cocci. Even after multiple re-streaks, both *rgpCD* and *erm* could still be amplified using single colony PCR (Fig. 3B). Therefore, a heterologous population seems unlikely, also considering that Δ*rgpCD**fast bacteria were always grown in the presence of Erm and are all morphologically uniform, but clearly different from both WT and Δ*rgpCD**slow (Fig. 2B and S1). Additional PCRs were performed to assure that *erm* had inserted at the correct place in the genome, i.e. replacing *rgpCD* (Fig. 3C). In addition, we confirmed that *rgpCD* was present at the same location between *rgpB* and *rgpE* in all Δ*rgpCD**fast strains, whereas only *erm* was presented at that position in the Δ*rgpCD**slow strains (Fig. 3C). These results suggest that both *erm* and *rgpCD* are located in the same and correct position on the genome of Δ*rgpCD**fast strains.

Southern blot analysis was applied to confirm the presence of both *erm* and *rgpCD* on the genome of two Δ*rgpCD**fast strains (F1.1 and F3.1). *S. mutans* WT and all mutants demonstrated the expected hybridization patterns (Fig. 4), indicating

that the slow growing mutant is a complete *rgpCD* deletion mutant, while the fast-growing mutants are merodiploid for this allele. Note that in Figure 4 the expected 518 band when stained with probe 1 is not clearly visible although it is present in the WT and Δ*rgpCD**fast mutants when overexposed, and not visible in the Δ*rgpCD**slow mutant. In addition, fragments of 5,302 and 1,743 bp are visible in the WT and Δ*rgpCD**fast mutants when stained with probe 1 due to partial HinDIII digestion. In the Δ*rgpCD**slow mutant, besides the 5,302 band, also a band of 7,735 is visible due to incomplete HinDIII digestion. When stained with probe 2, a fragment of 1,743 is visible in the WT and Δ*rgpCD**fast mutants due to partial HinDIII digestion.

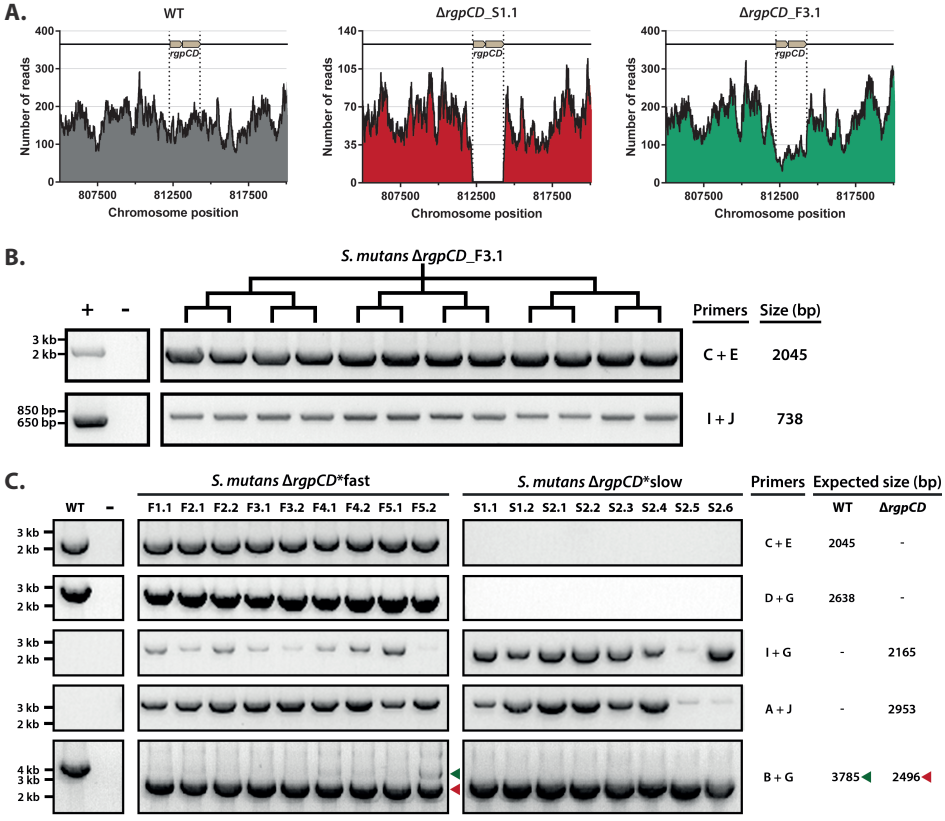


Figure 3. Δ*rgpCDfast strains still contain a copy of *rgpCD*. (A)** Coverage of the *rgpCD* sequence in wild type (WT; grey) and representative Δ*rgpCD**slow (S1.1; red) and Δ*rgpCD**fast (F3.1; green) strains. No *rgpCD* reads were obtained for Δ*rgpCD*_S1.1, indicating a complete deletion of *rgpCD*. Reads for *rgpCD* were still detected in Δ*rgpCD*_F3.1, albeit with a lower number compared to surrounding genes. **(B)** Amplification of *rgpCD* and *erm* in re-streaked colonies originating from Δ*rgpCD*_F3.1. Primer binding sites are indicated in Figure 1. Positive controls are WT gDNA (primers C+E) or pDC*erm* (primers I + J), negative control contains no template DNA. **(C)** Control PCRs to determine the presence and location of *rgpCD* and *erm* in all Δ*rgpCD**slow and Δ*rgpCD**fast strains. Primer binding sites are indicated in Figure 1. Negative control contains no template DNA.

Δ*rgpCD**fast is merodiploid

The presence of both *erm* and *rgpCD* suggest that *S. mutans* Δ*rgpCD**fast is either a full diploid or a merodiploid strain. To investigate these two possibilities, *S. mutans* Xc WT, Δ*rgpCD*_S1.1 and Δ*rgpCD*_F3.1 genomes were subjected to nanopore long read sequencing to generate closed genomes. A hybrid assembly of the Illumina and nanopore sequencing data, allowed us to close the genomes of *S. mutans* Xc WT and Δ*rgpCD*_S1.1. *S. mutans* Xc WT had a genome size of 2.0 Mb, with a G+C content of 37%. Prokka (version 1.13, [20]) was used for genome annotation and predicted a total of 1,851 coding sequences, 66 tRNA, 1 tmRNA and 15 rRNA. This strain has an average nucleotide identity of 99% compared to *S. mutans* UA159, a representative serotype c strain [21]. A total of 1,676 out of the 1,851 coding sequences present in *S. mutans* Xc (90.6%) are predicted (using Roary, [22]) to be common in the UA159 strain. The closed Δ*rgpCD*_S1.1 genome confirmed that *rgpCD* was replaced by *erm*.

Interestingly, the genome of Δ*rgpCD*_F3.1 could not be closed and consisted of six separate contigs, which could be assembled using the closed WT genome as a template and were ordered: 1-6-3-6-2-4/5-1 (Fig. S3). Contig 1 is the largest (1.92 Mb) and made up 96% of the total genome. Adjacent to contig 1 lays contig 6, which represents a small repeat region that encloses contig 3. Then contig 2 made up a small genomic region in between contig 6 and 4 or 5, the latter representing the genes *rgpCD* and *erm*. As predicted, *rgpCD* and *erm* shared the same location on the genome. Finally, both contig 4 and 5 could be linked to contig 1.

Mapping of Illumina reads from all strains against the closed *S. mutans* Xc WT genome, revealed a region of 621.7 kb with twice the number of reads compared to the rest of the genome in all five sequenced Δ*rgpCD**fast strains, but not in the WT or Δ*rgpCD**slow strains (Fig. 5A, B and S4). These results suggest that Δ*rgpCD**fast strains are merodiploid for about 30% of the genome. The genes *rgpCD* are located in this duplicated region, but show a normal read count compared to the rest of the genome, indicating that in only one of the duplicated regions, *rgpCD* is completely replaced by *erm* and that the other region still harbors the *rgpCD* genes (Fig. 3A, 5B, S2 and S4). Therefore, Δ*rgpCD**fast strains are designated as heterogenous merodiploid strains. Furthermore, the 621.7 kb duplicated region is bordered on both sites by identical genes, *orfB* and *orfA*, which were previously characterized as part of the insertion sequence (IS) IS*199* (Fig. 5 and S4) [23]. The presence of identical repeat sequences on the both sites of the duplicated region suggest a tandem duplication was generated during natural transformation via unequal crossing-over during chromosomal replication.

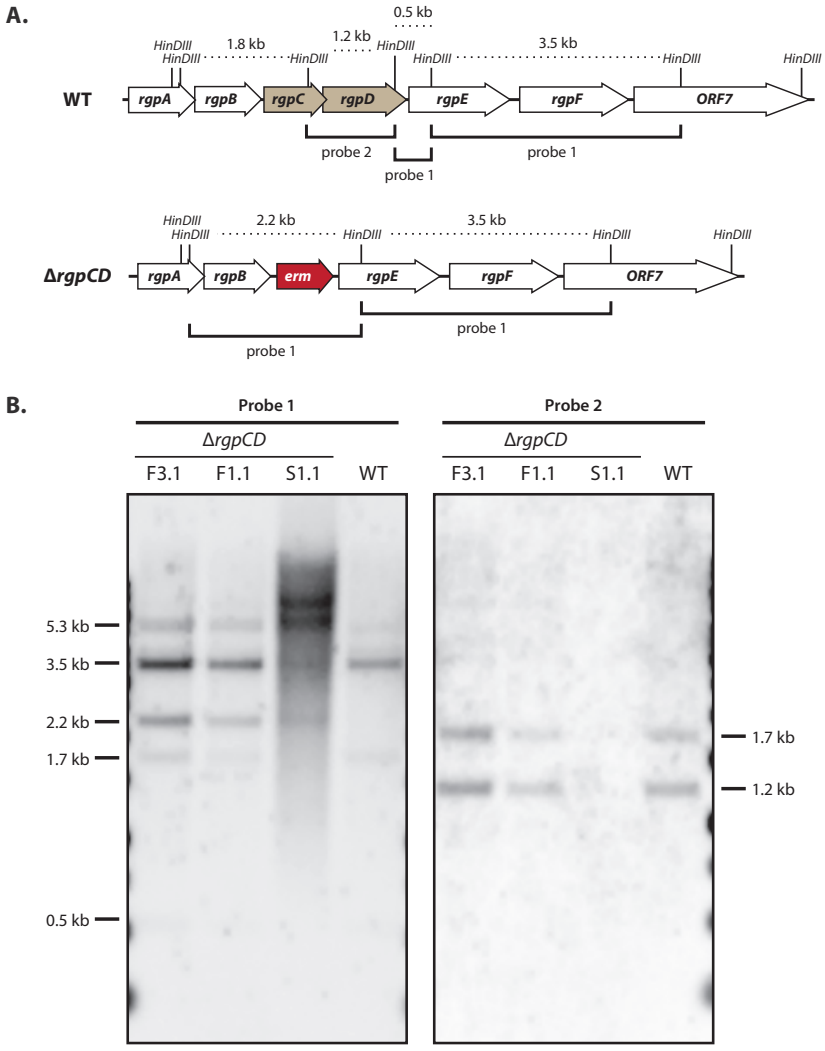


Figure 4. Southern blot analysis demonstrates that Δ*rgpCDfast strains are merodiploid. (A)** Schematic representation of the *rgpCD* locus in *Streptococcus mutans* and of the slow growing *rgpCD* mutant. **(B)** Genomic DNA of the wild type, two fast growing and one slow growing *rgpCD* replacement mutant was digested with HinDIII and hybridized using two separate probes as indicated in the figure. Each mutant displayed the anticipated restriction pattern for probe 1 and 2, indicating that the slow growing mutant is a complete *rgpCD* replacement mutant, while the fast-growing mutants are merodiploid for this allele.

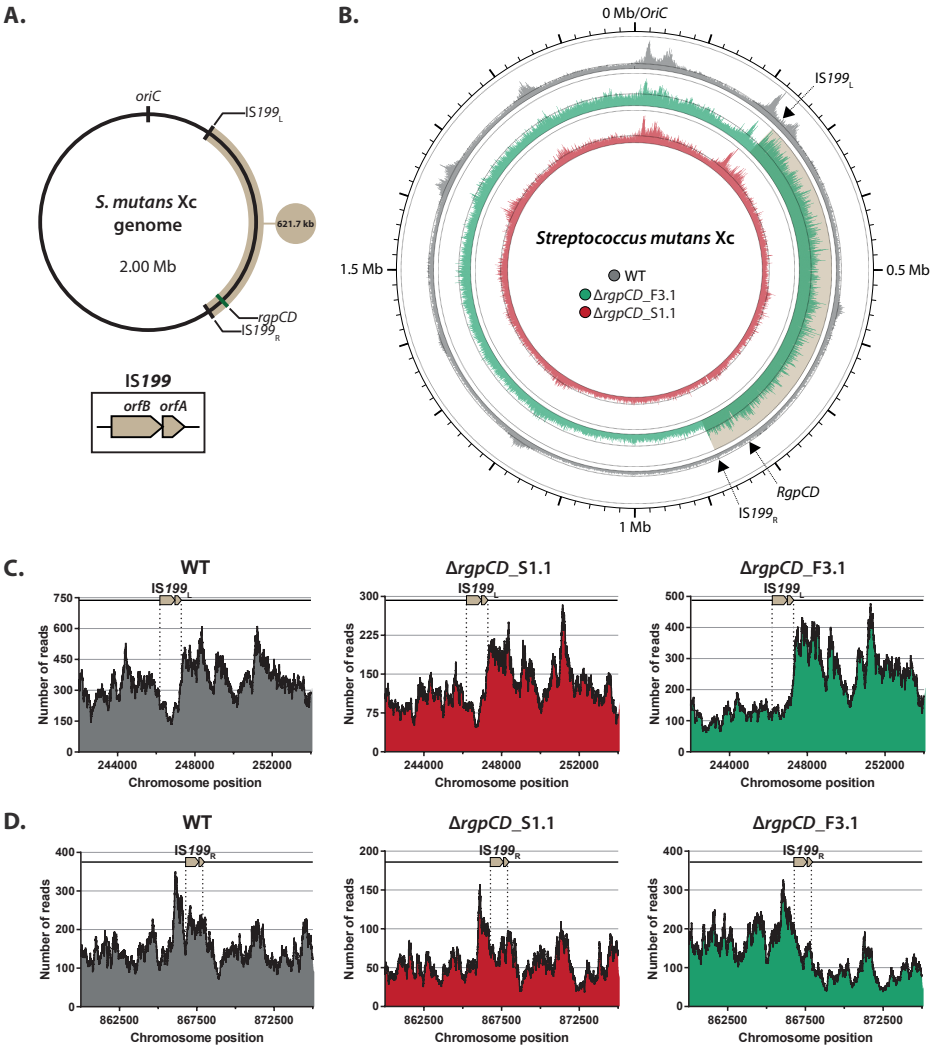


Figure 5. Coverage of *S. mutans* Xc WT, Δ*rgpCD*_S1.1 and Δ*rgpCD*_F3.1 genomes identifies duplication of 621.7 kB. **(A)** Schematic representation of *S. mutans* Xc WT genome. A duplication of 621.7 kb is observed in Δ*rgpCD*_F3.1 (shaded in beige). The duplicated region is bordered by the repeat sequence IS*199*, consisting of the genes *orfB* and *orfA*. The serotype c carbohydrate biosynthesis gene cluster, including *rgpCD*, is located inside the duplicated region. **(B)** Illumina sequencing reads of *S. mutans* Xc WT (grey), Δ*rgpCD*_S1.1 (red) and Δ*rgpCD*_F3.1 (green) genomes mapped against the closed genome of *S. mutans* Xc WT. The duplicated region (shaded in beige) is bordered by IS199_L on the left and by IS199_R on the right. **(C)** Coverage of the IS199_L sequence region and **(D)** the IS199_R sequence region.

DISCUSSION

In this study, we found that *S. mutans* can partially rescue the deletion of the critical serotype c carbohydrate transporter genes *rgpCD* by duplicating about 30% of its total genome. Merodiploidy in Δ*rgpCD**fast suppressor strains emerged during five independent transformations of *S. mutans* WT with a linear *rgpCD* knock-out construct. Furthermore, using colony PCR, Southern blotting and whole genome sequencing, we demonstrated that these merodiploid strains have integrated either *rgpCD* or *erm* between *rgpB* and *rgpE* in the duplicated serotype c carbohydrate biosynthesis gene cluster and are therefore heterogenous merodiploids. All five sequenced Δ*rgpCD**fast strains contained an identical duplication of 621.7 kb. Two hypotheses arise from this observation: 1. either our *S. mutans* Xc WT glycerol stock contains a small homogenous merodiploid population carrying the 621.7 kb duplication or 2. duplication happens during natural transformation of *S. mutans* WT with the 1.5 kb *rgpCD* knock-out construct (Fig. 6). We also wonder whether a homogenous merodiploid Δ*rgpCD* strain could exist as well upon double recombination of *rgpCD* with the Δ*rgpCD* knock-out construct in a merodiploid strain or whether this mutant will be lethal (Fig. 6). Sequencing of additional Δ*rgpCD**slow and Δ*rgpCD**fast strains, may help clarify this issue.

