Crystal Structure of a UDP-glucose-specific Glycosyltransferase from a Mycobacterium Species*

Zara Fulton, Adrian McAlister, Matthew C. J. Wilce, Rajini Brammananth, Leyla Zaker-Tabrizi, Matthew A. Perugini, Stephen P. Bottomley, Ross L. Coppel, Paul K. Crellin, Jamie Rossjohn, and Travis Beddoe

From the Protein Crystallography Unit, Department of Biochemistry and Molecular Biology, School of Biomedical Sciences, Monash University, Clayton, Victoria 3800, the Australian Research Council Centre of Excellence in Structural and Functional Microbial Genomics, Monash University, Clayton, Victoria 3800, the Department of Microbiology, School of Biomedical Sciences, Monash University, Clayton, Victoria 3800, the Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, and the Department of Biochemistry and Molecular Biology, School of Biomedical Sciences, Monash University, Clayton, Victoria 3800, Australia

Received for publication, March 6, 2008, and in revised form, July 23, 2008. Published, JBC Papers in Press, July 30, 2008, DOI 10.1074/jbc.M801853200

Glycosyltransferases (GTs) are a large and ubiquitous family of enzymes that specifically transfer sugar moieties to a range of substrates. Mycobacterium tuberculosis contains a large number of GTs, many of which are implicated in cell wall synthesis, yet the majority of these GTs remain poorly characterized. Here, we report the high resolution crystal structures of an essential GT (MAP2569c) from Mycobacterium avium subsp. paratuberculosis (a close homologue of Rv1208 from M. tuberculosis) in its apo- and ligand-bound forms. The structure adopted the GT-A fold and possessed the characteristic DXXD motif that coordinated an Mn2+ ion. Atypical of most GTs characterized to date, MAP2569c exhibited specificity toward the donor substrate, UDP-glucose. The structure of this ligated complex revealed an induced fit binding mechanism and provided a basis for this unique specificity. Collectively, the structural features suggested that MAP2569c may adopt a “retaining” enzymatic mechanism, which has implications for the classification of other GTs in this large superfamily.

Mycobacterium tuberculosis, the causative agent of tuberculosis, is a devastating bacterial pathogen, and infection by M. tuberculosis continues to be common particularly in developing countries, with particular strains becoming resistant to multiple frontline drugs. The pathogenicity of M. tuberculosis is partly attributable to its waxy cell wall that consists of covalently linked layers of peptidoglycan, arabinogalactan, and mycolic acids (for a review, see Crick et al. (1)), interspersed with lipids and glycolipids, and capped with polysaccharides (2). Despite the major importance to the biology of Mycobacterium sp. and other Corynebacteriaceae, the biosynthetic pathways for many of these carbohydrates are poorly understood, although it is established that the most prolific class of enzymes involved in these pathways are the glycosyltransferases (GTs)3. Indeed, there are 41 known or putative GTs in M. tuberculosis strain H37Rv listed in the CAZy (carbohydrate-active enzyme) data base, representing over 1% of the 3900 open reading frames in the M. tuberculosis H37Rv genome (4). Ethambutol, a front line anti-tuberculosis drug, targets the EmbA and EmbB GTs that synthesize the arabinogalactan cell wall layer, thereby marking GTs as promising new drug targets.

GTs transfer sugar moieties from the activated donor substrate, mostly in the form of a nucleoside-diphospho-sugar, to a myriad of specific acceptor substrates to generate innumerable oligosaccharide and glycoconjugate products that are often required for species- or cell specific processes. GTs are classified as either “inverting” or “retaining,” depending on whether the stereochemistry of the anomeric carbon is retained or inverted in the product relative to that in the donor substrate. Moreover, despite the sequence diversity within the GT superfamily, GT structures solved to date adopt one of two common folds, termed “GT-A” or “GT-B.” Both of these folds adopt the “Rossman-like fold,” with the GT-A fold containing a single Rossman-like or nucleotide-binding domain (5) and the GT-B fold containing two similar Rossman-like domains. Moreover, members within a particular GT family are predicted to share the same inverting or retaining mechanism and GT-A or GT-B fold. Furthermore, GTs of the GT-A fold that utilize a nucleoside-diphospho-sugar typically contain a “DXXD” (or XDD or EXD) motif that is involved in metal ion-mediated activated donor substrate coordination and is also required for catalytic activity (6).

In contrast to the abundance of GT sequences available, there is a lack of functional, mechanistic, and structural data on

* This work was supported by the Australian Research Council (ARC) Centre of Excellence in Structural and Functional Microbial Genomics and the National Health and Medical Research Council (NHMRC) of Australia. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† The on-line version of this article (available at http://www.jbc.org) contains supplemental Table 1 and Figs. 1–4.

The atomic coordinates and structure factors (codes 3CKJ, 3CKN, 3CKQ, 3CKV, and 3CKO) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

1 An ARC Federation Fellow. To whom correspondence may be addressed. Tel.: 613-9905-3736; Fax: 613-9905-4699; E-mail: jamie.rossjohn@med.monash.edu.au.

2 A NHMRC Career Development Award Fellow. To whom correspondence may be addressed. Tel.: 613-9905-3736; Fax: 613-9905-4699; E-mail: travis.beddoe@med.monash.edu.au.

