Tetrameric Structure of the GlfT2 Galactofuranosyltransferase Reveals a Scaffold for the Assembly of Mycobacterial Arabinogalactan

Received for publication, January 31, 2012, and in revised form, May 9, 2012 Published, JBC Papers in Press, June 15, 2012, DOI 10.1074/jbc.M112.347484

Robert W. Wheatley§, Ruixiang Blake Zheng¶, Michele R. Richards†, Todd L. Lowary‡¶, and Kenneth K. S. Ng‡¶1

From the §Alberta Glycomics Centre, ¶Department of Biological Sciences, University of Calgary, Calgary, Alberta T2N 1N4 and the †Department of Chemistry, Gunning-Lemieux Chemistry Centre, University of Alberta, Edmonton, Alberta T6G 2G2, Canada

Background: GlfT2 is a bifunctional galactofuranosyltransferase essential for mycobacterial cell wall biosynthesis.

Results: Crystal structures of free and UDP-bound GlfT2, along with mutagenesis and kinetic studies, reveal novel details underlying substrate binding and catalysis.

Conclusion: The homotetrameric architecture, distinctive nucleotide-binding site, and unprecedented structural features underlying the bifunctional polymerase activity of GlfT2 are revealed.

Significance: Novel insights into polymerizing glycosyltransferases and the design of anti-mycobacterial therapeutics are obtained.

Biosynthesis of the mycobacterial cell wall relies on the activities of many enzymes, including several glycosyltransferases (GTs). The polymerizing galactofuranosyltransferase GlfT2 (Rv3808c) synthesizes the bulk of the galactan portion of the mycol-arabinogalactan complex, which is the largest component of the mycobacterial cell wall. We used x-ray crystallography to determine the 2.45-Å resolution crystal structure of GlfT2, revealing an unprecedented multidomain structure in which an N-terminal β-barrel domain and two primarily α-helical C-terminal domains flank a central GT-A domain. The kidney-shaped protomers assemble into a C4-symmetric homotetramer with an open central core and a surface containing exposed hydrophobic and positively charged residues likely involved with membrane binding. The structure of a 3.1-Å resolution complex of GlfT2 with UDP reveals a distinctive mode of nucleotide recognition. In addition, models for the binding of UDP-galactofuranose and acceptor substrates in combination with site-directed mutagenesis and kinetic studies suggest a mechanism that explains the unique ability of GlfT2 to generate alternating β-(1→5) and β-(1→6) glycosidic linkages using a single active site. The topology imposed by docking a tetrameric assembly onto a membrane bilayer also provides novel insights into aspects of processivity and chain length regulation in this and possibly other polymerizing GTs.

Tuberculosis (TB),3 a disease that remains a significant worldwide health threat, is caused by infection with Mycobacterium tuberculosis. In common with all mycobacteria, M. tuberculosis possesses a unique glycolipid-rich cell wall structure that provides substantial protection from the environment, including preventing the passage of antibiotics into the organism (1). The standard TB drug regimen (2) therefore involves multiple antibiotics, including some that target cell wall biosynthesis and others that have intracellular targets (3). Concerns about drug resistance (4) have led to heightened interest in identifying novel therapeutic agents and developing a more detailed understanding of mycobacterial biochemical pathways to facilitate this task. In this regard, cell wall assembly has received particular attention (1).

The largest component of the mycobacterial cell wall is the mycol-arabinogalactan complex. This glycolipid has at its core a galactan domain composed of 30–35 galactofuranose (Galβ) residues attached in alternating β-(1→5) and β-(1→6) linkages, which, in turn, is covalently bound to cell wall peptidoglycan through a linker consisting of rhamnose and N-acetyl-glucosamine phosphate (1). The galactan serves as the attachment site for three arabinidomains, each containing ~30 α-(1→5)-, α-(1→3)-, and β-(1→2)-linked arabinofuranose residues. Esterified to the nonreducing termini of the arabinan are mycolic acids, branched long-chain (C90–C160) lipids that are characteristic to mycobacteria and related organisms such as nocardia and corynebacteria. Two clinically used anti-TB drugs target enzymes involved in the assembly of the mycol-arabinogalactan complex (1). Ethambutol inhibits at least one of the arabinosyltransferases responsible for arabinan formation (5), whereas isoniazid is a produg that targets an enoyl-acyl carrier protein reductase required for mycolic acid biosynthesis (6).

---

* This work was supported by funding from the Alberta Glycomics Centre, the Natural Sciences and Engineering Research Council Discovery grants (to T. L. L. and K. K. S. N.), and the Alberta Heritage Foundation for Medical Research Senior Scholar Award (to K. K. S. N.).

1 To whom correspondence may be addressed: Dept. of Chemistry, Gunning-Lemieux Chemistry Centre, University of Alberta, Edmonton, Alberta T6G 2G2, Canada. E-mail: tlowary@ualberta.ca.

2 To whom correspondence may be addressed: Dept. of Biological Sciences, University of Calgary, Calgary, Alberta T2N 1N4, Canada. Tel: 403-220-4320; Fax: 403-289-9311; E-mail: ngk@ucalgary.ca.