To gain more insight in the possible mechanism of duplication, we traced back the border sequences of the duplicated region in the Δ*rgpCD**fast strains. Interestingly, this duplicated region is bordered on both sites by the repeat sequence IS*199*. This IS was first identified in *S. mutans* V403 and contains two open reading frames *orfA* and *orfB* [23]. However, the order of these open reading frames was reversed (*orfB-orfA*) in *S. mutans* Xc. Nevertheless, the presence of a repeat sequence on both ends of the duplicated region, suggests a tandem duplication as a result of unequal crossing-over during chromosomal replication could have occurred. This type of duplication has been described previously in *S. mutans* 109cS [24]. However, instead of using a linear knock-out construct, Sato et al. used an integration plasmid to achieve random mutagenesis via a Campbell-like insertion in this strain. Multiple vector insertions created a repeat sequence in the genome, which via unequal crossing-over during chromosomal replication, resulted in a tandem duplication varying between 60 and 200 kb. Subsequently, two of the now three repeat sequences, consisting of the linearized vector with an *erm* selection marker, loop-out following Erm selection, resulting in a tandem duplication with a single inserted vector and no repeat region. In *S. pneumoniae*, deletion of the essential gene *codY* resulted in a tandem duplication of 107.4 kb and the duplicated region was bordered on both sites by IS*861*, in accordance to Δ*rgpCD**fast merodiploid strains [25]. Again, tandem duplication was

the result of unequal crossing-over during replication due to alternative pairing of a repeat sequence (IS*861*) with the repeat sequence on the daughter strand of the opposite mother strand resulting in interchromatid integration. Furthermore, it was shown that co-transformation with isogenic chromosomal DNA or with a donor fragment, containing the repeat as well as the associated non-repeated flanking sequence, stimulated merodiploid formation by 8-fold compared to transformation with a small *codY* knock-out construct alone. Moreover, co-transformation with an alternative donor fragment containing a repeat sequence with its associated non-repeated flanking sequence, trigged merodiploid formation with a duplicated region size of 935.6 kb, about 40% of the genome.

Similar to these results, we expect that the *S. mutans* Δ*rgpCD**fast merodiploid strains harbor a tandem duplication of 621.7 kb, which is triggered by unequal crossing-over of the repeat sequence IS*199* (Fig. 6). Tandem duplication can be confirmed by amplification of the unique junction between the duplicated regions by pcr using a reverse primer downstream of IS199_{left} and a forward primer upstream of IS199_{right}. Additionally, Southern blotting could be used when probes are designed to specifically target this junction. Finally, the genomic duplication could be confirmed with either pulsed-field gel electrophoresis or fluorescence in situ hybridization.

Other examples of potential merodiploid *S. mutans* strains can be found in literature, suggesting duplication events are common in this species. For example, attempts to insertionally inactivate the essential gene *gbpB,* encoding Glucan binding protein B (GbpB), in *S. mutans* UA130, yielded only merodiploid mutants where both *erm* and *gbpB* could be amplified by PCR [26]. Likewise, attempts to create a *psr brpA* double mutant in *S. mutans* UA159 probably yielded merodiploid strains [27]. Transformation of a *brpA* deletion mutant with a *psr* deletion construct was unsuccessful, while several colonies were obtained when a *psr* deletion mutant was transformed with a *brpA* deletion construct. In these strains, both *brpA* and *erm* could be amplified using PCR, suggesting merodiploidy. However, in both examples, neither the extend nor the mechanism of the duplication was identified and the possibility of a heterogenous population was not excluded.

The ability to create merodiploidy might harbor evolutionary benefits and provide bacteria with a mechanism to 'test' if foreign DNA is beneficial. If this hypothesis is true, bacteria would also have to be able to lose the least beneficial chromosomal duplication. Perhaps if a heterogenous merodiploid *S. mutans* Δ*rgpCD* strain is grown in the absence of the Erm selective pressure, another event of unequal crossing-over could reverse the duplication and re-create a haploid WT strain. In

general, merodiploidy could also serve as a stress response mechanism, similar to how genes are up- or down-regulated when exposed to external stimuli, and might serve as a potential antibiotic resistance or tolerance mechanism. Finally, better understanding of the exact mechanism of chromosomal duplication events and the circumstances under which duplication occurs, might provide a useful tool to study the physiological function of (essential) genes as well as to identify their interacting pathways [25, 28].

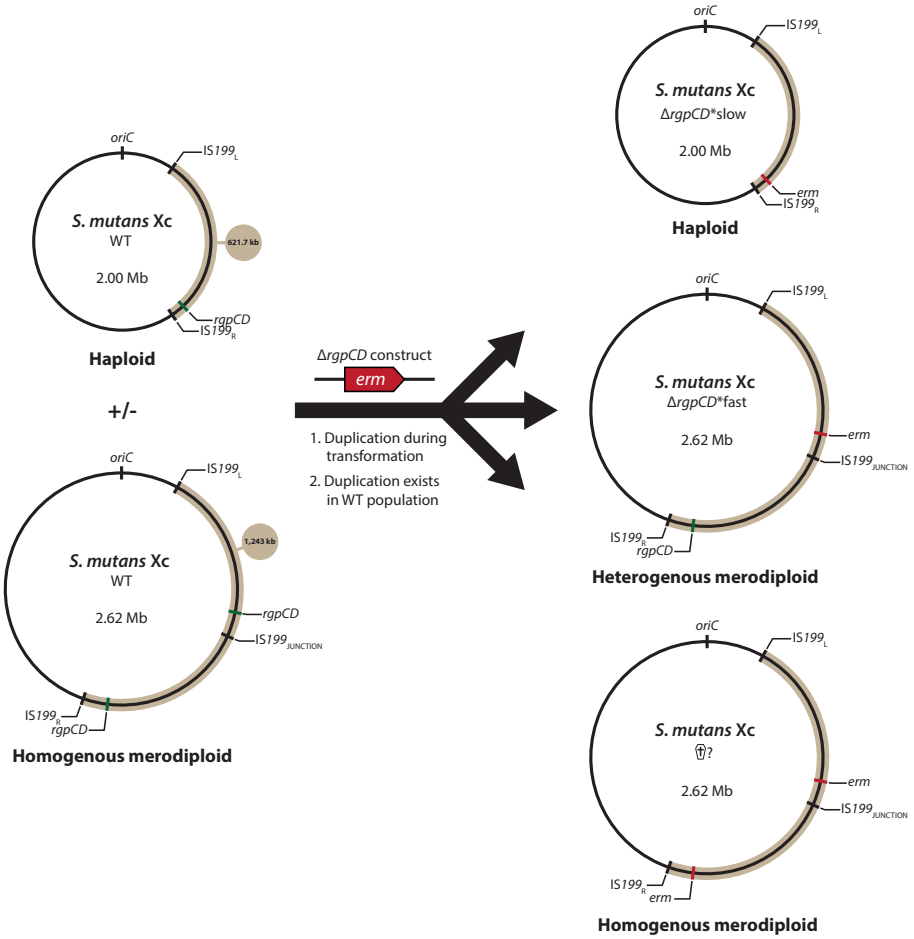


Fig 6 Hypothetical schematic representation of the creation of haploid and merodiploid *S. mutans* Δ*rgpCD* mutants. (1) The *S. mutans* WT population exist solely of haploid bacteria (1x *rgpCD*) and deletion of *rgpCD* results in either haploid Δ*rgpCD**slow mutants (1x *erm*) or in heterogenous merodiploid Δ*rgpCD**fast mutants (1x *erm* and 1x *rgpCD*). In the case of Δ*rgpCD**fast mutants, a tandem duplication occurs during transformation as a result of unequal crossing-over of the repeat insertion sequence IS*199*. **(2)** The *S. mutans* WT population already harbors a small homogenous merodiploid population (2x *rgpCD*) and no additional duplication occurs during transformation. It is unknown whether a homogenous merodiploid Δ*rgpCD* mutant (2x *erm*) is formed and if this mutant is viable.

ACKNOWLEDGEMENTS

The authors would like to thank Axel Janssen and Vincent van Hensbergen for their help with the SNP and indel analysis, Anita Schurch for her help with bioinformatics analysis and revision of the manuscript, and Mike Jennings and Malcolm Winkler for their helpful discussion and suggestion of merodiploidy.

REFERENCES

1. Loesche, W.J.; *Role of Streptococcus mutans in human dental decay.* Microbiol Rev, 1986. **50**(4): p. 353-80.
2. Koo, H., Falsetta, M.L., and Klein, M.I.; *The exopolysaccharide matrix: a virulence determinant of cariogenic biofilm.* J Dent Res, 2013. **92**(12): p. 1065-73.
3. Han, Y.W. and Wang, X.; *Mobile microbiome: oral bacteria in extra-oral infections and inflammation.* J Dent Res, 2013. **92**(6): p. 485-91.
4. Nagata, E., et al.; *Serotype-specific polysaccharide of Streptococcus mutans contributes to infectivity in endocarditis.* Oral Microbiol Immunol, 2006. **21**(6): p. 420-3.
5. St Michael, F., et al.; *Investigating the candidacy of the serotype specific rhamnan polysaccharide based glycoconjugates to prevent disease caused by the dental pathogen Streptococcus mutans.* Glycoconj J, 2017.
6. De, A., et al.; *Deficiency of RgpG Causes Major Defects in Cell Division and Biofilm Formation, and Deficiency of LytR-CpsA-Psr Family Proteins Leads to Accumulation of Cell Wall Antigens in Culture Medium by Streptococcus mutans.* Appl Environ Microbiol, 2017. **83**(17).
7. Kovacs, C.J., Faustoferri, R.C., and Quivey, R.G., Jr.; *RgpF Is Required for Maintenance of Stress Tolerance and Virulence in Streptococcus mutans.* J Bacteriol, 2017. **199**(24).
8. van der Beek, S.L., et al.; *GacA is Essential for Group A Streptococcus and Defines a New Class of Monomeric dTDP-4-dehydrorhamnose Reductases (RmlD).* Mol Microbiol, 2015.
9. Nakano, K., et al.; *Role of glucose side chains with serotype-specific polysaccharide in the cariogenicity of Streptococcus mutans.* Caries Res, 2005. **39**(4): p. 262-8.
10. Corrigan, R.M., et al.; *c-di-AMP is a new second messenger in Staphylococcus aureus with a role in controlling cell size and envelope stress.* PLoS Pathog, 2011. **7**(9): p. e1002217.
11. Grundling, A. and Schneewind, O.; *Synthesis of glycerol phosphate lipoteichoic acid in Staphylococcus aureus.* Proc Natl Acad Sci U S A, 2007. **104**(20): p. 8478-83.
12. Zheng, J.J., et al.; *Absence of the KhpA and KhpB (JAG/EloR) RNA-binding proteins suppresses the requirement for PBP2b by overproduction of FtsA in Streptococcus pneumoniae D39.* Mol Microbiol, 2017. **106**(5): p. 793-814.
13. Koga, T., et al.; *Effect of subculturing on expression of a cell-surface protein antigen by Streptococcus mutans.* J Gen Microbiol, 1989. **135**(12): p. 3199-207.
14. Southern, E.M.; *Detection of Specific Sequences among DNA Fragments Separated by Gel-Electrophoresis.* Journal of Molecular Biology, 1975. **98**(3): p. 503-+.
15. Wick, R.R., et al.; *Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads.* Plos Computational Biology, 2017. **13**(6).
16. Halaby, T., et al.; *Genomic Characterization of Colistin Heteroresistance in Klebsiella pneumoniae during a Nosocomial Outbreak.* Antimicrob Agents Chemother, 2016. **60**(11): p. 6837-6843.
17. van Mansfeld, R., et al.; *Within-Host Evolution of the Dutch High-Prevalent Pseudomonas aeruginosa Clone ST406 during Chronic Colonization of a Patient with Cystic Fibrosis.* PLoS One, 2016. **11**(6): p. e0158106.
18. Langmead, B. and Salzberg, S.L.; *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.
19. Crabtree, J., et al.; *Circleator: flexible circular visualization of genome-associated data with BioPerl and SVG.* Bioinformatics, 2014. **30**(21): p. 3125-7.
20. Seemann, T.; *Prokka: rapid prokaryotic genome annotation.* Bioinformatics, 2014. **30**(14): p. 2068-2069.
21. Ajdic, D., et al.; *Genome sequence of Streptococcus mutans UA159, a cariogenic dental pathogen.* Proc Natl Acad Sci U S A, 2002. **99**(22): p. 14434-9.
22. Page, A.J., et al.; *Roary: rapid large-scale prokaryote pan genome analysis.* Bioinformatics, 2015. **31**(22): p. 3691-3693.

23. Macrina, F.L., Jones, K.R., and Laloi, P.; *Characterization of IS199 from Streptococcus mutans V403.* Plasmid, 1996. **36**(1): p. 9-18.
24. Sato, Y., Yamamoto, Y., and Kizaki, H.; *Generation of duplication or deletion mutants in Streptococcus mutans following homologous recombination-mediated random mutagenesis by integrational vector pVA891.* Japanese Journal of Oral Biology, 1998. **40**(5): p. 506-514.
25. Johnston, C., et al.; *Natural genetic transformation generates a population of merodiploids in Streptococcus pneumoniae.* PLoS Genet, 2013. **9**(9): p. e1003819.
26. Mattos-Graner, R.O., et al.; *Functional analysis of glucan binding protein B from Streptococcus mutans.* J Bacteriol, 2006. **188**(11): p. 3813-25.
27. Bitoun, J.P., et al.; *Psr is involved in regulation of glucan production, and double deficiency of BrpA and Psr is lethal in Streptococcus mutans.* Microbiology, 2013. **159**(Pt 3): p. 493-506.
28. Sato, Y., Yamamoto, Y., and Kizaki, H.; *Construction of region-specific partial duplication mutants (merodiploid mutants) to identify the regulatory gene for the glucan-binding protein C gene in vivo in Streptococcus mutans.* FEMS Microbiol Lett, 2000. **186**(2): p. 187-91.

SUPPLEMENTAL MATERIAL

Table S1. Primers used during this study

Primer	Sequence (5′ → 3′)
A	GGCCTAATGACTGGCTTTTATAATGTTTTTTCTTTTCGCAAAATTCT
B	ACCACAGAATTATACCGCCAGC
C	CCGGAATTCATGGACTTTTTTAGTCGTAAAAATCGTAT
D	ATGACAAAAAATAATATTGCAGTCAAAGTT
E	CGCGGATCCCTATGAAAATTGCCACTGCCCA
F	GAATTATATTCCCTTTAGAGTCCACCAT
G	CGCGGATCCTCATTTTTTTCTCCTGAATAAGTTGATAAT
H	GTTTTGAGAATATTTTATATTTTTGTTCATTATTTTCTCCTATAACCAAATTTAGTAATA
I	ATGAACAAAAATATAAAATATTCTCAAAACTTTTTAACG
J	TTATTTCCTCCCGTTAAATAATAGATAACT
K	AGTTATCTATTATTTAACGGGAGGAAATAAGATATAGTGTCGGAGAATTAAAATGGTAAA
OVL1105	GATATTCTAGAACTTTCTTATAATCACTATC
OVL1106	TATAGTGTCGGAGAATTAAAATGGTAAAAG
OVL1107	CGATCAACATCTGTCAAAGAAAGAGC
OVL1108	AGAACTGACTGGCCGTGAAAATGTTTATATG

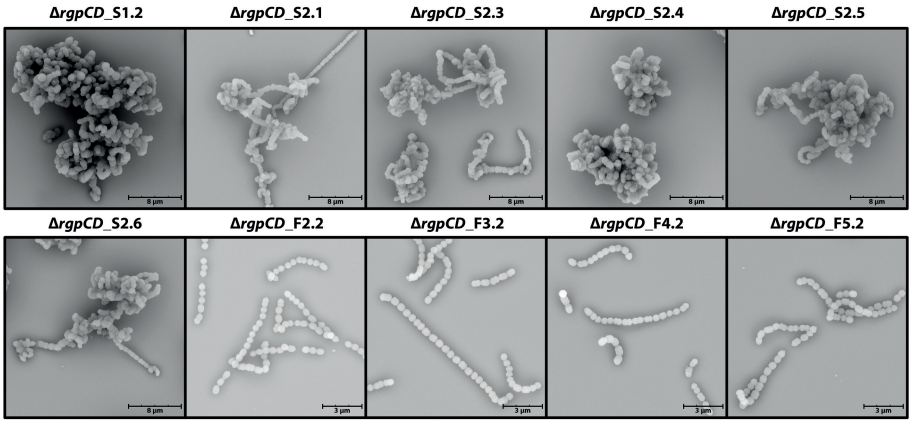


Figure S1. Morphological characterization of Δ*rgpCDslow and Δ*rgpCD**fast strains.** Scanning electron microscopy images of Δ*rgpCD**slow (S1.2, S2.2, S2.3, S2.4, S2.5 and S2.6) and Δ*rgpCD**fast (F2.2, F3.2, F4.2 and F5.2) strains collected from five independent transformations.