3 The abbreviations used are: GT, glycosyltransferase; MGS, mannosylglycerate synthase; MAD, multiple anomalous dispersion.
Crystal Structure of a Mycobacterial Glycosyltransferase of the GT-A Fold

these enzymes. This paucity of biochemical data reflects the difficulty in characterizing enzymes with countless possible combinations of donor and acceptor substrates and products. Indeed, only 29 of the 90 GT families have representative structures solved. The vast majority of sequences uncovered by genomic analyses that show distant sequence identity to known GTs are thus grouped into large “polyspecific” families. For example, the rapidly growing GT2 family is the largest such family, with over 10,000 known and candidate members, and is considered the ancestral inverting family from which all GTs that share the GT-A fold have evolved (7). The spore coat polysaccharide biosynthesis protein (SpsA) from Bacillus subtilis (8), which adopts the GT-A fold, remains the only representative of the GT2 family whose structure is solved. Due to the versatility of the GT-A fold, “rules” for discerning the substrate specificity and discriminating the mechanism of GTs that exhibit this fold remain elusive. To further our understanding of this family of GTs and to characterize new anti-tuberculosis drug targets, we investigated Rv1208, an essential enzyme (9) implicated in the biosynthesis of the oligosaccharide- and glycoconjugate-rich M. tuberculosis cell wall. As a candidate member of the GT2 family, Rv1208 is predicted to use an inverting mechanism, Mn2+ and a nucleoside-diphospho-sugar as its activated donor substrate. We have determined the crystal structure of MAP2569c from Mycobacterium avium subsp. paratuberculosis, a close homologue (83% sequence identity) to Rv1208. We reveal that MAP2569c possesses the GT-A fold and exhibits specificity toward the donor substrate, UDP-glucose.

EXPERIMENTAL PROCEDURES

Production and Crystallization of MAP2569c—Detailed methods used for the cloning, overexpression, purification, and crystallization of MAP2569c have been published previously (10). Briefly, selenomethionine-labeled protein was overexpressed in Escherichia coli B834 (DE3) grown in autoinducing medium, as described by Studier (11) at 25 °C for 24 h. The crystallized form of MAP2569c in this study comprised residues 5–329 of the native sequence. MAP2569c crystallized in space group P4₁2₁2 with unit cell dimensions a = b = 86.6 Å and c = 104.3 Å. There is one monomer in the asymmetric unit corresponding to a solvent content of ~54% (v/v).

The co-complex crystals were obtained by soaking the “native” crystals in 1 M ammonium sulfate, 0.1 M HEPES, pH 7.0, or 0.1 M sodium citrate, pH 5.5, 20 mM manganese chloride, 1 mM sodium thioulate, 5% (v/v) glycerol, and 100 mM UDP or UDP-sugar for 6 h. The crystals were prepared for x-ray data collection by the addition of 15% (v/v) glycerol prior to flash cooling directly in liquid nitrogen.

X-ray Diffraction Data Collection—X-ray diffraction data were collected from crystals at 100 K on IMCA beamline 1710-D at the Advanced Photon Source, Argonne National Laboratory (Chicago, IL), using a MAR CCD 165 detector, at General Medicine and Cancer Institutes Collaborative Access Team (GM/CA-CAT) beamline 23 ID-D at the Advanced Photon Source using a MAR CCD 300 detector, and in house using a Rigaku RU-H3RHB rotating anode generator and R-Axis IV++ detector (see Table I). X-ray data were processed and analyzed using the CCP4 program suite (12) (see Table 1).

Phase Determination, Model Building, and Refinement—The multiple anomalous dispersion (MAD) technique was used to obtain the initial phases from selenomethionine-substituted crystals of apo-MAP2569c. Three selenium atom sites were located, and initial phases were calculated using BnP (13). The figure of merit was 0.585 to 2.3 Å (see Table 1). The initial phases were extended to 1.8 Å resolution with automated density modification using DM (14). The starting model was built into the density-modified electron density map using ARP/wARP (15). The final model was obtained after iterative cycles of manual model building with COOT (16) and TLS and restrained refinement using REFMAC (14, 17). The final model consists of two contiguous polypeptide chains and comprised 303 residues (residues 15–173 and 189–329 of the native sequence) and 102 water molecules. The structure was refined to 1.85 Å to R<sub>factor</sub> and R<sub>free</sub> values of 18.4 and 19.8%, respectively (see Table 2).

For the co-complexation studies, difference Fourier analyses were used to evaluate nucleotide and nucleotide-sugar binding. Accordingly, unbiased features in the electron density maps revealed the location of the nucleotide or nucleotide-sugar molecules, and the subsequent structures were refined using similar protocols to those listed above (see Tables 1 and 2). For refinement of the ligated MAP2569c complexes, the same R<sub>free</sub> data set was used as selected in the apo crystal form.

Analytical Ultracentrifugation Analysis—Sedimentation velocity experiments were conducted in a Beckman model XL-1 analytical ultracentrifuge at a temperature of 20 °C. A sample of MAP2569c (380 µl, 1.5 mg/ml) solubilized in 20 mM Tris, 150 mM NaCl, pH 8.0, and reference (400-µl) solutions were loaded into a conventional double sector quartz cell and mounted in a Beckman four-hole An-60 Ti rotor. Data were collected in continuous mode at 295 nm using a rotor speed of 40,000 rpm, a time interval of 300 s, and a step size of 0.003 cm without averaging. Solvent density (1.005 g/ml at 20 °C) and viscosity (1.021 cp) as well as estimates of the partial specific volume of MAP2569c (0.74 ml/g) were computed using the program SEDNTERP (18). Sedimentation velocity data at multiple time points were fitted to a single discrete species or a continuous size distribution model (19) using the program SEDFIT, which is available on the World Wide Web.

