

Structure and mechanism of biosynthesis of *Streptococcus mutans* cell wall polysaccharide

Jeffrey S. Rush¹, Svetlana Zamakhaeva², Nicholas R. Murner², Pan Deng^{3, 6}, Andrew J. Morris^{3, 7}, Cameron W. Kenner², Ian Black⁴, Christian Heiss⁴, Parastoo Azadi⁴, Konstantin V. Korotkov¹, Göran Widmalm⁵, and Natalia Korotkova^{2, 1*}

¹Department of Molecular and Cellular Biochemistry, University of Kentucky, Lexington, Kentucky, USA

²Department of Microbiology, Immunology and Molecular Genetics, University of Kentucky, Lexington, Kentucky, USA

³Division of Cardiovascular Medicine and the Gill Heart Institute, University of Kentucky, Lexington, Kentucky, USA

⁴Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia, USA

⁵Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, Stockholm, Sweden

⁶Present address: Jiangsu Key Laboratory of Neuropsychiatric Diseases and College of Pharmaceutical Sciences, Soochow University, Suzhou, Jiangsu, China

⁷Present address: Department of Pharmacology and Toxicology, University of Arkansas for Medical Science and Central Arkansas Veterans Affairs Healthcare System, Little Rock, Arkansas, USA

*Correspondence and request for materials should be addressed to N.K. (email: nkorotkova@uky.edu)

Short title: Mechanism of SCC glucosylation in *Streptococcus mutans*

Keywords: *Streptococcus mutans*, SCC, glycosyltransferase, cell wall, peptidoglycan, rhamnose polysaccharide, glucose side-chain, glycerol phosphate, biofilm

Abstract

Streptococcus mutans, the causative agent of human dental caries, expresses a cell wall attached Serotype c-specific Carbohydrate (SCC) that is critical for cell viability. SCC consists of a repeating $\rightarrow 3)\alpha\text{-Rha}(1\rightarrow 2)\alpha\text{-Rha}(1\rightarrow$ polyrhamnose backbone, with glucose (Glc) side-chains and glycerol phosphate (GroP) decorations. This study reveals that SCC has one major and two minor Glc modifications. The major Glc modification, $\alpha\text{-Glc}$, attached to position 2 of 3-rhamnose, is installed by SccN and SccM glycosyltransferases and is the site of the GroP addition. The minor Glc modifications are $\beta\text{-Glc}$ linked to position 4 of 3-rhamnose installed by SccP and SccQ glycosyltransferases, and $\alpha\text{-Glc}$ attached to position 4 of 2- rhamnose installed by SccN working in tandem with an unknown enzyme. Both the major and the minor $\beta\text{-Glc}$ modifications control bacterial morphology, but only the GroP and major Glc modifications are critical for biofilm formation.

Introduction

The Gram-positive bacterium *Streptococcus mutans* is recognized as a major etiological agent of human dental caries. This organism colonizes tooth surfaces forming dental plaque, one of the most robust biofilms in nature. Similar to other streptococcal species, *S. mutans* strains decorate peptidoglycan with rhamnose (Rha)-containing cell wall polysaccharides, so called rhamnopolysaccharides, that are important for the viability of these bacteria¹. The polysaccharides are functional homologs of teichoic acid glycopolymers that are present in the cell walls of many Gram-positive bacteria^{2, 3, 4}. Streptococcal rhamnopolysaccharides have attracted significant attention as promising candidates for the development of glycoconjugate vaccines^{5, 6, 7, 8, 9, 10, 11, 12} and as receptors for bacteriophage-encoded endolysins^{13, 14}, a novel class of antimicrobials that kill bacteria by degrading peptidoglycan^{15, 16}. Furthermore, enzymatic steps in rhamnopolysaccharide synthesis are considered to be novel targets for antibacterial drug design^{17, 18}.

S. mutans strains are classified into four serotypes, *c*, *e*, *f* and *k*, based on variations in the molecular structures of their rhamnopolysaccharides¹⁹. Epidemiological surveys indicate that 70–80% of strains found in the oral cavity are

classified as serotype *c*, followed by serotypes *e* (about 20%) and *f* and *k* (less than 5% each)^{19, 20}. The major structural feature of all *S. mutans* rhamnopolysaccharides is a linear backbone structure, composed of repeating $\rightarrow 3)\alpha\text{-Rha}(1\rightarrow 2)\alpha\text{-Rha}(1\rightarrow$ disaccharides. In *S. mutans* serotype *c*, the rhamnopolysaccharide, called the serotype *c*-specific carbohydrate (SCC), is reported to contain α -glucose (Glc) side-chains attached to the 2-position of 3-linked Rha units⁷. Alternatively, serotype *e* has β -Glc attached to the corresponding hydroxyl group of Rha²¹. Serotype *f* contains α -Glc or (Glc)₂ side-chain attached to the 3-position of 2-linked Rha and serotype *k* has an α -galactose side-chains attached to the corresponding position on Rha⁷. In *S. pyogenes*, the homologous polysaccharide called the Group A Carbohydrate (GAC) displays a similar core polyrhamnose backbone structure modified with *N*-acetyl- β -glucosamine (GlcNAc) side-chains²². Our recent detailed compositional and structural analysis of GAC revealed that approximately 25% of the GlcNAc side-chains are decorated with glycerol phosphate (GroP) moieties at the C6 hydroxyl group³. This modification had gone unnoticed in the past decades, presumably due to loss of GroP during preparation. The presence of GroP was also detected on the Glc side-chains of *S. mutans* SCC, although the exact structure of the GroP-Glc linkage has not yet been reported^{3, 4}.

This study was undertaken to investigate the SCC structure and poorly understood steps of SCC biosynthesis. Most genes responsible for rhamnopolysaccharide biosynthesis in *S. mutans* are arranged in a single cluster that can be functionally divided into two regions (Fig. 1a). Region 1 is conserved among all *S. mutans* serotypes and many streptococcal species, and encodes the proteins for polyrhamnose biosynthesis and export^{1, 6, 23}. Region 2 is species and serotype-specific, and encodes proteins for side-chain attachment and GroP modification^{3, 6, 23}. Assembly of the polyrhamnose backbone is initiated on the cytoplasmic side of the plasma membrane by the transfer of GlcNAc-phosphate from UDP-GlcNAc to a carrier lipid, undecaprenyl-phosphate (Und-P), catalyzed by the TarO/TagO/WecA homolog²⁴, known as RpgG in *S. mutans*²⁵, followed by sequential addition of L-Rha units catalyzed by SccB, SccC and SccG^{26, 27}. The lipid-linked polyrhamnose intermediate is

then transported into the periplasm by an ABC transporter²⁷ where it is further decorated with the side-chains and GroP^{3,4} (Fig. 1b).

The molecular details of Glc side-chain addition to SCC have not yet been described in detail. We previously reported that both SccN and SccP participate in decoration of SCC with the Glc side-chains, although SccN is required for the addition of the majority of the Glc moieties⁴. Moreover, only the Glc side-chains provided by SccN are recognized by the GroP transferase, SccH, for the attachment of GroP moieties⁴. The Glc side-chains and GroP decorations are critically important in the biology and pathogenesis of *S. mutans*. The SccN and SccH-deficient mutants exhibit self-aggregation, aberrant cell morphology and increased autolysis⁴. These defects are due to mislocalization of the cell division proteins in these mutants. Furthermore, the SccN-deficient mutant shows a defect in biofilm formation and exopolysaccharide production, and attenuated virulence in a rat model of dental caries²⁸.

Here, we show that SccN and SccP are glucosyl phosphoryl undecaprenol (Glc-P-Und) synthases that catalyze the formation of β -Glc-P-Und and α -Glc-P-Und, respectively (Fig. 1b). SccM and SccQ are Glc-P-Und:polyrhamnose glucosyltransferases that transfer Glc from Glc-P-Und to the polyrhamnose chain to form the Glc side-chains (Fig. 1b). Glycosyl linkage analysis together with NMR analysis indicate that SccN and SccM are responsible for the formation of the most abundant α -Glc modification, whereas SccP and SccQ contribute to the formation of the minor β -Glc modification. Furthermore, the GroP decoration is critically important for *S. mutans* biofilm formation. It is found exclusively on the 6-OH of the α -Glc side-chains donated by SccN and SccM.

Results

SccN and SccP are UDP-Glc:Und-P glucosyl transferases

In bacteria, multi-component transmembrane glycosylation complexes catalyze the attachment of sugars to a variety of glycans. This machinery requires a GT-A fold glycosyltransferase to synthesize an Und-P linked glycosyl donor, a flippase to move the lipid intermediate across the cell membrane and a periplasmically-oriented GT-C fold glycosyltransferase to transfer the sugar from the lipid intermediate to the nascent

polysaccharide^{29, 30}. In *S. mutans* serotype *c*, Region 2 encodes two GT-A fold glycosyltransferases SccN and SccP, and two GT-C fold glycosyltransferases SccM and SccQ (Fig. 1a)⁴. We hypothesized that the Glc side-chains of SCC in *S. mutans* are donated to the polyrhamnose backbone from an undecaprenol-linked intermediate Glc-P-Und as has been reported for an analogous biosynthetic pathway in *S. pyogenes*³⁰. To test this idea, *sccN* and *sccP* were expressed on a plasmid in the *E. coli* JW2347 strain, which is Glc-P-Und synthase deficient. In membrane preparations of the recombinant strains, both synthases actively catalyzed the formation of a [³H]glucolipid when incubated with UDP-[³H]Glc and Und-P. In contrast, the activity was negligible in membrane fractions of *E. coli* JW2347 (Supplementary Table 1). A kinetic analysis of *E. coli* expressed SccN and SccP revealed that both enzymes show a high affinity for UDP-Glc and Und-P, whereas SccP possesses a slightly lower affinity for UDP-Glc, but higher affinity for Und-P than SccN. However, the apparent maximal rates for SccN under these reaction conditions are higher than that of SccP (Fig. 2a and Table 1). These data confirm that SccN and SccP are both Glc-P-Und synthases. To understand if SccN and SccP are active in *S. mutans*, the membrane fractions prepared from *S. mutans* Xc wild type (WT), the *sccN* and *sccP* deletion mutants ($\Delta sccN$ and $\Delta sccP$), and the double deletion mutant $\Delta sccN\Delta sccP$ were incubated with UDP-[³H]Glc and Und-P, *in vitro*. *S. mutans* membrane fractions synthesized two classes of glucolipids: the major glucolipid product with properties consistent with Glc-P-Und and an additional minor glucolipid which is most likely a glucosyldiglyceride. Subsequently, to eliminate the glycerolipid contaminant, organic extracts of *in vitro* reactions were subjected to mild alkaline de-acylation prior to analysis for [³H]-Glc-P-Und. The synthase activity in the $\Delta sccN$ membranes was dramatically reduced compared to WT cells (Supplementary Fig. 1a and Fig. 2b). In contrast, deletion of SccP had a modest effect on Glc-P-Und synthase activity. Membrane fractions from $\Delta sccN\Delta sccP$ did not catalyze the synthesis of Glc-P-Und. The [³H]glucolipid products formed in the reactions containing the $\Delta sccN$ and $\Delta sccP$ membranes were not distinguishable by thin layer chromatography (TLC) to that formed in the WT reactions (Fig. 2b). The activities were stimulated by the exogenous addition of Und-P as a suspension in CHAPS detergent (Supplementary Fig. 1a), and inhibited by amphotericin which is known to form an insoluble complex with

Und-P (Supplementary Fig. 1b, c)³¹. Further analysis showed that the [³H]glucolipids are anionic, sensitive to mild acid and resistant to mild alkali as expected for a glycosyl phosphoryl isoprenoid. In addition, a co-migrating compound, purified from the *S. mutans* membrane fraction by organic solvent extraction and preparative TLC, yielded a molecular ion, $m/z=1007.7$ by ESI-MS which is expected for Glc-P-Und (Supplementary Fig. 1d). Thus, our data indicate that both SccN and SccP catalyze the synthesis of Glc-P-Und.

SccN and SccP synthesize β -Glc-P-Und and α -Glc-P-Und, respectively

Although the enzymatic products formed by SccN and SccP from UDP-Glc and Und-P were chromatographically indistinguishable, and both glycosyltransferases are active in *S. mutans* membranes, only the Glc side-chains provided by SccN are decorated with GroP⁴. Glycosyltransferases producing products with a sugar 1-phosphate (like Glc-P-Und), are known to be readily reversible in the presence of the appropriate nucleoside diphosphate, re-synthesizing the pertinent NDP-sugar^{32, 33}. To investigate if the enzymatic products of SccN and SccP are interchangeably equivalent substrates in the reverse direction to reform UDP-Glc, we tested the purified products of each reaction for reverse synthesis by incubation with UDP in the presence of each of the enzymes. Interestingly, the enzymatic product of the SccN Glc-P-Und synthase reaction is efficiently discharged in a time-dependent reaction by SccN, but not by SccP (Fig. 2c and Supplementary Fig. 2). The enzymatic product formed by SccP is likewise discharged by SccP, but not by SccN (Fig. 2c and Supplementary Fig. 2). This experiment indicates that the Glc-P-undecaprenols synthesized by the SccN and SccP are most likely unique stereoisomers.