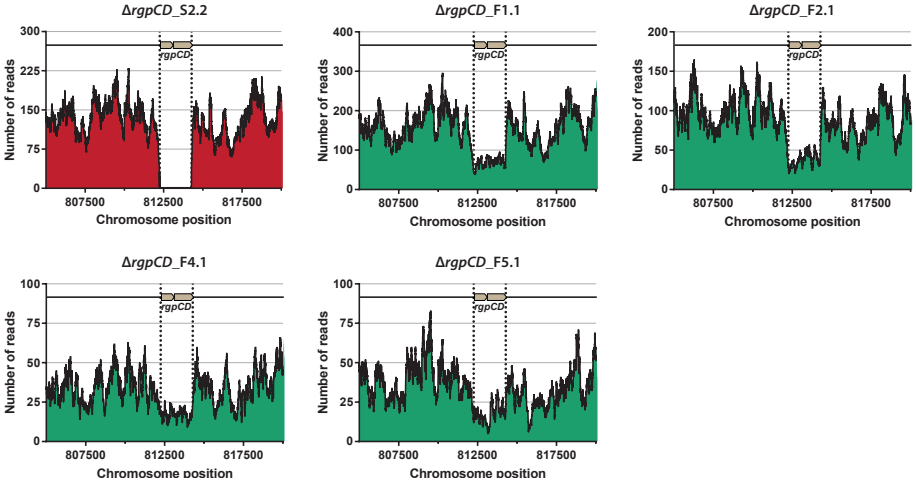


Figure S2. Δ*rgpCDfast strain still contain a copy of *rgpCD*.** Coverage of the *rgpCD* sequence in Δ*rgpCD**slow (S2.2; red) and ΔrgpCD*fast (F1.1, F2.1, F4.1 and F5.1; green) strains. No *rgpCD* reads were detected for Δ*rgpCD*_S2.2, indicating a complete deletion of *rgpCD*. Reads for *rgpCD* were still detected in Δ*rgpCD**fast strains, albeit with a lower number compared to surrounding genes.

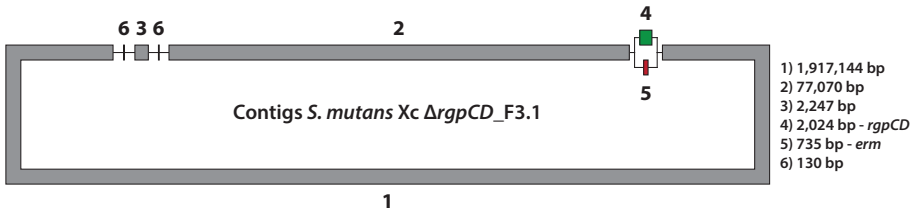


Figure S3. Bandage of contigs obtained by a hybrid assembly of Illumina and Nanopore sequencing of *S. mutans* Δ*rgpCD*_F3.1.

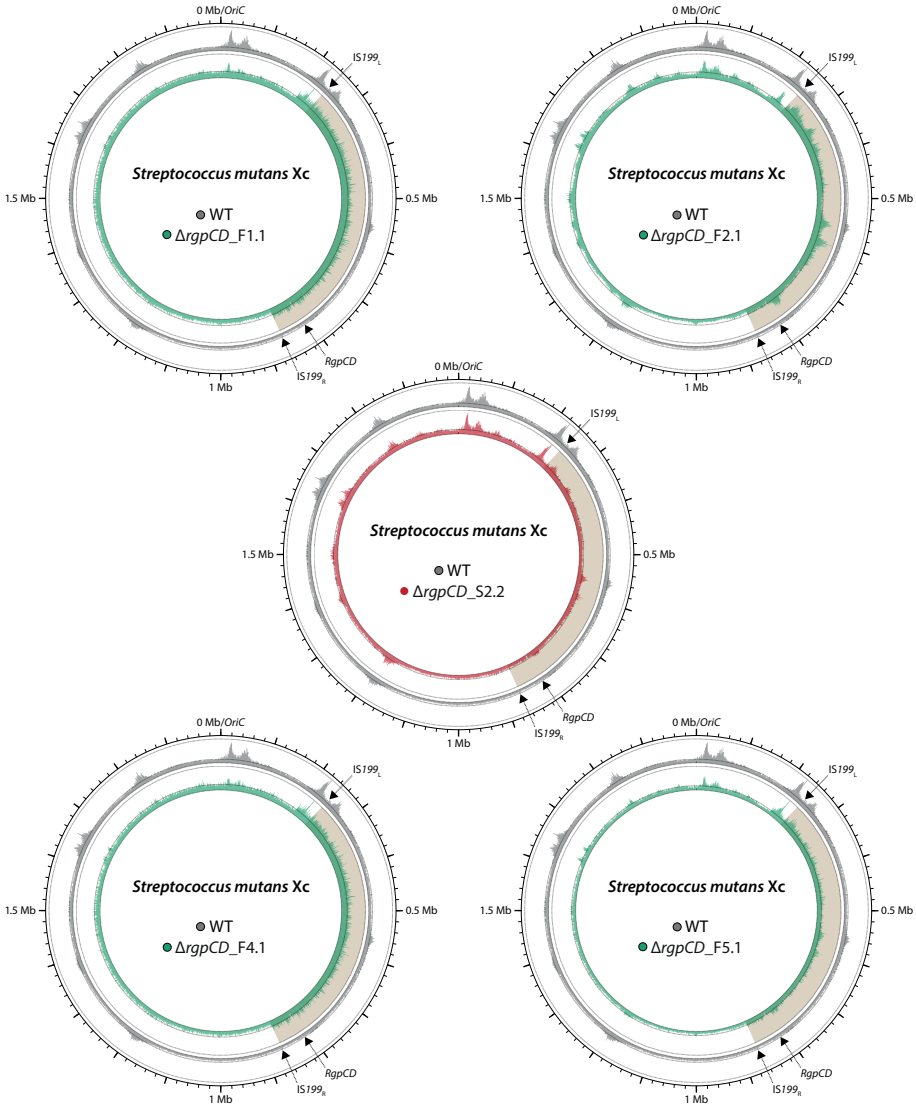


Figure S4. Coverage of *S. mutans* Δ*rgpCDslow and Δ*rgpCD**fast genomes.** Illumina sequencing reads of *S. mutans* Xc Δ*rgpCD**slow (S2.2; red) and Δ*rgpCD**fast (F1.1, F2.1, F4.1, F5.1; green) genomes mapped against the closed genome of WT (grey). A duplication of 621.7 kb (shaded in beige) is observed in all Δ*rgpCD**fast strains, but not in Δ*rgpCD*_S2.2.


6

Summarizing discussion

Samantha L. van der Beek

New insights into Streptococcal Rhamnose Polysaccharide biosynthesis and its therapeutic potential

Many streptococcal species such as *Streptococcus pyogenes*, *Streptococcus agalactiae*, *Streptococcus equi* subsp., *Streptococcus dysgalactiae* subsp., *Streptococcus mutans*, *Streptococcus sobrinus* and *Streptococcus uberis* are pathogenic for humans or animals. Some species can even cause disease in both and can be transmitted from animals to humans, making them zoonotic threats. Together, they can cause a wide range of disease from skin, throat and tooth infections to strangles, mastitis, endocarditis, meningitis, septicemia, necrotizing fasciitis, bacteremia and post-infectious streptococcal sequelae like rheumatic heart disease, streptococcal toxic shock syndrome and acute glomerulonephritis. Antibiotics that target these bacteria are readily available and still active, but are not sufficient to prevent the millions of deaths that result from streptococcal infections each year worldwide. Already for more than half a century, scientists have made efforts to develop vaccines that prevent streptococcal infections in humans and animals. Especially vaccines targeting *S. pyogenes* and *S. mutans* have received special attention. *S. pyogenes* ranks in the top ten of infection associated mortality worldwide, with a disproportionately great disease burden in low and middle-income countries [1, 2]. *S. mutans* is considered the main etiological agent of dental caries, causing a less severe, but widespread global burden in human health [3]. Current vaccine strategies mainly focus on protein antigens on the bacterial cell walls of these bacteria. However, there is incentive to use non-protein antigens for vaccine design as the introduction of glycoconjugate vaccines, which consist of isolated polysaccharides covalently attached to a protein carrier, has significantly reduced infection-related mortality of formidable pathogens such as *Streptococcus pneumoniae*. From this perspective, the use of Streptococcal Rhamnose Polysaccharides (SRPs) as vaccine antigens is of considerable interest. SRPs are glycopolymers that are covalently attached to peptidoglycan and are rich in the sugar rhamnose. The majority of the SRPs characterized to date are composed of an α-1,2/α-1,3-linked polyrhamnose (rhamnan) backbone and differ in the identity or linkage of the side chains to this backbone. In contrast to cell wall-anchored proteins, SRPs are much more abundant representing up to 60% of the cell wall mass. In addition, SRPs have essential roles in bacterial physiology, host colonization, pathogenesis and immune evasion (reviewed in **chapter 1**). Therefore, vaccine-induced antibodies could not only facilitate bacterial clearance by efficient opsonophagocytosis, they also interfere with bacterial virulence mechanisms of SRPs.

Antibodies directed against SRPs have been identified in humans following streptococcal infections as well as after active immunization of animals with protein-conjugated SRPs [4-11]. More importantly, active and passive immunizations of animals offered protection against subsequent streptococcal infections [8]. However, knowledge regarding SRP biosynthesis or structure is limited and SRP vaccine studies have been banned for several decades. Through this thesis, I hope to create awareness of the therapeutic potential of SRPs, not only as vaccine candidates, but also as targets for antibiotic development. Even though currently available antibiotics are still effective against most streptococcal species, this effectivity is not always sufficient in cases of severe fast-developing invasive disease and the rapid increase of antibiotic resistance observed in other bacteria is of big concern. In addition, the SRP biosynthesis pathway might provide a more specific antimicrobial therapy. Here, I will discuss the findings presented in this thesis in the context of the most recent developments in SRP biosynthesis characterization and its implications for the development of therapeutic agents, with special focus on dTDP-L-rhamnose biosynthesis (**chapter 2 and 3**), SRP translocation (**chapter 4**) and side chain attachment.

Production of dTDP-L-rhamnose

dTDP-L-rhamnose is the nucleotide-sugar precursor that is used as substrate by rhamnosyltransferases during SRP biosynthesis. Biosynthesis of dTDP-L-rhamnose occurs via a four-step enzymatic pathway by enzymes RmlABCD. These four enzymes are highly conserved among both Gram-positive and Gram-negative bacteria and the genes can often be found next to each other in the same operon. For unknown reasons however, in many streptococcal species, *rmlD* is separated from the *rmlABC* operon and is located in the SRP biosynthesis gene cluster. In *S. pyogenes*, *rmlD* is the first gene of the group A carbohydrate (GAC) biosynthesis gene cluster, which consists of 12 genes, and was previously renamed *gacA* [8]. In **chapter 2** of this thesis, I describe the structural and functional characterization of GacA, and confirm that it is indeed a dTDP-4-keto-L-rhamnose reductase. Furthermore, I discovered that in contrast to a previously characterized RmlD homolog from the Gram-negative bacterium *Salmonella enterica*, GacA is functional as a monomer [12]. *S. enterica* RmlD requires metal-dependent homo-dimerization for enzymatic function [13]. Further bioinformatics analysis of 213 Gram-positive and Gram-negative RmlD homologs revealed that all RmlD enzymes from Gram-positive species and some Gram-negative species lack the previously identified dimerization motif and hence are predicted to be functional as monomeric enzymes. The enzymes RmlB and RmlC have been structurally characterized in both Gram-positive and Gram-negative bacteria, including *S. enterica*, *Streptococcus suis* and *Mycobacterium tuberculosis*,

and both enzymes are functional as homo-dimers in these species [14-17]. Since no structural discrepancies were found, we did not structurally characterize RmlB and RmlC from *S. pyogenes*.

All four dTDP-L-rhamnose biosynthesis genes are essential for the viability of *S. pyogenes* in the competitive environment of a transposon mutant library and attempts to inactivate *gacA* in isolation were previously unsuccessful [8, 18, 19]. Nevertheless, we succeeded in making a conditional *gacA* deletion mutant in *S. pyogenes*, which displayed severe growth, morphology and cell separation defects when *gacA* expression was turned off (**chapter 2**). In contrast, dTDP-L-rhamnose biosynthesis genes are not essential for the viability in *S. mutans* under non-competitive conditions *in vitro*, although *rmlB, rmlC* and *rmlD* deletion mutants are severely attenuated in growth and show defects in both cell separation and cell division, resulting in an aberrant morphology (**chapter 2**). The non-essential nature of these genes in *S. mutans*, allowed us to confirm the function of *S. pyogenes* RmlB, a dTDP-D-glucose 4,6-dehydratase, RmlC, a dTDP-4-keto-6-deoxy-D-glucose 3,5 epimerase, and GacA by heterologous expression in *S. mutans* (**chapters 2 and 3**). Furthermore, *rmlABCD* are essential or at least very critical for both the viability and virulence of many other medically relevant human and animal pathogens, including, *S. agalactiae* [20, 21], *S. equi* subsp. *equi* [22], some serotypes of *Streptococcus pneumoniae* [23, 24], *Enterococcus faecalis* [25, 26], *Mycobacterium spp.* [27, 28], *Pseudomonas spp.* [29] and *S. enterica* serovar *Typhimurium* [30]. Recently, the World Health Organization has published a priority list for research and development of new antibiotics for antibiotic-resistant bacteria, including many of aforementioned species [31]. Indeed, the dTDP-L-rhamnose biosynthesis pathway has been previously studied as a new antibiotic target for *M. tuberculosis* resulting in the identification of several inhibitory lead compounds [32-36].

The fact that rhamnose biosynthesis genes are essential in *S. pyogenes* instigated us to investigate the druggability of RmlB, RmlC and GacA (**chapter 3**). Bio-layer interferometry using recombinant RmlB, RmlC and GacA against a commercially available library of ~1,000 compounds identified small molecule fragments that bound specifically to one or more of these enzymes in a concentration-dependent manner. One fragment in particular, Ri03, was considered a promising lead compound due to its ability to inhibit growth of *S. pyogenes, S. mutans* and *S. equi* with an IC_{50} of 120-240 μM, its minimal off-target effects in other bacteria and low toxicity towards eukaryotic cells. For future therapeutic application, further optimization of this fragment is certainly needed to increase the therapeutic window, by improving the affinity and specificity. Identification of the binding site of Ri03 on RmlB, RmlC

and GacA by structural characterization, will help further development of this lead fragment into an actual drug. Furthermore, future studies should be directed towards increasing the solubility of Ri03 and lead compounds identified against *M. tuberculosis* RmlABCD as most of these compounds are only soluble in DMSO. An additional challenge, is the accessibility of the dTDP-L-rhamnose biosynthesis enzymes as potential inhibitors should be able to cross the cell wall of both Gram-positive and Gram-negative bacteria. Nevertheless, lead inhibitory compounds identified by us and other groups, were able to inhibit the growth of *S. pyogenes* and *M. tuberculosis*, respectively [33-36]. Ultimately, it would be interesting to compare inhibitor efficiencies between several Gram-positive and Gram-negative bacteria, although the results will also be dependent on the essentiality of the dTDP-L-rhamnose biosynthesis pathway in these bacteria. Preliminary experiments to test the efficacy of Ri03 on *M. tuberculosis* revealed that the MIC was similar to the IC_{50} of Ri03 on *S. pyogenes* (unpublished data).