3 The abbreviations used are: TB, tuberculosis; CAZy, carbohydrate-active enzyme; Galβ, galactofuranose; GlcNAc, N-acetylglucosamine; GT, glycosyltransferase; Rha, rhamnopyranose; β-ME, β-mercapto-ethanol; TCEP, tris(2-carboxyethyl)phosphine; CBM, carbohydrate-binding module.
Although none of the drugs in use or in development for the treatment of TB are known to target enzymes responsible for galactan biosynthesis, these enzymes are likely suitable targets, as the galactan forms the foundation upon which both the arabinan and mycolate moieties are incorporated. This proposal is further supported by gene knock-out studies demonstrating that galactan formation is essential for mycobacterial viability (7). Because a clearer, more detailed picture of galactan biosynthesis is a necessary prerequisite for identifying novel therapeutics, this area is now receiving increasing attention (8).

Previous work has revealed that mycobacterial galactan biosynthesis requires two bifunctional GTs, GlfT1 and GlfT2, which use UDP-\(\alpha\)-D-galactofuranose (UDP-Galf) as the donor and a growing polyprenol-bound oligosaccharide as the acceptor (9, 10). The \(\alpha\)-stereochemistry in UDP-Galf and the \(\beta\)-stereochemistry of the newly synthesized glycosidic linkages in the products generated by GlfT1 and GlfT2 indicate that catalysis follows an inverting mechanism (9, 11). GlfT1 adds the first two Galf moieties, whereas GlfT2, a polymerizing glycosyltransferase that produces both \(\beta-(1\rightarrow5)\) and \(\beta-(1\rightarrow6)\) linkages between Galf residues (10–12), introduces the remaining ~30 monosaccharides (Fig. 1). GlfT2 has been more extensively studied than GlfT1. Of particular note are saturation-transfer difference-NMR studies suggesting that GlfT2 synthesizes both the \(\beta-(1\rightarrow5)\) and \(\beta-(1\rightarrow6)\) linkages using a single active site (13), and mass spectrometry studies indicate that the enzyme is processive (14). A very recent study examining GlfT2 mutants provides additional support for the presence of a single active site that carries out both glycosylations (15).

In addition to the importance of understanding how GlfT2 works as a key enzyme in mycobacterial cell wall synthesis, GlfT2 is a member of the Carbohydrate-Active enZyme (CAZy) GT-2 family (16, 17), which also contains other polymerizing GTs, including cellulose synthase (18) and chondroitin polymerase (19). GlfT2 has recently become of interest as a model for studying processivity and chain length control in a wide range of polymerizing GTs (14, 20).

**EXPERIMENTAL PROCEDURES**

**Protein Expression and Purification**—For crystallographic studies, GlfT2 was produced from *Escherichia coli* C41(DE3) transformed with pET-15b/Rv3808c (9). Cells stored in glycerol stocks were grown overnight on MDG-agar (21). Single colonies were added to MDG (100 ml) and grown for 36 h at 25 °C. These starter cultures were added to Overnight Express Instant TB medium (Novagen, 1 liter) and grown for 24 h at 25 °C. Cells
were harvested by centrifugation (5000 × g, 20 min) and stored at −70 °C. All purification steps were performed at 4 °C. Thawed cells (~16 g) were suspended in 50 ml of disruption buffer (50 mM NaH2PO4, pH 7.6, 400 mM NaCl, 15% (v/v) glycerol, 20 mM imidazole, 5 mM β-ME, and 0.2% (v/v) Triton X-100), disrupted by two passes through a Model 110 Microfluidizer (Microfluidics), and clarified by centrifugation (60 min, 100,000 × g). The supernatant was loaded onto nickel-Sepharose HP (7 ml, GE Healthcare) equilibrated with buffer A (50 mM NaH2PO4, pH 7.6, 400 mM NaCl, 15% (v/v) glycerol, 5 mM β-ME) plus 20 mM imidazole. The column was washed with 30 ml of buffer A plus 50 mM imidazole and eluted with a 90-ml gradient of buffer A plus 50–400 mM imidazole. Fractions containing GlfT2 were identified by SDS-PAGE and dialyzed overnight against 2 liters of buffer B (50 mM MOPS, pH 7.6, 100 mM NaCl, 10 mM TCEP, and no β-ME). The dialysate was centrifuged (12,000 × g, 15 min), and the supernatant was concentrated to ~3.5 mg/ml (Bradford assay, Bio-Rad) using Ultrafree-15 centrifugal filters (10,000 molecular weight cutoff, Millipore). Aliquots were flash-frozen in liquid nitrogen and stored at −80 °C.

Gel filtration chromatography was performed using a Sephacryl S-200 HR 16/60 column (GE Healthcare) equilibrated with buffer C (50 mM MOPS, pH 7.6, 100 mM NaCl, 10 mM MgCl2, 5 mM β-ME). GlfT2 (1 ml, 8 mg/ml) or a mixture of standard proteins, including blue dextran 2000, was eluted at 0.5 ml/min and monitored by absorbance at 280 nm.

Crystallization—Crystals were grown by hanging-drop vapor diffusion at room temperature by mixing GlfT2 (2 μl, 2.2 mg/ml, buffer B plus 0.2% (v/v) Triton X-100) with precipitant A (2 μl, 5–7% (w/v) PEG 4000, 100 mM sodium cacodylate, pH 6.5, 100 mM sodium acetate). Crystals (~350 × 70 × 70 μm) grew to full size in 7–14 days. Crystals were cryoprotected by transferring to precipitant A plus 1% (w/v) PEG 4000 and successively 5, 10, 15, and 20% (v/v) glycerol. Crystals were equilibrated at least 15 min between transfers and flash-cooled under nitrogen gas at 100 K.