Enzyme Activity—The nucleotide-sugar specificity of MAP2569c was determined by a continuous linked enzyme assay described by Grosselin et al. (20). The 200-µl reaction (carried out at 37 °C) consisted of 13 mM HEPES, pH 7.4, buffer containing 10 mM MnCl₂, 13 mM MgCl₂, 50 mM KCl, 13 mg/ml bovine serum albumin, 0.7 mM phosphoenolpyruvate, 0.6 mM NADH, 1 unit of pyruvate kinase, 1.23 units of lactate dehydrogenase, 10 mM 3-phosphoglycerate (the acceptor molecule), and 1 mM NDP-sugar. The amount of protein varied between 0.5 and 2.5 µg/ml. The decrease in NADH was continuously measured at 340 nm using a Fluostar Optima plate reader (BMG Labtech). Specific activity was calculated by the change of absorbance over time using an excitation coefficient of NADH of 6.22 mm cm⁻¹.
Crystal Structure of MAP2569c—Although Rv1208 was insoluble, the expression of its close homologue, MAP2569c, yielded soluble protein that was amenable for structural studies (10). We subsequently determined the crystal structure of apo-MAP2569c using the MAD technique (see “Experimental Procedures” and Table 1). The structure was refined to 1.85 Å to \( R_{\text{factor}} \) and \( R_{\text{free}} \) values of 18.4 and 19.8%, respectively (see Table 2).

The MAP2569c (329 amino acids, 34.8 kDa) protomer is globular (~65 \( \times \) 45 \( \times \) 40 Å) and rich in \( \alpha \)-helices (Fig. 1, A and B). At the core of MAP2569c is a mixed nine-stranded twisted \( \beta \)-sheet (sA) with topology \( \beta_5\beta_4\beta_3\beta_6\beta_8\beta_7\beta_9\beta_8\beta_1 \) (\( \beta_1 \) and \( \beta_7 \) are antiparallel), flanked by three helical segments on either side. As expected from its low sequence similarity to the GT2 family, MAP2569c contained the GT-A scaffold observed in SpS (8) (see Fig. 1, A and B) and can be subdivided into an N-terminal domain, “catalytic” domain, and C-terminal domain.

The catalytic domain of MAP2569c (residues 50–260) exhibited the characteristic \( \alpha/\beta/\alpha \) architecture of the GT-A fold, which is subdivided into “nucleotide-binding” (residues 50–138) and “acceptor-binding” (residues 142–260) subdomains. The nucleotide-binding subdomain of MAP2569c comprised the “Rossman-like” fold (5) with four parallel \( \beta \)-strands (\( \beta_5\beta_4\beta_3\beta_6 \)), interspersed with helices (Fig. 1, A and B). The acceptor-binding subdomain contained three antiparallel \( \beta \)-strands (\( \beta_8\beta_7\beta_9 \)), also separated by helices, and a disordered region (residues 174–188). The signature catalytic DXD motif of the GT-A fold is found, as expected, at the juncture of the two catalytic subdomains and comprised Asp-139–Ser-140–Asp-141. A second small antiparallel \( \beta \)-sheet (sB), perpendicular to the central \( \beta \)-sheet (sA), is formed by two short \( \beta \)-stands: \( \beta_6' \) at the N terminus of the acceptor-binding subdomain and \( \beta_9' \) within a short loop bridging the acceptor-binding subdomain and C-terminal domain. This loop (residues 261–267) sat adjacent to the DXD motif and traced the nucleotide-binding/acceptor-binding subdomain divide. The C-terminal domain (residues 268–329) contained a long helical region and a prominent antiparallel two-stranded twisted \( \beta \)-sheet (sC; \( \beta_1\beta_1\beta_1\beta_1 \)) rich in aromatic residues.

Although there was a monomer in the asymmetric unit, a crystallographic dimer was observed in which this unusual C-terminal tail wrapped around its 2-fold related partner (see Fig. 1C). The remainder of this dimeric interface is formed by residues in the long helical region at the start of the C-terminal domain and in the N-terminal domain and acceptor-binding subdomain. The buried surface area at this dimeric interface was quite extensive (~1900 Å²) and is therefore likely to represent a biological dimer. Supporting this assertion are AUC studies of MAP2569c, which was shown to be a dimer in aqueous solution (see supplemental Fig. 1 and supplemental Table 1). Given the location of the substrate-binding sites, it appears that the MAP2569c dimer is probably required for stability as opposed to catalysis. Accordingly, the dimeric MAP2569c exhibits a GT-A fold and DXD motif consistent with GT activity.

### RESULTS

#### Crystal Structure of MAP2569c

Although Rv1208 was insoluble, the expression of its close homologue, MAP2569c, yielded soluble protein that was amenable for structural studies (10). We subsequently determined the crystal structure of apo-MAP2569c using the MAD technique (see “Experimental Procedures” and Table 1). The structure was refined to 1.85 Å to \( R_{\text{factor}} \) and \( R_{\text{free}} \) values of 18.4 and 19.8%, respectively (see Table 2).

The MAP2569c (329 amino acids, 34.8 kDa) protomer is globular (~65 \( \times \) 45 \( \times \) 40 Å) and rich in \( \alpha \)-helices (Fig. 1, A and B). At the core of MAP2569c is a mixed nine-stranded twisted \( \beta \)-sheet (sA) with topology \( \beta_5\beta_4\beta_3\beta_6\beta_8\beta_7\beta_9\beta_8\beta_1 \) (\( \beta_1 \) and \( \beta_7 \) are antiparallel), flanked by three helical segments on either side. As expected from its low sequence similarity to the GT2 family, MAP2569c contained the GT-A scaffold observed in SpS (8) (see Fig. 1, A and B) and can be subdivided into an N-terminal domain, “catalytic” domain, and C-terminal domain.