For some bacteria, like *S. mutans,* rhamnose is not truly essential for viability, exemplified by the fact that we were able to delete the dTDP-L-rhamnose biosynthesis genes. However, in these cases bacteria become severely attenuated in growth, likely also hampering their survival *in vivo* under the pressure of immune defense mechanisms. Moreover, for some bacteria, rhamnose contributes to virulence, but is not essential for viability. In these bacteria, rhamnose inhibitors may act as anti-virulence drugs, shifting the immune balance in favor of the host, or in combination therapy to synergize with existing antibiotics.

Importantly, rhamnose is not expressed or used by humans, theoretically providing rhamnose targeting drugs with a large therapeutic window. However, blasting of the four dTDP-L- rhamnose biosynthesis proteins against the human proteome revealed a possible homolog of RmlB. This protein, named TGDS, was annotated as a dTDP-D-glucose 4,6-dehydratase, but could not functionally replace *S. mutans* RmlB in our heterologous expression model (**chapter 3**). Nevertheless, caution should be taken when developing RmlB inhibitors as these might cause TGDS-dependend toxicity. This is of particular interest since mutations in TGDS are associated with the rare developmental Catel-Manzke syndrome [37-39].

Translocation

Most glycosyltransferases are cytosolic proteins and it was assumed that SRPs were completely synthesized at the cytosolic side of the membrane, whereafterthe complete SRP will be transported across the membrane and coupled to peptidoglycan. However, knew insights into the GAC biosynthesis pathway in *S. pyogenes* have

shown that only the rhamnan backbone is translocated and GlcNAC side chains are added after translocation [40]. Based on protein homology of the GlcNAc side chain biosynthesis machinery, it seems likely that this mechanism is conserved among other SRP expressing streptococcal species. The transport of the rhamnan backbone is likely executed by an ABC-transporter consisting of a permease protein, which forms the pore in the membrane, and an ATP-binding protein, which provides the energy for translocation. Genes encoding these transporter proteins have been identified in all SRP biosynthesis gene clusters and lay next to the rhamnan biosynthesis genes. In **chapter 4**, I have performed an initial characterization of RgpC (permease) and RgpD (ATP-binding) as the putative rhamnan transporter in *S. mutans.* Indeed, phenotypic characteristics of an *rgpCD* deletion strain were very similar to a rhamnose-deficient mutant (**chapter 2, 3 and 4**). However, more evidence regarding the transporters function should be collected, for example by determination of the cell wall carbohydrate composition in the *rgpCD* deletion strain, either by isolation of these carbohydrates or by staining the serotype c carbohydrate (SCC) with rhamnan- or SCC-specific antibodies. Unfortunately, such antibodies are currently unavailable. Alternatively, bacterial phages that bind specifically to the glucose side chain of the SCC could be used as a tool to detect this carbohydrate in the *S. mutans* cell wall [41]. The characterization of SRP ABC transporters is likely also of interest in relation to virulence. We know that expression of the wall teichoic acid (WTA) transporter TarGH is tightly regulated in *Staphylococcus aureus*. Moreover, overexpression of the ATP-binding protein TarH increased WTA levels in the cell wall, which increased *S. aureus* virulence [42]. Similarly, SRP translocation is probably the rate limiting step in SRP biosynthesis. Deletion or inhibition of SRP transporters will lead to accumulation of the SRP (rhamnan backbone) on the intracellular site and is lethal or critical for the viability of many streptococcal species. Lethality is most likely caused by to lack of cell wall structure and depletion of undecaprenyl-phosphate (Und-P), the SRP membrane-anchor, which is also used for peptidoglycan biosynthesis.

Interestingly, the rhamnan ABC-transporter proteins of *S. pyogenes*, GacD (permease) and GacE (ATP-binding), could functionally replace RgpC and RgpD, even in a chimeric complex (**chapter 4**). These results suggest that the transporter interacts with a conserved part, most likely the rhamnan backbone of the SRP, or that the transporters are promiscuous regarding substrate interaction. Furthermore, the SRP transporter proteins RgpCD are highly conserved (99-100 % protein sequence identity) among different *S. mutans* serotypes, which differ in the identity and the linkage of the side chain to the rhamnan backbone, indicating that substrate specificity is indeed side chain-independent. It would be interesting to exchange

more diverse transporters to determine the substrate specificity even further. Nevertheless, dissimilar RgpCD homologs can also be identified in *S. mutans* strains of the same serotype, suggesting these transporters recognize structurally different substrate. Perhaps, the serotype-specific SRPs of *S. mutans* harbor yet unidentified additional modifications, which can cause SRP microheterogeneity within serotypes and require a different transporter. RgpCD sequence variation was most prominent in the C-terminal domain of RgpD. Compared to glycopolymer ABC-transporters in other bacteria, such as the WTA transporter in *S. aureus*, the ATP-binding proteins of all SRP transporters have an extended C-terminal domain of about 150 amino acids. While the N-terminal domain of RgpD, conferring the ATPase activity, is highly conserved among streptococcal species, the C-terminal domain is much more confined to *S. mutans* and has an unknown function. Interestingly, the C-terminal domain of RgpD from *S. mutans* Xc could readily be deleted without affecting bacterial growth or morphology, suggesting a non-essential role for this domain in rhamnan translocation.

Further investigations focused on the possible interaction of RgpD with RgpE, a putative glucosyltransferase encoded by a gene downstream of *rgpD*. The possible interaction was instigated by the observation that RgpD sequence heterogeneity is directly correlated with sequence variation in RgpE. Previous studies had implicated RgpE in glucose side chain attachment to the rhamnan backbone, since an *rgpE* insertion mutant contained less glucose in the cell wall [43]. We observe that deletion of *rgpE* results in cell separation and minor growth defects. Interestingly, additional deletion of the *rgpD* C-terminus had a pronounced impact on bacterial morphology resulting from cell division defects. These observations suggest a functional interaction between the two proteins (**chapter 4**). The serotype-independent sequence heterogeneity in RgpD and RgpE in combination with the functional analysis of *rgpE* and *rgpD* mutants has let us to hypothesize that these proteins may add structural modifications, resulting in SRP microheterogeneity in *S. mutans*. This hypothesis is inspired by observations for another glycopolymer ABC-transporter in *Escherichia coli*, which is responsible for the translocation of the lipopolysaccharide-linked O-antigen polysaccharide (O-PS). The ATP-binding protein Wzt contains an extended C-terminal domain that recognizes a serotype-specific terminal cap on the O-PS, which is added by WbdD, a gene downstream of *wzt* [44]. Such capping residues often play an important role in polymer chain length regulation. It is currently unknown how and if rhamnan chain length is regulated. Perhaps, RgpE places a serotype-independent capping residue on the rhamnan backbone, which is subsequently recognized by the C-terminus of RgpD, similar to WbdD and Wzt. SRP structure redetermination will be required to identify these putative additional modifications. Subsequent SRP structure

analysis from *S. mutans* deletion mutants in the *rgpD* C-terminal domain and *rgpE* may connect these genes to this termination process.

Despite their high potential for antibiotic development, glycopolymer ABC-transporters in general are relatively unexplored targets with the exception of the *S. aureus* WTA transporter TarGH. In *S. aureus*, inhibition of late, but not early, WTA biosynthesis steps, including translocation, is lethal [45]. Using high throughput WTA biosynthesis inhibition screens, several inhibitors have been identified to target TarG, the permease protein of the WTA transporter [46-49]. Moreover, one of these inhibitors that binds to the extracellular site of TarG, was even able to inhibit the ATPase activity of TarH intracellularly [46]. Unfortunately, antibiotic resistance against these inhibitors develops very quickly already *in vitro*, either by the gain of suppressor mutations in TarG or by acquisition of non-lethal null-mutations in early step WTA biosynthesis enzymes TarO or TarA [49, 50]. However, antibiotic resistance was suppressed when used in combination with β-lactams, a class of antibiotics that target peptidoglycan biosynthesis [49, 51]. Moreover, methicillin-resistant *S. aureus* could be resensitized to β-lactams when used in combination with TarG inhibitors. These results indicate that despite the development of antibiotic resistance mechanisms when *S. aureus* is treated with WTA transporter inhibitors alone, combination therapy might overcome these problems and might provide a useful tool for treating infections caused by antibiotic resistant strains. It is therefore advised to explore the druggability of SRP transporters as well, either for combination therapy or as anti-virulence agents.

Additional modifications

Despite the early structural characterization of the main repeating units of many SRPs in the 1970's and 1980's, follow up studies to determine anchor residues, linkage to peptidoglycan or additional modifications, such as capping residues, are lacking. Also functional characterization studies of SRP biosynthesis enzymes have been limited. In the past four years while writing this thesis, the handful of publications that have appeared on these topics, have revealed striking new insights in SRP side chain biosynthesis, but also describe the discovery of additional side chain modifications to the SRP side chains [9, 40, 52]. As shortly mentioned above, Rush et al. discovered that GlcNAc side chains are attached to the GAC rhamnan backbone on the extracellular site as opposed to intracellular completion of the GAC before translocation [40]. They propose a model where GacI and GacJ are responsible for the transfer of GlcNAc to a bactoprenol acceptor, most likely Und-P, which is subsequently translocated by the putative flippase GacK. Finally, membrane anchored GacL transfers a GlcNAc residue from Und-P-GlcNAc to the rhamnan backbone. Similar mechanisms have been described for the

glycosylation of lipopolysaccharide O-PS in *Shegella flexneri* [53, 54], lipoteichoic acid in *Streptococcus sanguis, Listeria monocytogenes*, *Bacillus* species and *S. aureus* [55-59], and arabinogalactan and lipoarabinomannan in *M. tuberculosis* [60]. This mechanism of extracellular glycosylation, utilizing an Und-P-sugar intermediate, thus seems to be common for the glycosylation of many bacterial cell wall glycopolymers. Many SRP biosynthesis gene clusters from different streptococcal species contain genes, which are annotated as bactoprenol glycosyltransferases similar to GacI and share similar predicted membrane topologies, suggesting extracellular glycosylation of SRP rhamnan backbones is well conserved among streptococcal species.

Besides new insights into SRP side chain biosynthesis, additional side chain modifications have been identified in both *S. mutans* and *S. pyogenes*. The serotype f carbohydrate (SFC) from *S. mutans* was initially reported to consist of an α-1,2/α-1,3-linked polyrhamnose backbone with alternating α-1,3-linked glucose side chains [61, 62]. However, St. Michaels et al. discovered that the SFC contains a di-glucose side chain, where the second glucose is linked in α-1,2 configuration to the first glucose [9]. It is noteworthy to mention, that the intermediate SRP products, unglucosylated and mono-glucosylated SFC, could also be identified from the *S. mutans* cell wall in addition to double glucosylated SFC. It is currently unknown how side chain addition is regulated and might be turned on or off depending on the growth conditions. Such growth dependent regulation may be the reason that this modification has previously gone unnoticed. St. Michael et al. also identified that the serotype k carbohydrate from *S. mutans* contains an α-1,3-linked galactose side chain instead of the assumed unsubstituted rhamnan glycopolymer [9, 63].

Except for the group B carbohydrate, SRPs were thought to be uncharged molecules and free of phosphatediesters. However, glycerolphosphate (GroP) had been observed in GAC preparations in *S. pyogenes* as early as 1967 by Heymann and coworkers [64, 65], but these studies had never been followed up. Recently, Edgar et al. confirmed that GroP is indeed incorporated in the GAC on the C6 of the GlcNAc side chain (work in preprint: [52]). They also show that the GroP transferase GacH is responsible for this modification and that GacH homologs are well conserved among streptococcal species. Moreover, they identified SccH, formerly called ORF7 in the serotype-specific carbohydrate biosynthesis gene clusters, as the GacH homolog in *S. mutans*. Deletion of *sccH* did not alter the rhamnose to glucose ratio in the cell wall of *S. mutans,* suggesting the GroP modification is attached to the glucose side chain. Interestingly, van Sorge et al. also reported the presence of small amounts of glucose and mannose in a GAC carbohydrate composition analysis [8]. Perhaps, these glycans function as anchor or capping residues in the GAC.

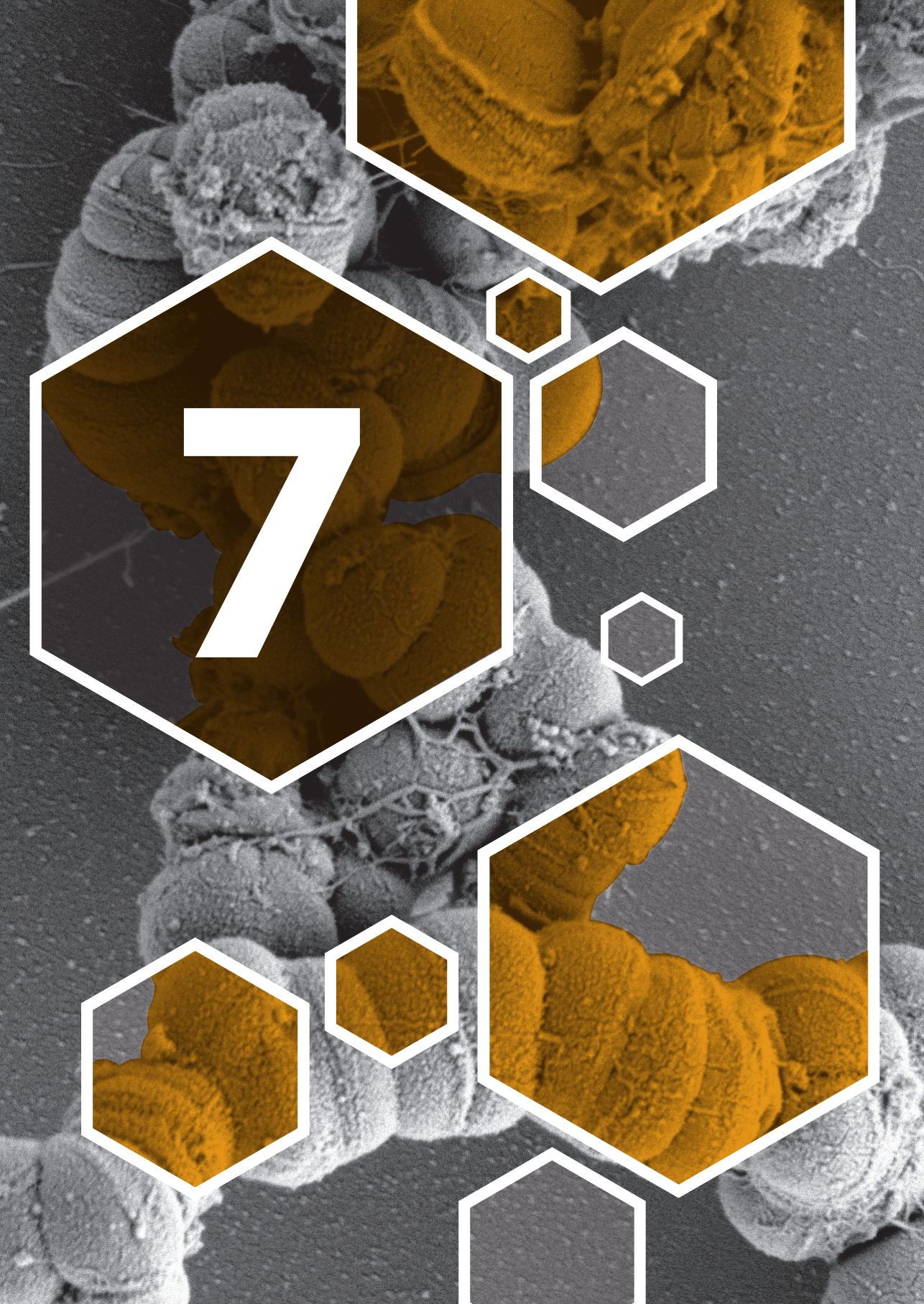
These newly identified SRP modifications might have important implications for the development of new antibiotics and vaccines. As reviewed in the general introduction of this thesis (**chapter 1**), antibodies directed against the GAC GlcNAc side chain are thought to cross-react with GlcNAc epitopes on heart tissue, contributing to the development of rheumatic fever and rheumatic heart disease. These findings have raised concern regarding the use of the GAC as vaccine to prevent *S. pyogenes* infections. However, the presence of GroP on the GlcNAc side chain may shields the GlcNAc-epitope from immune recognition and may represent a safe and effective vaccine antigen. Similarly, the identification of new SRP modifications, may provide new antibiotic targets as these modifications could contribute to the virulence of streptococcal pathogens.

