Structural and Functional Analysis of a New Subfamily of Glycosyltransferases Required for Glycosylation of Serine-rich Streptococcal Adhesins*\$\

Received for publication, December 2, 2010, and in revised form, May 25, 2011 Published, JBC Papers in Press, June 7, 2011, DOI 10.1074/jbc.M110.208629

Fan Zhu*, Heidi Erlandsen†, Lei Ding†, Jingzhi Li†, Ying Huang†, Meixian Zhou†, Xiaobo Liang‡, Jinhiao Ma§, and Hui Wu†1

From the Departments of† Pediatric Dentistry and Microbiology, § Periodontology, ¶ Biochemistry and Molecular Genetics, and ¶ Cell Biology, University of Alabama at Birmingham, Schools of Dentistry and Medicine, Birmingham, Alabama 35294

Serine-rich repeat glycoproteins (SRRPs) are a growing family of bacterial adhesins found in many streptococci and staphylococci; they play important roles in bacterial biofilm formation and pathogenesis. Glycosylation of this family of adhesins is essential for their biogenesis. A glucosyltransferase (Gtf3) catalyzes the second step of glycosylation of a SRRP (Fap1) from an oral streptococcus, Streptococcus parasanguinis. Although Gtf3 homologs are highly conserved in SRRP-containing streptococci, they share minimal homology with functionally known glycosyltransferases. We report here the 2.3 Å crystal structure of Gtf3. The structural analysis indicates that Gtf3 forms a tetramer and shares significant structural homology with glycosyltransferases from GT4, GT5, and GT20 subfamilies. Combining crystal structural analysis with site-directed mutagenesis and in vitro glycosyltransferase assays, we identified residues that are required for UDP- or UDP-glucose binding and for oligomerization of Gtf3 and determined their contribution to the enzymatic activity of Gtf3. Further in vivo studies revealed that the critical amino acid residues identified by the structural analysis are crucial for Fap1 glycosylation in S. parasanguinis in vivo. Moreover, Gtf3 homologs from other streptococci were able to rescue the gtf3 knock-out mutant of S. parasanguinis in vivo and catalyze the sugar transfer to the modified SRRP substrate in vitro, demonstrating the importance and conservation of the Gtf3 homologs in glycosylation of SRRPs. As the Gtf3 homologs only exist in SRRP-containing streptococci, we conclude that the Gtf3 homologs represent a unique subfamily of glycosyltransferases.

Serine-rich repeat glycoproteins (SRRPs) are a growing family of bacterial adhesins found in many streptococci, and other Gram-positive bacteria (1). They have been shown to be important for bacterial biofilm formation and pathogenesis. A number of SRRPs including GsbP (2) and Hsa (3) of Streptococcus gordonii and SrpP of Staphylococcus aureus (4) have been implicated in pathogenesis of infective endocarditis. This family of streptococcal adhesins also mediates other infectious processes. PsrP is associated with pathogenesis of invasive pneumococcal disease by Streptococcus pneumoniae (5). Srr-1 is involved in crossing the blood-brain barrier by Streptococcus agalactiae to induce meningitis (6); Srr-2 is required for neonatal infection by S. agalactiae (7). Furthermore, vaccination against Srr-2 prevents maternal disease transfer caused by S. agalactiae COH1 in a neonatal animal model (8); immunization of experimental animals by PsrP prevents invasive pneumonia induced by S. pneumoniae in a mouse model (5).

Fimbriae-associated protein 1 (Fap1) from an oral streptococcus, Streptococcus parasanguinis, was the first SRRP identified (9, 10). As its biological function and biosynthetic pathways have been studied extensively (1), Fap1 has emerged as a model to investigate the SRRP family of bacterial adhesins (11). Fap1 is required for fimbrial assembly, bacterial adhesion, and biofilm formation (1). A cluster of 11 genes flanking the fap1 gene locus is responsible for biogenesis of Fap1. A core region of seven genes located downstream of fap1 is responsible for Fap1 glycosylation and secretion (12). Gtf1 and Gtf2 catalyze the first step of glycosylation by transferring GlcNAc residues to the Fap1 polypeptide. These homologous core genes are highly conserved in every Gram-positive bacterium that harbors SRRP (1). A cluster of four genes located upstream of fap1 shares significant homology with genes involved in carbohydrate synthesis, mostly glycosyltransferases (13). For instance, Gtf3 is a glycosyltransferase (14). Gtf3 homologs have been found in many other streptococci shown to be important for glycosylation of GsbP (15) and Srr-1 (16) and are involved in bacterial adhesion and biofilm formation. The homologous gene from S. agalactiae has also been implicated in bacterial pathogenesis, albeit its biochemical properties are not characterized. Gtf3 homologs were initially designated as putative nucleotide sugar synthases (Nss) because they share some homology to Nss from Escherichia coli (17). Our recent studies have revealed that Gtf3...
is actually a glycosyltransferase, which catalyzes the transfer of glucose residues to the GlcNAc-modified Fap1 polypeptide and mediates the second step of the Fap1 glycosylation (14).

