Glycobiology vol. 17 no. 6 pp. 35R–56R, 2007
doi:10.1093/glycob/cwm010
Advance Access publication on January 29, 2007

The glycosyltransferases of *Mycobacterium tuberculosis*—roles in the synthesis of arabinogalactan, lipoarabinomannan, and other glycoconjugates

Stefan Berg[2], Devinder Kaur[3], Mary Jackson[4], and Patrick J Brennan[1,3]

[3]Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, CO 80523; and [4]Unité de Génétique Mycobactérienne, Institut Pasteur, 75724 Paris Cedex 15, France

Received on October 20, 2006; revised on January 22, 2007; accepted on January 23, 2007

Several human pathogens are to be found within the bacterial genus *Mycobacterium*, notably *Mycobacterium tuberculosis*, the causative agent of tuberculosis, one of the most threatening of human infectious diseases, with an annual lethality of about two million people. The characteristic mycobacterial cell envelope is the dominant feature of the biology of *M. tuberculosis* and other mycobacterial pathogens, based on sugars and lipids of exceptional structure. The cell wall consists of a peptidoglycan-arabinogalactan-mycolic acid complex beyond the plasma membrane. Free-standing lipids, lipoglycans, and proteins intercalate within this complex, complement the mycolic acid monolayer and may also appear in a capsular-like arrangement. The consequences of these structural oddities are an extremely robust and impermeable cell envelope. This review reflects on these entities from the perspective of their synthesis, particularly the structural and functional aspects of the glycosyltransferases (GTs) of *M. tuberculosis*, the dominating group of enzymes responsible for the terminal stages of their biosynthesis. Besides the many nucleotide-sugar dependent GTs with orthologs in prokaryotes and eukaryotes, *M. tuberculosis* and related species of the order *Actinomycetales,* in light of the highly lipophilic environment prevailing within the cell envelope, carry a significant number of GTs of the GT-C class dependent on polyprenyl-phosphate-linked sugars. These are of special emphasis in this review.

Key words: Mycobacterium/glycosyltransferase/classification/arabinogalactan/lipoarabinomannan

Introduction

Glycosylation events are among the most common and important enzymatic reactions in nature. Still, the responsible glycosyltransferases (GTs) are in general poorly understood. The gathering of information about GTs is a difficult task mainly because the enzymes are unstable, often membrane-associated, and present in the cell in very low concentrations. In bacteria, the majority of GTs is involved in the synthesis of, e.g., glycolipids, peptidoglycan (PG), and lipooligosaccharides (LOSs)—essential components of the cell envelope—and can thus be suitable targets for drug development against bacterial pathogens. In *Mycobacterium tuberculosis*, some of the first-line tuberculosis (TB) drugs target cell wall synthesis, but their specific targets and mechanisms of inhibition are not well defined. Together with the serious problem of drug resistance, particularly multidrug resistant-TB, a better understanding of mycobacterial cell wall biosynthesis is required in order to elucidate the targets of existing drugs and to find new ones. In this context, the many uncharacterized GTs of *M. tuberculosis* are of particular interest. This review summarizes current information on characterized and putative GTs in *Mycobacterium* spp. Extra attention has been given to a dozen open reading frames (ORFs) recently proposed as polyprenyl-dependent GTs (Liu and Mushegian 2003; VanderVen et al. 2005; Alderwick et al. 2006; Dinadayala et al. 2006; Kaur et al. 2006; Morita et al. 2006), including the Emb proteins that were the first GTs shown to be involved in arabinan biosynthesis (Belanger et al. 1996; Zhang et al. 2003; Berg et al. 2005). These ORFs are conserved in most mycobacterial genomes and are of particular interest in efforts to define the many unknown aspects of cell wall synthesis.