To identify the anomeric configuration of the Glc-P-Und lipids produced by *S. mutans*, and the roles of the glycosyltransferases encoded in Region 2 of the SCC gene operon in their synthesis, we constructed mutants deficient in *sccM* and *sccQ* ($\Delta sccM$ and $\Delta sccQ$). Preparative scale quantities of Glc-P-Und from the WT, $\Delta sccN$, $\Delta sccP$, $\Delta sccM$ and $\Delta sccQ$ cells were extracted in organic solvent, purified by multiple rounds of preparative TLC and analyzed by ESI-MS liquid chromatography. The mass spectra obtained by ESI-MS from each of the purified Glc-P-Und isolates contained high

mass ions derived from $[M-H]^-$ ($m/z=1007.7$), $[M-H_2O-H]^-$ ($m/z=988.7$), $[Und-PO_4-C_2H_3O]^-$ ($m/z=887.7$), $[Und-HPO_4]^-$ ($m/z=845.7$) and $[glucosyl-1-PO_3]^-$ ($m/z=241.7$), as expected (Supplementary Fig. 3). During fragmentation of sugar 1-phosphates by ESI-MS, if the phosphodiester at C1 and the hydroxyl group at C2 are oriented *trans* to the sugar ring then $[Und-HPO_4]^-$ ($m/z=845.7$) is released as the highest molecular weight fragment (Fig. 2d). However, if the relative orientation of the groups at C1 and C2 is arranged *cis*, then an additional cross-ring fragmentation product, $[Und-PO_4-C_2H_3O]^-$ ($m/z=887.7$) is released (Fig. 2d). Wolucka, et al. have reported that the ratio of the relative abundance of the $[Und-PO_4-C_2H_3O]^-$ ion to the relative abundance of the $[Und-HPO_4]^-$ ion is close to one for sugar 1-phosphates with C1-O and C2-OH in *cis* configuration, but close to zero for sugar 1-phosphates with a *trans* orientation of these hydroxyl groups³⁴. Importantly, the spectra derived from the $\Delta sccN$ glucolipid compared with the spectra of the $\Delta sccP$ glucolipid revealed dramatic differences in the relative ion abundances of the fragments arising from loss of $[Und-HPO_4]^-$ ($m/z=845.7$) and the cross-ring fragmentation pathway ($m/z=887.7$). Table 2 shows that the ratio of the $[Und-PO_4-C_2H_3O]^-$ ($m/z=887.7$) fragment to $[Und-HPO_4]^-$ ($m/z=845.7$) is approximately 0.14 in the Glc-P-Und isolated from the $\Delta sccP$ strain (indicating that the C1- and C2-groups are in *trans* orientation), whereas this ratio is approximately 1.02 in the Glc-P-Und isolated from the $\Delta sccN$ strain (C1- and C2-groups oriented *cis*). This result is consistent with the conclusion that the $\Delta sccN$ strain (expressing SccP) synthesizes an α -Glc-P-Und and that $\Delta sccP$ (expressing SccN) synthesizes a β -Glc-P-Und. The WT strain shows a relative abundance ratio of $[Und-PO_4-C_2H_3O]^-$ to $[Und-HPO_4]^-$ of 0.49, indicating that a mixture of the two stereoisomers is most likely present. Furthermore, it is reassuring that the spectra of the glucolipids from the $\Delta sccM$ and $\Delta sccQ$ strains, in which SccN and SccP expression is unaffected, give relative ion abundance ratios that are similar to that of the WT strain.

The functions of SccN, SccP, SccM and SccQ in SCC biosynthesis

To establish the exact roles of the SccN, SccP, SccM and SccQ glycosyltransferases in SCC glucosylation, we purified sodium borohydride reduced SCCs from the WT, $\Delta sccN$, $\Delta sccP$, $\Delta sccM$ and $\Delta sccQ$ strains, and analyzed them for glycosyl composition as TMS-

methyl glycosides following methanolysis. The analysis found primarily Rha and various amounts of Glc, along with *N*-acetyl-D-glucosaminitol (GlcNAcitol) derived from the reducing end GlcNAc of SCC³⁵ and lesser amounts of terminal, non-reducing end GlcNAc (Supplementary Fig. 4). Deletion of *sccN* dramatically reduces the Glc content of SCC, whereas deletion of *sccP* has a relatively minor affect, in agreement with our previous study⁴. It is worth noting that the reduction in Glc content observed in the $\Delta sccP$ SCC, is approximately the same as the residual Glc content found in the $\Delta sccN$ SCC, and the Glc content in the $\Delta sccN\Delta sccP$ SCC is reduced even further⁴.

Interestingly, the Glc contents of the SCCs isolated from $\Delta sccM$ and $\Delta sccQ$ are each reduced to approximately 1/3rd of the WT content. These surprising results suggest that both SccM and SccQ can contribute similarly to glucosylation of SCC.

To determine which Glc-P-Und stereoisomers are utilized by SccM and SccQ for the transfer of Glc to SCC and to identify the specific glycosidic products produced by these enzymes, the SCCs were subjected to glycosyl linkage analysis as partially methylated alditol acetates (PMAA). As expected from the glycosyl composition analysis (see above), linkage analysis identified Glc as the major non-reducing terminal sugar and minor amounts of non-reducing terminal Rha, GlcNAc and 4-substituted GlcNAcitol. The PMAA arising from 2-substituted Rha was the main component in all samples, but the quantity of 3-linked Rha was highly variable between the samples. Specifically, we found a major reduction in 2,3-disubstituted Rha and a dramatic concomitant increase in 3-linked Rha in the $\Delta sccN$ and $\Delta sccM$ SCCs (Fig. 3a and c and Supplementary Table 2). This is consistent with the function of these proteins in the transfer of Glc to the 2-OH of 3-substituted Rha. Surprisingly, a minor amount of 2,4-, 3,4- and 2,3,4-branched Rha was detected in some samples, suggesting that *S. mutans* possesses previously unreported 4-glucosyltransferase activities that can transfer Glc to either 2-linked, 3-linked, or 2,3-disubstituted Rha (Fig. 3a, b, c and d, and Supplementary Table 2).

Consistent with the glycosyl composition of the $\Delta sccN$ SCC, linkage analysis revealed a small residual amount of terminal Glc in this polysaccharide (Fig. 3e, Supplementary Table 2), which correlates very closely with the amount of 3,4-branched Rha. It is important to note that this Glc addition to the 4-position does not require SccN. The 2,3- and 2,3,4-substituted Rha branches are nearly absent in $\Delta sccN$ suggesting

that the Glc addition to the 2-position of these branches requires SccN. In contrast, the 3,4-Rha branch is significantly enhanced when *sccN* is deleted (Fig. 3a, Supplementary Table 2). Hence, these observations suggest that the Glc attached to the 4-OH of 3,4-Rha is derived from the SccP pathway and the Glc attached to the 2-OH of the 2,3,4-branched Rha originates from the SccN product.

The signal for 3,4-disubstituted Rha is also increased in $\Delta sccM$, at the expense of the 2,3,4-branched Rha component, and nearly absent in $\Delta sccQ$, confirming the role of SccM in glucosylation of the 2-OH and strongly suggesting that the formation of the branch at the 4-position requires SccQ in addition to SccP (Fig. 3c, Supplementary Table 2). It is important to note that there is a very small signal for 3,4-Rha in the WT SCC. The increase of this branch in the $\Delta sccN$ and $\Delta sccM$ SCCs may be a consequence of the deletion of *sccN* and *sccM*, respectively, allowing enhanced expression of the SccP/SccQ pathway. Lastly, of interest is the effect of these genetic deletions on the abundance of the 2,4-Rha branch. As we will discuss below, high resolution NMR analysis indicates that this side-branch is highly unusual in that it is formed by addition of Glc to the 4-position of 2-Rha, unlike the other branching residues. The 2,4-Rha signal is significantly reduced in $\Delta sccN$ and increased in $\Delta sccP$, suggesting that the SccN product is required for this modification. 2,4-Rha increases dramatically in the $\Delta sccM$ SCC (probably due to increased availability of the SccN Glc-P-Und product), but does not change significantly in the $\Delta sccQ$ SCC implying that neither of these GT-C glycosyltransferases are required to glucosylate the 4-OH of 2-Rha (Fig. 3a, b, c and d, Supplementary Table 2). Taken together, our data indicate that SccN and SccM are the major enzyme pair that mediate glucosylation of the 2-OH of the 3-substituted Rha. Furthermore, SccQ may utilize the SccP product, α -Glc-P-Und, to transfer Glc to the 4-position of 3-substituted and 2,3-disubstituted Rha to form the 3,4- and 2,3,4-Rha branches. It appears that the product of the SccN-mediated reaction, β -Glc-P-Und, may be utilized by an unknown glycosyltransferase to transfer Glc to the 4-position of 2-Rha to form the 2,4-branched Rha.

The Glc residues installed by SccN-SccM pair serve as GroP acceptors

To understand the roles of SccM and SccQ in the formation of the GroP acceptor side-chains of SCC, we compared the phosphate content of purified SCCs isolated from WT, $\Delta sccM$, $\Delta sccQ$, $\Delta sccN$ and $\Delta sccP$, and five double mutants, $\Delta sccN\Delta sccP$, $\Delta sccM\Delta sccQ$, $\Delta sccN\Delta sccQ$, $\Delta sccM\Delta sccN$ and $\Delta sccM\Delta sccP$. The $\Delta sccN$, $\Delta sccN\Delta sccP$, $\Delta sccN\Delta sccQ$, and $\Delta sccM\Delta sccN$ SCCs showed a dramatically lower phosphate content, consistent with our earlier report that the $\Delta sccN$ mutant is GroP-deficient^{3, 4}. Importantly, the phosphate content in the $\Delta sccM$, $\Delta sccM\Delta sccQ$ and $\Delta sccM\Delta sccP$ SCCs was also very low, similar to that of the SCCs from strains deleted for *sccN* (Fig. 3f), suggesting that SccM also participates in the formation of the GroP acceptor site. Interestingly, the $\Delta sccQ$ SCCs contain somewhat reduced levels of phosphate (Fig. 3f).

Additionally, we conjugated the isolated SCC variants with a fluorescent dye, 7-amino-1,3-naphthalenedisulfonic acid (ANDS) by reductive amination and examined the electrophoretic mobilities of ANDS-SCCs. The dye ANDS inserts a negative charge to SCCs providing electrophoretic mobility to neutral polysaccharides lacking the GroP moieties as we described in our earlier study⁴. Electrophoresis of ANDS-SCCs isolated from WT and $\Delta sccP$ revealed a distinctive “laddering” of SCC bands indicating a high content of negatively charged residues and heterogeneity in charge density. These data are consistent with the presence of GroP modification (Fig. 3g). In agreement with our previously reported analysis of ANDS-SCCs isolated from $\Delta sccN$ and $\Delta sccN\Delta sccP$ ⁴, these GroP-deficient mutants migrated as a single band. The $\Delta sccM$, $\Delta sccM\Delta sccQ$, $\Delta sccN\Delta sccQ$, $\Delta sccM\Delta sccN$ and $\Delta sccM\Delta sccP$ SCCs also migrated as a single band, but the $\Delta sccQ$ SCCs showed the characteristic “laddering” pattern similar to the WT and $\Delta sccP$ SCCs (Fig. 3g). Thus, these data indicate that only the Glc side-chains provided by SccM and SccN serve as the acceptors for GroP.

GroP is linked to position 6 on α -Glc located at position 2 on the 3-linked Rha

A previous analysis of SCC isolated from *S. mutans* after release by autoclaving was conducted by ¹H and ¹³C NMR spectroscopy revealing two polymers, a minor one containing the branched trisaccharide repeating unit $\rightarrow 2)-\alpha\text{-L-Rhap-(1}\rightarrow 3)[\alpha\text{-D-Glcp-(1}\rightarrow 2)]-\alpha\text{-L-Rhap-(1}\rightarrow$ and a major one containing the linear rhamnan backbone with

the structure $\rightarrow 2)\text{-}\alpha\text{-L-Rhap-(1}\rightarrow 3)\text{-}\alpha\text{-L-Rhap-(1}\rightarrow 7$. Significantly, this reported analysis did not detect the GroP modifications or the Glc side-chains linked to the 4-position of branched Rha residues. To investigate the site of attachment of GroP to SCC and characterize the chemical nature of the newly discovered 4-Rha glucosyl modifications, we purified undegraded SCCs from the cell walls of WT, $\Delta sccM$, $\Delta sccQ$, $\Delta sccN$ and $\Delta sccP$, and two double mutants, $\Delta sccN\Delta sccP$, $\Delta sccM\Delta sccQ$ followed by reduction with sodium borohydride, as we described in our earlier study³⁵, and analyzed them by 1D and 2D NMR spectroscopy. The anomeric region of the ^1H NMR spectrum of the WT SCC was similar to previously reported rhamnan polysaccharides⁷. The 2D ^1H , ^{13}C -HSQC NMR spectrum revealed, inter alia, a characteristic cross-peak at $\delta_{\text{H}}/\delta_{\text{C}}$ 5.06/98.27 originating from the anomeric atoms of the $\alpha\text{-(1}\rightarrow 2)\text{-linked}$ glucosyl group as a side-chain. A ^{31}P NMR spectrum of the WT SCC showed a resonance at 1.1 ppm, indicative of a phosphodiester linkage³⁶. This was also the case for the $\Delta sccP$ and $\Delta sccQ$ SCCs, whereas ^{31}P NMR signals were absent in spectra from the $\Delta sccM$, $\Delta sccN$ and $\Delta sccM\Delta sccQ$ SCCs. Additional information that the WT, $\Delta sccP$ and $\Delta sccQ$ SCCs contained a distinct ^{31}P NMR resonance came from 2D ^1H , ^{31}P -HMBC NMR experiments that showed heteronuclear correlations, presumably over three bonds, between the ^{31}P resonance at 1.1 ppm and protons resonating at ~ 4.23 , ~ 4.08 , ~ 3.96 and ~ 3.89 ppm (Fig. 4a), consistent with a phosphodiester linkage. In ^{13}C NMR spectra, the resonances at δ_{C} 71.6 and 67.3 were doublets with J_{CP} of 7.5 and 5.6 Hz, respectively, supporting the findings of a phosphodiester linked substituent. Thus, NMR analysis agrees with phosphate analysis indicating that a phosphodiester linkage requires the Glc side-chain provided by SccN and SccM.