The catalytic domain of MAP2569c (residues 50–260) exhibited the characteristic \( \alpha/\beta/\alpha \) architecture of the GT-A fold, which is subdivided into “nucleotide-binding” (residues 50–138) and “acceptor-binding” (residues 142–260) subdomains. The nucleotide-binding subdomain of MAP2569c comprised the “Rossman-like” fold (5) with four parallel \( \beta \)-strands (\( \beta_5\beta_4\beta_3\beta_6 \)), interspersed with helices (Fig. 1, A and B). The acceptor-binding subdomain contained three antiparallel \( \beta \)-strands (\( \beta_8\beta_7\beta_9 \)), also separated by helices, and a disordered region (residues 174–188). The signature catalytic DXD motif of the GT-A fold is found, as expected, at the juncture of the two catalytic subdomains and comprised Asp-139–Ser-140–Asp-141. A second small antiparallel \( \beta \)-sheet (sB), perpendicular to the central \( \beta \)-sheet (sA), is formed by two short \( \beta \)-stands: \( \beta_6' \) at the N terminus of the acceptor-binding subdomain and \( \beta_9' \) within a short loop bridging the acceptor-binding subdomain and C-terminal domain. This loop (residues 261–267) sat adjacent to the DXD motif and traced the nucleotide-binding/acceptor-binding subdomain divide. The C-terminal domain (residues 268–329) contained a long helical region and a prominent antiparallel two-stranded twisted \( \beta \)-sheet (sC; \( \beta_1\beta_1\beta_1\beta_1 \)) rich in aromatic residues.

Although there was a monomer in the asymmetric unit, a crystallographic dimer was observed in which this unusual C-terminal tail wrapped around its 2-fold related partner (see Fig. 1C). The remainder of this dimeric interface is formed by residues in the long helical region at the start of the C-terminal domain and in the N-terminal domain and acceptor-binding subdomain. The buried surface area at this dimeric interface was quite extensive (~1900 Å²) and is therefore likely to represent a biological dimer. Supporting this assertion are AUC studies of MAP2569c, which was shown to be a dimer in aqueous solution (see supplemental Fig. 1 and supplemental Table 1). Given the location of the substrate-binding sites, it appears that the MAP2569c dimer is probably required for stability as opposed to catalysis. Accordingly, the dimeric MAP2569c exhibits a GT-A fold and DXD motif consistent with GT activity.
Comparison of the GT-A Fold of MAP2569c with Other Glycosyltransferases—We next compared the GT-A fold of MAP2569c to other structures in the PDB using the DALI server (21) and identified a number of nucleotide-binding proteins with similar architecture. The closest structural match (with a Z-score of 18.6) was to the catalytic domain of mannosylglycerate synthase (MGS) from *Rhodothermus marinus* (22) (Protein Data Bank code 2bo4, 181 Ca atoms; root mean square deviation of 2.0 Å, and 22% sequence identity). Based on sequence similarity, MGS was previously classified as a member of the GT2 family (23); however, MGS was subsequently shown to use a retaining mechanism and was reclassified as the founding member of the GT78 family (22). The next two closest structural homologues from the GT superfamily (with Z-scores of 13.5 and 12.4, respectively) belong to SpsA from *B. subtilis* (8) (Protein Data Bank code 2iu8, 166 Ca atoms; root mean square deviation of 3.4 Å and 13% sequence identity) and polypeptide *N*-acetylglucosaminyltransferases from *Homo sapiens* (24) (Protein Data Bank code 2lq8, 171 Ca atoms; root mean square deviation of 3.0 Å, and 16% sequence identity). As with MGS, polypeptide *N*-acetylglucosaminyltransferase was initially grouped with SpsA in the GT2 family but has since been shown to use a retaining mechanism and was subsequently reclassified as a member of the GT27 family (19, 21). Of the 20 highest structural matches to the GT-A fold of MAP2569c, only six are GTs. Consequently, the GT-A fold of MAP2569c shows highest structural homology to the catalytic domain of a GT with a retaining mechanism and displays significant structural homology to other nucleotide-binding protein families. This structural similarity of MAP2569c to other nucleotide-binding protein families is indicative of both the structural diversity of and lack of structural data for this rapidly growing class of enzymes.

**Nucleotide Binding**—As a candidate member of the GT2 family, MAP2569c is predicted to use Mn$^{2+}$ and a nucleotide-sugar as its activated donor substrate (23). Accordingly, we sought to define the nucleotide specificity of this enzyme. First, we screened for nucleotide binding via a crystallographic approach. A crystallography-based approach has previously been applied with other GTs to identify the donor sugar (25). Namely, we soaked 10 mM CMP, GDP, and UDP nucleotides into the apo-crystal form. The difference Fourier electron density maps calculated using these data (statistics not shown) revealed evidence for UDP only and not for CMP or GDP at the predicted donor substrate site of MAP2569c and thus indicated that UDP is the likely nucleotide component of the donor substrate for MAP2569c.

We subsequently determined the 2.3 Å resolution structure of MAP2569c in complex with its donor substrate product, Mn$^{2+}$-UDP, at pH 7.0 and refined it to $R_{factor}$ and $R_{free}$ values of 19.3 and 21.7%, respectively (see “Experimental Procedures”) (Tables 1 and 2 and Fig. 2A). Overall, the structure of MAP2569c is unchanged on binding UDP, although a major conformational change was observed in the loop (residues 261–267) linking the acceptor-binding subdomain and C-terminal domain (see below). The α- and β-phosphates on UDP interact with Asp-141$^{\text{Oxidized}}$ of the DXD motif via the Mn$^{2+}$ ion (Fig. 3A), as observed previously in the SpsA structure (6). In addition, His-263$^{\text{Mutated}}$ on the loop linking the acceptor-binding subdomain and C-terminal domain also coordinates the Mn$^{2+}$ ion. The α-phosphate also hydrogen-bonds to Tyr-234$^{\text{D99}}$, and the β-phosphate forms a water-mediated hydrogen bond with Asp-139$^{\text{Mutated}}$ and is further within van der Waals contact of the side chains of Met-274 and Arg-266.