A large number of glycosyltransferases (GTs) have been deposited in the public databases, and the number continues to grow as several human microbiome projects release their sequencing data periodically. Compared with the wealthy genomic information, there is a lack of biochemical and gene expression studies for this large family of enzymes. For instance, it is largely unknown what sugar donors and substrate acceptors are required for the majority of deposited GTs. Only limited numbers of three-dimensional structures from different GT families have been reported. All solved GT structures belong to two subfamilies, GT-A and GT-B (18). GTs are functionally diverse enzymes. Among the GTs that have three-dimensional structures, some catalyze synthesis of secondary metabolites (19). Others mediate biogenesis of bacterial cell walls or polysaccharides (20, 21). There are limited numbers of reports documenting bacterial glycosyltransferases that are involved in protein glycosylation, albeit many structures for GTs that catalyze protein glycosylation in eukaryotes are available (22, 23). Primary sequence predictions suggest that Gtf3 belongs to the GT-B subfamily; however, an attempted homology modeling failed to construct a useful three-dimensional model due to the low sequence homology (less than 14% identity). This subfamily of glycosyltransferases is only present in streptococci that produce SRRP (14), suggesting Gtf3 represents a unique subfamily of GT-B. Structural analysis of Gtf3 will provide new insights into this important glycosyltransferase.

In this study, we report the x-ray crystal structure of Gtf3. Structure-based mutagenesis and functional studies have revealed key residues that are involved in UDP or UDP-glucose binding and also identified protein–protein interaction regions that are crucial for oligomerization and the enzymatic activity. We also demonstrated that these key residues and domains are critical for glycosylation of Fap1 in vivo and that the Gtf3 homologs from other streptococci are functionally conserved.

**MATERIALS AND METHODS**

**Bacterial Strains and Plasmids—**All strains and plasmids constructed and used in this study were listed in supplemental Table S1.

**Protein Expression, Purification, and Crystallization—**The full-length gene gtf3 was amplified from genomic DNA of *S. parasanguinis* FW213 using primer set Gtf3-HindIII-1F and Gtf3-Xhol-987R (supplemental Table S2). The PCR product was purified and cloned into pET-SUMO. The resulting plasmid pET-SUMO-gtf3 was transformed into *E. coli* BL21 Gold (DE3) cells. The recombinant strain grown to A600 0.6 in LB medium was induced with 0.1 mM isopropyl β-D-1-thiogalactopyranoside at 18 °C overnight. Selenomethionine (Se-Met)-substituted Gtf3 was produced using a similar protocol except that the complete amino acid medium with Se-Met substituted the LB medium and that the induction was carried out at 25 °C overnight. Bacterial cells were harvested by centrifugation, resuspended in binding buffer (20 mM Tris–HCl pH 8.0, 500 mM NaCl, and 25 mM imidazole), and lysed under high pressure using an Emulsiflex C3 high pressure homogenizer (Avestin). The clear cell lysates obtained after centrifugation were subjected to protein purification using HiTrap™ Column (Ni2+ affinity). Proteins were eluted from the affinity resin by elution buffer (20 mM Tris–HCl, pH 8.0, 500 mM NaCl, and 500 mM imidazole). The N-terminal His-SUMO tag was cleaved by incubating the elution fractions with SUMO protease, ULP (ubiquitin-like protein protease), during overnight dialysis at 4 °C against 20 mM Tris–HCl, pH 8.0, 500 mM NaCl. Diazyme protein samples were reapplied to HiTrap™ Column (Ni2+ affinity) to remove ULP, the cleaved His tag, and uncleaved proteins. Flow-through was collected and applied for anion exchange chromatography on a Resource Q column equilibrated with Q buffer A (20 mM Tris–HCl, pH 8.0, 100 mM NaCl, 0.5 mM Tris(2-carboxyethyl)phosphine). The protein samples were eluted with Q buffer B (20 mM Tris–HCl, pH 8.0, 1 mM NaCl, 0.3 mM Tris(2-carboxyethyl)phosphine) using a linear gradient. Fractions containing the target proteins were pooled and concentrated for gel filtration. The gel filtration experiments were carried out on a HiLoad 16/60 Superdex™ 75 preparation grade column connected to an AKTA-purifier FPLC system (GE Healthcare). Before protein loading, the column was equilibrated with buffer G (10 mM Tris–HCl, pH 8.0, 100 mM NaCl, 0.2 mM DTT). Protein purity was analyzed by SDS-PAGE. Peak fractions were collected and concentrated to 30 mg/ml for crystallization screen.

Se-Met Gtf3 was crystallized at 20 °C by using a hanging-drop vapor-diffusion method. Each drop contains 1 μl of protein solution (15 mg/ml protein and 10 mM UDP-glucose) with 1 μl of reservoir (0.1 M succinic acid, pH 7.0, 13% polyethylene glycol 3350, 10% glycerol). For native Gtf3, crystals were obtained using the same condition except for the components of the reservoir (0.1 M succinic acid, pH 7.0, 11% polyethylene glycol 3350, 10% glycerol). The crystals were cryoprotected by the addition of 25% glycerol and flash-frozen in liquid nitrogen.

**Data Collection, Phasing, and Structure Determination—**Single-wavelength anomalous diffraction data were collected on a Se-Met/native crystal at the Argonne National Laboratory beam line SER-CAT ID-22. The diffraction data were processed and scaled with the HKL package (24). The data processing statistics are summarized in Table 1.