A lutetium derivative was grown by mixing GlfT2 (1.8 μl, 2.2 mg/ml, buffer C plus 1.5 mM LuCl3, 0.2% (v/v) Triton X-100, 5 mM TCEP, and no β-ME) with precipitant B (1.8 μl, 3–5% (w/v) PEG 4000, 100 mM sodium acetate, 5 mM LuCl3, and 5 mM TCEP). Crystals (~200 × 100 × 25 μm) grew to full size in 7–10 days and were adapted to precipitant B plus 1% (w/v) PEG 4000 and 20% (v/v) glycerol, as described for native crystals.

A complex of GlfT2 and UDP was prepared by mixing GlfT2 (2 μl, 2.2 mg/ml, buffer B plus 5 mM TCEP, 0.2% (v/v) Triton X-100) with precipitant C (2 μl, 6–7% (w/v) PEG 4000, 100 mM sodium acetate, 10% (v/v) glycerol, 5 mM TCEP). Crystals (~500 × 50 × 50 μm) grew to full size in 3–7 days. Crystals were adapted to precipitant C plus 10 mM MgCl2, 1% (w/v) PEG 4000, and successively 12, 16, and 20% (v/v) glycerol. Crystals were allowed to equilibrate at least 15 min between transfers. Immediately before flash cooling, crystals were soaked for 10 min in the final cryoprotectant solution plus 5 mM UDP.

**Structure Determination and Refinement**—Diffraction intensities were measured at 100 K at Stanford Synchrotron Research Laboratory (SSRL), beamline 9-2, using a MAR 325 CCD detector. Data were processed and scaled using HKL-2000 (22). All crystals belonged to space group P4212 with two protein molecules per asymmetric unit. Data collection and refinement statistics are presented in Table 1. The data were somewhat anisotropic, which is reflected by the higher value of Rsym in the high resolution shell, and the high resolution cutoff

---

**TABLE 1**

| Crystallographic statistics | Data collection | Apoenzyme | UDP complex |
|-----------------------------|----------------|-----------|-------------|
| Unit cell dimensions        |                |           |             |
| a = b = c                  | 150.2, 145.2 Å | 150.8, 148.0 Å | 150.4, 147.4 Å |
| Wavelength                  | 1.3406 Å       | 1.0442 Å  | 0.9791 Å    |
| Resolution                  | 40 to 3.0 Å    | 40.0 to 2.45 Å | 40.0 to 3.1 Å |
| Rsym                        | 0.147 (0.837)  | 0.104 (0.942) | 0.125 (1.00) |
| Lc/s                        | 18.2 (1.2)     | 27.2 (2.0)  | 220 (1.9)   |
| Completeness                | 98.0% (92.3%)  | 99.9% (99.4%) | 99.8% (99.2%) |
| Redundancy                  | 6.7 (3.1)      | 6.2 (4.9)  | 8.8 (7.0)   |

**Refinement**

| Resolution                  | 40.0 to 2.45 Å | 40.0 to 3.1 Å |
| Unique reflections          | 59,863         | 30,138       |
| Rwork/Rfree                 | 0.186/0.222    | 0.204/0.279  |
| No. of atoms                | 414            | 14           |
| Protein                     | 41.3           | 49.9         |
| Glycerol                    | 72.6           | 66.1         |
| Mn+                         | 74.1           | 108.4        |
| UDP                         | 4153           | 113.9        |
| Water                       | 38.7           | 31.2         |
| Root mean square deviation  |                |              |
| from ideal geometry         |                |              |
| Bond lengths                | 0.008          | 0.008        |
| Bond angles                 | 1.06           | 1.08         |

---

**Values from the outermost resolution shell are given in parentheses.**

**Values of reflection data used in refinement.**

**Values of reflection data excluded from refinement.**
Crystal Structure of GlfT2

was chosen as a compromise that indicates the average difference limit in all directions without sacrificing significant data in the direction of stronger high resolution diffraction. Three lutetium sites were located, and multiple wavelength anomalous dispersion phases were determined to 3.1 Å using SOLVE (23). The initial figure of merit was only 0.37, but nearly the entire model could be built using PHENIX (24) after density modification, including 2-fold noncrystallographic averaging and phase extension to 2.45 Å using amplitudes from a higher resolution native data set. The final model of the native structure was refined by iterative cycles of model building with COOT (25) and reciprocal space refinement with REFMAC (26). Non-crystallographic restraints were not applied during refinement. Structural comparisons were performed using DALI (27) and structural alignments using LSQMAN (28). PyMOL (Schrödinger, LLC) was used for graphical structural analysis and to generate figures.

**Mutagenesis and Kinetics**—GlfT2 mutants were expressed from a chemically synthesized gene with codon usage optimized for *E. coli* (Genscript). Mutagenesis was performed using the QuikChange II XL kit (Stratagene) with the oligonucleotide primer sequences given in supplemental Table 1. GlfT2 mutants were expressed and purified as described previously (29). Enzyme activity was measured using a coupled spectrophotometric assay (29) and a radiochemical assay (9) as described previously.