The known and putative roles of GTs in mycobacterial cell wall biosynthesis

The envelope of *Mycobacterium* spp. is a source of unique carbohydrates. A complex, consisting of mycolic acids, the heteropolysaccharide arabinogalactan (AG), and PG, constitutes "the core" of the cell wall (Figure 1) (reviewed in Crick et al. 2001). This covalently linked structure is intercalated with numerous glycolipids such as lipoarabinomannan (LAM), the phosphatidylinositol-containing mannosides (PIMs), the phthiocerol-containing the so-called phenolic glycolipids (PGLs), trehalose-dimycolate (TDM; the so-called "cord factor"), trehalose-monomycolate (TMM), etc. In addition, the proposed capsular-like material located outside the characteristic mycolic acid layer consists mostly of polysaccharides such as arabinomannan and α-glucans (Daffé and Draper 1998). Most of these carbohydrates are unique to the order *Actinomycetales*, and important from the perspective as the source of potential targets for new drug development against TB. For instance, the arabinans and galactans of mycobacterial cell walls consist of D-arabinofuranose (D-Ara*f*) and D-galactofuranose (D-Gal*f*) (McNeil et al. 1987), and the synthesis of these rare sugars alone presents opportunities for new chemotherapeutics;

[1]To whom correspondence should be addressed; e-mail: Tel: +1-970 491 6700; Fax: 970 491 1815; e-mail: patrick.brennan@colostate.edu
[2]Present address: Centre for Molecular Microbiology and Infection, Imperial College London, London SW7 2AZ, UK

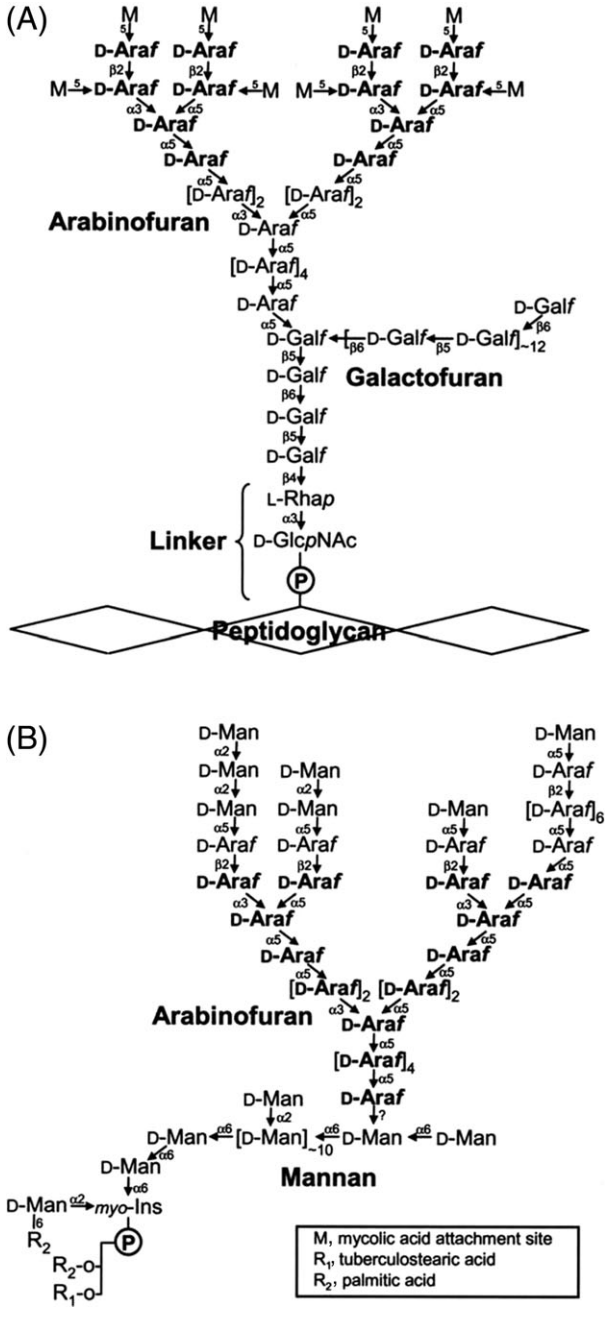

Fig. 1. Characteristic structures of AG (**A**) and ManLAM (**B**) of *M. tuberculosis*. A galactofuran of approximately 30 Gal*f* residues and three arabinofuran chains ($3 \times Ara f_{22}$ of which only one is shown here) form AG that is connected to PG via a linker unit at the reducing end and to mycolic acids at the nonreducing ends, all together forming the mycobacterial cell wall complex mAGP. The characteristic Ara_6 motif of AG-arabinan is shown in boldface (Figure 1A). The mannan of LAM has been defined with approximately 20–25 Man residues while the number and length of the attached arabinan chain(s) are uncertain. An internal Ara_{18} motif (in boldface in Figure 1B) was recently proposed for LAM of *M. smegmatis*. This motif, also part of AG-arabinan, may constitute a precursor arabinan, subsequently substituted at the nonreducing ends with different extensions in the formation of native LAM-arabinan (Shi et al. 2006).

indeed, the mode of action of ethambutol (EMB), a first-line drug in the treatment of TB, involves inhibition of the synthesis of the D-arabinans (Mikušová et al. 1995).