The uracil base of UDP stacked against the side chains of Leu-57 and of Lys-119 on the α-helix hF and made further van der Waals contacts with the side chains of Pro-55 and Ser-86. The O2 on the uracil hydrogen-bonds to one of the two alternate conformations modeled for Ser-86$^{\text{W84}}$. In addition to these contacts with the nucleotide-binding subdomain, the uracil is also in van der Waals contact with Tyr-234 on the 3$_{10}$-helix hL.

The ribose ring of UDP formed van der Waals contacts with the side chains of Leu-57 and of Lys-119 on the α-helix hF and made further van der Waals contacts with the side chains of Pro-55 and Ser-86. The O2 on the uracil hydrogen-bonds to one of the two alternate conformations modeled for Ser-86$^{\text{W84}}$. In addition to these contacts with the nucleotide-binding subdomain, the uracil is also in van der Waals contact with Tyr-234 on the 3$_{10}$-helix hL.

The ribose ring of UDP formed van der Waals contacts with the side chains of Pro-55, Lys-119, and Tyr-234. The O2 on the uracil hydrogen-bonds to one of the two alternate conformations modeled for Ser-86$^{\text{W84}}$. In addition to these contacts with the nucleotide-binding subdomain, the uracil is also in van der Waals contact with Tyr-234 on the 3$_{10}$-helix hL.
MAP2569c Complexed with UDP-glucose—To identify what UDP-sugar MAP2569c could bind, a similar co-crystallization approach was employed. UDP-activated sugars that act as natural donor substrates for GTs include UDP-GalNAc, UDP-GlcNAc, UDP-\(\alpha\)-L-arabinose, UDP-Gal, UDP-Glc, UDP-\(\alpha\)-D-glucuronic acid, and UDP-\(\alpha\)-D-xylose (although only UDP-GlcNAc, UDP-Gal, and UDP-Glc have been found in mycobacteria, all natural UDP-sugar donor substrates were investigated for comparison). For all crystal structures solved using these UDP-sugars, there was electron density for at least the uracil ring of the UDP-sugar at the predicted donor substrate site of MAP2569c. For example, the complex with UDP-GlcNAc was refined to 2.2 Å to \(R_{\text{factor}}\) and \(R_{\text{free}}\) values of 18.8 and 20.5%, respectively (see “Experimental Procedures” and Tables 1 and 2) and revealed evidence for the uracil and ribose rings and \(\alpha\)-phosphate of the UDP-sugar (see supplemental Fig. 2). However, only in the complex with UDP-Glc was electron density also observed for Mn\(^{2+}\), the \(\beta\)-phosphate, and the sugar moiety (see below) (Fig. 2B) at the predicted donor substrate site. This suggests specificity for UDP-Glc over all other natural, including chemically related, UDP-sugars.

Second, to further verify this specificity, the enzyme activity of MAP2569c was analyzed with the UDP-sugars found in mycobacteria, UDP-GlcNAc, UDP-Gal, and UDP-Glc. The highest activity was observed with UDP-Glc, with no activity toward UDP-Gal or UDP-GlcNAc (Fig. 3), suggesting that MAP2569c uses UDP-Glc as its natural donor sugar. There was some residual activity observed with GDP-Glc (20% of that observed with UDP-Glc), suggesting that there is some plasticity in donor sugar binding (Fig. 3).

UDP-Glc sat in a “folded back” conformation at the putative donor site, reminiscent of the conformation first observed in a donor substrate analogue in complex with the retaining GT LgtC (26), with a buried surface area of \(\sim 50\) Å\(^2\). The donor substrate traverses the cleft between the nucleotide- and acceptor-binding subdomains (see Fig. 1B). The interactions with UDP are similar to those observed in the complex of MAP2569c with Mn\(^{2+}\)-UDP.

The Glc moiety of the donor substrate is within hydrogen bonding distance of several residues within MAP2569c (see below) (Figs. 2C and 4). The O4’ and O6’ OH moieties hydrogen-bond to Glu-237\(^{\text{O4d}}\) on the \(\alpha\)-helix hP; the O3’-OH and O4’-OH hydrogen-bond to Lys-119\(^{\text{N4}}\), and the O3’-OH further hydrogen-bonds to Asp-139\(^{\text{G62}}\). In addition, Asp-139\(^{\text{G62}}\) forms a water-mediated hydrogen bond with the O2’-OH. The O5’-OH hydrogen-bonds to Leu-214\(^{\text{O3}}\), and the O6’-OH fur-
ther forms water-mediated contacts with Tyr-234^O and Ile-238^N. Given that the O4'-OH is the only distinction between UDP-Glc and UDP-Gal, the interactions with this oxygen are probably important for discriminating the natural donor substrate for MAP2569c. The C6' and O6' make van der Waals contacts with the side chain of Leu-214, and the C5' and O5' also make van der Waals contacts with the side chain of Met-274, which adopts a different conformation to accommodate the Glc moiety. Consequently, the sugar moiety of the donor substrate for MAP2569c interacts with residues from the catalytic and C-terminal domains of MAP2569c, and the large number of contacts dictates the specificity for UDP-Glc.

**Induced Fit**—The crystal structures of GTs of the GT-A fold have often indicated that a flexible loop in the vicinity of the nucleotide-binding site plays a critical role in the catalytic mechanism of the enzyme (for a review, see Qasba et al. (27)). To investigate whether induced fit plays a role in MAP2569c, we compared the structures of the apo and ligated forms. The comparative analyses revealed that upon ligation, the loop (residues 261–267) linking the catalytic and C-terminal domains in MAP2569c changes conformation.