Further analysis of ^1H , ^{13}C -HSQC and ^1H , ^{13}C -HMBC NMR spectra of the WT SCC identified a glycerol residue with NMR chemical shifts δ_{H1} 3.95 and 3.89, δ_{C1} 67.3, δ_{H2} 3.92, δ_{C2} 71.6, δ_{H3} 3.69 and 3.62, and δ_{C3} 63.0, similar to those found in GAC³. In the ^1H , ^{13}C -HSQC spectrum, cross-peaks from a methylene group were observed at $\delta_{\text{H}}/\delta_{\text{C}}$ 4.22/64.8 and 4.07/64.8, i.e., at the ^1H NMR chemical shifts identified by correlations in the ^1H , ^{31}P -HMBC spectrum. From a ^1H , ^{13}C -HSQC-TOCSY NMR spectrum with a mixing time of 200 ms, a complete spin-system of seven resonances could be identified from δ_{C} 64.8, viz., at 5.05, 4.22, 4.07, 3.94, 3.77, 3.64 and 3.57 ppm

(Fig. 4b), on the one hand, and from δ_{H1} 5.06 at 98.3, 73.4, 73.0, 72.1, 70.3 and 61.4 ppm, on the other. Thus, the former spin-system identifies the glycerol residue as being linked to position 6 of the Glc side-chain residue and the latter as the non-substituted Glc side-chain, where the correlations in the latter spin-system are in agreement with the earlier analysis of SCC⁷ as well as an NMR chemical shift prediction by CASPER³⁷ of the trisaccharide repeating unit of the WT SCC. Furthermore, the downfield NMR chemical shift displacement at C6 from 61.4 to 64.8 ppm is consistent with phosphorylation at this position³⁸. The trisaccharide repeating unit of the WT SCC was further supported by correlations in the $^1H, ^{13}C$ -HMBC NMR spectrum, viz., between δ_C 98.3 and δ_H 4.18, δ_H 5.06 and δ_C 76.6, 73.4, 73.0, as well as between δ_H 5.32 and δ_C 75.4⁷. Small chemical shift displacements were present between residues in the most abundant trisaccharide repeating unit of the WT SCC, lacking GroP, and the minor repeat unit, containing GroP at position 6 of Glc. These can be observed, for example, in the $^1H, ^1H$ -NOESY NMR spectrum with a mixing time of 250 ms where correlations were detected from the anomeric proton of the glucosyl residue at 5.06 ppm to 4.18 and 3.54 ppm (major repeat unit), whereas from 5.05 ppm correlations were observed to 4.16 and 3.57 ppm (minor repeat unit). Furthermore, from the anomeric proton of the branched Rha residue at 5.12 ppm, an intra-residual NOE correlation to H2 at 4.18 ppm (major) was detected, whereas from 5.10 ppm the corresponding correlation was observed to 4.16 ppm (Fig. 4c). Thus, the Glc side-chain residue is partially substituted at position 6 by a GroP residue, which results in a repeating unit having the structure $\rightarrow 2)\text{-}\alpha\text{-L-Rhap-(1}\rightarrow 3)[\alpha\text{-D-Glcp6P(S)Gro-(1}\rightarrow 2)]\text{-}\alpha\text{-L-Rhap-(1}\rightarrow$ (Fig. 4d), the absolute configuration of the substituent is defined by *sn*-Gro-1-*P* (vide supra), as found in the GAC³.

Minor glucosyl substitutions of SCC repeat unit

Our glycosyl linkage analysis of the WT SCC revealed the presence of minor alternative Rha branch points formed by the addition of Glc to the 4-position of Rha. These novel, unexpected modifications could be attributed to the expression of specific glycosyltransferases encoded by the SCC genetic locus. To confirm the presence of these branching residues and to investigate the chemical nature of these minor

modifications, high resolution 2D ^1H , ^{13}C -HSQC NMR analysis was conducted using highly-enriched, borohydride-reduced, SCCs from strains deleted for *ScsN* and *ScsM*, which mediate the synthesis of the dominant $\rightarrow 2$)- α -L-Rhap-(1 \rightarrow 3)[α -D-Glcp-(1 \rightarrow 2)]- α -L-Rhap-(1 \rightarrow repeat unit. NMR analysis of the $\Delta scsN$ SCC confirmed that the major form of the polymer is a polysaccharide with a linear repeating $\rightarrow 2$)- α -L-Rhap-(1 \rightarrow 3)- α -L-Rhap-(1 \rightarrow backbone. Importantly, NMR spectral analysis of the $\Delta scsM$ SCC revealed that a β -linked Glc residue is present (Fig. 4e) as a side-chain substituting the 3-linked Rha backbone residue at position 4, cf. NOE from H1 in **C** to H4 in **A** (Fig. 4f, Table 3, non-primed residues). Interestingly, an additional NOE between H1 in **C** and H2 in **B** of the 2-linked Rha is consistent with a three-dimensional oligosaccharide model (Supplementary Fig. 5) of this structural region of the polysaccharide, further corroborating the proposed trisaccharide structure as a constituent of the polysaccharide (Fig. 4g). Moreover, the ^{13}C NMR chemical shifts for C3 and C4 of the 3,4-disubstituted residue **A** are in excellent agreement with those of a corresponding O-methyl rhamnosyl residue substituted in position 3 by α -L-Rhap and in position 4 by β -D-Glcp³⁹, like in the $\Delta scsN$ SCC. In summary, 2D NMR analysis of the $\Delta scsN$ and $\Delta scsM$ SCC variants confirm and establish the presence of β -Glc at the 4-position of 3-Rha in the polymer. Although it is not possible to directly determine the chemical nature of the 4-Glc substituent of the 2,3,4-substituted Rha (since the 2-substituent is not present in either the $\Delta scsN$ or $\Delta scsM$ SCCs), it has very likely a β -anomeric configuration (Fig. 4g).

Because glycosyl linkage analysis of the $\Delta scsM$ SCC variant showed a prominent 2,4-Rha branch (Fig. 3c), 2D NMR analysis was employed to characterize this unexpected branched Rha in the $\Delta scsM$ and $\Delta scsM\Delta scsQ$ SCCs. In the ^1H , ^{13}C -HSQC spectrum of the $\Delta scsM\Delta scsQ$ SCC, the spectral region for anomeric resonances showed cross-peaks, inter alia, at $\delta_{\text{H}}/\delta_{\text{C}}$ 5.20/101.6 and 4.98/102.7 (major) and, inter alia, at $\delta_{\text{H}}/\delta_{\text{C}}$ 5.21/101.2 and 5.06/100.3 (minor), similar to those in the $\Delta scsM$ SCC for which the NMR resonance assignments of the linear rhamnan backbone (Table 3, primed residues) were in good agreement with those reported in the literature⁷. Notably, the minor form was shown to originate from a trisaccharide-containing repeating unit (RU) in which the rhamnan backbone carries an α -Glc residue that substitutes the position 4 of the 2-linked Rha in the backbone of the polysaccharide

(Fig. 4h, Table 3, double primed residues). The assignments of the NMR resonances from this trisaccharide RU are in excellent agreement with the ^1H and ^{13}C NMR chemical shifts predicted by the CASPER program ³⁷ (Supplementary Fig. 6) substantiating the proposed structure (Fig. 4i).

Further analysis of NMR spectra from the WT SCC, aided by spectral assignments obtained from the ΔsccM SCC (Table 3, double primed residues), revealed key resonances such as those traced from H6 (1.42 ppm) in the side-chain substituted backbone residue $\rightarrow 2,4)\text{-}\alpha\text{-L-Rhap-(1}\rightarrow 3)$. In addition, similar connectivities were observed from a peak at 1.44 ppm in the ^1H NMR spectrum, with correlations to 17.8 ppm in the $^1\text{H}, ^{13}\text{C}$ -HSQC spectrum, to 69.1 ppm in the $^1\text{H}, ^{13}\text{C}$ -H2BC spectrum, and to 82.7 ppm in the $^1\text{H}, ^{13}\text{C}$ -HMBC spectrum. The latter ^{13}C resonance also showed a correlation to a ^1H resonance at 5.03 ppm in the same spectrum. These results indicate that adjacent to the trisaccharide structural element (Fig. 4i), the neighboring RUs of the rhamnan may also, besides being non-substituted, be substituted by the 2-linked Glc residue and/or substituted by the $\alpha\text{-D-Glcp6P(S)Gro}$ entity. Among the analyzed mutants, the NMR spectra of the ΔsccP and ΔsccQ SCCs were similar to those of the WT SCC. However, in the ΔsccQ SCC, the linear rhamnan backbone polysaccharide was the major form, and the branched trisaccharide RU was the minor form. Importantly, the ^1H NMR spectrum analysis did not detect the presence of the Glc side-chains in the $\Delta\text{sccN}\Delta\text{sccP}$ SCC, indicating the function of SccN in providing Glc for the transfer to the 4-position of 2-Rha.

Lastly, a glycosyl linkage analysis of the $\Delta\text{sccM}\Delta\text{sccQ}$ SCC revealed that this polysaccharide contains a significant increase in the 2,4-Rha branch (Supplementary Fig. 7) in agreement with the structural entity identified by the NMR analysis. Since this strain lacks both of the canonical GT-C type transferases, present in the SCC genetic locus, these data support a hypothesis that *S. mutans* has an additional unknown GT-C type transferase, which together with SccN participates in the synthesis of the 2,4-branch.

Major and minor side-chain decorations are important for *S. mutans* morphology

We observed that planktonic cultures of the WT and $\Delta sccP$ bacteria remain in suspension. In contrast, the $\Delta sccM$, $\Delta sccQ$ and $\Delta sccM\Delta sccQ$ bacteria sediment rapidly (Fig. 5a). Previously, we observed this phenotype in $\Delta sccN$ and $\Delta sccH$ which we attributed to alterations of the physical properties of the cell wall due to increased intermolecular association between unmodified SCC chains⁴. These data indicate that the minor Glc decorations installed by SccQ also contribute to these intermolecular interactions.

The Glc side-chains provided by SccN and the GroP moieties added by SccH were shown to regulate cell division of *S. mutans*⁴. We compared the morphology of exponentially grown cells of WT, $\Delta sccP$, $\Delta sccM$ and $\Delta sccQ$ using scanning electron microscopy (SEM) (Fig. 5b). The $\Delta sccM$ mutant displayed an irregular cell shape and size. In many cells, the division septa were not perpendicular to the long axis of the cell (Fig. 5b). The morphological phenotype of this mutant was similar to the phenotypes reported for $\Delta sccH$ and $\Delta sccN$ ⁴ which is in agreement with the role of SccM/SccN enzyme pair in installing α -Glc side-chains for GroP decorations. The deletion of *sccP* did not affect the cell shape of *S. mutans*. Surprisingly, the $\Delta sccQ$ mutant also showed irregular cell shape, and the cells were significantly smaller and wider than the WT cells (Fig. 5b, c, and Supplementary Table 3). Thus, our data indicate that the major and minor side-chain decorations added by SccM and SccQ are important to support the proper morphology of *S. mutans* cells.

GroP modification of SCC is critical for biofilm synthesis

S. mutans causes dental caries by adhering to teeth surfaces and producing biofilm. An exopolysaccharide-based biofilm is developed when sucrose is present in the growth medium. In the absence of sucrose, the bacterium forms a protein-based biofilm mediated by protein–protein interactions of the surface proteins⁴⁰. To understand the role of SCC decorations in the pathogenesis of *S. mutans*, we analyzed the formation of the exopolysaccharide- and protein-based biofilms in WT and the mutants defective in Region 2 genes of the SCC gene operon. All mutants produced the exopolysaccharide-based biofilms similar to WT (Supplementary Fig. 8). However, $\Delta sccH$, $\Delta sccN$ and $\Delta sccM$ were deficient in the protein-based biofilms (Fig. 5d). The phenotypes of $\Delta sccH$

and $\Delta sccN$ were restored to WT in mutants complemented with the WT copies of *sccH* and *sccN*, respectively. These data indicate that the GroP moieties, and the major Glc side-chains that provide the sites for GroP residues, are essential for biofilm formation.

The side-chains of a homologous polysaccharide GAC are GlcNAc decorated with GroP, and they are encoded by the *gacHIJKL* genes³⁰. To replace the major Glc side-chains of SCC with the GAC side-chains, $\Delta sccN$ was complemented with the *gacHIJKL* genes on a plasmid. The production of the protein-based biofilm by $\Delta sccN$ was partially restored by this complementation (Fig. 5d) indicating that *S. mutans* biofilm formation is GroP-dependent and independent of the specific glycosyl side-chain.

Discussion

The biosynthetic pathways of cell wall polysaccharides represent areas of active research due to their importance in bacterial physiology. Based on initial structural studies of streptococcal rhamnopolysaccharides it was assumed that the cell walls of most streptococci do not carry the net negative charge attributed to the phosphate-rich polymers (wall teichoic acids) linked to peptidoglycan in Gram-positive bacteria. Our recent analyses of the *S. mutans* and *S. pyogenes* rhamnopolysaccharides, SCC and GAC, established that the side-chain branches on these polymers carry negatively charged GroP moieties^{3,4}. In this study, we re-examine the structure of *S. mutans* SCC revealing that the polysaccharide has major and minor Glc side-chains attached to the polyrhamnose backbone. The major side-chains are α -Glc linked to position 2 of the 3-substituted Rha and partially modified with GroP at the C6 position. Minor side-chains are β -Glc linked to position 4 of the 3- and 2,3-disubstituted Rha and α -Glc linked to position 4 of the 2-substituted Rha. This new structural information is critical for a functional description of the enzymes of the SCC biosynthetic pathway and understanding the function of GroP in streptococci.