Details of the structures of the two dominating heteropolysaccharides of mycobacterial cell wall, AG and LAM, are shown in Figure 1. It is well established that the reducing end of AG consists of a Rha-α1,3-GlcNAc disaccharide, which is attached in phosphodiester linkage to some of the muramic acids of PG (McNeil et al. 1990; Hancock et al. 2002). The linear galactan of AG is attached to this unit by a β1,4 anomeric binding while the linkages within the galactan chain are alternating β1,5 and β1,6, to form a galactofuran with a size of about a 30-mer (McNeil et al. 1987; Daffé et al. 1990). The arabinan of AG consists of two to three apparently similar branched chains attached to 5-positions of some of the Gal*f* residues of the galactan chain, close to its reducing end; each D-arabinan chain consists of 22 Ara*f* residues (Besra et al. 1995). A backbone built of α1,5-linked Ara*f* with several α1,3-linked branch points has been suggested as the core structure (Daffé et al. 1990), and the nonreducing ends are always terminated by β1,2-Ara*f*. This assembly leads to characteristic hexa-arabinoside (Ara_6) motifs at the nonreducing ends of AG (McNeil et al. 1994), of which the dimers [β-D-Ara*f*-1,2-α-D-Ara*f*] constitute attachment sites for mycolic acids. Together with PG, AG forms a substantial covalently linked network located between the plasma membrane and the mycolic acid layer. These barriers in concert make the mycobacterial cell wall extremely robust and difficult to penetrate.

Unlike AG, LAM is a noncovalently linked component of the cell envelope and may be anchored in the plasma membrane and/or in the mycolic acid layer, or both, via its phosphatidyl-*myo*-inositol (PI) unit. The reducing end of LAM (Figure 1) shares structural similarities with the PI-mannosides (PIMs; Figure 2) in that the inositol residues of the PI of both the PIMs and LAM are mannosylated at the 2- and the 6-positions. The mannan of LAM is an extension of the PIMs and it is composed of a linear α1,6-linked mannan backbone, frequently branched with single α1,2-linked mannoses, leading to a mannan of about 20–25 residues (Chatterjee et al. 1991; Khoo et al. 1996). The arabinan of LAM is endowed with a D-arabinan structure more variable than that in AG. A recent study of the LAM-arabinan of *M. smegmatis* suggested the occurence of an Ara_{18} motif (Figure 1B) resembling the internal structure of AG-arabinan. However, the length of the terminal extensions linked at the nonreducing ends of this motif seem to vary in LAM-arabinan (Shi et al. 2006). Each LAM carries about 50–80 Ara*f* residues (Khoo et al. 1996), but little is known about the number of arabinan chains attached to the mannan core of LAM. An important feature of LAM of pathogenic species of mycobacteria (members of the *M. tuberculosis* complex, *M. leprae*, and *M. avium*) is the presence of "Man-caps" (hence the name Man-LAM) consisting of single α-Man or short mannooligosaccharides attached to the expected β1,2-Ara*f* termini of the D-arabinan; these units were shown to promote the binding and entry of these mycobacteria into antigen-presenting cells through the C-type lectins, mannose receptor and DC-SIGN (Schlesinger et al. 1994; Geijtenbeek et al. 2003; Maeda et al. 2003; Koppel et al. 2004). In most other mycobacteria, the nonreducing termini of LAM either lack Man-caps or are sporadically modified by inositol-phosphate units, as is the case in *M. smegmatis* (Nigou et al. 2003).