REFERENCES

1. Carapetis, J.R., et al.; *The global burden of group A streptococcal diseases.* Lancet Infect Dis, 2005. **5**(11): p. 685-94.
2. Ralph, A.P. and Carapetis, J.R.; *Group a streptococcal diseases and their global burden.* Curr Top Microbiol Immunol, 2013. **368**: p. 1-27.
3. Loesche, W.J.; *Role of Streptococcus mutans in human dental decay.* Microbiol Rev, 1986. **50**(4): p. 353-80.
4. Karakawa, W.W., Osterland, C.K., and Krause, R.; *Detection of Streptococcal Group-Specific Antibodies in Human Sera.* J Exp Med, 1965. **122**: p. 195-205.
5. Ayoub, E.M., Hawthorne, T., and Miller, J.; *Assay for antibodies to group C and G streptococcal carbohydrate by enzyme-linked immunosorbent assay.* J Lab Clin Med, 1986. **107**(3): p. 204-9.
6. Araj, G.F., et al.; *Immunoglobulin isotype response to the group-A streptococcal carbohydrate in humans.* APMIS Suppl, 1988. **3**: p. 8-12.
7. Rizkallah, M.F., Hoffler, E., and Ayoub, E.M.; *Serological confirmation of group C streptococcal pneumonia.* J Infect Dis, 1988. **158**(5): p. 1092-4.
8. van Sorge, N.M., et al.; *The classical lancefield antigen of group a streptococcus is a virulence determinant with implications for vaccine design.* Cell Host Microbe, 2014. **15**(6): p. 729-40.
9. St Michael, F., et al.; *Investigating the candidacy of the serotype specific rhamnan polysaccharide based glycoconjugates to prevent disease caused by the dental pathogen Streptococcus mutans.* Glycoconj J, 2017.
10. Bruyere, T., et al.; *Local response in rat to liposome-associated Streptococcus mutans polysaccharide-protein conjugate.* Vaccine, 1987. **5**(1): p. 39-42.
11. Lett, E., et al.; *Mucosal Immunogenicity of Polysaccharides Conjugated to a Peptide or Multiple-Antigen Peptide-Containing T-Cell and B-Cell Epitopes.* Infection and Immunity, 1995. **63**(7): p. 2645-2651.
12. van der Beek, S.L., et al.; *GacA is Essential for Group A Streptococcus and Defines a New Class of Monomeric dTDP-4-dehydrorhamnose Reductases (RmlD).* Mol Microbiol, 2015.
13. Blankenfeldt, W., et al.; *Variation on a theme of SDR. dTDP-6-deoxy-L- lyxo-4-hexulose reductase (RmlD) shows a new Mg2+-dependent dimerization mode.* Structure, 2002. **10**(6): p. 773-86.
14. Allard, S.T., et al.; *Toward a structural understanding of the dehydratase mechanism.* Structure, 2002. **10**(1): p. 81-92.
15. Giraud, M.F., et al.; *RmlC, the third enzyme of dTDP-L-rhamnose pathway, is a new class of epimerase.* Nat Struct Biol, 2000. **7**(5): p. 398-402.
16. Dong, C., et al.; *High-resolution structures of RmlC from Streptococcus suis in complex with substrate analogs locate the active site of this class of enzyme.* Structure, 2003. **11**(6): p. 715-23.
17. Kantardjieff, K.A., et al.; *Mycobacterium tuberculosis RmlC epimerase (Rv3465): a promising drug-target structure in the rhamnose pathway.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 5): p. 895-902.
18. Le Breton, Y., et al.; *Essential Genes in the Core Genome of the Human Pathogen Streptococcus pyogenes.* Sci Rep, 2015. **5**: p. 9838.
19. Le Breton, Y., et al.; *Genome-wide discovery of novel M1T1 group A streptococcal determinants important for fitness and virulence during soft-tissue infection.* PLoS Pathog, 2017. **13**(8): p. e1006584.
20. Hooven, T.A., et al.; *The essential genome of Streptococcus agalactiae.* BMC Genomics, 2016. **17**: p. 406.
21. Caliot, E., et al.; *Role of the Group B antigen of Streptococcus agalactiae: a peptidoglycan-anchored polysaccharide involved in cell wall biogenesis.* PLoS Pathog, 2012. **8**(6): p. e1002756.
22. Charbonneau, A.R.L., et al.; *Defining the ABC of gene essentiality in streptococci.* BMC Genomics, 2017. **18**(1): p. 426.

23. Kim, J.O., et al.; *Relationship between cell surface carbohydrates and intrastrain variation on opsonophagocytosis of Streptococcus pneumoniae.* Infect Immun, 1999. **67**(5): p. 2327-33.
24. Magee, A.D. and Yother, J.; *Requirement for capsule in colonization by Streptococcus pneumoniae.* Infect Immun, 2001. **69**(6): p. 3755-61.
25. Xu, Y., et al.; *Analysis of a gene cluster of Enterococcus faecalis involved in polysaccharide biosynthesis.* Infect Immun, 2000. **68**(2): p. 815-23.
26. Rigottier-Gois, L., et al.; *The surface rhamnopolysaccharide epa of Enterococcus faecalis is a key determinant of intestinal colonization.* J Infect Dis, 2015. **211**(1): p. 62-71.
27. Li, W., et al.; *rmlB and rmlC genes are essential for growth of mycobacteria.* Biochem Biophys Res Commun, 2006. **342**(1): p. 170-8.
28. Ma, Y., Pan, F., and McNeil, M.; *Formation of dTDP-rhamnose is essential for growth of mycobacteria.* J Bacteriol, 2002. **184**(12): p. 3392-5.
29. Engels, W., et al.; *Role of lipopolysaccharide in opsonization and phagocytosis of Pseudomonas aeruginosa.* Infect Immun, 1985. **49**(1): p. 182-9.
30. Joiner, K.A.; *Complement evasion by bacteria and parasites.* Annu Rev Microbiol, 1988. **42**: p. 201-30.
31. Tacconelli, E., et al.; *Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis.* Lancet Infect Dis, 2018. **18**(3): p. 318-327.
32. Ren, J.X., et al.; *Virtual screening for the identification of novel inhibitors of Mycobacterium tuberculosis cell wall synthesis: inhibitors targeting RmlB and RmlC.* Comput Biol Med, 2015. **58**: p. 110-7.
33. Babaoglu, K., et al.; *Novel inhibitors of an emerging target in Mycobacterium tuberculosis; substituted thiazolidinones as inhibitors of dTDP-rhamnose synthesis.* Bioorg Med Chem Lett, 2003. **13**(19): p. 3227-30.
34. Ma, Y., et al.; *Drug targeting Mycobacterium tuberculosis cell wall synthesis: genetics of dTDP-rhamnose synthetic enzymes and development of a microtiter plate-based screen for inhibitors of conversion of dTDP-glucose to dTDP-rhamnose.* Antimicrob Agents Chemother, 2001. **45**(5): p. 1407-16.
35. Sivendran, S., et al.; *Identification of triazinoindol-benzimidazolones as nanomolar inhibitors of the Mycobacterium tuberculosis enzyme TDP-6-deoxy-d-xylo-4-hexopyranosid-4-ulose 3,5-epimerase (RmlC).* Bioorg Med Chem, 2010. **18**(2): p. 896-908.
36. Wang, Y., et al.; *Novel inhibitors of Mycobacterium tuberculosis dTDP-6-deoxy-L-lyxo-4-hexulose reductase (RmlD) identified by virtual screening.* Bioorg Med Chem Lett, 2011. **21**(23): p. 7064-7.
37. Ehmke, N., et al.; *Homozygous and compound-heterozygous mutations in TGDS cause Catel-Manzke syndrome.* Am J Hum Genet, 2014. **95**(6): p. 763-70.
38. Pferdehirt, R., et al.; *Catel-Manzke Syndrome: Further Delineation of the Phenotype Associated with Pathogenic Variants in TGDS.* Mol Genet Metab Rep, 2015. **4**: p. 89-91.
39. Schoner, K., et al.; *Mutations in TGDS associated with additional malformations of the middle fingers and halluces: Atypical Catel-Manzke syndrome in a fetus.* Am J Med Genet A, 2017. **173**(6): p. 1694-1697.
40. Rush, J.S., et al.; *The molecular mechanism of N-acetylglucosamine side-chain attachment to the Lancefield group A carbohydrate in Streptococcus pyogenes.* J Biol Chem, 2017. **292**(47): p. 19441-19457.
41. Shibata, Y., Yamashita, Y., and van der Ploeg, J.R.; *The serotype-specific glucose side chain of rhamnose-glucose polysaccharides is essential for adsorption of bacteriophage M102 to Streptococcus mutans.* FEMS Microbiol Lett, 2009. **294**(1): p. 68-73.
42. Wanner, S., et al.; *Wall teichoic acids mediate increased virulence in Staphylococcus aureus.* Nat Microbiol, 2017. **2**: p. 16257.
43. Yamashita, Y., et al.; *Genes involved in cell wall localization and side chain formation of rhamnose-glucose polysaccharide in Streptococcus mutans.* J Bacteriol, 1998. **180**(21): p. 5803-7.
44. Cuthbertson, L., Powers, J., and Whitfield, C.; *The C-terminal domain of the nucleotide-binding domain protein Wzt determines substrate specificity in the ATP-binding cassette transporter for the lipopolysaccharide O-antigens in Escherichia coli serotypes O8 and O9a.* J Biol Chem, 2005. **280**(34): p. 30310-9.

45. Swoboda, J.G., et al.; *Wall teichoic acid function, biosynthesis, and inhibition.* Chembiochem, 2010. **11**(1): p. 35-45.
46. Matano, L.M., et al.; *Antibiotic That Inhibits the ATPase Activity of an ATP-Binding Cassette Transporter by Binding to a Remote Extracellular Site.* Journal of the American Chemical Society, 2017. **139**(31): p. 10597-10600.
47. Swoboda, J.G., et al.; *Discovery of a small molecule that blocks wall teichoic acid biosynthesis in Staphylococcus aureus.* ACS Chem Biol, 2009. **4**(10): p. 875-83.
48. Lee, K., et al.; *Development of improved inhibitors of wall teichoic acid biosynthesis with potent activity against Staphylococcus aureus.* Bioorg Med Chem Lett, 2010. **20**(5): p. 1767-70.
49. Wang, H., et al.; *Discovery of wall teichoic acid inhibitors as potential anti-MRSA beta-lactam combination agents.* Chem Biol, 2013. **20**(2): p. 272-84.
50. Sewell, E.W. and Brown, E.D.; *Taking aim at wall teichoic acid synthesis: new biology and new leads for antibiotics.* J Antibiot (Tokyo), 2014. **67**(1): p. 43-51.
51. Campbell, J., et al.; *An Antibiotic That Inhibits a Late Step in Wall Teichoic Acid Biosynthesis Induces the Cell Wall Stress Stimulon in Staphylococcus aureus.* Antimicrobial Agents and Chemotherapy, 2012. **56**(4): p. 1810-1820.
52. Edgar, R.J., et al., *Genetic insight into zinc and antimicrobial toxicity uncovers a glycerol phosphate modification on streptococcal rhamnose polysaccharides*. 2018: BioRxiv.
53. Korres, H., et al.; *Topological analysis of GtrA and GtrB proteins encoded by the serotype-converting cassette of Shigella flexneri.* Biochemical and Biophysical Research Communications, 2005. **328**(4): p. 1252-1260.
54. Lehane, A.M., Korres, H., and Verma, N.K.; *Bacteriophage-encoded glucosyltransferase GtrII of Shigella flexneri: membrane topology and identification of critical residues.* Biochemical Journal, 2005. **389**: p. 137-143.
55. Mancuso, D.J. and Chiu, T.H.; *Biosynthesis of glucosyl monophosphoryl undecaprenol and its role in lipoteichoic acid biosynthesis.* J Bacteriol, 1982. **152**(2): p. 616-25.
56. Fischer, W.; *Lipoteichoic acid and lipids in the membrane of Staphylococcus aureus.* Med Microbiol Immunol, 1994. **183**(2): p. 61-76.
57. Iwasaki, H., et al.; *Structure and glycosylation of lipoteichoic acids in Bacillus strains.* J Bacteriol, 1989. **171**(1): p. 424-9.
58. Yokoyama, K., Araki, Y., and Ito, E.; *The function of galactosyl phosphorylpolyprenol in biosynthesis of lipoteichoic acid in Bacillus coagulans.* Eur J Biochem, 1988. **173**(2): p. 453-8.
59. Percy, M.G., et al.; *Identification of a Lipoteichoic Acid Glycosyltransferase Enzyme Reveals that GW-Domain-Containing Proteins Can Be Retained in the Cell Wall of Listeria monocytogenes in the Absence of Lipoteichoic Acid or Its Modifications.* Journal of Bacteriology, 2016. **198**(15): p. 2029-2042.
60. Berg, S., et al.; *The glycosyltransferases of Mycobacterium tuberculosis - roles in the synthesis of arabinogalactan, lipoarabinomannan, and other glycoconjugates.* Glycobiology, 2007. **17**(6): p. 35-56R.
61. Linzer, R., Reddy, M.S., and Levine, M.J.; *Structural studies of the serotype-f polysaccharide antigen from Streptococcus mutans OMZ175.* Infect Immun, 1987. **55**(12): p. 3006-10.
62. Pritchard, D.G., et al.; *Structure of the serotype f polysaccharide antigen of Streptococcus mutans.* Carbohydr Res, 1987. **166**(1): p. 123-31.
63. Nakano, K. and Ooshima, T.; *Serotype classification of Streptococcus mutans and its detection outside the oral cavity.* Future Microbiol, 2009. **4**(7): p. 891-902.
64. Heymann, H., Manniello, J.M., and Barkulis, S.S.; *Structure of streptococcal cell walls. V. Phosphate esters in the walls of group A Streptococcus pyogenes.* Biochem Biophys Res Commun, 1967. **26**(4): p. 486-91.
65. Munoz, E., Ghuysen, J.M., and Heymann, H.; *Cell walls of Streptococcus pyogenes, type 14. C polysaccharide-peptidoglycan and G polysaccharide-peptidoglycan complexes.* Biochemistry, 1967. **6**(12): p. 3659-70.


7

Nederlandse samenvatting
Dankwoord
Curriculum Vitae
List of publications

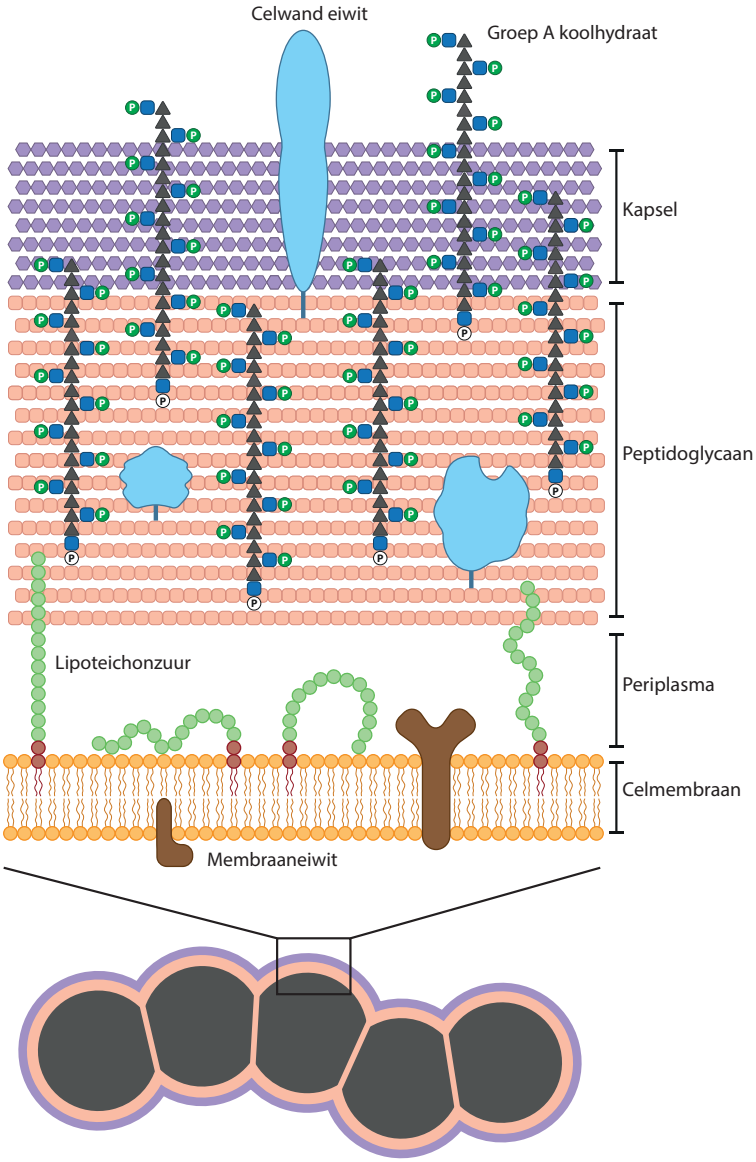
Nederlandse samenvatting

In het kort

Wereldwijd vallen er jaarlijks miljoenen doden door een infectie van de groep A streptokokkenbacterie, vooral in ontwikkelingslanden. Ondanks dat de momenteel beschikbare antibiotica nog steeds effectief zijn tegen de groep A streptokokkenbacterie, is deze effectiviteit niet altijd voldoende in gevallen van snel ontwikkelende, invasieve infecties zoals het vleesetend syndroom. Bovendien is er een opvallende toename van groep A streptokokkeninfecties zoals roodvonk in Westerse landen. Daarnaast is de waargenomen trend van snelle toename van antibioticaresistentie in andere bacteriën zorgwekkend. Wetenschappers richten zich daarom op het ontwikkelen van een vaccin ter voorkoming van groep A streptokokkeninfecties. Een potentieel interessant doelwit voor de ontwikkeling van zowel vaccins als nieuwe antibiotica is een karakteristiek suiker wat zich op het oppervlak van deze bacterie bevindt en essentieel is voor het overleven en ziekmakend vermogen van de bacterie.