Se-Met structure was determined through single-wavelength anomalous diffraction method using Phenix system. For Se-Met Gtf3 there are two molecules in an asymmetric unit in the C222₁ crystal form. The native Gtf3 structure in the orthorhombic space group P2$_1$2$_1$2$_1$ was determined to 2.3 Å using molecular replacement. The Matthews coefficient $V_M$ for 2.4 Å$^3$ Da$^{-1}$ corresponds to a solvent content of 48.2%.

**Refinement and Model Building—**All refinement was carried out with Refmac5 in CCP4 (25). Five percent of the data were randomly selected and set aside for free R-factor cross validation calculations. Rigid-body refinement was followed by restrained, isotropic refinement. In between refinement cycles, the $2F_{o} - F_{c}$ and the $F_{o} - F_{c}$ maps were used for manual rebuilding of the peptide chain as well as for the addition of solvent molecules in COOT (26). The refinement was continued until convergence of the free R-factor. Details of the final refinement are given in Table 1. Atomic coordinates and struc-
Structural Insights into a New Subfamily of GTs

| TABLE 1 | Crystallographic data collection, processing statistics, structure refinement, and model statistics |
|------------------|--------------------------------------------------|
| **Se-Met (C2221)-UDP** | **Native (P2,2,21)-UDP** |
| **Data collection** | **Se-Met (C2221)-UDP** | **Native (P2,2,21)-UDP** |
| Wavelength (Å) | 0.979 | 1.033 |
| Unit cell parameters (Å) | | |
| a | 77.18 | 77.22 |
| b | 193.86 | 99.24 |
| c | 96.25 | 188.19 |
| α | 90 | 90 |
| β | 90 | 90 |
| γ | 90 | 90 |
| Resolution range (Å)* | 30.00–1.90 (1.97–1.90) | 30.00–2.30 (2.34–2.30) |
| No. of unique observations | 57280 | 63333 |
| Completeness (%)* | 99.9 (97.6) | 96.3 (91.5) |
| Rmerge (%)* | 13.2 (96.3) | 6.8 (52.0) |
| I/σI | 16.8 (2.5) | 28.9 (2.4) |
| Rwork | 7.4 (7.3) | 5.7 (4.9) |

| **Refinement** | **Se-Met (C2221)-UDP** | **Native (P2,2,21)-UDP** |
| Resolution | 30.00–1.90 (1.97–1.90) | 42.5–2.3 (2.34–2.30) |
| Rwork* | 0.181 (0.241) | 0.187 (0.259) |
| Rfree* | 0.223 (0.292) | 0.250 (0.335) |
| No. of reflections | 57262 (5662) | 63229 (5227) |

| **Model quality** | **Se-Met (C2221)-UDP** | **Native (P2,2,21)-UDP** |
| Estimated coordinate error (Å) | 0.22 | 0.37 |
| r.m.s.d. bonds (Å) | 0.007 | 0.008 |
| r.m.s.d. angles (°) | 1.107 | 1.145 |
| Mean B value (protein/UDP/waters) | 23.23 | 41.82 |
| Mean B value for UDP | 41.95 | 46.0 |
| Ramachandran plot | | |
| No. of residues in total | 656 | 1507 |
| No. of residues in most favored region | 547 (93.7%) | 1074 (92.0%) |
| No. of residues in additionally and generously allowed regions | 35 (6%) | 89 (7.7%) |
| No. of residues in outliers | 2 (0.3%) | 4 (0.3%) |

* Values in parentheses are for the highest resolution shell.

Gtf3 Variants and In Vitro Glycosylation Assays—The mutated Gtf3 protein variants were purified using glutathione-Sepharose 4B beads (Amersham Biosciences) and used as enzymes for in vitro glycosyltransferase assays as described previously (14).

To examine the importance of the dimer interface located at the C terminus of Gtf3, a truncated Gtf3A314–330 was created using pGEX-gtf3 as a template, with a primer pair, Gtf3-BamHI-1F and Gtf3-XhoI-939R (supplemental Table S2). The amplified PCR product was then digested with the restriction enzymes BamHI and Xhol and inserted into the same restriction sites of pGEX-6P-1 to produce a fusion plasmid, pGEX-gtf3A314–330. The GST-Gtf3A314–330 protein was expressed in E. coli Top10 and purified as described above.

To examine the functional complementation of Gtf3, gtf3 homologs from S. agalactiae COH1, S. pneumoniae TIGR4, and Streptococcus sanguinis SK36 were amplified from the genomic DNA of each strain using primer sets with BamHI and Xhol sites (supplemental Table S2), respectively. The amplified PCR products were then ligated into pGEX-6P-1 to produce fusion plasmids, pGEX-gtf3 (GBS), pGEX-gtf3 (SK36), and pGEX-gtf3 (TIGR4). The recombinant proteins were expressed from each fusion strain and purified and used as an enzyme source for in vitro glycosyltransferase assays.