The spectrophotometric assay was performed in a 384-well microtitrator plate with an incubation mixture containing 50 mM MOPS, pH 7.6, 50 mM KCl, 20 mM MgCl₂, 1.1 mM NADH, 3.5 mM phosphoenolpyruvate, 7.5 units of pyruvate kinase (PK, EC 2.7.1.40), 16.8 units of lactate dehydrogenase (EC 1.1.1.27), acceptor 1 (Fig. 1B), and UDP-Gal. Each reaction was initiated by the addition of 15 μg of GlfT2 protein or the related mutant to the assay mixture. Assays were performed at 37 °C in a total volume of 20 μl and continuously monitored at 37 °C using a Spectra Max 340PC microplate reader in the kinetics mode. Each reaction was monitored at 12–15-s intervals for up to 15 min at 340 nm. The rate of NADH oxidation was converted to picomoles by using a path length of 0.37 cm and an extinction coefficient of 6300 (M⁻¹ cm⁻¹).

The radiochemical assay mixtures consisted of 50 mM MOPS, pH 7.6, 5 mM β-mercaptoethanol, 10 mM MgCl₂, 62.5 μM ATP, 10 mM NADH, 5 μM UDP-Galp, 4 μl of purified UDP-Galp mutase (4 mg/ml), 6 μl of purified E371S or E372S mutant (0.5 mg/ml), and uridine-diphosphate-galactose[6-3H] (American Radiolabeled Chemicals, Inc., 20 Ci/mmol, 0.1 μl) in a total reaction volume of 20 μl. The reaction mixtures were incubated at 37 °C for 1 and 16 h, respectively. Radiolabeled products were isolated by reverse-phase chromatography on SepPak C18 cartridges (Waters) and eluted with methanol (3.5 ml). The reaction products in the eluants were quantified by liquid scintillation counted on a Beckman LS6500 Scintillation Counter using 10 ml of Ecolite mixture.

**Molecular Modeling of Donor and Acceptor Substrate Conformations**—The starting point for the modeling of donor and acceptor molecules was to generate conformations consistent with the preferences expected in solution. Low energy conformations of methyl β-D-galactofuranoside from DFT calculations (B3LYP/6-31+G** (30–33)) in the gas phase indicated a preference for the 4E ring pucker. As a result, two conformers of 1 and 2 were generated by restraining the galactofuranose rings to 4E in minimizations with the MM2 forcefield (34). Comparisons between the structure of GlfT2 and related GT enzymes containing bound acceptor molecules suggested a way to place 1 and 2 into the putative acceptor-binding site of GlfT2. LSQMAN (28) was used to superimpose the GT domains of related enzymes and to superimpose parts of 1 and 2 onto acceptor molecules bound to these related enzymes. The torsion angles for the glycosidic linkages of 1 and 2 were adjusted to avoid steric clashes within the glycan and between the glycan and the protein. The hydroxyl group expected to react with the anomeric carbon of UDP-Galp was positioned near the predicted general base catalyst (Asp-372) and the anomeric carbon of UDP-Galp.

The conformation adopted by UDP-Galp was modeled by adding a galactofuranose ring onto the end of the UDP molecule seen in the product complex crystal structure. The conformation of the pyrophosphate moiety and general position of the Galp ring were modeled to be consistent with crystal structures determined at high resolution, as well as NMR and molecular mechanics calculations (35). The conformation adopted by the galactofuranose ring was modeled in the 2T₁, conformation at the energy minimum from molecular mechanics calculations and near the broad minimum consistent with NMR and molecular dynamics simulations of α-D-galactofuranosyl phosphate and methyl α-D-galactofuranoside. The torsion angle of the glycosidic bond was manually adjusted to minimize steric clashes with the protein, and the position and torsion angles of the exocyclic moiety at C-4 were rotated to minimize steric clashes and to form favorable hydrogen bonds with the protein.

**RESULTS**

**Structure of the GlfT2 Protomer**—Crystal structures were determined for the unliganded form of GlfT2 as well as for the enzyme in complex with the product UDP. Crystals of GlfT2 contain two protein chains per asymmetric unit. Apart from the N-terminal histidine tag and the final eight (native) or nine (UDP complex) residues, the entire polypeptide backbone of both chains was clearly defined by electron density. The structures of the two chains in both the unliganded and product-bound enzyme are nearly identical (root mean square deviation of 0.18 and 0.29 Å, respectively), even though noncrystallographic symmetry restraints were not imposed during refinement. Each protein chain forms a kidney-shaped structure with four domains (Fig. 2A). A central GT family A (GT-A) (36) domain is preceded by an N-terminal β-sandwich domain and followed by an α-helical domain and a C-terminal mixed α + β domain (Fig. 2A and supplemental Fig. 1).