Our previous work established that decoration of SCC with the GroP moieties requires the *SccN* glycosyltransferase which is essential for the formation of the Glc side-chain acceptor sites for GroP transfer catalyzed by *SccH*^{3,4}. Here we demonstrate that *SccN* is a β -Glc-P-Und synthase which works in pair with the *SccM* glucosyltransferase to transfer α -Glc from β -Glc-P-Und to polyrhamnose. This inverting

mechanism, in which a sugar moiety is transferred with inversion of stereochemistry at the anomeric carbon atom of the donor substrate, is a common feature of all GT-C fold enzymes identified to date ⁴¹. Furthermore, we identify SccP as an α -Glc-P-Und synthase which works in pair with the SccQ glucosyltransferase to add β -Glc to the alternative position 4 on the 3-substituted Rha forming the 3,4-Rha branch (and on 2,3-substituted Rha, to form a 2,3,4-Rha branch). Analysis of the *S. mutans* gene clusters involved in the synthesis of the serotype *f* and *e* rhamnopolysaccharides supports the proposed enzymatic activities of SccN, SccP, and SccQ. The serotype *f* and *e* rhamnopolysaccharides are decorated with α -Glc and β -Glc side-chains, respectively ^{7, 19, 21}. An SccN homolog (99% identity) is present in the serotype *f*. In contrast, serotype *e* expresses the homologs of SccP (84% identity) and SccQ (98% identity) (Fig. 1a).

Surprisingly, the SccN enzyme also provides β -Glc-P-Und for an unknown GT-C-fold glycosyltransferase to add α -Glc to position 4 of the 2-substituted Rha forming the 2,4-Rha branch. The side-chains installed by the alternative mechanisms present as minor decorations on SCCs possibly because the dominant α -Glc side-chains attached by SccM may prevent SccQ from adding β -Glc to an alternative site on the same Rha. Furthermore, addition of α -Glc at position 4 of the 2-linked Rha might be repressed due to the competition of SccM and the unknown glycosyltransferase for β -Glc-P-Und. Importantly, when SCCs from various mutant strains are labelled at the reducing end with a fluorescent tag (ANDS) and separated by SDS-PAGE, only strains containing active SccN and SccM show the distinctive 'laddering' pattern attributed to the presence of the GroP modifications. This observation correlates with NMR analysis of SCCs indicating that the major side-chains attached to 2,3-Rha serve as the acceptors for GroP.

As established in our earlier study, defects in GroP decorations lead to severe cell shape alterations and increased autolysis ⁴. These phenotypes are attributed to mislocalization of the major autolysin AtlA involved in the separation of daughter cells, and the cell division protein MapZ, which regulates the correct positioning of the Z-ring ⁴. Here, we report that the SccM-deficient mutant resembles the *S. mutans* SccH and SccN mutants ⁴ producing irregularly shaped cells with misplaced division septa. This observation is in line with our conclusion that SccM works in pair with SccN to add α -Glc

residues to SCC that are specifically recognized by SccH for GroP transfer.

Interestingly, SEM experiments revealed that the SccQ-deficient mutant produces less elongated cells than the parental strain. Given that this mutant has a reduced Glc content in the SCCs, the underlying mechanisms of morphological abnormalities may be related to alteration in cell wall density due to absence of the SCC side-chain substituents. It is possible that the Glc side-chains regulate cell shape by limiting the access of cell wall synthetic and hydrolytic enzymes to peptidoglycan. Another important finding of our study is that the GroP decorations are necessary for development of the protein-based *S. mutans* biofilm. Interestingly, this phenotype does not require the minor Glc side-chains and is not specific to a glycosyl modification providing sites for GroP attachment. How GroP regulates *S. mutans* biofilm is still unclear and warrants future studies.

Taken together, we propose a revised generic structure of SCC containing additional side-chain modifications and a mechanistic model of SCC biosynthesis which identifies the glucosyltransferases involved in the major and alternative pathways of glucosylation (Fig. 6). In this model, polyrhamnose is synthesized on the cytoplasmic face of the plasma membrane and exported to the exoplasmic surface by an ABC transporter complex, as reported for GAC synthesis in *S. pyogenes*³. Glc residues are then transferred from β -Glc-P-Und and α -Glc-P-Und synthesized by SccN and SccP on the cytoplasmic surface, respectively, to the 2- and 4-positions of Rha by periplasmically oriented GT-C type transferases, SccM and SccQ (Supplementary Fig. 9), and a prospective third, currently unidentified, glucosyltransferase. Further studies will be required to identify flippase(s)/scramblase(s) involved in the transport of the α -Glc-P-Und and β -Glc-P-Und lipids to the exoplasmic leaflet of the membrane and the glucosyltransferase involved in the transfer of the minor α -Glc side-chains. Because SccN, SccM and SccH are important for biofilm formation, our characterization of SCC biosynthesis may be helpful in guiding the design of anti-caries agents targeting the activities of this pathway in *S. mutans*.

Methods

Bacterial strains, growth conditions and media

All plasmids, strains and primers used in this study are listed in Supplementary Tables 4 and 5. *S. mutans* strains were grown in BD Bacto Todd-Hewitt broth supplemented with 1% yeast extract (THY) without aeration at 37 °C. The agar plates were grown with 5% CO₂. To study biofilm formation, *S. mutans* were grown in UFTYE (2.5% tryptone and 1.5% yeast extract) containing either 1% (wt/vol) D-sucrose or 1% (wt/vol) D-glucose. *E. coli* strains were grown in Lysogeny Broth (LB) medium or on LB agar plates at 37 °C. When required, antibiotics were included at the following concentrations: chloramphenicol at 10 µg mL⁻¹ for *E. coli*; spectinomycin at 500 µg mL⁻¹, erythromycin at 5 µg mL⁻¹ and kanamycin at 300 µg mL⁻¹ for *S. mutans*.

Construction of mutant strains

To delete *sccM* and *sccQ* in *S. mutans* Xc we used a PCR overlapping mutagenesis approach, as previously described³. Briefly, 600-700 bp fragments both upstream and downstream of the gene of interest were amplified with designed primers that contained 16-20 bp extensions complementary to the nonpolar antibiotic resistance cassette. The nonpolar spectinomycin, kanamycin, or erythromycin resistance cassettes were PCR-amplified from pLR16T, pOSKAR, and pPHY304 (Supplementary Table 4), respectively. The two fragments of the gene of interest and the fragment with the antibiotic resistance cassette were purified using the QIAquick PCR purification kit (Qiagen) and fused by Gibson Assembly (SGA-DNA) using primers listed in Supplementary Table 5. The assembled DNA fragments were directly transformed into the *S. mutans* Xc cells by electroporation. The transformants were selected on THY agar containing the corresponding antibiotic. Double-crossover recombination was confirmed by PCR and Sanger sequencing using the primers listed in Supplementary Table 5. The double deletion mutants, $\Delta sccM\Delta sccQ$, $\Delta sccM\Delta sccN$ $\Delta sccN\Delta sccQ$ and $\Delta sccM\Delta sccP$, were constructed using a similar approach by deleting *sccM* or *sccQ* in the mutant backgrounds. All deletion mutants were verified by whole-genome sequencing (Supplementary Table 6).

Construction of the plasmids for *E. coli* expression of SccN and SccP

To create vectors for expression of SccN and SccP, the genes were amplified from *S. mutans* Xc chromosomal DNA using the primer pair listed in Supplementary Table 5. The PCR products were digested by the corresponding restriction enzymes and subcloned into a pBAD33 vector. The resultant plasmids, pSccN and pSccP were transferred into competent *E. coli* JW2347 strain that has a deletion of the *gtrB* gene⁴².

Isolation of SCCs from *S. mutans* cell wall

S. mutans cell wall was isolated from late exponential phase cultures (OD₆₀₀~0.8) by the SDS-boiling procedure as described for *S. pneumoniae* and *S. pyogenes*^{30, 43}. Purified cell wall samples were lyophilized and stored at -20 °C before the analysis.

Preparation of SCCs

SCCs were released from highly enriched cell wall preparations by mild acid hydrolysis, following chemical *N*-acetylation, as described previously⁴, and partially purified by size exclusion chromatography on a column of Bio Gel P150 (Bio Rad, 1×18 cm) equilibrated in 0.2 N Na acetate, pH 3.7, 0.15 M NaCl³⁵.

NMR spectroscopy

SCCs isolated from WT and the various mutant strains were dissolved in D₂O. NMR experiments were conducted in 5 mm outer diameter NMR tubes on Bruker NMR spectrometers operating at ¹H frequencies of 400 or 700 MHz at temperatures of 23 °C or 50 °C, respectively, using experiments suitable for resonance assignments of glycans^{36, 44}. ¹H NMR chemical shifts were referenced to internal sodium 3-trimethylsilyl-(2,2,3,3-²H₄)-propanoate (δ_H 0.0), ¹³C chemical shifts were referenced to external dioxane in D₂O (δ_C 67.4) and ³¹P chemical shifts were referenced to external 2% H₃PO₄ in D₂O (δ_P 0.0). Acquired NMR data were processed and analyzed using the TopSpin® software from Bruker.

Phosphate assays

Total phosphate content in SCCs was determined by the malachite green method following digestion with perchloric acid. Fractions containing 10 to 80 μL were heated to 110 °C with 20 μL 70% perchloric acid (Fisher Scientific) in 13 \times 100 mm borosilicate screw-cap disposable culture tubes for 1 h. The reactions were diluted to 160 μL with water and 100 μL was transferred to a flat-bottom 96-well culture plate. Malachite Green reagent (0.2 mL) was added and the absorbance at 620 nm was read after 10 min at room temperature. Malachite Green reagent contained one vol 4.2% ammonium molybdate tetrahydrate (by weight) in 4 M HCl, 3 vol 0.045% malachite green (by weight) in water and 0.01% Tween 20. Phosphate concentrations were estimated using a phosphate standard curve.

Modified anthrone assay

Total Rha content was estimated using a minor modification of the anthrone procedure, as described elsewhere ⁴.

PAGE analysis of SCCs derivatized with ANDS

SCCs were derivatized with 7-amino-1,3-naphthalenedisulfonic acid (ANDS) by reductive amination of the reducing end of the polysaccharide as described previously ⁴. Derivatized SCCs were further purified by SEC over a Superdex 75 10/300 GL column (GE Healthcare Bio-Sciences AB). ANDS-labeled SCCs were resolved on Bolt™ Bis-Tris 4-12% Gel (ThermoFisher) and visualized by fluorescence.

Preparation of undecaprenyl phosphate (Und-P)

Und-P was synthesized from undecaprenol (ARC, ARCD 0126) by the trichloroacetimidate procedure as described in ⁴⁵ and purified by preparative TLC on silica gel G.

Solubilization and partial purification of *S. mutans* Glc-P-Und synthases

For expression of SccN and SccP, *E. coli* JW2347 cells carrying pSccN and pSccP plasmids were grown to an OD₆₀₀ of 0.8 and induced with 13 mM L-arabinose at 25 °C for approximately 3 hours. The cells were lysed in 20 mM Tris-HCl pH 7.5, 300 mM

NaCl with two passes through a microfluidizer cell disrupter. The lysate was centrifuged (1000 g, 15 minutes, 4 °C) followed by centrifugation of the supernatant (40,000 g, 60 min, 4 °C) to isolate the membrane fraction. The SccN and SccP enzymes were solubilized from the pellet and partially purified. Solubilization mixtures contained 50 mM HEPES, pH 7.4, 1× bacterial protease inhibitor cocktail (Pierce Chemical Co.), 10 mM DTT, 0.5% CHAPS (3-[(3-cholamidopropyl)dimethyl-ammonio]-1-propane sulfonate) and *E. coli* membrane proteins (5 mg/ml). After 1 h on ice, solubilization mixtures were centrifuged (100,000 g, 30 min) and the supernatant liquid removed and loaded onto a 30 mL column of DEAE 650M (Toso Haas), equilibrated in the cold in 10 mM HEPES, pH 8.0, 20% glycerol, 5 mM DTT, 0.1% CHAPS. The column was eluted with 2 column volumes of buffer and then with a 40 mL linear gradient of NaCl (0 - 1 M). Fractions of 3 mL were collected and analyzed for protein and Glc-P-Und synthase activity. Fractions containing Glc-P-Und synthase activity were combined, concentrated by ultrafiltration (Amicon, 10,000 MWCO), snap-frozen on dry-ice ethanol in aliquots and stored at -20 °C until use.

***In vitro* analysis of Glc-P-Und synthase activity**

Reactions for the *in vitro* analysis of Glc-P-Und synthase activity contained 50 mM HEPES-OH pH 7.4, 20 mM MgCl₂, 10 mM DTT, 0.5 mM sodium orthovanadate, the indicated amount of UDP-[³H]Glc (50-500 cpm/pmol) and bacterial enzyme (either *S. mutans* membrane suspension or solubilized *E. coli* membrane proteins) in a total volume of 0.02 mL. Where indicated, reactions were supplemented with the indicated amount of Und-P, dispersed in 1% CHAPS by bath sonication, and with a final concentration of 0.35% CHAPS. Following incubation at 37 °C for 3-30 min, the reaction was stopped by the addition of 2 mL CHCl₃/CH₃OH (2:1). Reactions were freed of unincorporated radioactivity as described previously³⁰ and analyzed for radioactivity by scintillation spectrometry or by TLC on silica gel G developed in CHCl₃/CH₃OH/H₂O/NH₄OH (65:30:4:1). Reactions with membrane fractions from *S. mutans* were subjected to mild alkaline de-acetylation with 0.1 N KOH in methanol/toluene (3:1) for 60 min at 0 °C, to destroy glycerolipids and freed of released radioactivity by

partitioning as described³⁰. Degradation of Glc-lipids by mild acid and mild alkaline treatment were conducted as previously described³⁰.

Discharge of synthetic [³H]Glc-P-Und by reversal of Glc-P-Und synthase

Large amounts of the [³H]Glc-P-Und_{ScCN} and [³H]Glc-P-Und_{ScCP} were synthesized using Und-P, UDP-[³H]Glc and the appropriate partially purified Glc-P-Und synthase purified by Folch partition⁴⁶ and preparative TLC on silica gel G developed in CHCl₃/CH₃OH/H₂O/NH₄OH (65:30:4:1) as described below. These glycolipids were dispersed by sonication in 1% CHAPS, for 5 min in a bath sonicator, and used as substrates in the discharge reactions. Reaction mixtures were identical to reactions for the *in vitro* analysis of Glc-P-Und synthase activity, except that Und-P and UDP-Glc were replaced by the indicated purified [³H]Glc-P-Und (10,000 cpm) and UDP (5 mM) and contained the appropriate partially purified Glc-P-Und synthase (430 pmol/min ScCN Glc-P-Und synthase activity, ~0.7 µg protein; ~555 pmol/min ScCP Glc-P-Und synthase activity, ~0.13 µg protein).