In Vivo Genetic Complementation and Fap1 Glycosylation—Gtf3 and mutated Gtf3 alleles were amplified with primers Gtf3-Sall 5’ and Gtf3-KpnI 3’ (supplemental Table S2) using pGEX-gtf3 or mutated pGEX-gtf3 constructs as templates. Amplified PCR products were digested with Sall and KpnI and...
cloned in-frame into *E. coli*-streptococcal shuttle vector pVPT-CHSV. The *gtf3* homologs from *S. agalactiae* COH1, *S. pneumoniae* TIGR4, and *S. sanguinis* SK36 were amplified from each strain using primer sets engineered with Sall and BamHI sites (supplemental Table S2). Amplified PCR products were digested with Sall and BamHI and cloned in-frame into pVPT-CHSV. Plasmids carrying correct fusion constructs were transformed into the *gtf3* mutant of *S. parasanguinis*. The ability of Gtf3 and Gtf3 derivatives to restore production of mature Fap1 was examined using Western blotting analysis with Fap1 peptide-specific monoclonal antibody E42, glycan-specific antibody D10, and mature Fap1-specific antibody F51.

**RESULTS AND DISCUSSIONS**

The Overall Structure of Gtf3—Two different space groups C222₁, and P2₁2₁2₁ were obtained for the Gtf3 protein. Se-Met Gtf3 and native Gtf3 protein both form a similar tetramer. In this paper we used the native Gtf3 tetrameric structure as the final structure for study.

The Gtf3 structure presented here consists of a tetramer with two very similar dimers (the r.m.s.d. value is only 0.30 Å for the superimposition of the dimer consisting of molecules AD onto the BC dimer in the Gtf3 tetramer) (Fig. 1A). Pairwise comparisons of the individual subunits within the Gtf3 tetramer result in r.m.s.d. values of between 0.25 Å (molecule A compared with molecule B) and 0.58 Å (molecule B compared with molecule C) (Fig. 1C).

For each monomer it has the canonical GT-B-fold with two Rossmann-like β-α-β domains that have a central β-sheet of 6 strands and surrounding helices and loops in each of the N- and C-terminal domains (18, 31). The N-terminal domain (comprising residues 1–154) and the C-terminal domain (comprising residues 155–330) are separated by a crevice in the interdomain region (Fig. 1B). The C terminus forms a long, kinked, α-helix (residues 293–330) that folds over onto the N-terminal domain (Fig. 1B). We detected well defined density for one molecule of UDP in each monomer of the tetramer (Fig. 1D). However, there is no detectable electron density for the glucose moiety of the UDP-glucose substrate in any of the four Gtf3 molecules in the tetramer, indicating that the UDP-glucose substrate may have been turned over. The structure presented here is, therefore, of the binary Gtf3-UDP (product) complex. The UDP molecule is found in the same pocket as GDP-mannose in the structure of PimA (PDB code 2GEJ) (32), albeit the two molecules do not superimpose completely (Fig. 1, E and F). This structure was, therefore, used as a model to show approximately how the active site residues will interact with the intact substrate (Fig. 2, A and B).

A Ramachandran plot shows that 99.4% (1163/1507) of all residues in the tetramer are in the favored regions (supplemental Fig. S1). Thr-77 in all molecules was in the disallowed regions on the Ramachandran plot (same as Se-Met supplemental Fig. S2). Thr-77 is located in a tight loop just before an α-helix and is thus forced into a disallowed conformation due to both crystal-packing interactions.

A DALI search using molecule A from Gtf3 revealed numerous hits for structurally similar proteins, mainly glycosyltransferases, as well as other sugar-converting enzymes (see supplemental Fig. S3 and Table S3). Together these search results indicate that Gtf3 shares significant structural homology to GT-B glycosyltransferases of GT-4, GT-5, or GT-20 subfamily.

There is no metal or divalent cation observed in the active site of the structure close to the α- or β-phosphate groups of the UDP. Furthermore, Gtf3 still possesses its full enzymatic activity without any metal in its *in vitro* glycosyltransferase assay (data not shown), indicating that it is indeed not dependent upon metal for its glucose transfer activity. Structural analysis of a number of GT-B family glycosyltransferases failed to reveal a UDP-Glc binary complex (19, 33). In fact, T4 phage βGT has been shown experimentally to cleave UDP-Glc (34). Future studies using an inactive UDP-Glc analog to determine a complex structure would help to define the sugar donor binding site and provide further insights into the glucose transfer mechanism.