Rhamnose is een suikerbouwsteen welke een essentieel onderdeel is van het oppervlaktesuiker van de groep A streptokok. In dit proefschrift beschrijf ik de functie van de enzymen die betrokken zijn bij de aanmaak van rhamnose. Bovendien heb ik chemische stoffen getest die de aanmaak van rhamnose remmen. Eén stof in het bijzonder was veelbelovend en heeft de potentie om doorontwikkeld te worden tot een nieuw soort antibioticum. Dit antibioticum zou potentieel niet alleen groep A streptokokken kunnen bestrijden, maar ook andere ziekmakende bacteriën waarvoor rhamnose essentieel is om te overleven of om ziek te maken. Een voorbeeld hiervan is de meervoudig antibioticaresistente bacterie *Mycobacterium tuberculosis*, de veroorzaker van tuberculose. Een andere voorbeeld is de bacterie *Streptococcus mutans* die een oppervlaktesuiker maakt welke erg lijkt op die van groep A streptokokken. *Streptococcus mutans* wordt beschouwd als de belangrijkste veroorzaker van tandcariës, wat leidt tot een veel minder ernstige, maar wereldwijdverspreide aantasting van de menselijke gezondheid.

Verder heb ik bestudeerd hoe het oppervlaktesuiker van zowel groep A streptokokken als *Streptococcus mutans* naar de buitenkant van de cel wordt getransporteerd over het celmembraan, nadat deze is aangemaakt aan de binnenkant van de cel. Dit gebeurt door middel van twee eiwitten die samen een transportsysteem vormen. De een maakt een gat in het celmembraan, terwijl de andere de energie levert voor het transport. Ook dit proces zou een goed doelwit kunnen zijn voor



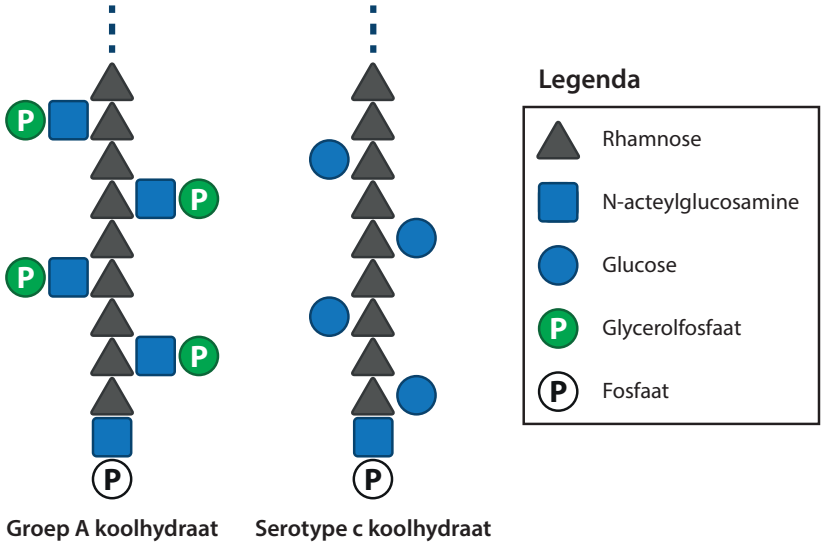
Figuur 1. Schematisch overzicht van de celwand van *Streptococcus pyogenes*. De celwand van een streptokok bestaat grofweg uit 4 lagen: **1)** het celmembraan, waarin zowel eiwitten als het polymeer lipoteichonzuur zitten verankerd; **2)** het periplasma, een ruimte tussen het celmembraan en het peptidoglycaan; **3)** een dikke laag peptidoglycaan, waaraan zowel eiwitten als SRPs, in dit geval het Groep A koolhydraat, gekoppeld zitten; en **4)** het kapsel.

de ontwikkeling van nieuwe antibiotica. Tijdens de studies naar de functie van deze transporter eiwitten, heb ik ontdekt dat *Streptococcus mutans* 25% van zijn genoom kan dupliceren als stress reducerend mechanisme. Tot slot hoop ik met dit proefschrift bewustzijn te creëren voor het therapeutisch potentieel van deze rhamnose-bevattende oppervlaktesuikers, niet alleen als vaccin kandidaat, maar ook als doelwit voor de ontwikkeling van nieuwe antibiotica.

Introductie

Het menselijk lichaam is gekoloniseerd met meer dan honderdduizend miljard bacteriën die in symbiose met ons samenleven. Als deze balans verstoord wordt of als het immuunsysteem verzwakt is, kunnen slechte bacteriën de overhand krijgen met ziekte als gevolg. Deze ziekmakende bacteriën worden ook wel pathogene bacteriën genoemd. Hieronder vallen ook veel streptokokken soorten, zoals *Streptococcus pyogenes* (beter bekend als groep A streptokokken), *Streptococcus agalactiae* (groep B streptokokken), *Streptococcus equi* en *Streptococcus dysgalactiae* subsp. (groep C en G streptokokken), *Streptococcus mutans*, *Streptococcus sobrinus* en *Streptococcus uberis*, die pathogeen kunnen zijn voor mens of dier. Sommige van deze soorten kunnen 'overspringen' van dier op mens en daardoor ziekte veroorzaken; dit noemen we zoönotische infecties. Streptokokken veroorzaken een breed scala aan ziektebeelden zoals huid-, keel- en tandinfecties, maar ook droes, mastitis, endocarditis, hersenvliesontsteking, bloedvergiftiging, vleesetend syndroom, toxisch shock syndroom en postinfectieuze ziekten zoals reumatische hartziekte en nierfilterontsteking. Ondanks dat antibiotica nog altijd effectief zijn tegen streptokokken, vallen er jaarlijks wereldwijd nog steeds miljoenen doden als gevolg van streptokokken infecties. Al meer dan een halve eeuw proberen wetenschappers vaccins te ontwikkelen die deze infecties kunnen voorkomen in mens en dier. Er is met name onderzoek gedaan naar *S. pyogenes* en *S. mutans*. *S. pyogenes* staat namelijk in de top 10 van infectie-geassocieerde doodsoorzaken wereldwijd, met een grote ziektelast vooral in landen met lage en middelhoge inkomens. *S. mutans* daarentegen wordt beschouwd als de belangrijkste veroorzaker van tandcariës, wat leidt tot een veel minder ernstige maar wijdverspreide wereldwijde belasting van de menselijke gezondheid.

Bacteriën zijn eencellige organismen die onder andere verschillen in hun vorm (rondjes, staafjes of spiraaltjes) en de opbouw van de celwand. Streptokokken zijn Gram-positieve bacteriën, die korte ketens van ronde cellen vormen. De celwand is opgebouwd uit eiwitten en suikerstructuren die verankerd zitten in het



Figuur 2. Schematische weergave van het Groep A koolhydraat uit *Streptococcus pyogenes* en het serotype c koolhydraat uit Streptococcus mutans. Beide koolhydraten bestaan uit een polyrhamnose hoofdketen, maar verschillen in de identiteit van de zijketens en de manier waarop deze gekoppeld zijn aan de polyrhamnose hoofdketen. Naar voorspelling zijn de koolhydraten gekoppeld aan het peptidoglycaan door middel van een fosfaat en N-acetylglucosamine molecuul, zie figuur 1. De lengte van de polyrhamnose keten is niet exact bekend (stippellijn).

celmembraan of aan de dikke laag peptidoglycaan daaromheen (**Figuur 1**). Een van de suikerstructuren die aan het peptidoglycaan vastzitten, zijn de Streptokokken Rhamnose Polysachariden (SRPs). SRPs zijn lange suikerketens, zogenaamde polymeren die vooral het suiker rhamnose bevatten. De meerderheid van de SRPs die tot op heden gekarakteriseerd zijn, bestaan uit een α-1,2/α-1,3-gelinkte polyrhamnose keten en verschillen in de identiteit van de suikers in de zijketens of in de manier waarop deze suikers zijn gekoppeld aan de polyrhamnose hoofdketen (**Figuur 2**). Huidige vaccinstrategieën om streptokokken infecties te voorkomen richten zich voornamelijk op de eiwitten in de celwand van deze bacteriën. Echter, ook niet-eiwit antigenen kunnen gebruikt worden voor vaccinontwerp. Dit berust met name op het succes van de zogenaamde glycoconjugaatvaccins, die bestaan uit geïsoleerde celwand polysachariden (suiker polymeerketens) covalent gehecht aan een eiwit. Deze vaccins hebben de infectie-gerelateerde sterfte door pathogenen als *Streptococcus pneumoniae* significant verminderd. Vanuit dit perspectief is het gebruik van SRPs als vaccin antigen erg interessant. Ook is de dichtheid van SRPs in de bacteriële celwand veel groter in vergelijking met eiwitten wat positief is voor vaccin ontwikkeling. Daarnaast zijn SRPs essentieel voor het overleven van de

bacterie en spelen ze een belangrijke rol in de bacteriële fysiologie, de kolonisatie van de gastheer, de pathogenese en het ontwijken van het immuunsysteem (review in **hoofdstuk 1**). Bovendien worden er tijdens een streptokokken infectie antilichamen opgewekt in mensen welke gericht zijn tegen SRPs, maar ook in dieren na actieve immunisatie met een SRP gekoppeld aan een eiwit. Belangrijker nog, deze antilichamen die het gevolg zijn van actieve en passieve immunisatie kan in proefdieren bescherming bieden tegen streptokokken infecties. Echter, kennis over de biosynthese van SRPs, dus hoe ze gemaakt worden door de bacterie, of over de structuur en samenstelling van SRPs is erg beperkt. Daarnaast waren studies voor het testen van een streptokokken vaccin in mensen zelfs verboden van 1979 tot 2006. Met dit proefschrift hoop ik bewustzijn te creëren voor het therapeutisch potentieel van SRPs, niet alleen als vaccin kandidaat, maar ook als doelwit voor de ontwikkeling van nieuwe antibiotica. Hoewel de momenteel beschikbare antibiotica nog steeds effectief zijn tegen de meeste streptokokken soorten, is deze effectiviteit niet altijd voldoende in gevallen van ernstige, snel ontwikkelende, invasieve ziekte. Daarnaast is de snelle toename van antibioticumresistentie, waargenomen in andere bacteriën, zorgwekkend. In **hoofdstukken 2 en 3** van dit proefschrift beschrijf ik de aanmaak van dTDP-L-rhamnose, de geactiveerde suikerbouwsteen van rhamnose, in *S. pyogenes* en *S. mutans* en test ik potentieel nieuwe antibiotica die gericht zijn op het remmen van deze biosynthese route. Nadat deze suikerbouwstenen zijn gebruikt voor het maken van (een deel van) de SRPs aan de binnenkant van de cel, moeten de SRPs getransporteerd worden over het celmembraan naar de buitenkant van de cel. In **hoofdstuk 4** bestudeer ik welke eiwitten hierbij betrokken zijn en wat hun specificiteit is. Tijdens de studies naar de functie van de SRP transporter eiwitten, heb ik ontdekt dat *S. mutans* 25% van zijn genoom kan dupliceren als stress reducerend mechanisme (**hoofdstuk 5)**.

Productie van dTDP-L-rhamnose

dTDP-L-rhamnose is een suikerbouwsteen dat door enzymen, rhamnosyltransferases, gebruikt kan worden om rhamnose in te bouwen in SRPs. De productie van dTDP-L-rhamnose verloopt in meerdere stappen waarbij 4 enzymen betrokken zijn; RmlA, RmlB, RmlC en RmlD/GacA (**Figuur 3**). Hoewel de genen die coderen voor deze rhamnose biosynthese genen vaak geclusterd liggen op het genoom, ligt het gen *rmlD* in de meeste streptokokken elders op het genoom als onderdeel van een SRP biosynthese genen cluster. In **hoofdstukken 2 en 3** heb ik onder andere de functie bestudeerd van RmlB, RmlC en RmlD uit *S. pyogenes*. Deze bacterie maakt het karakteristieke Groep A koolhydraat, een SRP dat bestaat uit een polyrhamnose

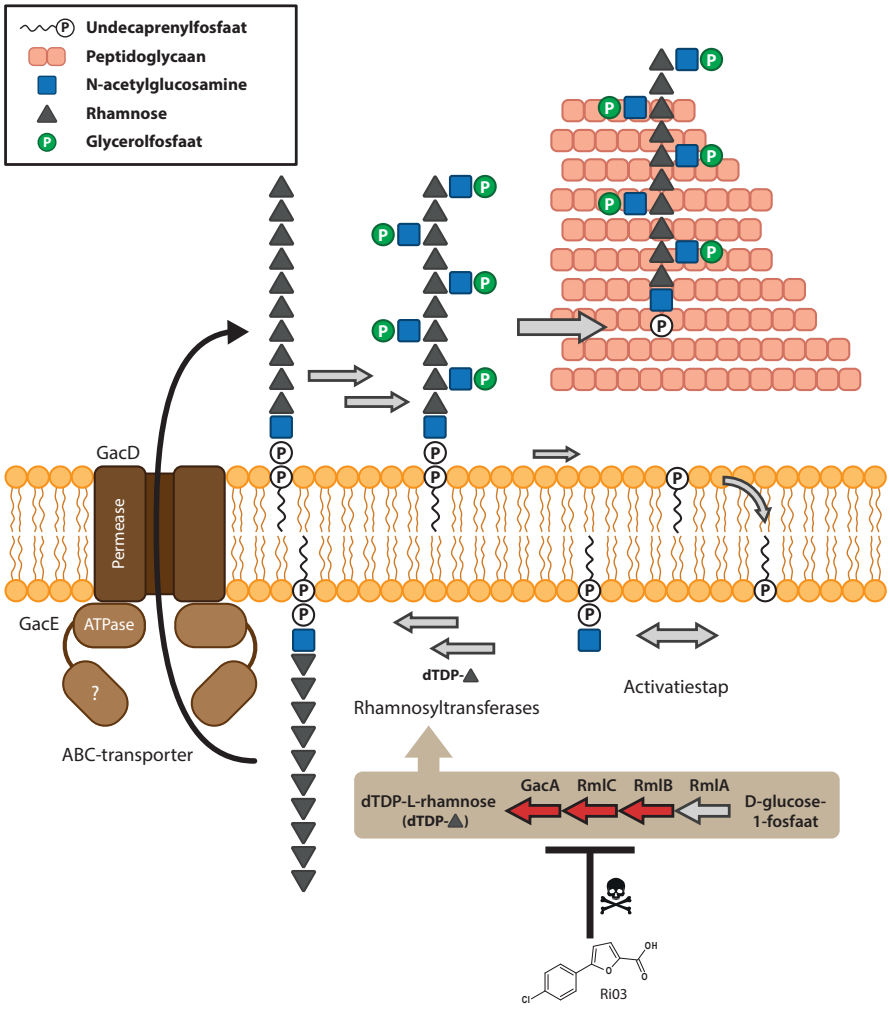
hoofdketen met daaraan om en om N-acetylglucosamine (GlcNAc) en glycerolfosfaat zijketens (**Figuur 2**). Het gen *rmlD* ligt in deze bacterie aan het begin van het Groep A koolhydraat biosynthese cluster en is daarom *gacA* genoemd. Bepaling van de eiwitstructuur van GacA en functionele biochemische analyse heeft geleid tot nieuwe inzichten over de werking van dit enzym. Het was al bekend dat het RmlD enzym in de Gram-negatieve bacterie *Salmonella enterica* functioneel is als homodimeer. Dit betekent dat er 2 dezelfde moleculen nodig zijn voor een goed functionerend enzym. Wij kwamen echter tot de ontdekking dat het RmlD enzym van *S. pyogenes* functioneel is als een monomeer (enkel molecuul) door het ontbreken van een dimerisatie motief. Verdere bio-informatica analyse van 213 verschillende RmlD enzymen toonde aan dat ook RmlD enzymen uit alle Gram-positieve en een aantal Gram-negatieve bacteriën naar voorspelling functioneel zijn als monomeer.