In the apo form of MAP2569c, residues 262–263 on the flexible loop form a short β-strand (β9') of the "transient" β-sheet (sb, β6'-β9') and residues 268–270 form a 3_10-helix (hO) at the start of the C-terminal domain (see Fig. 5). However, in the MAP2569c:Mn^{2+}-UDP-Glc complex, this flexible loop adopted a different conformation, driven by the participation of the imidazole group on His-263 in a Mn^{2+}-mediated interaction with the α- and β-phosphates on UDP-Glc (see Fig. 5). To accommodate this, the β9' and β6' strands are restructured into loops, and residues 268–270 form an α-helix.

In the complex of MAP2569c: Mn^{2+}-UDP, the loop adopted an "intermediate" conformation (root mean square deviation of ~0.4 Å) to that observed in the apo and "donor substrate" liganded forms of MAP2569c. Moreover, the β-phosphate on UDP interacted less with the C-terminal domain than in the complex with Mn^{2+}-UDP-Glc. Instead, the β-phosphate made van der Waals contacts with the side chain of Arg-266 (disordered in the apo and Mn^{2+}-UDP-Glc liganded forms of MAP2569c) and is within water-mediated hydrogen bond-distant distance of Asp-139^Glu^2. Consequently, the conformational changes observed in the flexible loop containing His-263 and associated changes in adjacent secondary structural elements are important for coordination of the donor substrate and for stabilization of the donor substrate product of MAP2569c.

**The Role of His-263**—To further investigate the involvement of His-263 in Mn^{2+} coordination, we determined the crystal structure of MAP2569c in complex with Mn^{2+}-UDP-Glc at a pH value below the pK_a of histidine (pH 5.5 was used in this study). At such a pH value, histidine acts as an acid rather than as a base (as under physiological conditions) and thus cannot participate in Mn^{2+} coordination. The complex with Mn^{2+}-UDP-Glc at pH 5.5 was refined to 2.6 Å to R_\text{factor} and R_\text{free} values of 18.0 and 21.9%, respectively (see "Experimental Procedures") (Tables 1 and 2). The structure revealed evidence for the uracil and ribose rings and α-phosphate of the UDP-sugar only (see Fig. 2C), indicating the requirement for the ionizable

---

**FIGURE 2.** **Unbiased electron density for the putative donor substrate product and donor substrate of MAP2569c.** A, Mn^{2+}-UDP; B, Mn^{2+}-UDP-Glc at pH 7.0; C, UDP-Glc at pH 5.5. A–C, the electron density maps shown are F_o - F_c, simulated annealing omit maps contoured at 3σ. The Mn^{2+} ion and water molecules are represented by purple and red spheres, respectively. Hydrogen bonds and Mn^{2+}-mediated interactions are represented by dashed black and purple lines, respectively.
imidazole group on His-263 for Mn$^{2+}$ coordination, and thus also for β-phosphate and sugar coordination, in the complex of MAP2569c with Mn$^{2+}$-UDP-Glc at pH 7.0.

Comparison with Other GTs—Of the inverting GTs of the GT-A fold solved in complex with a nucleotide-sugar donor substrate, the catalytic domain of MAP2569c showed the highest structural homology to N-acetylglucosaminyltransferase I from Oryctolagus cuniculus (28) (Z-score 12.0, Protein Data Bank code 1foa, 164 Ca atoms superimpose with a root mean square deviation of 3.2 Å and a sequence identity of 13%) (21). Although the flexible loop in N-acetylglucosaminyltransferase I also undergoes conformational change on activated donor substrate coordination, no residue on this loop is directly involved in divalent metal coordination. In addition, rather than trace the nucleotide-binding/acceptor-binding subdomain divide, this loop in N-acetylglucosaminyltransferase I forms a “flap” over its nucleotide-sugar substrate (see Fig. 6A).

Next we compared the residues involved in donor substrate coordination in MAP2569c with those found in other GTs of the GT-A fold and identified structurally homologous segments in the “retaining” GTs MGS (22) and polypeptide N-acetylgalactosaminyltransferase (24). In all three structures, a histidine (as part of an Asp/His motif) within this loop is required for divalent metal ion coordination. In addition, there is a similar long helical region at the start of the C-terminal domain. In MAP2569c and MGS, structurally homologous methionine side chains in this region (Met-274 in MAP2569c and Met-229 in MGS) have also been shown (here) to play analogous roles in sugar coordination (see Fig. 6A).

The folded back conformation of UDP-Glc in complex with MAP2569c also resembled the conformation of GDP-Man in complex with MGS, and the mode of sugar coordination in these structures is similar (see Fig. 4) The side chains of Leu-163, Trp-189, and Met-229 in MGS overlay with those of Leu-214, Tyr-234, and Met-274 in MAP2569c and similarly form hydrophobic contacts with the sugar moiety in these structures. The side chains of Asp-192 in MGS and Glu-237 in MAP2569c are both within hydrogen bonding distance of the O4’ and O6’
Crystal Structure of a Mycobacterial Glycosyltransferase of the GT-A Fold

A

B

FIGURE 6. Comparison of the flexible loop of MAP2569c in complex with its putative donor substrate and those of inverting and retaining GTs of the GT-A fold. A, N-acetylglucosaminyltransferase I (in purple) in complex with \( \text{Mn}^{2+} \)-UDP-GlcNAc (donor substrate not shown) superposed on MAP2569c (in green) in complex with \( \text{Mn}^{2+} \)-UDP-Glc (shown in a ball and stick representation). B, MGS (in pink) in complex with \( \text{Mn}^{2+} \)-GDP-Man (donor substrate not shown) superposed on MAP2569c (green) in complex with \( \text{Mn}^{2+} \)-UDP-Glc (shown in a ball and stick representation). The \( \text{Mn}^{2+} \)-mediated interactions between the N61 on His-263 of MAP2569c and the \( \alpha \)- and \( \beta \)-phosphates on UDP-Glc are represented by dashed purple lines.