Purification of *S. mutans* phosphoglycolipids for ESI-MS analysis

To purify glycolipids for ESI-MS analysis, *S. mutans* cells (2 L) from late exponential phase cultures (OD₆₀₀~0.8) were recovered by sedimentation (10,000 g, 30 min) and washed twice with ice-cold phosphate-buffered saline (PBS). Cells were resuspended in 15 ml PBS, sensitized by incubation with mutanolysin (0.2 M Na acetate, pH 5.2, 200 U/ml, 1 h, 37 °C) and stirred vigorously with two volumes of CH₃OH and one volume of CHCl₃ for 30 min at room temperature. The mixture was supplemented with additional CH₃OH (~ 5 mL) to achieve a single phase, divided into 5 mL aliquots in 12 mL glass centrifuge tubes and sedimented (200 g, 20 min) to remove insoluble material. The organic extract was transferred to a 300 mL separatory funnel and the insoluble residue was further extracted with CHCl₃/CH₃OH (2:1), 1 mL per tube, two times. The extracts were combined with the previous organic phase and supplemented with CHCl₃ and 0.9% NaCl to give a final composition of CHCl₃/CH₃OH/0.9% NaCl (3:2:1), mixed vigorously, and allowed to stand in the cold until phase separation was achieved. The organic phase was drained off into a second separatory funnel and the organic layer

was washed with ~1/3 volume of CHCl₃/CH₃OH/0.9% saline (3:48:47), two times. The aqueous layers were discarded. The organic extract was dried on a vacuum rotary evaporator, dissolved in a small volume of CHCl₃/CH₃OH (2:1), transferred quantitatively to a 12.5 × 100 mm screw cap glass tube (with Teflon lined cap) and partitioned with 1/5th volume of water to remove residual salts. The desalted organic extract was dried under a stream of nitrogen gas (at ~ 30 °C) and the glycerolipids were destroyed by deacylation in 4 mL 0.1 M KOH in toluene/CH₃OH (1:3) at 0 °C, 60 min. Following deacylation, the reactions were neutralized with acetic acid, diluted with CHCl₃ and 0.9% NaCl/10 mM EDTA to give a final composition of CHCl₃/CH₃OH/0.9% NaCl (3:2:1). The two-phase mixture was mixed vigorously and centrifuged to separate the phases. The aqueous phase was discarded and the organic phase was washed with ~1/3 volume of CHCl₃/CH₃OH/0.9% saline (3:48:47), two times. The organic phase was dried under a stream of N₂ gas, dissolved in 1 mL CHCl₃/CH₃OH (2:1), partitioned with 1/5th volume of water and dried again. The dried glycolipid fraction was redissolved in 1 mL of CHCl₃/CH₃OH (19:1) and applied to a column of silica gel (BioRad) equilibrated in CHCl₃. The column was washed with 5 column volumes of CHCl₃ and 5 column volumes of acetone. The glycolipids were eluted with 5 column volumes of CHCl₃/CH₃OH (2:1), partitioned with 1/5th volume of water and dried under a stream of N₂ gas. Fractions recovered from silica gel chromatography were analyzed by TLC on silica gel, developed in CHCl₃/CH₃OH/H₂O/NH₄OH (65:25:4:1) and visualized by exposure to iodine vapors. Comparison of mobility on silica with that of authentic [³H]Glc-P-Und, synthesized separately, confirmed that the glycolipids were quantitatively recovered in the CHCl₃/CH₃OH (2:1) fraction following silica gel chromatography. The glycolipid fraction was dried and purified further by preparative thin layer chromatography on silica gel G developed in CHCl₃/CH₃OH/H₂O/NH₄OH (65:25:4:1). Glc-P-Unds were visualized by exposure to iodine vapors and located by reference to [³H]Glc-P-Und standard in adjacent lanes detected by a Bioscan 2000 radiochromatoscanner. Glc-P-Unds were recovered from the silica gel plate by elution with CHCl₃/CH₃OH (2:1), dried under a stream of nitrogen gas and reserved for analysis by LC/MS/MS.

Mass spectrometry analysis of phospholipids isolated from *S. mutans* strains

The glycolipids were analyzed by LC-MS using a Q-exactive mass spectrometer, employing polarity switching and data dependent acquisition, and an Ultimate 3000 ultra high performance liquid chromatography system (Thermo Fisher Scientific, San Jose, CA) on a Kinetex C18 reversed-phase column (2.1 mm × 100 mm, 2.6 μm, Phenomenex, USA) as described previously³⁰.

Analysis of SCCs for composition and glycosyl linkages

Composition of sodium borohydride-reduced SCCs was determined as trimethylsilyl (TMS) derivatives of O-methyl glycosides following methanolysis in methanol/1 N HCl (100 °C, 3 h) either, in house as described previously³⁵, or at the Complex Carbohydrate Research Center, University of Georgia (CCRC). Quantities of component sugars were calculated from gas-chromatographic (GC) areas of component TMS-methylglycosides after normalization to the internal standard and response factor corrections.

Response factors (Rf) were calculated from standard mixtures of sugars and internal standard using the formula: $Rf = (A_x/A_{is})/(C_x/C_{is})$ in which, A_x is GC area of analyte, A_{is} is GC area of internal standard, C_x is amount of analyte and C_{is} is amount of internal standard. Amounts of sugars were calculated using the formula: $C_x = C_{is} \times (A_x/A_{is})/Rf$. For illustrative purposes, the amounts of sugars are expressed as nmol of each sugar in the sample normalized to 50 nmol Rha (the proposed size of the polyrhamnose backbone). In some instances, an exponential increase in response factor associated with increasing amounts of analyte relative to internal standard was observed. To correct for this change in response factor, standard curves of increasing amounts of analytes to a constant quantity of internal standard were prepared and used to estimate the appropriate response factor, when necessary.

Linkage analysis of the glycans was determined at the CCRC as partially methylated alditol acetates as described⁴⁷. Quantitative estimates of specific linkages per polymer were calculated by multiplying the nmol sugar per 50 nmol Rha in each sample, by the relative area percent of the GC area for each linkage type.

Scanning electron microscopy (SEM)

Exponentially growing bacteria ($OD_{600} \sim 0.7$) were fixed with paraformaldehyde (4% final concentration), and then pipetted onto microscope slide cover glasses coated with poly-L-lysine. Following one hour incubation, the cover glasses were washed three times with PBS. Bacteria were dehydrated stepwise in a gradient series of ethanol (35%, 50%, 70%, 80% and 96% for 20 min each at room temperature and then 100% overnight at -20°C), followed by critical point drying with liquid CO_2 in a Leica EM CPD300.

Samples were coated with about 5 nm of platinum controlled by a film-thickness monitor. SEM images were performed in the immersion mode of an FEI Helios Nanolab 660 dual beam system.

Differential interference contrast (DIC) microscopy

Exponentially growing bacteria ($OD_{600} \sim 0.7$) were fixed with paraformaldehyde (4% final concentration), pipetted onto microscope slide cover glasses (high performance, $D=0.17$ mm, Zeiss) coated with poly-L-lysine, and allowed to settle for one hour at room temperature. The samples were washed three times with PBS and mounted on a microscope slide with ProLong Glass Antifade (Invitrogen). Samples were imaged on a Leica SP8 equipped with 100X, 1.44 N.A. objective, DIC optics. ImageJ software ObjectJ plugin was used to measure the sizes of cells.

Biofilm assay

UFTYE supplemented with 1% (wt/vol) sucrose or glucose was inoculated with the 17 h overnight cultures of *S. mutans* creating a 1:100 fold inoculation. Bacterial biofilms were grown in 24-well plates (Corning, NY) at 37°C in the presence of 5% CO_2 . After 24 h of growth, cell suspension in each well was removed and biofilm was washed three times with water. The biofilm was stained with crystal violet (0.2%) followed by washes and de-staining in 10% methanol, 7.5% acetic acid solution. Finally, the absorbance of the de-staining solution was measured at 540 nm.

Statistical analysis

Unless otherwise indicated, statistical analysis was carried out on pooled data from at least three independent biological repeats. Statistical analyses were performed using Graph Pad Prism version 9.2.0. Quantitative data was analyzed using one-way ANOVA, 2-way ANOVA, and unpaired t-test as described for individual experiments. A *P*-value equal to or less than 0.05 was considered statistically significant.

Data availability

All data generated during this study are included in the article and supplementary information files.

Acknowledgments

The authors thank Dr. Catherine Chaton (University of Kentucky) for the help with figure preparation and Dr. Christopher D. Radka (University of Kentucky) for helpful discussion of ESI-MS analysis of phosphoglycolipids.

This work was supported by NIH grants R01 DE028916 from the NIDCR and R01 AI143690 from the NIAID (to NK), and the Swedish Research Council (no. 2022-03014) and The Knut and Alice Wallenberg Foundation (to GW). Scanning electron microscopy was performed at the Electron Microscopy Center, which belongs to the National Science Foundation NNCI Kentucky Multiscale Manufacturing and Nano Integration Node, supported by ECCS-1542174. Carbohydrate composition/linkage analysis at the Complex Carbohydrate Research Center was supported by the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, U.S. Department of Energy grant DE-SC0015662 (to PA), as well as by NIH grant R24 GM137782 from the NIGMS (to PA). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions

JSR, GW, CH, PA, and NK designed the experiments. JSR, SZ, NRM, CWK and IB performed functional and biochemical experiments. SZ performed microscopy analysis. PD performed MS analysis. GW performed NMR studies. NK, KVK and SZ constructed

plasmids and isolated mutants. JSR, SZ, PD, IB, CH AJM, KVK, GW and NK analyzed the data. NK, JSR and GW wrote the manuscript with contributions from all authors. All authors reviewed the results and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Mistou MY, Sutcliffe IC, van Sorge NM. Bacterial glycobiology: rhamnose-containing cell wall polysaccharides in Gram-positive bacteria. *FEMS Microbiol Rev* **40**, 464-479 (2016).
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* **8**, e1002756 (2012).
3. Edgar RJ, *et al.* Discovery of glycerol phosphate modification on streptococcal rhamnose polysaccharides. *Nature chemical biology* **15**, 463-471 (2019).
4. Zamakhaeva S, *et al.* Modification of cell wall polysaccharide guides cell division in *Streptococcus mutans*. *Nature chemical biology* **17**, 878-887 (2021).
5. Kabanova A, *et al.* Evaluation of a Group A *Streptococcus* synthetic oligosaccharide as vaccine candidate. *Vaccine* **29**, 104-114 (2010).
6. van Sorge NM, *et al.* The classical Lancefield antigen of Group A *Streptococcus* is a virulence determinant with implications for vaccine design. *Cell host & microbe* **15**, 729-740 (2014).
7. 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*. *Glycoconjugate journal* **35**, 53-64 (2018).
8. Sabharwal H, *et al.* Group A streptococcus (GAS) carbohydrate as an immunogen for protection against GAS infection. *J Infect Dis* **193**, 129-135 (2006).
9. Di Benedetto R, *et al.* Rational Design of a Glycoconjugate Vaccine against Group A *Streptococcus*. *Int J Mol Sci* **21**, (2020).
10. Wang S, Zhao Y, Wang G, Feng S, Guo Z, Gu G. Group A *Streptococcus* Cell Wall Oligosaccharide-Streptococcal C5a Peptidase Conjugates as Effective Antibacterial Vaccines. *ACS infectious diseases* **6**, 281-290 (2020).
11. Khatun F, *et al.* Immunogenicity Assessment of Cell Wall Carbohydrates of Group A *Streptococcus* via Self-Adjuvanted Glyco-lipopeptides. *ACS infectious diseases* **7**, 390-405 (2021).
12. Pitirillo O, *et al.* Elucidating the role of N-acetylglucosamine in Group A Carbohydrate for the development of an effective glycoconjugate vaccine against Group A *Streptococcus*. *Carbohydr Polym* **311**, 120736 (2023).

13. Broendum SS, *et al.* High avidity drives the interaction between the streptococcal C1 phage endolysin, PlyC, with the cell surface carbohydrates of Group A Streptococcus. *Mol Microbiol* **116**, 397-415 (2021).
14. King H, *et al.* Molecular basis for recognition of the Group A Carbohydrate backbone by the PlyC streptococcal bacteriophage endolysin. *Biochem J* **478**, 2385-2397 (2021).
15. Shen Y, *et al.* A bacteriophage endolysin that eliminates intracellular streptococci. *eLife* **5**, (2016).
16. Shen Y, Koller T, Kreikemeyer B, Nelson DC. Rapid degradation of Streptococcus pyogenes biofilms by PlyC, a bacteriophage-encoded endolysin. *The Journal of antimicrobial chemotherapy* **68**, 1818-1824 (2013).
17. van der Beek SL, *et al.* Streptococcal dTDP-L-rhamnose biosynthesis enzymes: functional characterization and lead compound identification. *Mol Microbiol* **111**, 951-964 (2019).
18. Wang Y, Hess TN, Jones V, Zhou JZ, McNeil MR, McCammon JA. Novel inhibitors of Mycobacterium tuberculosis dTDP-6-deoxy-L-lyxo-4-hexulose reductase (RmlD) identified by virtual screening. *Bioorganic & medicinal chemistry letters* **21**, 7064-7067 (2011).
19. Nakano K, Ooshima T. Serotype classification of Streptococcus mutans and its detection outside the oral cavity. *Future microbiology* **4**, 891-902 (2009).
20. Nakano K, Nomura R, Nakagawa I, Hamada S, Ooshima T. Demonstration of Streptococcus mutans with a cell wall polysaccharide specific to a new serotype, k, in the human oral cavity. *J Clin Microbiol* **42**, 198-202 (2004).
21. Pritchard DG, Gregory RL, Michalek SM, McGhee JR. Characterization of the serotype e polysaccharide antigen of Streptococcus mutans. *Molecular immunology* **23**, 141-145 (1986).
22. Huang DH, Rama Krishna N, Pritchard DG. Characterization of the group A streptococcal polysaccharide by two-dimensional ¹H-nuclear-magnetic-resonance spectroscopy. *Carbohydrate research* **155**, 193-199 (1986).
23. Shibata Y, *et al.* Analysis of loci required for determination of serotype antigenicity in Streptococcus mutans and its clinical utilization. *J Clin Microbiol* **41**, 4107-4112 (2003).
24. Swoboda JG, Campbell J, Meredith TC, Walker S. Wall teichoic acid function, biosynthesis, and inhibition. *ChemBiochem* **11**, 35-45 (2010).