Although the chemical composition of mycobacterial cell walls and the chemistry of individual components are well

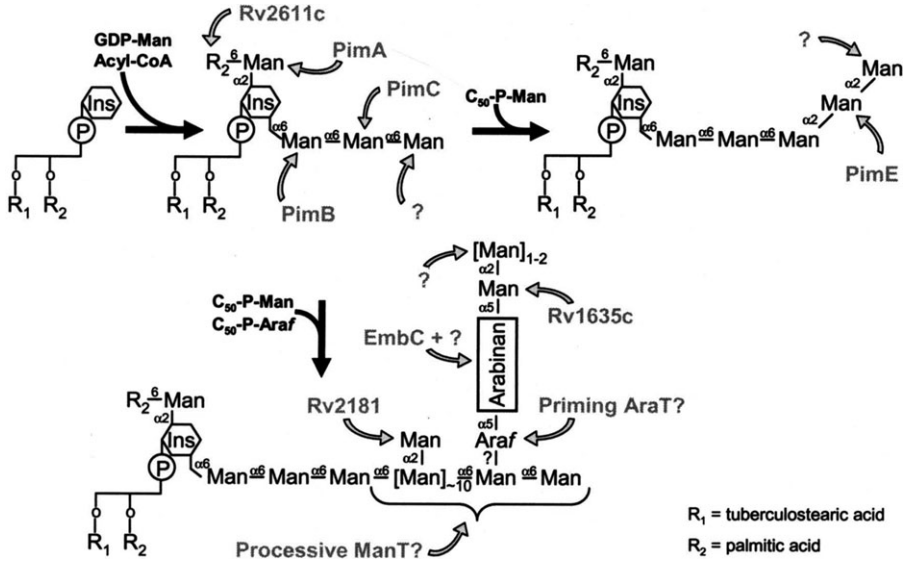

Fig. 2. Schematic figure for biosynthesis of PIM_4, PIM_6, and ManLAM. PIM_3 is formed from the precursor PI and the donor GDP-Man, used by the three consecutive GDP-Man-utilizing PimA-C. The ManT responsible for PIM_4 synthesis is still unknown and may be C_{50}-P-Man dependent. PIM_1 and PIM_2 are acylated with palmitic acid in the 6-position of the Man residue linked to position 2 of *myo*-inositol in a reaction catalyzed by the acyltransferase Rv2611c (Kordulakova et al. 2003). The subsequent catalytic steps for synthesis of polar PIM_5-PIM_6, and native LM and ManLAM have a proposed requirement for C_{50}-P-Man and C_{50}-P-Ara*f* (*Polyprenyl-dependent GTs of* M. tuberculosis). Enzymes indicated with names have been characterized.

understood, biosynthetic pathways are only now being defined, aided considerably by the sequences of mycobacterial genomes (Cole et al. 1998). Recent (Crick et al. 2001; Morita et al. 2004) and present knowledge of the biogenesis of the mycolate-AG-PG (mAGP) complex and of LAM of *M. tuberculosis* is summarized in Figure 3. Current knowledge of the many glycosylation steps that must be involved in these pathways are largely limited to the synthesis of precursors and intermediates. GTs implicated in the early stages of AG and LAM synthesis have been shown to utilize nucleotide-diphosphate (NDP)-linked sugar donors, while many glycosylations in the subsequent steps are believed to be dependent on decaprenyl-phosphate (C_{50}-P)-linked sugars (Besra et al. 1997).

Biosynthesis of sugar donors

The biosynthetic pathways for most of the nucleotide-sugar donors in *M. tuberculosis* have strong analogy to those of other prokaryotic species or have been characterized (Figure 4), notably UDP-α-D-Glc*p*NAc, UDP-α-D-Glc*p*, UDP-α-D-Gal*p*, and GDP-α-D-Man*p*; UDP-α-D-Gal*f* which is an isomeric product of UDP-α-D-Gal*p* is synthesized by the well characterized UDP-α-D-Gal*p* mutase (Glf) (Weston et al. 1997), and dTDP-β-Rha*p* is synthesized by the well characterized Rml-pathway (Ma et al. 1997). The known polyprenyl-based sugar donors of *Mycobacterium* spp. are β-D-glucosyl-1-monophosphoryl-decaprenol (C_{50}-P-Glc; Schultz and Elbein 1974), β-D-mannosyl-1-monophosphoryl-decaprenol (C_{50}-P-Man; Yokoyama and Ballou 1989), and β-D-arabinofuranosyl-1-monophosphoryl-decaprenol (C_{50}-P-Ara*f*) (Wolucka et al. 1994). Attempts to find a NDP-arabinose in mycobacteria have been reported (Takayama and Kilburn 1989; Singh and Hogan 1994), but no solid evidence on either the type of nucleotide or the form of arabinose was presented. Therefore, C_{50}-P-Ara*f* remains as the only proven donor of the Ara*f* units of mycobacterial LAM and AG. In support, a recent study of arabinan synthesis in the related *Corynebacterium glutamicum*, demonstrated C_{50}-P-Ara*f* as the only arabinose donor (Alderwick et al. 2005). As substrates for GTs, it is important to note that the anomeric configuration of the sugars in dTDP-Rha and the polyprenyl-based sugar donors is in the β-form, whereas it is α-form in the remaining NDP-sugars listed above.