Gtf3 is most likely a retaining glycosyltransferase, as it shares high structural similarity to retaining GT-B-fold family GT-4, GT-5, and GT-20 glycosyltransferases based on a DALI search (supplemental Table S3) and information in the CAZy database (35). In the Gtf3 structure we observe an interaction between Arg-179 (both NH1 and NH2) and both the α- and β-phosphate oxygen atoms of UDP at a distance of 3.2 Å and 3.0 Å, respectively (Fig. 2A). The side chain density of Arg-179 is high in molecules B and D but is very low in molecules A and C in the Gtf3 tetramer (no density for Cβ to C8 and not well defined NH1 and NH2 densities), likely a result of movement during catalysis, which may have occurred in the substrate-co-crystallized crystals. This type of interaction between an arginine side chain and the phosphates is considered a conserved feature of the retaining glycosyltransferase (33, 36, 37). The arginine side chain presumably neutralizes the negative charge that develops on the phosphate during catalysis. Indeed, upon mutation of Arg-179 to alanine, the enzyme loses all glucose transfer activity (Fig. 2C). In addition, Lys-246 (Nε atom) is also observed close to the α-phosphate oxygen of the UDP moiety at a distance of ~3.3 Å, close enough to possibly interact with the UDP molecule. Similar to the mutation of Arg-179, the mutation of Lys-246 to an alanine renders Gtf3 catalytically inactive (Fig. 2C). These data suggested that both residues mediate neutralization of the negative charge on the α- and β-phosphate groups of UDP during catalysis and also may help UDP to leave the active site after catalysis. Sequence alignment among Gtf3 homologs revealed that Arg-179 and Lys-246 are invariant (Fig. 2D), supporting their importance in potential binding to UDP or UPD-Glc. The Tyr-211 phenyl ring is stacked parallel to the uracil ring of UDP (3.4 Å distance) (Fig. 2A). A mutation to alanine (Y221A) inhibited 75% of glycosyltransferase activity (Fig. 2C), presumably because it abolishes stacking interactions and, therefore, lowers Gtf3 binding affinity to the UDP-glucose substrate. The uracil ring of UDP is buttressed by Tyr-211 on one side and Leu-217 on the other side by a distance of 4.1 Å. Interestingly, Ser-249 hydrogen bonds to UDP through the sugar moiety (atom O3’) at a distance of 2.7 Å. Although Ser-249 may form an H-bond with His-156 (2.9 Å) and pack next to the UDP diphosphates, the S249A mutant did not affect Gtf3 *in vitro* glycosyltransferase activity (Fig. 2C), indicating a more important catalytic involve-
ment observed for the other residues (Arg-179, Lys-246, and Tyr-211). It is worth noting that Ser-249 is not conserved at all in Gtf3 homologs, Tyr-211 is semi-conserved, and only Arg-179 and Lys-246 are absolutely conserved (Fig. 2D). The conservation of amino acid residues further confirms the structural prediction and is consistent with in vitro glycosyltransferase activities as well.

Most glycosyltransferases (both GT-A and GT-B-fold) have at least one flexible region/loop close to the sugar-nucleotide binding site that plays an important role during catalysis (18, 31). In the known glycosyltransferase sugar-nucleotide-bound structures, this flexible loop is either in a closed conformation, covering the bound donor sugars, or is found disordered in the vicinity of the sugar-nucleotide binding site. This loop has been suggested to act as a lid covering the bound sugar-nucleotide to aid in catalysis. In the Gtf3 structure we observed lower electron density in some side chains (Arg-179, Phe-180, and Met-231) as well as two longer loop regions (Met-106 to Phe-111 and Met-231 to Glu-237) in the region covering the active site and near where the glucose moiety in the UDP-glucose substrate...
should bind. This disorder may be due to the fact that the crystals were co-crystallized with UDP-glucose substrate that has been converted into UDP and perhaps the glucose has been released during crystal growth. There are a number of differences in the active site for these disordered loops when comparing the four molecules in the Gtf3 tetramer; most notably, molecule D has an extra “blob” of density close to UDP, which could indicate a presence of UDP-glucose remnants in some of the molecules in the crystal. However, the density is not high enough to model in glucose. The active site in our Gtf3 structure is very open and solvent-accessible, which could indicate that the presented structure is in its “open” conformation, and the conformational changes might occur upon UDP-glucose binding, which involves closing of the active site. An open active site is in sharp contrast with the narrow cavity of the structurally similar glycosyltransferase OtsA from family GT-20 (38). This difference could be due to the fact that OtsA utilizes a single sugar as an acceptor molecule, whereas Gtf3 must accommodate the larger acceptor GlcNAc-modified Fap1 protein to gain access to the UDP-glucose molecule for glucose transfer to occur during catalysis. Our future studies will explore how the acceptor, GlcNAc-modified Fap1 polypeptide, interacts with the active site of Gtf3 and will identify crucial residues for the interaction.

Oligomerization Interfaces in the Gtf3 Tetramer — The structure of Gtf3 with UDP bound is a tetramer consisting of two dimers (Fig. 1). The dimer interactions consist of mostly non-bonded contacts but also hydrogen bonds and a few salt bridges. AD and BC interfaces share similar features with the buried interface of 1100 Å². Structural analysis revealed that the C-terminal domain is involved in the dimer interface (Fig. 3, A and B); thus, it may play a role in Gtf3 function or activity. The C terminus has been shown to be important for the activity of some glycosyltransferases (39). A glycosyltransferase Waal from E. coli (GT-A-fold, family 8) displays a dramatic reduction in enzymatic activity upon deletion of only 5–12 residues at the C terminus (40). More interestingly, a homodimer human 1,3-glucuronyltransferase (GlcAT) (PDB code 3CU0) (GT-B-fold, family 43) dimerizes through the C-terminal end, specifically the C-terminal loop and last helix. However, none of these enzymes has the same involvement of the C-terminal long kinked helix as seen in Gtf3.

To test whether the C terminus is important for Gtf3 enzymatic activity, a Gtf3-deficient mutant that is devoid of the C-terminal region (314–330) was constructed and examined for in vitro enzymatic activity. Gtf3 without the interactive domain failed to transfer glucosyl residues to GlcNAc-modified Fap1 sub-
strate in vitro (Fig. 3C), suggesting the importance of the interface in Gtf3 enzymatic activity.