25. Yamashita Y, *et al.* A novel gene required for rhamnose-glucose polysaccharide synthesis in *Streptococcus mutans*. *J Bacteriol* **181**, 6556-6559 (1999).
26. Zorzoli A, *et al.* Group A, B, C, and G *Streptococcus* Lancefield antigen biosynthesis is initiated by a conserved alpha-d-GlcNAc-beta-1,4-l-rhamnosyltransferase. *J Biol Chem* **294**, 15237-15256 (2019).
27. Shibata Y, Yamashita Y, Ozaki K, Nakano Y, Koga T. Expression and characterization of streptococcal *rgp* genes required for rhamnan synthesis in *Escherichia coli*. *Infect Immun* **70**, 2891-2898 (2002).
28. Rainey K, Michalek SM, Wen ZT, Wu H. Glycosyltransferase-Mediated Biofilm Matrix Dynamics and Virulence of *Streptococcus mutans*. *Appl Environ Microbiol* **85**, (2019).
29. Rismondo J, Gillis A, Grundling A. Modifications of cell wall polymers in Gram-positive bacteria by multi-component transmembrane glycosylation systems. *Current opinion in microbiology* **60**, 24-33 (2021).
30. Rush JS, *et al.* The molecular mechanism of N-acetylglucosamine side-chain attachment to the Lancefield group A carbohydrate in *Streptococcus pyogenes*. *J Biol Chem* **292**, 19441-19457 (2017).
31. Banerjee DK, Scher MG, Waechter CJ. Amphomycin: effect of the lipopeptide antibiotic on the glycosylation and extraction of dolichyl monophosphate in calf brain membranes. *Biochemistry* **20**, 1561-1568 (1981).
32. Zhang C, *et al.* Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions. *Science* **313**, 1291-1294 (2006).
33. Mestrom L, *et al.* Leloir Glycosyltransferases in Applied Biocatalysis: A Multidisciplinary Approach. *Int J Mol Sci* **20**, (2019).
34. Wolucka BA, de Hoffmann E, Rush JS, Waechter CJ. Determination of the anomeric configuration of glycosyl esters of nucleoside pyrophosphates and polyisoprenyl phosphates by fast-atom bombardment tandem mass spectrometry. *J Am Soc Mass Spectrom* **7**, 541-549 (1996).
35. Rush JS, *et al.* PpID is a de-N-acetylase of the cell wall linkage unit of streptococcal rhamnopolysaccharides. *Nat Commun* **13**, 590 (2022).
36. Widmalm G. 1.14 - General NMR Spectroscopy of Carbohydrates and Conformational Analysis in Solution. In: *Comprehensive Glycoscience (Second Edition)* (ed Barchi Jr. JJ). Elsevier (2021).

37. Dorst KM, Widmalm G. NMR chemical shift prediction and structural elucidation of linker-containing oligo- and polysaccharides using the computer program CASPER. *Carbohydrate research* **533**, 108937 (2023).
38. Bernlind CO, S.; Widmalm, G. Synthesis, NMR, and conformational studies of methyl α -d-mannopyranoside 2-, 3-, 4-, and 6-monophosphates. *Carbohydrate research* **263**, 173-180 (1994).
39. Khatuntseva EA, Shashkov AS, Nifant'ev NE. ^1H and ^{13}C NMR Data for 3-O-, 4-O- and 3,4-Di- 1O-glycosylated Methyl α -L-Rhamnopyranosides. *Magnetic Resonance in Chemistry* **35**, 414-419 (1997).
40. Zheng T, *et al.* Regulatory mechanisms of exopolysaccharide synthesis and biofilm formation in *Streptococcus mutans*. *J Oral Microbiol* **15**, 2225257 (2023).
41. Liang DM, Liu JH, Wu H, Wang BB, Zhu HJ, Qiao JJ. Glycosyltransferases: mechanisms and applications in natural product development. *Chem Soc Rev* **44**, 8350-8374 (2015).
42. Baba T, *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006 0008 (2006).
43. Bui NK, *et al.* Isolation and analysis of cell wall components from *Streptococcus pneumoniae*. *Analytical biochemistry* **421**, 657-666 (2012).
44. Fontana C, Widmalm G. Primary Structure of Glycans by NMR Spectroscopy. *Chem Rev* **123**, 1040-1102 (2023).
45. Danilov LL, Druzhinina TN, Kalinchuk NA, Maltsev SD, Shibaev VN. Polyprenyl phosphates: synthesis and structure-activity relationship for a biosynthetic system of *Salmonella anatum* O-specific polysaccharide. *Chemistry and physics of lipids* **51**, 191-203 (1989).
46. Folch J, Lees M, Sloane Stanley GH. A simple method for the isolation and purification of total lipides from animal tissues. *J Biol Chem* **226**, 497-509 (1957).
47. Black I, Heiss C, Carlson RW, Azadi P. Linkage Analysis of Oligosaccharides and Polysaccharides: A Tutorial. *Methods Mol Biol* **2271**, 249-271 (2021).

Table 1. Kinetic parameters of Glc-P-Und synthases in membrane fractions from *E. coli* JW2347 expressing *sccN* or *sccP* on a plasmid

Enzyme Substrate	SccN		SccP	
	Und-P	UDP-Glc	Und-P	UDP-Glc
K _m , app (μmol)	5.85	1.61	2.56	2.50
V _{max} , app (μmol·min ⁻¹ ·mg ⁻¹)	5.42	3.16	0.96	0.93

Table 2. Relative abundances of fragment ions obtained by ESI MS/MS from Glc-P-Unds purified from various strains of *S. mutans* expressed as a percentage of the [M-H]⁻ ion ^a

<i>S. mutans</i> strains	Active Glc-P-Und Synthase	Relative abundance of ions (% [M-H] ⁻)		Ratio	Glucolipid
		[Und-PO ₄ - C ₂ H ₃ O] ⁻	[Und-HPO ₄] ⁻		
		<i>A</i>	<i>B</i>	<i>A/B</i>	
WT	SccN/SccP	40.2	82	0.49	α/β-Glc-P-Und
Δ <i>sccN</i>	SccP	65.4	64.2	1.02	α-Glc-P-Und
Δ <i>sccP</i>	SccN	12.7	89.4	0.14	β-Glc-P-Und
Δ <i>sccQ</i>	SccN/SccP	38.1	89.5	0.43	α/β-Glc-P-Und
Δ <i>sccM</i>	SccN/SccP	42.1	94.1	0.45	α/β-Glc-P-Und

^a Total lipids from the various strains were extracted with chloroform/methanol, deacylated in KOH/methanol, purified by preparative TLC, and analyzed by Q-Exactive Orbitrap LC/MS under conditions described in Methods. Relative abundances of fragment ions are expressed as a percentage of the pertinent ion fragment relative to the parent ion ([M-H]⁻). The observed *m/z* value for [M-H]⁻ was 1007.709, [Und-HPO₄]⁻ was 845.657 and [Und-PO₄-C₂H₃O]⁻ was 887.664.

Table 3. ^1H and ^{13}C NMR chemical shifts (ppm) of SCC RPS in D_2O at 323 K referenced to TSP (δ_{H} 0.00) and dioxane in D_2O (δ_{C} 67.40); $^1J_{\text{C},\text{H}1}$ in parenthesis are given in hertz. Repeating unit residues of the minor form of ΔsccN are non-primed (**A** - **C**). Repeating unit residues of the major form of ΔsccM (rhamnan backbone) are primed (**A'** and **B'**) and those from the minor form are double primed (**A''** □ **C''**).

Residue	Label	1	$^1J_{\text{CH}}$	2	3	4	5	6	6
$\rightarrow 3,4\text{-}\alpha\text{-L-Rhap-(1}\rightarrow 2\text{)}$	A	4.97	172	4.19	4.04	3.83	3.84	1.34	
		102.6		70.6	80.9	78.0	68.9	17.9	
$\rightarrow 2\text{-}\alpha\text{-L-Rhap-(1}\rightarrow 3\text{)}$	B	5.16	172	4.20	3.95	3.53	3.89	1.32	
		101.8		79.5	70.8	73.1	70.2	17.5	
$\beta\text{-D-Glcp-(1}\rightarrow 4\text{)}$	C	4.63	163	3.32	3.52	3.43	3.43	3.75	3.92
		103.7		74.1	76.8	70.5	76.7	61.6	
$\rightarrow 3\text{-}\alpha\text{-L-Rhap-(1}\rightarrow 2\text{)}$	A'	4.98	173	4.16	3.86	3.58	3.78	1.29	
		102.7		70.7	78.4	72.4	70.1	17.4	
$\rightarrow 2\text{-}\alpha\text{-L-Rhap-(1}\rightarrow 3\text{)}$	B'	5.19	173	4.09	3.96	3.51	3.83	1.33	
		101.6		78.8	70.8	73.1	70.0	17.5	
$\rightarrow 3\text{-}\alpha\text{-L-Rhap-(1}\rightarrow 2\text{)}$	A''	4.98	172	4.19	3.87	3.58	3.78	1.29	
		102.8		70.7	78.4	72.4	70.1	17.4	
$\rightarrow 2,4\text{-}\alpha\text{-L-Rhap-(1}\rightarrow 3\text{)}$	B''	5.21	174	4.09	4.08	3.57	3.98	1.42	
		101.2		79.2	69.4	82.2	69.1	17.9	
$\alpha\text{-D-Glcp-(1}\rightarrow 4\text{)}$	C''	5.06	170	3.58	3.70	3.46	4.00	3.78	3.83
		100.4		72.4	73.6	70.3	72.7	61.3	

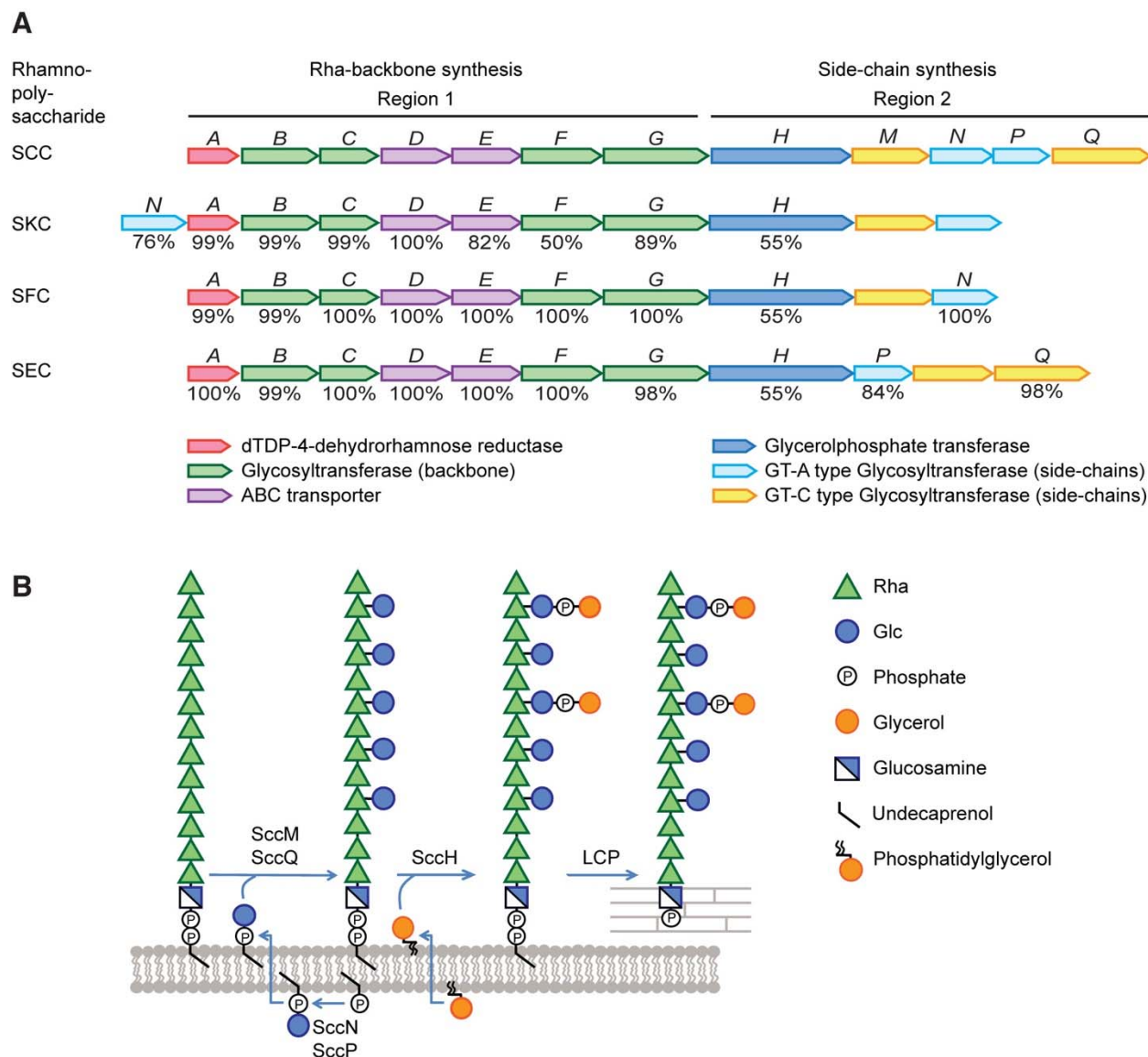


Fig. 1. Schematic representation of the *S. mutans* rhamnopolysaccharide biosynthetic gene clusters and the proposed mechanism of SCC modification with the Glc side-chains and GroP