C_{50}-P, not the usual undecaprenyl-phosphate (Mahapatra et al. 2005), plays the central role as carrier lipid in all known mycobacterial cell wall biosynthetic processes (Figure 3). It appears, for instance, in C_{50}-P-Ara*f;* the Ara*f* unit originates in 5-phosphoribosyl-pyrophosphate (pRpp) by a novel mechanism (Figure 4B) (Scherman et al. 1996; Mikušová et al. 2005). C_{50}-P is also the acceptor in the formation of C_{50}-P-Man and C_{50}-P-Glc; synthesis of the former is known to be catalyzed by Ppm1 (Rv2051c), a GDP-Man dependent mannosyltransferase (ManT) (Gurcha et al. 2002).

GTs involved in PG synthesis

PG synthesis in *M. tuberculosis* has been assumed to be similar to that of *Escherichia coli* (van Heijenoort 2001). However, PG of mycobacteria carries a variety of novel modifications (Mahapatra et al. 2005, and references therein). For instance, in *M. tuberculosis* and *M. smegmatis*, the muramic acid residues contain a mixture of the *N*-acetyl and *N*-glycolyl derivatives, a modification suggested to take place after the synthesis of the UDP-muramyl-NAc but before the formation of the UDP-muramyl-pentapeptides. The subsequent syntheses of Lipid I and Lipid II have been defined in *E. coli*. MraY transfers MurNAc-pentapeptide to undecaprenyl-phosphate and the product, Lipid I, is then further glycosylated by MurG through the use of UDP-GlcNAc to form Lipid II (Figure 3) (Ikeda et al. 1991). Proteins with strong homology to MraY and MurG of *E. coli* have been annotated in the