The N-terminal domain contains two short helices preceding a 10-stranded β-sandwich with jelly roll topology. Structural alignments using DALLI (37) indicate similarities with carbohydrate-binding modules (CBM) from a variety of glycosidases, with the CBM from ManA (38) scoring the highest at
Crystal Structure of GlfT2

9.9, despite sharing only ~11% sequence identity. Although the arrangement of a CBM domain N-terminal to a GT-A domain has not yet been described, enzymes from the GT family 27 (GT-27) have trefoil-fold CBMs appended C-terminal to GT-A domains (39). By binding to peptide-linked glycans, the CBMs in some GT-27 enzymes direct the GT domain to specific sites of glycosylation (40). In contrast, the N-terminal β-sandwich domain in GlfT2 has not yet been shown to act as a CBM. Indeed, the most commonly conserved binding site for carbohydrates in CBMs, formed by one of the faces of the β-sandwich (41), is part of the interface between protomers in the GlfT2 tetramer. A second, less commonly observed binding site in CBMs is formed by long loops at one of the edges of the β-sandwich (42). In GlfT2, the analogous loops are short, lack homology with known carbohydrate-binding sites, and form the N-terminal face of the protein, which is distant from the active site.

The GT-A domain in GlfT2 is most similar to enzymes in CAZy families 2, 27, and 78 and is less similar to those from families 13 and 64. Although sequence identity is less than 15% to other GTs with known three-dimensional structures, the GT-A fold is conserved. Common features of the GT-A domain that are found in GlfT2 include the 3214657 topology of the parallel β-sheet core and a “DXD” motif following β4 (43, 44). GT-A domain proteins commonly contain a less highly conserved “variable region” following β5. In GlfT2, this region consists of a β-hairpin followed by a long loop and a short α-helix. Although this region differs from other GT-2 proteins, a similar β-hairpin structure is also found in GT-27 (39) and GT-64 proteins. The β-hairpin extends away from the GT-A core to contact a helix-loop-helix motif at the beginning of domain 3. The β-hairpin also forms a key interface with domain 1. As a result, this entire variable region forms key interactions between domain 2 and both domains 1 and 3.

Immediately following the GT-A domain is a long loop (loop 1, residues 397–407), which, in various GT-A enzymes, can adopt alternate conformations to cover and uncover the active site (45). Although loop 1 is often poorly ordered in GT-A structures, the complete chain can be traced in GlfT2, and it adopts a similar conformation in both the presence and absence of UDP. Loop 1 is proposed to have an important role in the binding of acceptor substrates (below).

The third domain of GlfT2 consists of eight α-helices, five of which immediately follow the GT domain and form the central portion of domain 3. Three helices at the C terminus of the polypeptide chain wrap around these core helices to complete the domain. The first helix in domain 3 anchors loop 1 and is conserved in many GT-A proteins (43). However, for the rest of the domain, structural comparisons using DALI indicate only minor similarities (Z < 5.3) to proteins unrelated in overall structure and function. The fourth and last domain consists of a three-stranded β-sheet surrounded by three α-helices. The DALI algorithm fails to find any structures with significant similarity to this fold.

The accessory domains flanking the central GT domain in GlfT2 appear to be conserved in a small family of closely related proteins, but they differ significantly from all other GT-A proteins of known structure. Sequence comparisons indicate that the flanking domains and many hydrophobic residues found at the interface between domains 2 and 3 are conserved in >100 proteins from mycobacteria, related Actinobacteria, and even more distantly related species, thus suggesting conserved structure and function relationships (supplemental Figs. 2 and 3). Outside this family, however, there appear to be less closely related enzymes, such as the nonpolymerizing mycobacterial GT GlfT1, which contain a homologous GT-A domain but no flanking regions (10).

Structure of the GlfT2 Homotetramer—Each GlfT2 protomer is 90 × 50 × 50 Å in size, and each of the two protomers in the asymmetric unit forms a separate C4-symmetric homotetramer in the crystal (Fig. 3). This 295-kDa oligomer is consistent with the elution profile of the protein in gel filtration chromatography. Approximately 10.4% of the accessible surface area (2740 Å2 out of 26,400 Å2) for each protomer is buried in the tetramer. The amount and percentage of total buried surface area are both lower than the average (3900 Å2) seen in homodimeric proteins (46) and the average percentage of buried surface area seen in multimeric proteins (18%) (47). However, each of the two independent protomers in the asymmetric unit forms a separate but essentially identical tetramer, thus arguing that the tetramer is likely the form seen in solution and not a crystallographic artifact.

The dimensions of the tetramer are 100 × 100 × 75 Å. A hollow funnel-shaped pore with a diameter of less than 10 Å at the face formed primarily by the N-terminal portion of each protomer (N-face, Fig. 3A) expands to over 40 Å at the opposing
face (C-face, Fig. 3B). The C-face contains large hydrophobic patches (Fig. 3D), and the charged residues exposed are predominantly positive (Fig. 3E). Notably, the hydrophobic and positively charged residues on the C-face of GlfT2 are quite highly conserved in GlfT2 homologues predicted to have flanking C-terminal domains. For example, Trp-555 is perfectly conserved in the 110 most closely related sequences in GenBank™, whereas Arg-554 and Lys-445 are positively charged in nearly all of the ~80 most closely related sequences. Many surface-exposed leucine, alanine, isoleucine, and tryptophan residues are also generally conserved in the ~80 most closely related sequences. These features suggest that the C-face associates with the hydrophobic acyl chains and negatively charged head groups of membrane phospholipids. This inference is consistent with the observation that GlfT2 is localized in the membrane fractions of mycobacteria and heterologous hosts (11, 12). Assuming this orientation of the tetramer relative to the cell membrane, the N-face contains a small pore at its center, as well as grooves between adjacent subunits (Fig. 3C). Each of these grooves provides an opening toward the active site, which allows for the
Crystal Structure of GlfT2