Daarnaast bleek dat het niet mogelijk was om *gacA* in *S. pyogenes* te inactiveren, mits er gebruik werd gemaakt van een conditioneel inactivatiesysteem. Wanneer de expressie van *gacA* op deze gereguleerde manier werd uitgeschakeld, had dit een negatief effect op groei, morfologie en cel separatie van de bacterie. Overeenkomstig met *S. pyogenes*, gebruikt *S. mutans* de dTDP-rhamnose voor het synthetiseren van een SRP, de serotype c koolhydraat. Dit koolhydraat is qua structuur vergelijkbaar met de Groep A koolhydraat maar verschilt in de identiteit van de zijketens; in plaats van GlcNAc bevat het serotype c koolhydraat glucose zijketens aan de polyrhamnose hoofdketen (**Figuur 2**). In tegenstelling tot de dTDP-L-rhamnose biosynthese genen in *S. pyogenes*, is expressie van deze 4 genen in *S. mutans* niet essentieel om te overleven in laboratoriumomstandigheden. Desondanks resulteert inactivatie van elk van deze genen in extreem slechte groei en defecten in zowel celdeling als cel separatie, waardoor deze mutanten eruitzien als klompjes van gezwollen bacteriën die in meerdere richtingen delen. Eerder is al aangetoond dat ondanks dat zulke mutanten levensvatbaar zijn in isolatie, ze niet kunnen overleven onder de competitieve omstandigheden zoals die zich voordoen in het menselijk lichaam. Om de voorspelde functie van RmlB, RmlC, en GacA uit *S. pyogenes* te kunnen bevestigen als respectievelijk dTDP-D-glucose 4,6-dehydratase, dTDP-4-keto-6-deoxy-D-glucose 3,5 epimerase en dTDP-4-keto-L-rhamnose reductase, heb ik gebruik gemaakt van een heteroloog expressie systeem in *S. mutans*. Hiermee kon ik bevestigen dat alle drie de *S. pyogenes* enzymen de functie van hun homoloog in *S. mutans* kunnen overnemen en daarmee de groei en morfologische defecten, evenals de afwezigheid van rhamnose in de celwand, herstellen. Ook zijn op deze wijze de voorspelde katalytische residuen van RmlB en RmlC uit *S. pyogenes* bevestigd.

De dTDP-L-rhamnose biosynthese genen, *rmlABCD*, zijn ook kritisch of zelfs essentieel voor vele andere medisch-relevante humane en dierlijke pathogenen om te overleven, zoals *S. agalactiae*, *S. equi* subsp. *equi*, bepaalde serotypen van *Streptococcus pneumoniae*, *Enterococcus faecalis*, *Mycobacterium spp.*, *Pseudomonas spp.* en *S. enterica* serovar *Typhimurium*. Recentelijk heeft de Wereld Gezondheidsorganisatie een lijst gepubliceerd van antibioticumresistente bacteriën waar met prioriteit nieuwe antibiotica tegen ontwikkeld moet worden. Hieronder vallen een aantal van de bovenstaande bacteriën. De dTDP-L-rhamnose biosynthese route lijkt hiervoor een potentieel geschikt nieuw doelwit. Dit heeft ons ertoe aangezet om op zoek te gaan naar moleculen die de functie van RmlB, RmlC en GacA uit *S. pyogenes* kunnen remmen (**hoofdstuk 3**). Door middel van een screen van ongeveer 1000 chemische stoffen hebben wij een aantal moleculen geïdentificeerd die specifiek bonden aan een of meerdere van deze enzymen op een concentratie-afhankelijke manier. Eén stof in het bijzonder, Ri03, heeft de potentie om verder doorontwikkeld te worden tot een bruikbaar antibioticum, omdat het in staat is de groei van *S. pyogenes, S. mutans* en *S. equi* te remmen met een IC_{50} van 120-240 μM. Daarnaast heeft Ri03 slechts minimale bijeffecten in bacteriën die geen rhamnose tot expressie brengen en een lage toxiciteit op menselijke cellen. Verdere optimalisatie is echter nodig voor een therapeutische toepassing bereikt kan worden. Hiervoor moeten onder andere bijeffecten nog verder verminderd worden door de affiniteit en specificiteit voor de dTDP-L-rhamnose biosynthese enzymen te verhogen. Het identificeren van de bindingsplaats van Ri03 op deze enzymen zal hier zeker bij helpen. Daarnaast kan gekeken worden of Ri03 of aangepaste versies hiervan ook werkzaam zijn tegen bacteriën anders dan streptokokken, zoals de meervoudig antibioticumresistente en moeilijk te behandelen bacterie *M. tuberculosis*.

Translocatie van SRPs over het membraan

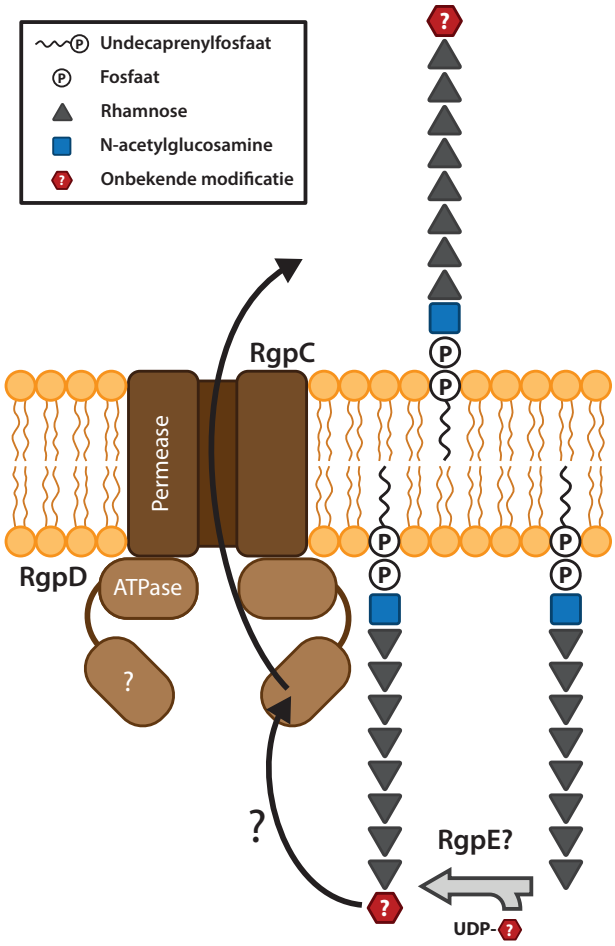
SRP glycosyltransferases zijn enzymen die de suikerbouwstenen aan elkaar koppelen. Aangezien de meeste SRP glycosyltransferases zich bevinden in het cytosol, werd aangenomen dat SRPs volledig aan de binnenkant van de cel worden opgebouwd voordat zij over het membraan worden getransporteerd. Echter, een recente studie heeft uitgewezen dat in het geval van de Groep A koolhydraat van *S. pyogenes*, de polyrhamnose hoofdketen en de bouwstenen voor de GlcNAc zijketens afzonderlijk over het membraan worden getransporteerd. Pas na deze translocatie worden de zijketens aan de hoofdketen bevestigd. Op basis van de homologie van de betrokken eiwitten, is het waarschijnlijk dat dit mechanisme geconserveerd is in andere streptokokken soorten die een SRP tot expressie brengen. De translocatie van de



Figuur 3. Schematisch overzicht van de biosynthese route van het Groep A koolhydraat in *Streptococcus pyogenes*. De aanmaak van het Groep A koolhydraat begint aan de binnenkant van de cel (in het cytoplasma) aan het celmembraan. De koppeling van N-acetylglucosamine aan undecaprenylfosfaat activeert de biosynthese, waarna rhamnosyltransferases dTDP-L-rhamnose gebruiken om de polyrhamnose keten te maken en te verlengen. De aanmaak van dTDP-L-rhamnose, wordt verzorgd door de 4 enzymen RmlA, RmlB, RmlC en GacA en kan worden geremd met het molecuul Ri03, wat dodelijk is voor de bacterie. De polyrhamnose keten wordt vervolgens getransporteerd over het celmembraan door een ABC-transporter, waarna de zijketens worden bevestigd. Tot slot wordt het voltooide Groep A koolhydraat gekoppeld aan het peptidoglycaan en zal undecaprenylfosfaat worden gerecycled.

polyrhamnose hoofdketen wordt waarschijnlijk verzorgd door een ABC-transporter die bestaat uit een permease eiwit, welke een porie vormt in het membraan, en een ATP-bindend eiwit, welke de energie levert voor de translocatie (**Figuur 3**). De genen die coderen voor deze eiwitten zijn geïdentificeerd in alle SRP biosynthese clusters en liggen naast de genen die verantwoordelijk zijn voor het maken van de polyrhamnose hoofdketen. In **hoofdstuk 4** beschrijf ik de karakterisatie van RgpC (permease) en RgpD (ATP-bindend) als de vermeende ABC-transporter van polyrhamnose in *S. mutans*. Overeenkomstig met de dTDP-L-rhamnose biosynthese genen, konden ook de transporter genen *rgpC* en *rgpD* uitgeschakeld worden, wat resulteerde in zeer slechte groei en morfologische afwijkingen ten gevolge van celdeling en cel separatie defecten. Ook heb ik aangetoond dat de polyrhamnose ABC-transporter eiwitten uit *S. pyogenes*, GacD (permease) en GacE (ATP-bindend), de functie van RgpC en RgpD kunnen overnemen in een heteroloog expressie systeem, hetgeen resulteerde in normale groei en morfologie. Deze data suggereren dat ook in *S. mutans* alleen de polyrhamnose hoofdketen wordt getransporteerd. Een alternatieve verklaring is dat GacD en GacE niet specifiek zijn voor de polyrhamnose keten, maar ook polyrhamnose ketens herkennen en transporteren waar glucose zijketens aan zitten. Meer onderzoek is echter nodig om dit definitief te kunnen bewijzen.

Uit bio-informatica onderzoek blijkt dat de eiwitsequenties van RgpC en RgpD uit verschillende *S. mutans* stammen over het algemeen sterk geconserveerd zijn (99-100%). Echter, er zijn er een aantal stammen met een afwijkende transporter eiwitsequentie, ondanks het feit dat ze dezelfde serotype c koolhydraat tot expressie brengen. Aangezien verschillen in eiwitsequentie een indicatie zijn voor verschil in herkenning of functie, suggereren deze bevindingen dat deze transporters SRPs herkennen die structureel van elkaar verschillen. Wij vermoeden derhalve dat er nog andere, tot nog toe onontdekte, modificaties aanwezig zijn op het serotype c koolhydraat die plaatsvinden voor de translocatie, waardoor er zogenaamde microheterogeniteit ontstaat binnen de serotype c koolhydraten van *S. mutans*. Dit verschil in eiwitsequentie is het meest prominent in het C-terminale domein van RgpD, terwijl het N-terminale domein, wat het ATP-bindende vermogen heeft, vrijwel identiek is. De functie van het C-terminale domein van RgpD is tot op heden onbekend en kan zonder zichtbaar effect op de groei en morfologie van *S. mutans* gemuteerd worden. Het lijkt er dus op dat dit domein niet essentieel is voor de translocatie van de polyrhamnose hoofdketen. Variatie in de sequentie van het C-terminale domein van *rgpD* was geassocieerd met variatie in de sequentie van *rgpE*, een gen dat naast *rgpD* op het genoom ligt en vermoedelijk codeert voor een enzym dat betrokken is bij het koppelen van de glucose zijketens aan



Figuur 4. Hypothetisch model voor het transport van de polyrhamnose hoofdketen van het serotype c koolhydraat in *Streptococcus mutans* over het mebraan. Eerst plaatst RgpE een onbekende terminale groep aan de polyrhamnose hoofdketen om de verlenging ervan te stoppen. Deze groep wordt vervolgens herkend door het C-terminale domein van RgpD wat het transport over het celmembraan initieert.

polyrhamnose. Een alternatieve hypothesis die wij hebben onderzocht is dat RgpE een andere functie heeft en een interactie aangaat met het C-terminale domein van RgpD. Deletie van *rgpE* resulteert in cel separatie problemen en een kleine groeiachterstand van *S. mutans*. Echter, als het C-terminale deel van *rgpD* samen met *rgpE* wordt uitgeschakeld, heeft dit ernstige gevolgen voor de morfologie van de bacterie wat zich uit in sterke celdelingsdefecten. Deze observaties doen vermoeden dat er inderdaad een functionele interactie bestaat tussen RgpD en RgpE. Hierdoor

zijn wij tot de hypothese gekomen dat deze eiwitten verantwoordelijk zijn voor nieuwe additionele modificaties aan het serotype c koolhydraat, wat resulteert in microheterogeniteit van deze SRP in *S. mutans* en invloed heeft op translocatie. Zo zou het mogelijk kunnen zijn dat RgpE een terminerende groep op de polyrhamnose hoofdketen plaatst, welke door het C-terminale domein van RgpD herkend wordt en fungeert als signaal om het verlengen van de hoofdketen te stoppen en de translocatie te starten (**Figuur 4**). Herbepaling van de SRP structuur is noodzakelijk om deze additionele voorspelde modificaties te kunnen identificeren en analyse van SRPs geïsoleerd uit voornoemde *rgpD* en *rgpE* mutanten zou vervolgens deze genen kunnen linken aan deze modificaties.

Het duidelijk is dat er meer onderzoek nodig is om de exacte functie van deze SRP transporters te kunnen bevestigen. Ondanks dat is het onmiskenbaar dat SRP transporters potentieel interessante targets zijn voor nieuwe antibiotica, aangezien deze net als de dTDP-L-rhamnose biosynthese route essentieel zijn voor de virulentie en de overlevingskansen van streptokokken. Vooruitlopend op de ontwikkeling van zulke nieuwe antibiotica, beschrijf ik in **hoofdstuk 5** een potentieel resistentie mechanisme van *S. mutans*. Deletie van *rgpC* en *rgpD* resulteerde namelijk niet alleen in slecht groeiende mutanten met morfologische defecten, maar er ontstonden ook mutanten die intermediaire groei en morfologie vertoonden. Na het bepalen en analyseren van de complete genoomsequenties bleek dat deze mutanten 25% van hun genoom hadden verdubbeld, waaronder het SRP biosynthese cluster met daarop *rgpC* en *rgpD*. In deze gevallen bleek dat slechts in een van de twee genen clusters *rgpC* en *rgpD* uitgeschakeld waren. Vergelijkbaar kan een verdubbeling van het genoom in een wild type bacterie leiden tot overexpressie van de SRP transporter, wat gevolgen zou kunnen hebben voor de dosering van potentiele antibiotica gericht op het remmen van deze transporters.


7

Dankwoord

Ik weet het nog goed: “Ik ga echt nooit promoveren!” zei ik altijd vol overtuiging. Maar toen was daar Nina met een toch wel heel leuk project. En dan ligt hier toch zomaar opeens een proefschrift voor mijn neus. Nou ja, zomaar opeens… Het heeft toch ruim vier jaar geduurd en ik had het zeker niet alleen gekund. Daarom wil ik graag iedereen bedanken die mij heeft geholpen, maar toch zeker een aantal mensen in het bijzonder.

Nina, al bij onze eerste ontmoeting wist ik dat ik goed zat. We zaten uren te kletsen over het project, maar ook over onze gemeenschappelijke voorliefde voor San Diego. Hoogzwanger kwam je terug uit je verlof om mij aan te nemen. Het was duidelijk, dit is een vrouw die weet wat ze wil en er alles aan doet om dat voor elkaar te krijgen. En jij wilde niets liever dan dat ik ging promoveren. Wat een geluk! Mede door jouw vele connecties zorgde je ervoor dat ik niets te kort kwam; studiereisjes naar Zweden, bijbaantjes in Florida, congressen en cursussen over de hele wereld en vele samenwerkingsverbanden waardoor geen experiment te gek was. Ik weet niet hoe je het doet, maar ik heb enorm veel respect voor hoe jij privé en werk weet te combineren. Je bent een harde werker en hebt alles in huis om het te maken in de wetenschap, daar heb ik alle vertrouwen in! Nina, ik heb ontzettend veel van je geleerd en ik had me geen betere begeleider kunnen wensen. Ik kan terugkijken op een mooie tijd. Bedankt voor alles.