(a) Schematic representation of the genetic loci encoding enzymes involved in the biosynthesis of rhamnopolysaccharides expressed by *S. mutans* serotypes *c*, *k*, *f*, and *e*: Serotype c Carbohydrate (SCC), Serotype k Carbohydrate (SKS), Serotype f Carbohydrate (SFC) and Serotype e Carbohydrate (SEC), respectively. Genes in Region 1 participate in the synthesis of the polyrhamnose backbone. Genes in Region 2 participate in modification of the polyrhamnose backbone with the side-chains and GroP. The SCC gene cluster contains genes designated *sccABCDEFGHIHMPNQ*

corresponding to smu.824-835 in *S. mutans* UA159 (GenBank: AE014133.2). The SKC gene cluster represents the SMULJ23_1190-SMULJ23_1180 gene locus in *S. mutans* LJ23 (GenBank: AP012336.1). The SFC gene cluster represents the K2F51_02855-K2F51_02900 gene locus in *S. mutans* OMZ175 (GenBank: CP082153.1). The SEC gene cluster represents the CO204_06335-CO204_06385 gene locus in *S. mutans* LAR01 (GenBank: CP023477.1). The percentage of amino acid identity to proteins encoded by the SCC gene cluster are noted below each gene. **(b)** Model illustrating SCC side-chain assembly and transfer to peptidoglycan in *S. mutans* serotype *c*. Side-chain addition starts with the synthesis of Glc-P-Und, on the cytosolic surface of the plasma membrane, catalyzed by two independent Glc-P-Und synthases (SccN and SccP) using UDP-Glc and Und-P as substrates. Following translocation of Glc-P-Und to the periplasmic surface, the GT-C type transferases, SccM and SccQ, transfer Glc from Glc-P-Und to the polyrhamnose backbone at the 2- and 4-positions of 3-Rha components, respectively. In addition, transfer of Glc from Glc-P-Und to the 4-position of 2-Rha takes place, but only to a limited extent (glycosyltransferase unknown). Following glucosylation, SccH transfers GroP from phosphatidylglycerol to the 6-OH of Glc residues found only at the 2-position of 3-Rha. After side-chain assembly has been completed, SCC is transferred to peptidoglycan by LytR-Cps2A-Psr (LCP) family proteins.

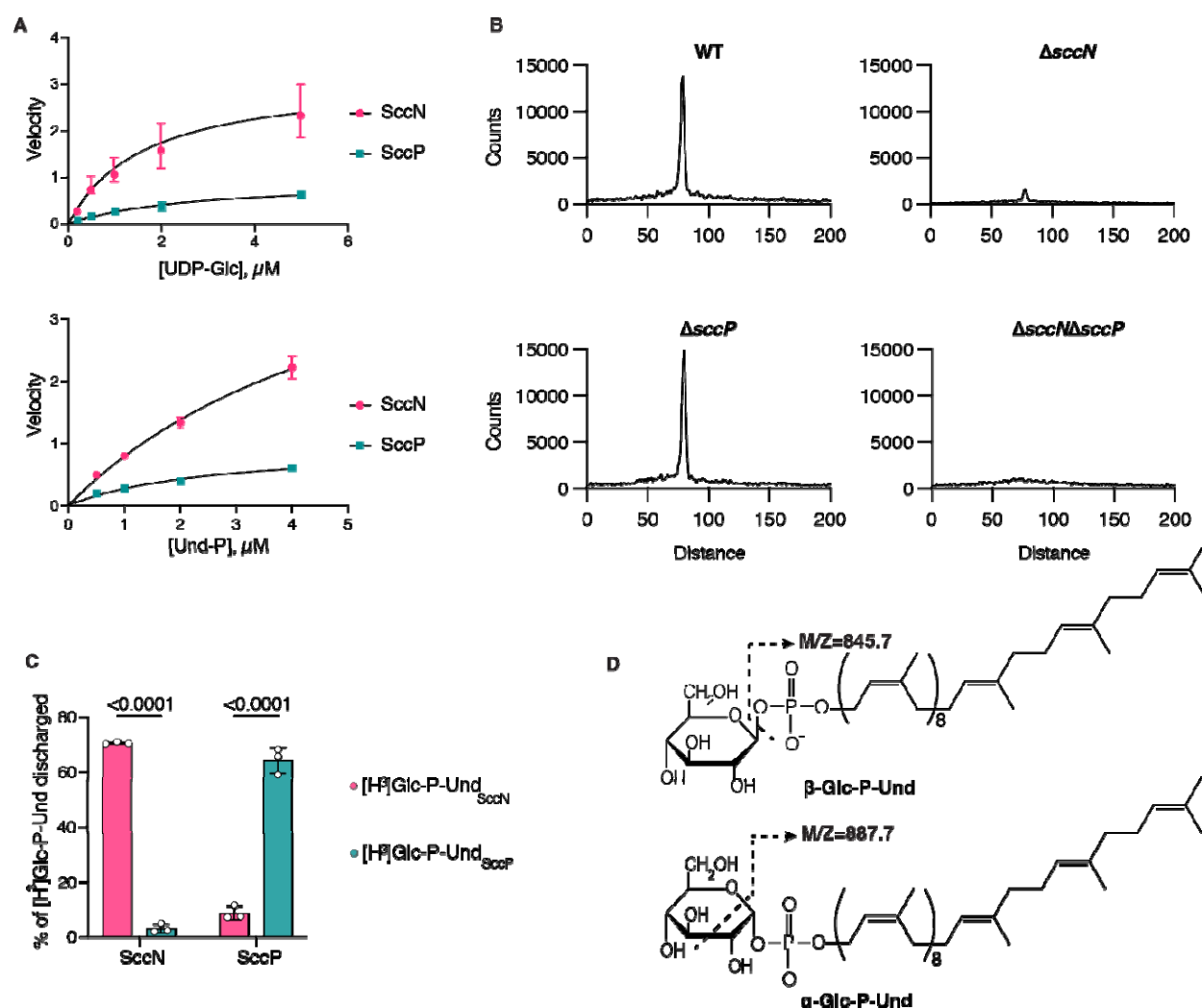


Fig. 2. SccN and SccP synthesize alternative stereoisomers of Glc-P-Und

(a) Kinetic analysis of Glc-lipid synthases in membrane fractions from the *E. coli* JW2347 strain expressing *sccN* or *sccP* on a plasmid. Membrane fractions were assayed for the formation of [^3H]Glc-lipids in the presence of increasing amounts of UDP-[^3H]Glc or Und-P, added as a dispersion in 1% CHAPS. Apparent kinetic parameters for UDP-Glc and Und-P were calculated using GraphPad Prism 9.3 software, Michaelis-Menten enzyme kinetics tool. K_m and V_{max} for UDP-Glc and Und-P are presented in the Table 1. The results are representative of three separate experiments. (b) Thin layer chromatography (TLC) of alkali-stable [^3H]Glc-lipids from *in vitro* incubations of the membranes of *S. mutans* mutants with UDP-[^3H]Glc. Membrane fractions from *S. mutans* WT, ΔsccN , ΔsccP , or $\Delta\text{sccN}\Delta\text{sccP}$ were incubated with UDP-[^3H]Glc and analyzed for [^3H]Glc-lipid synthesis by TLC following mild-alkaline

methanolysis as described in Methods. [^3H]Glc-lipids were detected by scanning with an AR2000 Bioscan Radiochromatoscanner. The results are representative of three independent experiments. (c) Reverse reactions of SccN and SccP to reform UDP-Glc from their respective enzymatic products and UDP. [^3H]Glc-P-Und_{SccN}, synthesized by SccN, and [^3H]Glc-P-Und_{SccP}, synthesized by SccP, were tested as substrates in the discharge reactions, containing solubilized, partially-purified SccN or SccP, as described in Methods. Columns and error bars represent the mean and S.D., respectively (biologically independent replicates). *P* values were calculated by two-way ANOVA with Šídák's multiple comparison test. (d) Origin of diagnostic ion fragments used to deduce the anomeric configuration of Glc-P-Und. If the stereochemistries of the phosphodiester at C1 of glucose and the 2-OH are configured *trans* (β -anomeric configuration) then [Und-HPO₄]⁻ (m/z =845.7) is the primary high molecular weight ion detected in the mass spectrum. However, if the stereochemistries of glucose at C1 and the 2-OH group are configured *cis* (α -anomeric configuration) then an additional ion fragment with the composition of [Und-PO₄-C₂H₃O]⁻ (m/z =887.7), formed by 'cross-ring' fragmentation as shown, is found (34).

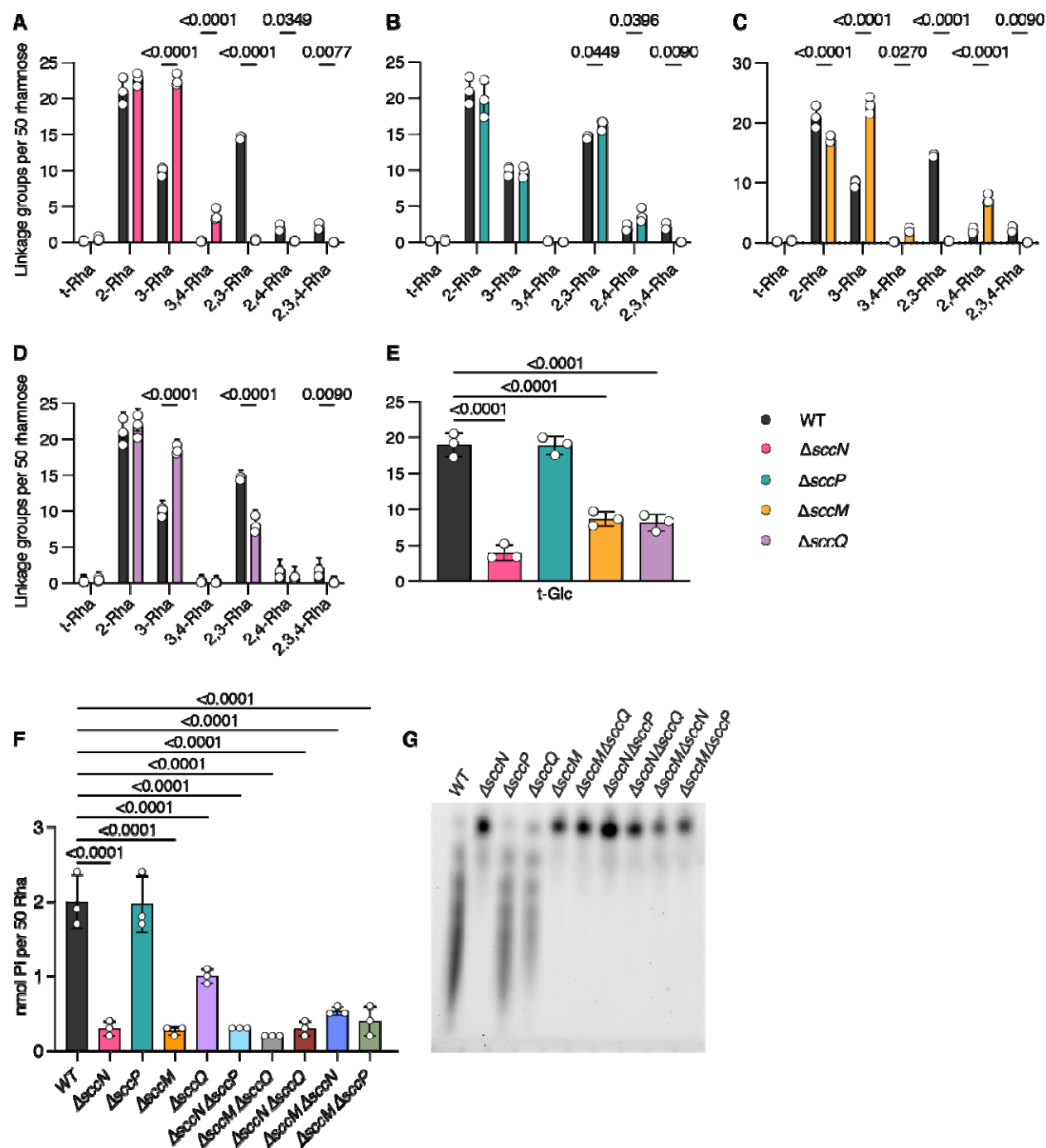


Fig. 3. Analysis of SCCs purified from *S. mutans* strains

Quantitative comparison of the linkage groups in WT SCC with (a) $\Delta sccN$ SCC, (b) $\Delta sccP$ SCC, (c) $\Delta sccM$ SCC and (d) $\Delta sccQ$ SCC. (e) Recovery of non-reducing terminal Glc in SCCs purified from *S. mutans* strains. For this illustration data are expressed as number of molecules normalized to a total of 50 Rha per polymer (the proposed average size of the polyrhamnose backbone). (f) Phosphate content of SCCs isolated

from *S. mutans* strains. Phosphate concentration was determined by the malachite green method following digestion with perchloric acid. Rha content was determined by a modified anthrone assay, as described in Methods. Data are expressed as nmol of phosphate in the sample normalized to 50 nmol Rha per molecule of SCC. In panels **a**, **b**, **c**, **d**, **e** and **f**, columns and error bars represent the mean and S.D., respectively (n = 3 independent replicates). *P* values were calculated by two-way ANOVA with Dunnett's multiple comparison test. Data used for plots shown in panels **a**, **b**, **c**, **d** and **e** are presented in Supplementary Table 2. (**g**) Fluorescence imaging analysis of SDS-PAGE of ANDS-labeled SCCs prepared from the indicated *S. mutans* strains. Representative image from at least three independent experiments is shown.