Based on structural prediction and sequence alignment, further site-directed mutagenesis within the C-terminal region was carried out to define residues that are critical for the Gtf3 enzymatic activity. Two mutations, F315A and L320A, completely inactivated the enzymatic activity of Gtf3, whereas two others, R318A and F314A, significantly reduced the activity (Fig. 3C). These data further support the idea that the C terminus is important for the Gtf3 activity. Because the structure shows that the C terminus mediates some of the protein oligomerization interactions in the Gtf3 tetramer, we next determined whether the site-directed mutants within the C-terminal region altered the oligomeric status of Gtf3. Native PAGE analysis was used to determine the oligomerization state of Gtf3. Wild-type Gtf3 mainly migrated at an apparent molecular mass
of 152 kDa (Fig. 3D), which corresponds to the tetramer. The formation of the tetramer was not dependent on the protein concentrations (data not shown), suggesting that the tetrameric structure is physiologically relevant for Gtf3. Interestingly, two mutants, R318A and F314A, that had reduced enzymatic activity mainly existed as a dimer (around 76 kDa). Furthermore, the other two mutants, F315A and L320A, which completely lost the enzymatic activity, did not form dimer or tetramer; they only existed as a monomer, suggesting that the tetramer formation is crucial for Gtf3 enzymatic activity. Both Phe-314 and Phe-315 are located at the hinge between the two domains in one monomer, and are both involved in interactions at the AD dimer interface (Fig. 3B). Phe-315 (in molecule A) π-stacks with Phe-325 (in molecule D), and the two form a sort of triad with Arg-318 (in molecule A), whereas Phe-314 stacks along Arg-317 (in molecule A) and is close to Val-3 and Val-324 (in molecule D) (Fig. 3B). Mutation of both Phe-314 and Phe-315 to alanine resulted in a loss of Gtf3 tetramer formation, which could be due to a loss of mediation of conformational changes that might occur between molecules in the dimer/tetramer, as needed for catalysis (a potential positive cooperativity). F315A resulted in a complete loss of activity, whereas F314A still retained some activity, indicating that loss of Phe-315 interactions at the dimer interface is more detrimental than loss of Phe-314 interactions. A possible reason is that Phe-315 (molecule A) has a phenyl-phenyl π-stacking interaction with Phe-325 (molecule D); the loss of this interaction might lead to less AD dimer association and, therefore, a loss of catalytic activity.

Arg-318 is located on the surface of the AD dimer. As mentioned above, Arg-318 π-stacks with Phe-325 but also forms a salt bridge with Glu-322 in another molecule in the AD and BC dimer (Fig. 3B). The R318A mutation potentially disrupts H-bonding and salt bridge network at the dimer interface, thereby affecting the oligomeric status of Gtf3. Leu-320 is located to a hydrophobic core in between the two domains of the monomer. It may be important for proper active site formation; thus the L320A mutation could lead to improper formation of an active site due to loss of conformational changes needed upon substrate binding.

Sequence alignment of the C-terminal motifs of Gtf3 homologs from diverse streptococci revealed that Phe-315 and Leu-320 are absolutely conserved, whereas Phe-314 and Arg-318 of Gtf3 can be substituted with similar amino acid residues Tyr and Lys, respectively (Fig. 3E). It is difficult to explain the complete loss of activity of the F315A and L320A mutants concurrently with the partial loss of activity from the F314A and R318A mutants based on the three-dimensional structure; the conservation of amino acid residues Phe-315 and Leu-320 may account for the discrepancy. Further enzyme kinetic studies of those Gtf3 variants may shed more mechanistic lights on how the mutations impair Gtf3 enzyme kinetics. As we have demonstrated, oligomerization plays an important role in Gtf3 enzymatic activity. Next, we wanted to determine whether the UDP binding deficiency would have any impact on oligomeric status. Native PAGE analysis of putative UDP binding mutant K246A revealed that K246A still existed as a tetramer (Fig. 3D), albeit the K246A mutation completely abolished the enzymatic activity (Fig. 2C). These data indicated that the UDP binding by Gtf3 plays a critical and distinct role in the Gtf3 function. Taken these together, we conclude that both UDP binding and oligomerization are essential for the Gtf3 enzymatic activity.

Key Amino Acid Residues for the Gtf3 Activity in Vitro Are Important for Fap1 Glycosylation in Vivo in S. parasanguinis—It is known that Gtf3 mediates the second step of Fap1 glycosylation (14). To examine the impact of Gtf3 mutations on Fap1 glycosylation in vivo, we performed in vivo complementation experiments in S. parasanguinis. Western blotting analysis revealed that the Gtf3 mutants R179A, K246A, and L320A (which lost Gtf3 enzymatic activity in vitro) failed to complement the gtf3 mutant. Like the gtf3 null mutant, they only produced a higher molecular mass protein when probed with mAbE42 (Fig. 4, top panel). This higher molecular mass protein did not react with mature Fap1-specific antibody F51 and gly-
can-specific antibody D10 (Fig. 4, middle and bottom panels), demonstrating that these three mutants are defective in Fap1 glycosylation. Two mutants, Y211A and R318A (which had a reduced enzymatic activity), partially complemented the gtf3 mutant. They produced a small amount of mature Fap1 (Fig. 4, lanes 7 and 10) and the higher molecular mass protein. By contrast, the D214A mutant (which retained Gtf3 activity in vitro) completely restored the Fap1 glycosylation. It displayed the same phenotype as wild-type S. parasanguinis and the fully complemented gtf3 mutant (Fig. 4, lane 8). These data demonstrate that the conserved amino acid residues critical for the Gtf3 function in vitro are also important for Fap1 glycosylation in S. parasanguinis in vivo.