Jos, wat was jij een leuke promotor! Altijd vrolijk en betrokken, maar serieus als het nodig was. Ik zie je nog lachend een rondje maken over het lab, altijd iedereen vragen hoe het gaat. Ik kon ook altijd bij je aankloppen om mijn zorgen te delen. Na elke meeting met jou wist ik precies waar ik aan toe was en wat ik moest doen. Je was als een soort kompas voor mij, omgekeerd moest ik jou nog wel eens wegwijs maken in die grote enzymatische lettersoep. “Dit is nog ingewikkelder dan het complement systeem!”, zei je dan klagend. Wat heb ik gelachen. Lieve Jos, bedankt dat jij mijn promotor wilde zijn!

Beste paranimfen, dankjewel dat jullie naast mij willen staan tijdens misschien wel een van de mooiste en belangrijkste momenten van mijn leven. Lieve **Marjolein**, tien jaar geleden raakten we bevriend tijdens de introductie van onze studie en zijn we samen aan ons eerste avontuur begonnen op de universiteit. Wie had toen gedacht dat jij in ‘mijn’ Rotterdam en ik in ‘jouw’ Utrecht terecht zou komen om het PhD avontuur aan te gaan. Dat was wel grappig. ;-) Maar Lein, gelukkig ben je weer wat dichterbij komen werken en ik hoop dat je ook snel een huisje vind in de buurt

van Utrecht. Ik heb gehoord dat je in Soest ook heel mooi kunt wonen… ;-) Ook onze roadtrip in 2013 door de VS zal ik nooit vergeten. Twee maanden zonder ruzie, dat is best bijzonder. We snappen elkaar helemaal en bij jou voel ik me volledig op mijn gemak. Voor jou dan ook de belangrijke taak om eventuele zenuwen te temmen tijdens de grote verdediging. Lieve Lein, ik had niemand liever naast me hebben willen staan. Beste **Rob**, naast dat je een uitstekende wetenschappelijke sparringpartner was op onze kamer, bleken we ook best veel gemeen te hebben. We hebben op dezelfde plek stage gelopen, we zijn beiden gek op spelletjes en je wist me zelfs aan te steken met het Brandon Sanderson virus. Duizend maal dank voor al zijn boeken die ik heb mogen lenen, ze waren een welkome afleiding van het soms toch stressvolle schrijven van dit proefschrift. Bedankt ook voor het verdragen van al mijn zuchten en klaaggeluiden, het is fijn om soms even je hart te kunnen luchten. Beste Rob, ik vind het fijn dat jij mijn paranimf wilt zijn.

Dear **Helge**, thank you so much for teaming up with me. Our collaboration resulted in the best start I could wish for with a first-author paper within my first two years of my PhD. It was fun to have you in our lab for a couple of weeks and I always looked forward to meeting up with you during conferences, especially because of our mutual interest in solving the group A carbohydrate biosynthesis pathway. Luckily for me, we can go all out next time we meet during the defence. ;-) I guess four years are a bit too short for me to solve the whole pathway, so now I leave it up to you to complete this important mission. Hopefully you can make the rhamnose inhibitors into a success!

Zonder één persoon was ik nooit bij de medische microbiologie terecht gekomen en dat ben jij **Marien**. Als mijn stagebegeleider in Nijmegen introduceerde je mij in de wonderlijke wereld van bacteriën. Je enthousiasme en passie voor het vak zijn aanstekelijk. Ik heb niet alleen vakinhoudelijk veel van je geleerd. Toen ik mijn eerste studenten moest begeleiden, heb ik jou als rolmodel gebruikt. Ook je eindeloze optimisme en veerkrachtigheid kon ik erg waarderen: "Ga naar de stad, pak een biertje en probeer het morgen nog eens." Dat het tien uur 's ochtends was, deed er even niet toe. Marien, heel erg bedankt voor alles wat je voor me hebt gedaan en vooral dat je me hebt voorgesteld en aanbevolen aan Nina. Ik ben vereerd dat jij in mijn beoordelingscommissie wil zitten en kijk uit naar onze discussie tijdens de verdediging.

Research involves a lot of trial and error. I was therefore very lucky to have so many wonderful and smart colleagues around me to discuss my work with and help me out on the lab. Therefore, **everyone from our department MMB**, thank you for your

help. I still remember some of the slightly confused or horrified faces during some of my first Monday and Tuesday morning meetings. Oh no, SUGARS! I'm sure the letter festival (GacA=RmlD, RgpC=GacD, etc) didn't help either. Sorry for that. Still, over the past four years, I hope I have convinced everyone that sugars can be very important and fun!

Anita, Malbert, Axel en **Vincent H**, zonder jullie hulp bij het verkrijgen en analyseren van de sequencing data had ik mijn laatste research hoofdstuk niet kunnen schrijven. Het was een behoorlijk ingewikkelde puzzel, maar mede dankzij jullie is deze nu bijna opgelost. Also thanks to **Jan-Willem** and **Anne-Stéphanie** for helping with the genomic puzzle and making chapter five into a success. I hope that together you will find the final evidence to publish this exciting story.

Job en **Antoni**, dankzij jullie hulp met de scanning elektronen microscoop werd de wereld van bacteriën nog mooier. Wat heb ik genoten tijdens alle microscopie sessies. Antoni, bedankt voor je 'schatkist', ik heb er inmiddels al twee gevuld. Job, bedankt voor alle tijd die je voor me hebt genomen, terwijl je zelf al druk zat was met het afronden van je eigen boekje. Zonder jou hadden mijn cover en hoofdstukpagina's er letterlijk heel anders uitgezien. Succes nog met de laatste loodjes!

Ik wil ook de studenten bedanken die ik heb mogen begeleiden. **Ebru**, **Ap** en **Hidde**, ook jullie hebben een belangrijke bijdrage geleverd aan dit proefschrift. Ik hoop dat ik jullie veel heb kunnen leren en wens jullie het allerbeste voor de toekomst. Jullie waren drie totaal verschillende studenten en ik heb daardoor ook veel van jullie kunnen leren.

Angelino, bedankt voor al je hulp met de celkweek en cytotoxische assays. **Kok**, wat moeten we toch zonder jou? Je weet vaak dingen te lokaliseren in het lab waar wij niet eens het bestaan van afwisten! Onder jouw begeleiding heb ik de eerste maanden het lab beter leren kennen, waardoor ik extra snel van start kon toen mijn PhD project begon. Bedankt ook voor alle hulp met FACS en andere machines op het lab. **Rob W**, bedankt voor de tussentijdse evaluaties en al je goede advies de afgelopen jaren.

Daarnaast zijn er nog zoveel andere mensen betrokken geweest bij werk dat dit proefschrift niet heeft gehaald. **Eefjan**, **Lisette, Sabine** en iedereen van het **Breukink lab** die mij geholpen heeft, ontzettend bedankt voor al jullie hulp bij het zuiveren van GacO en OleD. Helaas bleek dit niet heel makkelijk te zijn en waren we door tijdgebrek genoodzaakt het project stil te leggen. Ondanks dat, vond ik het

heel leuk om iets meer chemie te kunnen doen. Ook moet ik natuurlijk **Piet** en **Carla** bedanken voor alle hulp met de eiwitzuiveringen.

I would also like to thank everyone from the **Malmström lab** in Lund, Sweden. I felt very welcome during my stay and not only in the lab, but also in your homes. **Johan** and **Emma**, thank you for teaching me all about mass-spectrometry. If I ever need a mass-spectrometry expert, I know where to go to. Even though, this knowledge did not end up being used for this thesis, I made sure to pass it along to my colleagues.

Ook buiten het werk om heb ik een geweldige 4 jaar gehad, dankzij mijn collega's, vrienden en (schoon)familie. Jullie gaven mij afleiding van mijn werk, een hoop gezelligheid, mentale steun, goed gezelschap en een hoop liefde.

Zo heb ik heb het enorm getroffen met de afdeling MMB en ik ging altijd met veel plezier naar mijn werk. Ik heb dan ook genoten van de vele verkleedpartijtjes op de Koepel en PhD retreat, de buitendag activiteiten, de Sinterklaasspelletjes, de gezellige borrels en natuurlijk alle ICEA activiteiten. (Ex-)roomies, sorry for all the funny noises, the many sights and the distracting chats. Thanks for all fun moments, good conversations and delicious snacks and room dinners. **Anouk**, buurvrouw, bedankt voor alle goede, fijne en leuke gesprekken en natuurlijk bedankt voor het oppassen op Sjakie en Coco! Ik kom snel eens op bezoek om jullie parkietjes te ontmoeten in jullie nieuwe huis. **Yuxi**, you are such a sweet girl and your cooking is absolutely amazing. Vincent is one lucky guy to have you. **Rob**, wat gaaf dat we gewoon samen in Florida tussen de zeekoeien hebben gezwommen. De volgende keer toch maar iets meer zonnebrand op doen… ;-) **Gosia**, you crazy plant lover, thank you for the tomato plants and all the Polish snacks. **Vincent M**, bedankt voor je hulp met het transposon experiment en voor het bewaken van het ICEA geld. **Kirsten**, het leven is inderdaad soms zóóó zwaar en zóóó ontzettend moeilijk, ;-) maar met jou op de kamer werd het allemaal een heel stuk dragelijker. **Manouk**, je was een hele fijne en grappige buurvrouw. Ik vond het erg jammer dat je wegging. **Steven**, met jou in de buurt was het altijd gezellig. Niet alleen op de kamer, maar ook naast me op het lab. **Wouter** en **Kobus**, ik heb erg gelachen om jullie als komisch duo. **Mike**, je droge humor en slechte woordgrappen kon ik wel waarderen. Ook je gekraakte software kwam goed van pas. **Dennis** en **Lisanne**, ook jullie zijn een geweldig duo en waren fijne buren. **Stephanie**, ik wist niet dat er zoveel suiker opgelost kon worden in chocomelk! **Jery**, je was dan wel geen kamergenoot, maar wel een toffe bovenbuurman. Bedankt voor alle fietstochtjes naar Zeist, het watergeven van de plantjes tijdens mijn vakantie en niet te vergeten de gedownloade series. **Vincent H**, eindelijk iemand die snapt hoe cool de Group A Carbohydrate biosynthesis pathway

wel niet is! Het was niet alleen ontzettend fijn om samen met jou te werken, maar ook erg gezellig daaromheen. Ik wens je heel veel plezier en vooral succes straks in het verre China! Daarnaast wil ik alle ICEA-commissieleden die samen met mij in het bestuur hebben gezeten en al mijn opvolgers (**Anne, Nienke, Mandy, Lisette, Axel, Seline, Manouk, Tim, Wouter, Gosia, Stephanie, Vincent M, Elena, Paul, Jery** en **Elise**) en natuurlijk ook alle andere comités ontzettend bedanken voor de leuke tijd en coole activiteiten!

Marjolein, Tom en **Luc**, of we echt miljonairs worden betwijfel ik, maar wat ben ik blij dat het ons gewoon gelukt is om na onze studie nog bijna maandelijks af te spreken. Niet alleen voor de leuke uitjes en gezelligheid, maar ook dat we dan onze PhD ervaringen konden uitwisselen. De frequentie zal iets minder worden nu Tom in Zweden zit, maar ik hoop dat we onze uitjes nog vele jaren blijven doorzetten. Tom, echt ontzettend gaaf dat je speciaal voor mijn promotie terug wilt komen! Lieve **Maaike**, wat begon met samen vloeken op de bus die weer eens niet op kwam dagen, resulteerde niet alleen in een gezellige busbuddy, maar zorgde er ook nog eens voor dat ik weer ging sporten in de zomer. Ik hoop dat jij nog vele jaren mijn sportbuddy zal blijven. Ook jullie **Suzanne, Wietske, Denise** en **Marijke**, mijn allerliefste vriendinnen, wil ik nooit meer kwijt. Al 16 jaar heb ik lief en leed met jullie gedeeld. Dat is meer dan de helft van mijn leven! Dankjewel dat jullie altijd voor me klaar staan.

Lieve **papa** en **mama**, zonder jullie had ik hier natuurlijk nooit gestaan. Dankzij jullie ben ik nooit iets te kort gekomen. Bedankt dat jullie altijd voor me klaar staan. Ik hoop dat jullie trots op me zijn. **Kevin, Muriel, kleine Dax, Roos, Alexander** en **Amanda**, het is altijd fijn om jullie te zien en ik geniet er altijd van om samen te komen als gezin. Arigato! Lieve **schoonfamilie**, dank jullie wel hoe jullie mij hebben opgenomen in de familie en voor alle betrokkenheid bij mijn proefschrift.

Lieve **Matthijs**, ik weet dat je wel eens gek wordt van die eeuwige keuzestress van mij. Maar als ik ergens zeker van ben, dan is het wel dat jij en ik bij elkaar horen. Verhuizen, verbouwen, promoveren en competitie spelen tegelijk was misschien niet de allerbeste combinatie en het waren daardoor ook een paar zware maanden. Ondanks dat hebben we ons er samen doorheen geslagen. Dankjewel voor al je goede adviezen en inzichten, voor je slechte, maar stiekem best wel leuke, humor en natuurlijk voor al je liefde. Ik weet dat jij altijd voor mij klaar staat wanneer ik je nodig heb. Lieverd, ik wil je nooit meer kwijt!


7

Curriculum vitae

Samantha van der Beek was born on the 16th of January, 1990 in Dordrecht, The Netherlands. She graduated from secondary school (gymnasium) in 2008 at the RSG Hoeksche Waard in Oud-Beijerland and immediately started her Bachelor's degree studies Molecular Life Sciences at the Radboud University Nijmegen. She obtained her degree in 2011 after a short internship of 15 weeks at the department of Bio-organic Chemistry at the Radboud University Nijmegen where she synthesized and optimized enzyme containing polymersomes under supervision of Prof. Dr. Jan van Hest and Dr. Ruud Peters. She continued her studies with a Master Molecular Life Sciences at the same university. During her first internship of 8 months at the department of Pediatric Infectious Diseases at the Radboud University Medical Center in Nijmegen she functionally characterized a virulence-associated bacterial protein in *Streptococcus pneumoniae* under the supervision of Dr. Marien de Jonge. Seeking a new challenge, she traveled to San Diego (CA) in the USA to perform her second internship of 6 months at the Skaggs School of Pharmacy at the University of California San Diego in the Handel lab. Here she studied the chemokine receptor CCR1 under supervision of Prof. Dr. Tracy Handel and Dr. Taylor Gilliland. In 2013 she obtained her Master of Science degree and worked for 2 months as a research technician at the department of Medical Microbiology at the University Medical Center Utrecht in the immune evasion group under supervision of Prof. Dr. Jos van Strijp and Dr. Kok van Kessel. Following this, she obtained a PhD position at the same department in the bacterial glycobiology group under supervision of Prof. Dr. Jos van Strijp and Dr. Nina van Sorge. Her research focused on the characterization of rhamnose-containing cell wall polysaccharides in streptococci and the development of new antimicrobial agents. The results of her PhD research are presented in this thesis and published in peer-reviewed scientific journals.

List of publications

Related to this thesis

van der Beek, S. L., Le Breton, Y., Ferenbach, A. T., Chapman, R. N., van Aalten, D. M., Navratilova, I., Boons, G. J., McIver, K. S., van Sorge, N. M. and Dorfmueller, H. C.; *GacA is essential for Group A Streptococcus and defines a new class of monomeric dTDP-4-dehydrorhamnose reductases (RmlD)*. (2015) Mol Microbiol, **98**(5): p. 946-962.

van der Beek, S. L., Zorzoli, A., Çanak, E., Chapman, R. N., Meyer, B. H., Boons, G. J., Dorfmueller, H. C. and van Sorge, N. M.; *Identification of lead compounds targeting the dTDP-L-rhamnose biosynthesis pathway using Streptococcus pyogenes.* (2018) BioRxiv; under revision for publication in Mol Microbiol.

Other publications

Kuipers, K., Gallay, C., Martinek, V., Rohde, M., Martinkova, M., van der Beek, S. L., Jong, W. S., Venselaar, H., Zomer, A., Bootsma, H., Veening, J. W. and de Jonge, M. I.; *Highly conserved nucleotide phosphatase essential for membrane lipid homeostasis in Streptococcus pneumoniae*. (2016) Mol Microbiol, **101**(1): p. 12-26.

van den Kieboom, C. H., van der Beek, S. L., Meszaros, T., Gyurcsanyi, R. E., Ferwerda, G. and de Jonge, M. I.; *Aptasensors for viral diagnostics*. (2015) Trac-Trends in Analytical Chemistry, **74**: p. 58-67.