However, additional residues on the flexible loop in MGS, which is notably partially disordered in the complex of MGS with \( \text{Co}^{2+} \)-GDP, participate in \( \text{Mn}^{2+} \)-GDP-Man coordination, with Arg-218 and Tyr-220 also interacting with the \( \alpha \)- and \( \beta \)-phosphates on GDP-Man, respectively (see Fig. 6B). The structural homology observed between MAP2569c and retaining GTs of the GT-A fold extends beyond the catalytic domains of these enzymes to the flexible loops and adjacent secondary structural elements that undergo significant conformational change on donor substrate coordination.

The Acceptor Binding Site of MAP2569c—A bound organic molecule often indicates an important (functional) site in a crystal structure. Indeed, the binding site of a citrate molecule in the apo form of MAP2569c is similarly located to the binding site observed for a citrate in the apo form of MGS (see Fig. 7, A and B), with Thr-192 and Arg-261 in MAP2569c and Thr-139 and Arg-131 in MGS playing analogous hydrogen-bonding roles in coordinating the citrate in these structures. Indeed, the citrate in MGS notably makes interactions with MGS similar to that of the natural acceptor substrate of MGS, \( \alpha \)-glycerate, indicating the adaptability of this binding site. Although the threonines within hydrogen bonding distance of the citrate are from structurally homologous segments of MAP2569c and MGS, the arginines are from different regions. To illustrate, Arg-261 in MAP2569c is from the flexible loop involved in “substrate donor” coordination, whereas Arg-131 in MGS is from a loop that structurally corresponds to the disordered region (residues 174–188) in MAP2569c. Consequently, although the citrate binding sites in the apo form of MAP2569c and MGS are similarly located, the arginines within hydrogen bonding of the citrate in MAP2569c and MGS are from non-structurally homologous segments of the polypeptide chain, suggesting that the mode of natural acceptor substrate binding may differ between MAP2569c and MGS.

Sequence Similarity of MAP2569c to Putative Orthologues from the Corynebacterineae—To gain further insight into the function of MAP2569c, we compared the sequence of
MAP2569c with those of its putative orthologues from closely related *Mycobacterium* species, including *Mycobacterium ulcerans*, *M. tuberculosis*, and *Mycobacterium leprae*, and from other representatives of the *Corynebacterineae*, including *Nocardia farcinica* and *Corynebacterium glutamicum* (see supplemental Fig. 3). The sequence identity to MAP2569c, calculated using ClustalW (29), of these orthologues ranges from 47 to 85%, with a high level of identity for all sequences throughout the catalytic domain (residues 50–260), flexible loop (residues 261–267), and long helical region at the start of the C-terminal domain (residues 268–285) of MAP2569c, suggesting that these orthologues from the *Corynebacterineae* share the same architecture for these structural modules involved in donor substrate coordination in MAP2569c.

In addition, apart from Ser-86 and Ser-140, all of the MAP2569c residues observed to form (direct or indirect) hydrogen-bonding, Mn$^{2+}$-mediated, or van der Waals interactions with UDP-Glc at pH 7.0 are conserved in these orthologues. Moreover, the citrate-binding site in MAP2569c is well conserved throughout the orthologues. This essentially invariant nature of the donor and potential acceptor substrate binding sites in MAP2569c across the orthologues from the *Corynebacterineae* suggests that each of these candidate GTs performs a similar role in the synthesis of the unique cell wall of this suborder.

**DISCUSSION**

The ubiquitous nature of oligosaccharides and glycoconjugates and the rapidly growing number of candidate GT sequences uncovered by genomic analyses are indicative of a wealth of untapped knowledge on carbohydrate function. Due to the distant sequence similarities between some GT families, crystal structure analysis, as exemplified here, presents a powerful tool for establishing the relatedness between GTs and helps to bridge the chasm between the vast amount of sequence data and the paucity of functional and mechanistic knowledge about this important class of enzymes.

We have determined the structure of MAP2569c, which possesses a GT-A fold and a DXD motif associated with GT2 family activity. MAP2569c contains a flexible loop, which undergoes conformational change on activated donor substrate binding and houses a histidine residue required for divalent metal ion coordination, as has also been observed in the inverting GT, β4-galactosyltransferase (30). However, the overall fold and catalytic core of MAP2569c exhibits highest structural similarity to MGS, a GT that has been shown to use a retaining mechanism and represents the archetype of the GT78 family (22). Moreover, the structural homologies observed between MAP2569c and retaining GTs of the GT-A fold extend beyond the flexible loop in the vicinity of the activated donor substrate binding site to the adjacent secondary structural elements that also undergo conformational change on activated donor substrate coordination. Through crystallographic and enzymatic analyses, we revealed MAP2569c preferentially binds the activated donor substrate Mn$^{2+}$-UDP-Glc. The interactions between MAP2569c and Mn$^{2+}$-UDP-Glc are also reminiscent of those observed between MGS and its activated donor substrate (22). This level of specificity of MAP2569c is rather unusual in comparison with the GTs in general.