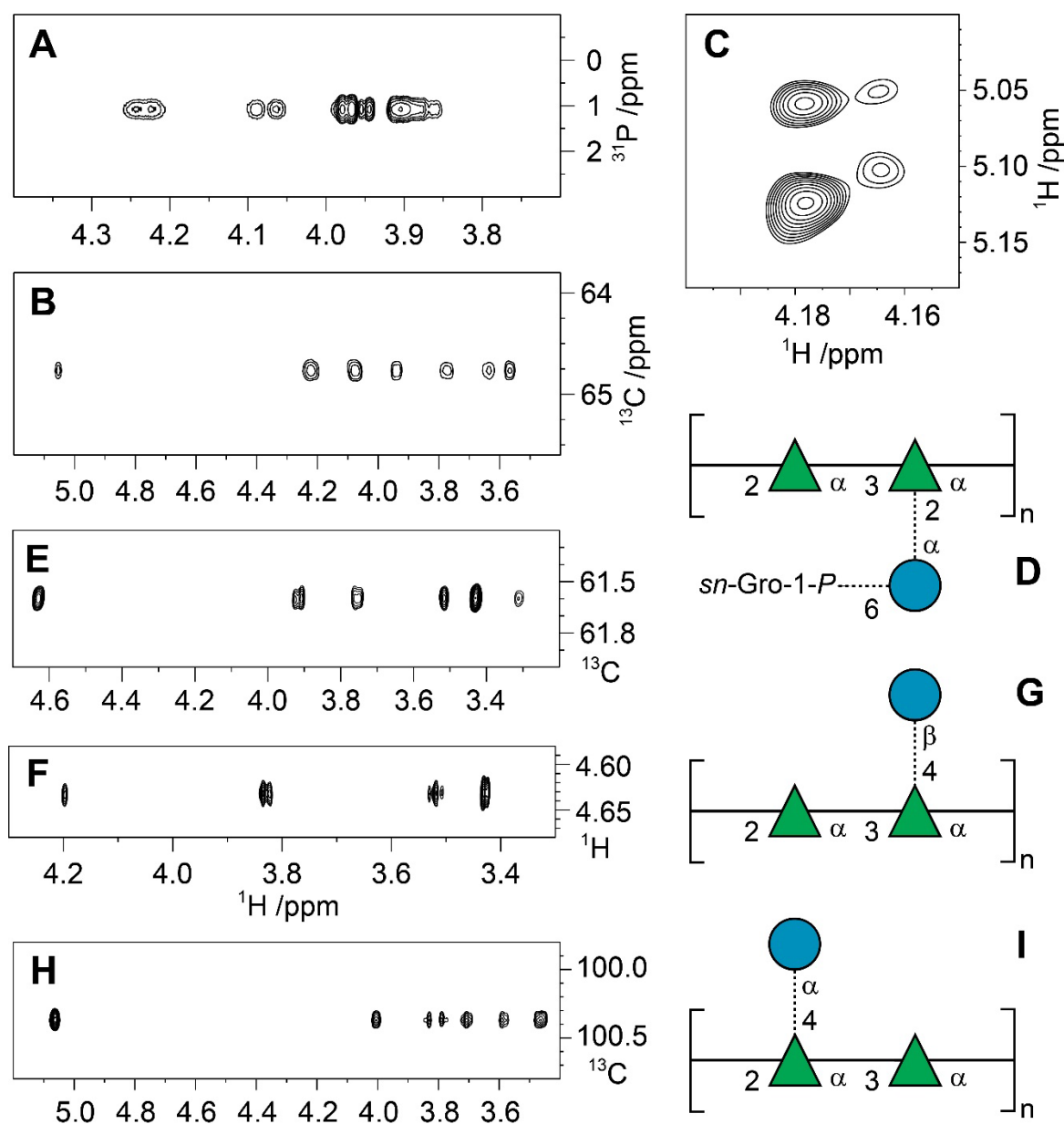


Fig. 4. Selected regions of NMR spectra and molecular models of alternative glucosyl-containing trisaccharide repeating units (RU) of SCCs

(a) ^1H , ^{31}P -HMBC NMR spectrum of the WT SCC identifying phosphodiester-linked entities. (b) Spectral region from a ^1H , ^{13}C -HSQC-TOCSY NMR experiment (mixing time of 200 ms) showing the seven-proton glucosyl spin-system of the α -D-Glcp6P(S)Gro-(1 \rightarrow 2)-linked side-chain residue in the WT SCC. (c) Spectral region from a ^1H , ^1H -

NOESY NMR experiment at 700 MHz (mixing time of 250 ms) depicting cross-peaks from anomeric protons in the trisaccharide repeating unit of WT SCC to H2 at ~4.18 ppm of the branched rhamnosyl residue (major) and the corresponding ones at ~4.16 ppm (minor) in which the RU contains an α -D-Glcp6P(S)Gro-(1→2)-linked side-chain residue. **(d)** Canonical structure of the major RU of WT SCC, where the dashed lines denote partial substitution by side-chain entities. **(e)** $^1\text{H}, ^{13}\text{C}$ -HSQC-TOCSY NMR spectrum (τ_{mix} 200 ms) showing the spin-system of the β -D-Glcp-(1→4)-linked side-chain residue in the ΔsccN SCC. **(f)** Spectral region from a $^1\text{H}, ^1\text{H}$ -NOESY NMR experiment at 700 MHz (τ_{mix} 200 ms) with cross-peaks from anomeric proton of the β -D-Glcp-(1→4)-linked side-chain residue in the ΔsccN SCC. The cross-peak at 4.20 ppm originates from H2 of the 2-linked rhamnosyl residue in the backbone of the SCC (cf. the 3D model in Figure S5). **(g)** Canonical structure of the RU containing the β -D-Glcp-(1→4)-linked side-chain residue, detected in the ΔsccN SCC. **(h)** Spectral region from a $^1\text{H}, ^{13}\text{C}$ -HSQC-TOCSY NMR experiment (τ_{mix} 200 ms) showing the seven-proton glucosyl spin-system of the α -D-Glcp-(1→4)-linked side-chain residue in the ΔsccM SCC. **(i)** Canonical structure of the RU containing the α -D-Glcp-(1→4)-linked side-chain residue, detected in the ΔsccM SCC. Sugar residues are depicted in SNFG-format with Rha as a green triangle and Glc as a blue circle.

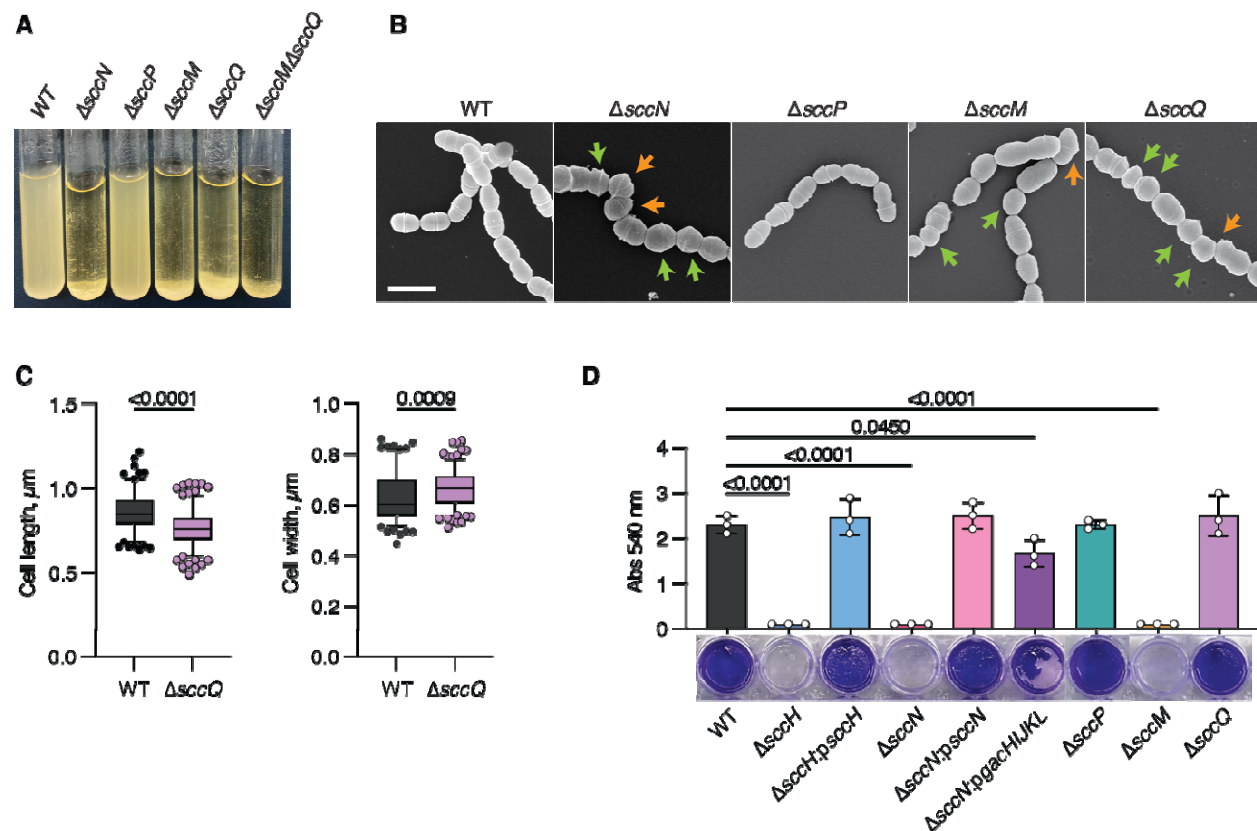


Fig. 5. Phenotypes of *S. mutans* strains defective in the SCC side-chain decorations

(a) Sedimentation phenotype of *S. mutans* WT and specific mutants after overnight growth in THY broth. (b) Scanning electron micrographs of *S. mutans* WT, $\Delta sccN$, $\Delta sccP$, $\Delta sccM$ and $\Delta sccQ$. Exponentially growing bacteria were fixed, dehydrated stepwise, and viewed by scanning electron microscopy (SEM). Orange arrows denote the cells with skewed division planes. Green arrows denote cells with irregular size. Scale bar is 1 μm . Representative images from at least three independent experiments are shown in **a** and **b**. (c) Distribution of cell length of cell population at mid-log growth phase for WT and $\Delta sccQ$. DIC images of bacterial cells were analyzed using the software ImageJ, ObjectJ plugin to quantify cell length and width. Total number of cells was $n = 144$ for WT, $n = 198$ for $\Delta sccQ$. Box plots show the median value (middle line), and 25%, 75% quartiles (boxes), and whiskers represent 5 – 95 percentile. *P* values were determined by unpaired two-tailed t-test with Welch correction. (d) Protein-based biofilm formation by *S. mutans* strains. Biofilms were incubated in UFTYE medium supplemented with 1% D-glucose for 24 h at 37 °C in presence of 5% CO₂ and analyzed

as outlined in Methods by crystal violet assay (Abs 540 nm). Column and error bars represent the mean \pm S.D. of three independent biological experiments. Significant differences ($P < 0.05$) are determined by one-way ANOVA with Dunnett's multiple comparisons test. Image of biofilms formed by *S. mutans* strains after crystal violet staining is representative image of three independent biological experiments.

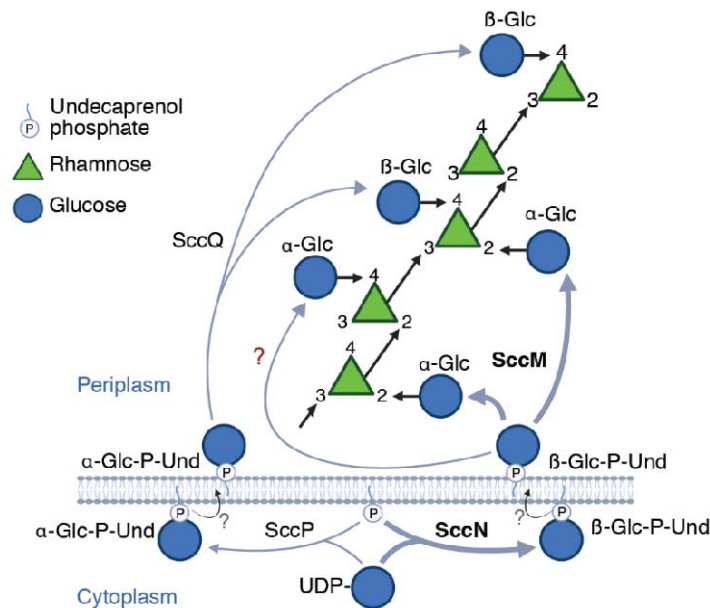


Fig. 6. Illustration of the proposed functions of SccN, SccP, SccM, SccQ and a hypothetical glucosyltransferase in modification of the polyrhamnose backbone with Glc side-chains

In this proposed model, α -Glc-P-Und and β -Glc-P-Und are synthesized on the cytoplasmic surface of the plasma membrane from Und-P and UDP-Glc, catalyzed by GT-A type glycosyltransferases encoded by SccP and SccN, respectively. Following synthesis, the Und-P linked intermediates diffuse transversely to the exoplasmic surface of the plasma membrane, mediated by hypothetical flippase(s)/scramblase(s), where they function as glucosyl donors in glucosylation reactions catalyzed by GT-C type glycosyltransferases, SccQ, SccM and an additional unidentified transferase. SccM uses β -Glc-P-Und to catalyze α -glucosylation of the 2-position of 3-Rha; SccQ uses α -Glc-P-Und to catalyze β -glucosylation of the 4-position of 3-Rha and of 2,3-Rha. α -glucosylation of the 4-position of 2-Rha appears to be catalyzed by a glycosyltransferase encoded outside of the SCC genetic locus, using β -Glc-P-Und, synthesized by SccN, as glucosyl donor.