**A**

![Image of UDP-Galf complex](image)

**B**

![Image of UDP-Galf complex](image)

**FIGURE 4.** Binding site for UDP-Galf donor. **A,** structure of UDP bound to GlfT2 and showing interactions with highly conserved residues and Mn\(^{2+}\) (gray sphere), as well as the coordinating side chains. The difference electron density map (contoured at 3σ) was calculated with phases after the model of UDP was removed, and the structure of GlfT2 was subjected to 20 cycles of positional and temperature factor refinement. **B,** model of the structure of UDP-Galf (magenta) and a Galf acceptor residue (gray) showing possible hydrogen bonding interactions (dashed red lines) with highly conserved residues. In this model, the 6-OH of the acceptor sugar is poised for interaction with the side chain of Asp-372 (general base catalysis) and the anomeric carbon of the donor sugar-nucleotide.

entry of UDP-Galf and the release of UDP after glycosyl transfer.

**Substrate Binding and Catalysis**—Similar to other GT-A enzymes, the active site of GlfT2 contains a divalent metal ion (Fig. 4A). Anomalous difference Fourier electron density maps indicate that Mn\(^{2+}\) is bound to the active site, even though Mg\(^{2+}\) and not Mn\(^{2+}\) was present in purification and crystallization solutions. The Mn\(^{2+}\)-binding site is formed by Asp-256 and Asp-258 from the “DXD” motif found in the loop following β4, as well as His-396 at the C-terminal end of the GT domain. This arrangement of coordinating residues resembles that seen in many GT-A enzymes, especially those from families 2, 27, and 78, although the details of metal coordination vary substantially among these enzymes.

The structure of the GlfT2-UDP complex also reveals a mode of UDP recognition with similarities to other GT-A enzymes, as well as some unique features (Fig. 4A). The side chain –NH\(_2\) and =O groups of Asn-229 donate and accept hydrogen bonds from the O4 and N3 atoms of the uracil base, and the side chain –NH\(_2\) group of Gln-200 donates a hydrogen bond to the O2 atom. This hydrogen-bonding arrangement resembles the interaction between the side chain of Asn-130 and the uracil base in the human GT-64 enzyme EXTL2 (48), and these two amino acids occupy a similar position in the loop preceding α3 in the GT-A fold. However, this hydrogen-bonding arrangement differs from the one most commonly seen in enzymes from the GT-2, -13, and -27 families, where Asp takes the place of Gln-200 and accepts a hydrogen bond from the uracil ring nitrogen. The aromatic rings of Phe-173 and Phe-367 flank the two faces of the uracil ring, forming face-to-face and edge-to-face stacking interactions, respectively. The former aromatic residue is highly conserved in other GTs, whereas the latter is more variable. In addition to these contacts, the uracil base as well as the ribose C1 and O4 atoms also contact the main chain of Gly-231 and Gly-232. The ribose moiety only forms a single hydrogen bond with the protein; the ring oxygen accepts an H-bond from the backbone –NH group of Gly-232. These Gly residues lie at the N-terminal end of α3 in the GT domain and are part of a highly conserved motif found only in the family of actinobacterial GTs to which GlfT2 and GlfT1 belong. The lack of side chains at these key positions provides a larger space for accommodating the ribose moiety than seen in most other GTs of known three-dimensional structure. In most other GTs, such as SpsA and EXTL2, for example, the other side of the ribose interacts with Asp residues equivalent to Asp-257, the middle X residue in the DXD motif. Although Asp-257 in GlfT2 is positioned similarly to the corresponding residue in these other GTs, the ribose moiety is too distant to interact (over 5.5 Å). Saturation-transfer difference-NMR studies confirm the distinctive binding mode seen in the GlfT2-UDP complex, as the UDP protons making the most intimate contacts with the protein are attached to the C1 atom of the ribose and the uracil base (49). Each phosphate group in the diphosphate moiety coordinates to Mn\(^{2+}\) through a single oxygen atom, as seen for UDP in solution and when bound to enzymes (35).

The conformation of the diphosphate group allows for the construction of a model for UDP-Galf by placing a furanose ring adopting a low energy \(^{2}\)T\(_1\) conformer in a “folded-back” conformation relative to the diphosphate group (Fig. 4B). This conformation allows hydroxyl groups on the Galf residue to form hydrogen bonds with the side chains of His-396, Asp-371, and Asp-372. In this model, the plane of the Galf ring is roughly perpendicular to the axis formed by the diphosphate moiety. A similar conformation is seen in nearly all structures of GT-donor complexes (36, 45). In other inverting GTs, a carboxylate group in the active site is proposed to act as a general base to activate an oxygen atom of the acceptor sugar for reaction with the anomeric carbon of the sugar nucleotide (36). In GlfT2, Asp-372 is positioned appropriately for this function (Fig. 4B).