Functional Conservation of Gtf3 Homologs in Other Streptococci—Gtf3 homologs are highly conserved in streptococci. To determine whether they are functionally conserved as well, we cloned genes coding for the Gtf3 homologs from S. agalactiae COH1, S. sanguinis SK36, and S. pneumoniae TIGR4 and tested their ability to complement the gtf3 mutant of S. parasanguinis. Western blotting analysis indicated that the Gtf3 homolog from S. agalactiae COH1 restored Fap1 glycosylation (Fig. 5A, lane 6). However, the Gtf3 homologs from S. pneumoniae TIGR4 and S. sanguinis SK36 failed to rescue Fap1 glycosylation (Fig. 5A, lanes 7 and 8). To further determine the function of these Gtf3 homologs, we performed in vitro glycosyltransferase assays. The Gtf3 homolog from S. agalactiae COH1 was able to transfer glucosyl residues to GlcNAc-modified Fap1 substrate. The Gtf3 homolog from S. sanguinis SK36 exhibited lower but significant enzymatic activity toward the recombinant Fap1 substrate. However, the Gtf3 homolog from S. pneumoniae TIGR4 did not exhibit any significant enzymatic activity at all (Fig. 5B). To determine whether this is due to the substrate specificity as the SRRP from S. pneumoniae is quite different from Fap1 of S. parasanguinis, we generated a recombinant SRRP, PsrP from S. pneumoniae TIGR4 and modified the recombinant protein by the Gtf1 and Gtf2 homologs of S. pneumoniae. The modified substrate of S. pneumoniae was used to assay the glycosyltransferase activity of the Gtf3 homolog from S. pneumoniae. Indeed, the Gtf3 homolog was capable of the glucosyltransferase activities.
syl transfer (Fig. 5C). These data demonstrate that the Gtf3 homologs are functionally conserved.

Much structural information is available for glycosyltransferases that catalyze the transfer of sugars to antibiotics, glycolipids, and oligosaccharides; however, little is known on atomic structure of bacterial glycosyltransferases, which catalyze the transfer the sugars to glycoproteins. Here we report a high resolution atomic structure of Gtf3, which is responsible for the second step of glycosylation of a SRPP from S. parasanguinis. Our structural analysis suggests that Gtf3 shares high homology with diverse subfamily of glycosyltransferases including GT-4, GT-5, and GT-20, albeit it is most similar to a GT-4 family, mannosyltransferase PimA. The classification of CAZY families of glycosyltransferase is based on sequence similarity. Gtf3 shares less than 13% identity with PimA and other structurally similar glycosyltransferases. Thus, structure comparison via experimentally determined atomic structures is a powerful tool to define structure and functional relationship of glycosyltransferases.

Intriguingly, sequence alignment analysis revealed that the Gtf3 homologs share 40–65% identity and are only found in streptococci that possess SRPP. Functional analysis by in vitro glycosyltransferase assays and in vivo complementation demonstrated that the Gtf3 homologs possess a conserved glycosyltransferase activity that catalyzes the transfer of Glc from UDP-Glc to GlcNAc modified SRPPs. As the glycosylated protein substrates are much larger than substrates by glycan chains, oligosaccharide, or lipid intermediates, the Gtf3 homologous enzymes may be evolved to adapt to an open active enzymatic site(s) in the Gtf3 structure to accommodate the large acceptor GlcNAc-modified Fap1 to catalyze enzymatic reactions. Furthermore, a C-terminal domain of Gtf3 mediates intersubunit interactions in the Gtf3 tetramer, which modulates the glycosyltransferase activity. Taken all the results together, we conclude here that Gtf3 represents a unique new subfamily of glycosyltransferases. The structure presented here adds to the emerging knowledge regarding this family of glycosyltransferases. SRPPs have been implicated in bacterial virulence. As the glycosylation is required for biogenesis of SRPPs, it is thus a potential new therapeutic target. The Gtf3 structure reported here lays the foundation for future mechanistic studies, which would provide new insights into the design of novel therapeutics agents against SRPP-mediated bacterial infection.

Acknowledgment—We especially thank Dr. Bingdong Sha from University of Alabama at Birmingham for advice and assistance.