GTs use a carboxylate residue to bind the acceptor substrate and initiate transfer of the sugar moiety (27). The acidic residue in the retaining GT MGS (Asp-192) is structurally conserved in MAP2569c (Glu-237). Moreover, these residues are involved in similar hydrogen-bonding interactions at the donor substrate sites of MAP2569c and MGS. Nevertheless, as noted by Flint *et al.* (22), the acidic residue is similarly located in GTs of the GT-A fold that use both inverting and retaining mechanisms, possibly reflecting the evolution of retaining GTs of the GT-A fold from the inverting GT2 family (31). Flint *et al.* (22) also suggested that a change in the angle of the corresponding α-helices containing the acidic group in inverting versus retaining GTs of the GT-A fold could define the mechanism of these enzymes. However, when comparing MAP2569c and other GTs, we could find no such correlation (data not shown). Rather, one shared feature observed in retaining GTs of the GT-A fold is the position of the side chain of the acidic residue on the nucleoside-proximal side of the sugar moiety, as opposed to the nucleoside distal side of the sugar moiety, as observed in inverting GTs of the GT-A fold (see supplemental Fig. 4) (32). Although there are no clear structural features that define the mechanism of GTs, based on the accumulative structural homologies between MAP2569c and MGS, we suggest that MAP2569c possesses the same retaining mechanism of MGS, but this contention will require further experimentation.

We also provide evidence that the orthologues of MAP2569c from other *Mycobacterium* sp. and from related *Corynebacterineae* share a common architecture. Moreover, given that the residues involved in donor and acceptor substrate recognition are also conserved across the orthologues, we suggest that these also exhibit similar substrate specificities and hence perform a similar role in the biosynthesis of the unique cell wall of these bacteria. Although all known and characterized GTs of the GT-A fold found in *Mycobacterium* species are members of the inverting GT2 family (3), our analyses indicate that some GTs may need to be reclassified.

A recent report describes glycosyl-3-phosphoglycerate synthase from *Mycobacterium smegmatis* and *Mycobacterium bovis* that has homology (~25% sequence identity) to the *M. tuberculosis* H$_{37}$Rv gene Rv1208. Glucosyl-(1→2)-glycerate is found at the reducing end of methylglucose lipopolysaccharide, which is involved in regulating fatty acid synthesis in *Mycobacterium* (33). The formation of glucosylglycerate is performed by glycosyl-3-phosphoglycerate synthase, in which it transfers glucose from NDP-glucose to glucosyl-3-phosphoglycerate. Recombinant glycosyl-3-phosphoglycerate synthase from *M. bovis* showed optimal activity with UDP-Glc and strictly required Mg$^{2+}$. Glycosyl-3-phosphoglycerate synthase enzymes are classified into the retaining family 81 of GTs. These observations lend support to MAP2569c (and Rv1208) being a UDP-Glc specific GT that uses a retaining mechanism.

Given the unique complex carbohydrate content of the *Mycobacterium* cell wall, GTs involved in its biosynthesis present promising targets for new drugs in the treatment of tuberculosis and other diseases caused by *Mycobacterium* species. Previous studies using saturation mutagenesis screening have
suggested that the *M. tuberculosis* homolog Rv1208 is essential for survival and growth (9). The function of Rv1208 is not yet known, but its most likely role is as a putative GT in *Mycobacterium* cell wall biosynthesis. Its high specificity for its candidate acceptor substrate that is probably exclusively found in *Mycobacterium* makes the orthologues of MAP2569c promising targets for new antimicrobials.

Acknowledgments—We thank the IMCA and GM/CA-CAT staff for assistance in data collection at the Advanced Photon Source (Chicago, IL).

REFERENCES

1. Crick, D. C., Mahapatra, S., and Brennan, P. J. (2001) *Glycobiology* 11, 107R–118R
2. Daffe, M., and Draper, P. (1998) *Adv. Microb. Physiol.* 39, 131–203
3. Berg, S., Kaur, D., Jackson, M., and Brennan, P. J. (2007) *Glycobiology* 17, 35R–56R
4. Wiggins, C. A., and Munro, S. (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 7945–7950
5. Rossmann, M. G., Moras, D., and Olsen, K. W. (1974) *Nature* 250, 194–199
6. Wiggins, C. A., and Munro, S. (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 7945–7950
7. Martinez-Fleites, C., Proctor, M., Roberts, S., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2006) *Acta Crystallogr. Sect. D Biol. Crystallogr.* 62, 743–753
8. Sassetti, C. M., Boyd, D. H., and Rubin, E. J. (2003) *Proc. Natl. Acad. Sci. U. S. A.* 100, 15307–15312
9. Zhang, Y., Xiang, Y., Van Etten, J. L., and Rossmann, M. G. (2007) *Structure* 15, 1031–1039
10. Persson, K., Ly, H. D., Dieckelmann, M., Wakarchuk, W. W., Withers, S. G., and Strynadka, N. C. (2001) *Nat. Struct. Biol.* 8, 166–175
11. Qasba, P. K., Ramakrishnan, B., and Bogegman, E. (2005) *Trends Biochem. Sci.* 30, 53–62
12. Gordon, R. D., Sivarajah, P., Satkunarajah, M., Ma, D., Lan, J. A., and Tenenbaum, J. A. (2006) *J. Mol. Biol.* 360, 67–79
13. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., and Thompson, J. D. (2003) *Nucleic Acids Res.* 31, 3497–3500
14. Ramakrishnan, B., Balaji, P. V., and Qasba, P. K. (2002) *J. Mol. Biol.* 318, 491–502
15. Tarbouriech, N., Charnock, S. J., and Davies, G. J. (2001) *J. Mol. Biol.* 314, 655–661
16. Ramakrishnan, B., Bogegman, E., Ramasamy, V., and Qasba, P. K. (2004) *Curr. Opin. Struct. Biol.* 14, 593–600
17. Empadinhas, N., Albuquerque, L., Mendes, V., Macedo-Ribeiro, S., and da Costa, M. S. (2008) *FEBS Microbiol. Lett.* 280, 195–202
18. Kabash, W., and Sander, C. (1983) *Biopolymers* 22, 2577–2637
19. DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, Palo Alto, CA