To evaluate the role of the carboxylate side chain of Asp-372, we replaced this residue by Ser using site-directed mutagenesis. Using both a coupled spectrophotometric assay (29) and a more sensitive radiochemical assay (9), the D372S mutant showed no detectable activity, even following an overnight incubation (Table 2). In comparison, replacing the nearby Asp-371 residue...
TABLE 2

Kinetic analysis

| UDP-Galf | Acceptor 1 | k_{cat} |
|----------|------------|--------|
| Wild type | 0.38 ± 0.06 | 0.60 ± 0.20 | 430 ± 35 |
| E371S    | 0.14 ± 0.06 | 0.28 ± 0.12 | 150 ± 30 |
| E372S    | 0.13 ± 0.05 | 0.22 ± 0.10 | 100 ± 25 |
| E300S    | 0.42 ± 0.06 | 1.78 ± 0.12 | 0.30 ± 0.01 |
| W399S    | 0.41 ± 0.04 | 1.78 ± 0.10 | 0.35 ± 0.04 |
| H413S    | 0.43 ± 0.05 | 1.54 ± 0.14 | 0.25 ± 0.01 |

* For E371S and E372S, activity was not detectable using the spectrophotometric assay.

DISCUSSION

One of the key puzzles central to the biological function of GlfT2 is its ability to generate alternating \( \beta-(1\rightarrow5) \) and \( \beta-(1\rightarrow6) \) glycosidic linkages through a single active site. Although the structures of GlfT2 bound to acceptor substrate molecules have not yet been determined, the structure of the GlfT2-UDP complex reveals a narrow channel near loop 1, which can accommodate low energy conformers of trisaccha-
Crystal Structure of GlfT2

A, B, and C represent different views of the GlfT2 tetramer, with key structural components highlighted. A and C show the central cavity and nascent glycan chains, while B focuses on the C-face and lipid bilayer interactions. The tetramer structure is overlaid with carbohydrate residues and lipid molecules to illustrate interactions.

The structure of GlfT2 suggests how interactions with the lipid bilayer membrane may affect the enzyme's activity. As mentioned above, the C-face of the GlfT2 tetramer contains an abundance of exposed hydrophobic and positively charged residues, as highlighted in A and B. These interactions are crucial for enzyme function in vitro, where the enzyme is soluble in the absence of a phospholipid bilayer. In vivo, due to the presence of a phospholipid bilayer, the enzyme is likely buried in the membrane, facilitating reaction with the acyl groups of membrane phospholipids.

The role of GlfT2 in galactan biosynthesis is to add ~30 Gal residues to a β-D-Galf-(1→5)-β-D-Galf-(1→4)-α-L-Rhap-(1→3)-α-D-GlcNAc-decaprenyl-pyrophosphate acceptor substrate (10). Models of the lipid-linked tetrasaccharide substrate and longer nascent glycan chains indicate that the growing polymer would be located in the hollow core of the GlfT2 tetramer, anchored to the membrane by the decaprenol group (Figs. 3 and 6; supplemental Fig. 3). The structure of GlfT2 indicates a distance of ~30 Å from the location of the hydrophobic residues in the C-face to the active site. The pyrophosphate group and carbohydrate residues of the nascent glycan chain span most of this distance. Some distortions in the plasma membrane from protein binding or transient interactions between the acceptor decaprenyl-pyrophosphate moiety and the protein may also be present. Specific interactions between the carbohydrate residues of the acceptor and the protein were not detected in an earlier study (20), which is consistent with the apparent lack of a well formed interaction between the acceptor and protein.
Crystal Structure of GlfT2

carbohydrate-binding site with highly conserved residues in the hollow core.

The molecular mechanisms underlying the polymerase activity of GlfT2 have been the subject of recent investigations (15, 20). Considered in aggregate, the model outlined above is generally consistent with some features of the tethering model proposed by Kiessling and co-workers (20). However, the structure of GlfT2 suggests that instead of a direct interaction between the decaprenol group of the natural acceptor substrate and the enzyme, the lipid-linked acceptor may diffuse freely in the portion of the membrane bilayer underlying the hollow central core of GlfT2 while still being confined by the surrounding protein. This topology suggests that GlfT2 acts to promote polymerization processivity through a mechanism akin to metabolic channeling.

In addition, the geometric restrictions of the central core on the growing polysaccharide may assist in limiting the extent of polymerization to the lengths of glycans observed in vivo (~30 Galf residues). The volume of the central cavity depends on assumptions about the location of the tetramer relative to the membrane bilayer, something that is not well defined at present. However, calculations using VOIDOO (51) estimate a volume of ~60,000 Å³, which is sufficient to accommodate at least 100–150 residues of Galf. Thus, the dimensions of the internal cavity appear to provide only sufficient space to accommodate four nascent chains of ~30 Galf residues each. Further studies are clearly needed to define the specific mechanisms underlying the polymerization activity of GlfT2 and possibly other GTs using membrane-embedded, lipid-linked acceptor substrates. Notably, the structure of GlfT2 indicates for the first time that the topology imposed by docking a tetrameric assembly onto a membrane bilayer is likely important for the function of the enzyme as a polymerase in vivo.