REFERENCES

1. Zhou, M., and Wu, H. (2009) Microbiology 155, 317–327
2. Bensing, B. A., and Sullam, P. M. (2002) Mol. Microbiol. 44, 1081–1094
3. Takahashi, Y., Konishi, K., Cisar, J. O., and Yoshikawa, M. (2002) Infect. Immun. 70, 1209–1218
4. Siboo, I. R., Chambers, H. F., and Sullam, P. M. (2005) Infect. Immun. 73, 2273–2280
5. Shivshankar, P., Sanchez, C., Rose, L. F., and Orihuela, C. J. (2009) Mol. Microbiol. 73, 663–679
6. von Sorge, N. M., Quach, D., Gurney, M. A., Sullam, P. M., Nizet, V., and Doran, K. S. (2009) J. Infect. Dis. 199, 1479–1487
7. Seifert, K. N., Adderson, E. E., Whiting, A. B., Bohnsack, J. F., Crowley, P. J., and Brady, L. J. (2006) Microbiology 152, 1029–1040
8. Doro, F., Liberatori, S., Rodriguez-Ortega, M. I., Rinaudo, C. D., Rosini, R., Mora, M., Scarselli, M., Altindis, E., D’Aurizio, R., Stella, M., Margaritt, L., Maione, D., Telford, I. L., Norais, N., and Grandi, G. (2009) Mol. Cell. Proteomics 8, 1728–1737
9. Wu, H., and Fives-Taylor, P. M. (1999) Mol. Microbiol. 34, 1070–1081
10. Wu, H., Mintz, K. P., Ladma, M., and Fives-Taylor, P. M. (1998) Mol. Microbiol. 28, 487–500
11. Rigel, N. W., and Braunstein, M. (2008) Mol. Microbiol. 69, 291–302
12. Bu, S., Li, Y., Zhou, M., Azadin, P., Zeng, M., Fives-Taylor, P., and Wu, H. (2008) J. Bacteriol. 190, 1256–1266
13. Wu, H., Zeng, M., and Fives-Taylor, P. (2007) Infect. Immun. 75, 2181–2188
14. Zhou, M., Zhu, F., Dong, S., Pritchard, D. G., and Wu, H. (2010) J. Biol. Chem. 285, 12140–12148
15. Takamatsu, D., Bensing, B. A., and Sullam, P. M. (2004) J. Bacteriol. 186, 7090–7111
16. Mistou, M. Y., Dramsi, S., Brega, S., Poyart, C., and Trieu-Cuot, P. (2009) J. Bacteriol. 191, 4195–4206
17. Takamatsu, D., Bensing, B. A., and Sullam, P. M. (2004) Mol. Microbiol. 52, 189–203
18. Lairson, L. L., Henrijs, B., Davies, G. J., and Withers, S. G. (2008) Annu. Rev. Biochem. 77, 521–555
19. Mullichak, A. M., Lossy, H. C., Walsh, C. T., and Garavito, R. M. (2001) Structure 9, 547–557
20. Hu, Y., Chen, L., Ha, S., Gross, B., Falcone, B., Walker, D., Mokhtarzadeh, M., and Walker, S. (2003) Proc. Natl. Acad. Sci. U.S.A. 100, 845–849
21. Yuan, Y., Barrett, D., Zhang, X., Cahne, D., Sli2, P., and Walker, S. (2007) Proc. Natl. Acad. Sci. U.S.A. 104, 5348–5353
22. Clarke, A. J., Hurtado-Guerrero, R., Pathak, S., Schittelkopf, A. W., Borodkin, V., Shepherd, S. M., Ibrahim, A. F., and van Aalten, D. M. (2008) EMBO J. 27, 2780–2788
23. Fritz, T. A., Hurley, J. H., Trinh, L. B., Shiloach, J., and Tabak, L. A. (2004) Proc. Natl. Acad. Sci. U.S.A. 101, 15307–15312
24. Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006) Acta Crystallogr. D Biol. Crystallogr. 62, 859–866
25. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Acta Crystallogr. D Biol. Crystallogr. 53, 240–255
26. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Acta Crystallogr. D Biol. Crystallogr. 66, 486–501
27. Zhang, Z., Lindstam, M., Unge, J., Peterson, C., and Lu, G. (2003) J. Mol. Biol. 332, 127–142
28. Krissinel, E., and Henrick, K. (2007) J. Mol. Biol. 372, 774–797
29. Li, Y., Chen, Y., Huang, X., Zhou, M., Wu, R., Dong, S., Pritchard, D. G., Fives-Taylor, P., and Wu, H. (2008) Mol. Microbiol. 70, 1094–1104
30. Schägger, H., and von Jagow, G. (1991) Anal. Biochem. 199, 223–231
31. Qasba, P. K., Ramakrishnan, B., and Boegeman, E. (2005) Trends Biochem. Sci. 30, 53–62
32. Guerin, M. E., Kordulakova, I., Schaeffer, F., Svetlikova, Z., Buschiazzo, A., Giganti, D., Gicquel, B., Mikusova, K., Jackson, M., and Alzari, P. M. (2007) J. Biol. Chem. 282, 20705–20714
33. Vetting, M. W., Frantom, P. A., and Blanchard, J. S. (2008) J. Biol. Chem. 283, 15834–15844
34. Morera, S., Imberty, A., Aschke-Sonnenborn, U., Rührer, W., and Free, P. S. (1999) J. Mol. Biol. 292, 717–730
35. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrijsat, B. (2009) Nucleic Acids Res. 37, D233–D238
36. Goedl, C., and Nidetzky, B. (2008) FEBS J. 275, 903–913
37. Lairson, L. L., and Withers, S. G. (2004) Chem. Commun. (Camb) 2243–2248
38. Gibson, R. P., Turkenburg, J. P., Charnock, S. J., Lloyd, R., and Davies, G. J. (2002) Chem. Biol. 9, 1337–1346
39. Hashimoto, K., Madej, T., Bryant, S. H., and Panchenko, A. R. (2010) J. Mol. Biol. 399, 196–206
40. Leipold, M. D., Kaniuk, N. A., and Whitfield, C. (2007) J. Biol. Chem. 282, 1257–1264