The efficacy of ethambutol, an arabinosyltransferase inhibitor, for the treatment of TB has prompted the search for other inhibitors of polysaccharide biosynthetic GTs as therapeutics for mycobacterial diseases (16). Knock-out studies also indicate that GlfT2 is essential for growth (7), and the structure of the enzyme reported here suggests at least three novel approaches for designing inhibitors that may be suitable lead compounds for therapeutic development. First, the recognition of uracil by Asn-229 in GlfT2 is unusual, as a negatively charged Asp residue occupies this position in most other GTs using UDP donors, with the exception of GT64 enzyme EXTL2, which has Gln (48). This feature could be combined with modifications of the uridine base to take advantage of a potential binding site provided by surface-exposed hydrophobic residues (e.g. Trp-408 from loop 1, which packs against Leu-480 in domain 3) adjacent to C5 in the uridine base to generate specificity and affinity (Fig. 3B). Related to this idea, modifications of the uridine base at C5 were recently shown to produce novel GT inhibitors with allosteric effects (52). Second, the larger binding pocket surrounding the ribose moiety of the UDP donor may be a distinctive feature of GlfT2 that could be exploited for inhibitor design. Increased specificity for GlfT2 could possibly be generated by incorporating a bulky functional group that would occupy this distinctive binding pocket in GlfT2 but lead to steric clashes with the smaller pocket found in many other GTs (45). Finally, the structure suggests that patches of surface-exposed hydrophobic and positively charged residues may also provide a target for inhibitor design. If these regions on the C-face of the GlfT2 tetramer are important for membrane attachment and confining the acceptor substrate within the central cavity, an inhibitor that bound to these regions with high affinity could prevent GlfT2 from forming some of the critical binding interactions required for galactan synthesis.

Acknowledgments—We acknowledge the help of the staff at the Stanford Synchrotron Radiation Laboratory and the Canadian Light Source for providing access to the synchrotron radiation beamlines used for crystal screening and data collection at various stages of this project. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource, a Directorate of SLAC National Accelerator Laboratory and an Office of Science User Facility operated for the United States Department of Energy Office of Science by Stanford University. The SSRl Structural Molecular Biology Program is supported by the Department of Energy Office of Biological and Environmental Research, and by National Institutes of Health Grant P41GM103393 from NIGMS and Grant P41RR001209 from NCRR. Portions of the research described in this paper were performed using beamline 08ID-1 at the Canadian Light Source, which is supported by the Natural Sciences and Engineering Research Council of Canada, the National Research Council Canada, the Canadian Institutes of Health Research, the Province of Saskatchewan, Western Economic Diversification Canada, and the University of Saskatchewan.

REFERENCES

1. Kaur, D., Guerin, M. E., Skovierová, H., Brennan, P. J., and Jackson, M. (2009) Chapter 2. Biogenesis of the cell wall and other glycoconjugates of Mycobacterium tuberculosis. Adv. Appl. Microbiol. 69, 23–78.
2. Bass, J. B., Jr., Farer, L. S., Hopewell, P. C., O’Brien, R., Jacobs, R. F., Ruben, F., Snider, D. E., Jr., and Thornton, G. (1994) Treatment of tuberculosis and tuberculosis infection in adults and children. American Thoracic Society and The Centers for Disease Control and Prevention. Am. J. Respir. Crit. Care Med. 149, 1359–1374.
3. Dover, L. G., and Coxon, G. D. (2011) Current status and research strategies in tuberculosis drug development. J. Med. Chem. 54, 6157–6165.
4. Nachega, J. B., and Chaisson, R. E. (2003) Tuberculosis drug resistance: a global threat. Clin. Infect. Dis. 36, S24–S30.
5. Belanger, A. E., Besra, G. S., Ford, M. E., Mikusová, K., Belisle, J. T., Brennan, P. J., and Inamine, J. M. (1996) The embAB genes of Mycobacterium avium encode an arabinosyltransferase involved in cell wall arabinogalactan biosynthesis that is the target for the antimycobacterial drug ethambutol. Proc. Natl. Acad. Sci. U.S.A. 93, 11919–11924.
6. Banerjee, A., Dubnau, E., Quemard, A., Balasubramanian, V., Um, K. S., Wilson, T., Collins, D., de Lisle, G., and Jacobs, W. R., Jr. (1994) inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis. Science 263, 227–230.
7. Pan, F., Jackson, M., Ma, Y., and McNeil, M. (2001) Cell wall core galactofuran synthesis is essential for growth of mycobacteria. J. Bacteriol. 183, 3991–3998.
8. Richards, M. R., and Lowry, T. L. (2009) Chemistry and biology of galactofuran-containing polysaccharides. ChemBioChem 10, 1920–1938.
9. Rose, N. L., Competto, G. C., Lin, S. J., McNeil, M., Palcic, M. M., and Lowry, T. L. (2006) Expression, purification, and characterization of a galactofuranosyltransferase involved in Mycobacterium tuberculosis arabinogalactan biosynthesis. J. Am. Chem. Soc. 128, 6721–6729.
10. Belánová, M., Dianisková, P., Brennan, P. J., Competto, G. C., Rose, N. L.,
51. Kleywegt, G. J., and Jones, T. A. (1994) Detection, delineation, measurement, and display of cavities in macromolecular structures. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 178–185

52. Pesnot, T., Jørgensen, R., Palcic, M. M., and Wagner, G. K. (2010) Structural and mechanistic basis for a new mode of glycosyltransferase inhibition. *Nat. Chem. Biol.* **6**, 321–323

53. Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* **179**, 125–142

54. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I., and Lomize, A. L. (2012) OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376