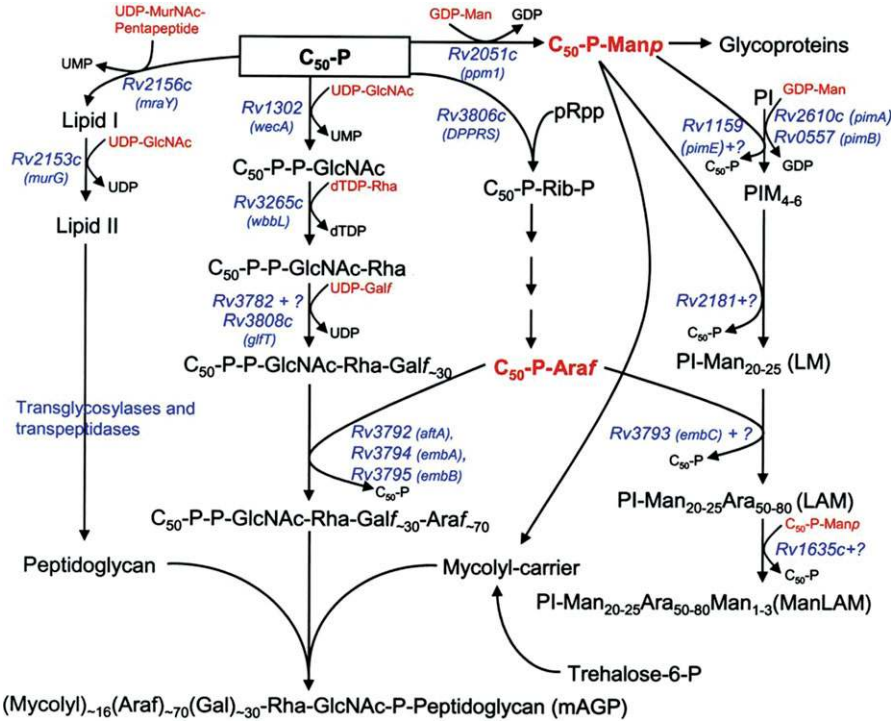

Fig. 3. Proposed biosynthetic pathways for mAGP and ManLAM of *M. tuberculosis*. The characterized and proposed H_{37}Rv genes responsible for the glycosylation steps in these pathways are indicated in blue. Sugar donors are shown in red.

M. tuberculosis genome (Cole et al. 1998) but their enzymatic activity has not yet been confirmed. It is believed that Lipid II is translocated to the periplasmic side of the plasma membrane where the glycopeptide moiety of Lipid II is polymerized in the final assembly of PG by activities of two putative penicillin-binding proteins (Rv0050 and Rv3682). These bifunctional proteins carry transglycosylase and transpeptidase activities and are, beside MurG, the only implicated GTs in the PG pathway (Figure 3) (Bhakta and Basu 2002).

GTs involved in linkage unit and galactan synthesis

The synthesis of the "linkage unit", on which the AG is assembled, is initiated by a transfer of GlcNAc-phosphate from UDP-GlcNAc to the acceptor C_{50}-P (Figure 3). This activity has been associated with Rv1302, due to its significant homology to WecA of *E. coli*, a well characterized GlcNAc-phosphotransferase (Amer and Valvano 2002). A Rha residue is then added in an α1,3 configuration by the recently described rhamnosyltransferase, WbbL (Rv3265c), to complete the linkage unit (Mills et al. 2004). Subsequent galactan synthesis involves at least two galactosyltransferases (GalTs) with specificity for the Gal*f* donor, UDP-Gal*f*. The transfer of the first Gal*f* is most probably performed by a GalT that is specifically designed to recognize the Rha residue in the linker unit and to create a β1,4 linkage. Rv3782 is the best-known candidate for this initial catalytic step (Mikušová et al. 2006). Several studies strongly support the principle that GlfT (Rv3808c) is a bifunctional GalT catalyzing the arrangement of the two differently linked Gal*f* in the formation of linear galactan (Mikušová et al. 2000; Kremer et al. 2001; Rose et al. 2006).

GTs involved in synthesis of the arabinans of AG and LAM

The variety of glycosidic linkages in both types of D-arabinans implies that the biosynthetic pathways should involve several arabinosyltransferases (AraTs). Realizing that the arabinans of AG and LAM are composed of D-Ara*f* with its origin solely in C_{50}-P-Ara*f* (Wolucka et al. 1994), those AraTs have to be dependent on this lipid-linked sugar donor. Thus far, the only candidates shown to be involved in arabinan synthesis are the Emb proteins (EmbA, EmbB, and EmbC) (Belanger et al. 1996) and AftA (Rv3792) (Alderwick et al. 2006). The Emb proteins play a key but largely undefined role in the synthesis of the arabinan components of both AG and LAM, with EmbA and EmbB contributing to AG synthesis (Escuyer et al. 2001), whereas EmbC is involved in the synthesis of LAM (Zhang et al. 2003). A more comprehensive discussion on the Emb proteins and their function in arabinan biosynthesis follows in section *The Emb proteins and their relationship to other GTs*. Disruption of the ortholog of *aftA* in *C. glutamicum* and use of a cell-free assay based on the recombinant *M. tuberculosis* enzyme provided evidence that AftA catalyzes the addition of the first Ara*f* residue from C_{50}-P-Ara*f* to the galactan domain of AG (Alderwick et al. 2006). Additional potential AraTs may be found among the putative polyprenyl-dependent GTs listed in Table II (section *Polyprenyl-dependent GTs of* M. tuberculosis).

ManTs involved in synthesis of LAM

The structural description of the PI-mannosides, spanning from PIM_1 to PIM_6 in different acylated forms (Lee and Ballou 1965; Khoo et al. 1995), and later on, the characterization of lipomannan (LM) and LAM (Hunter and Brennan

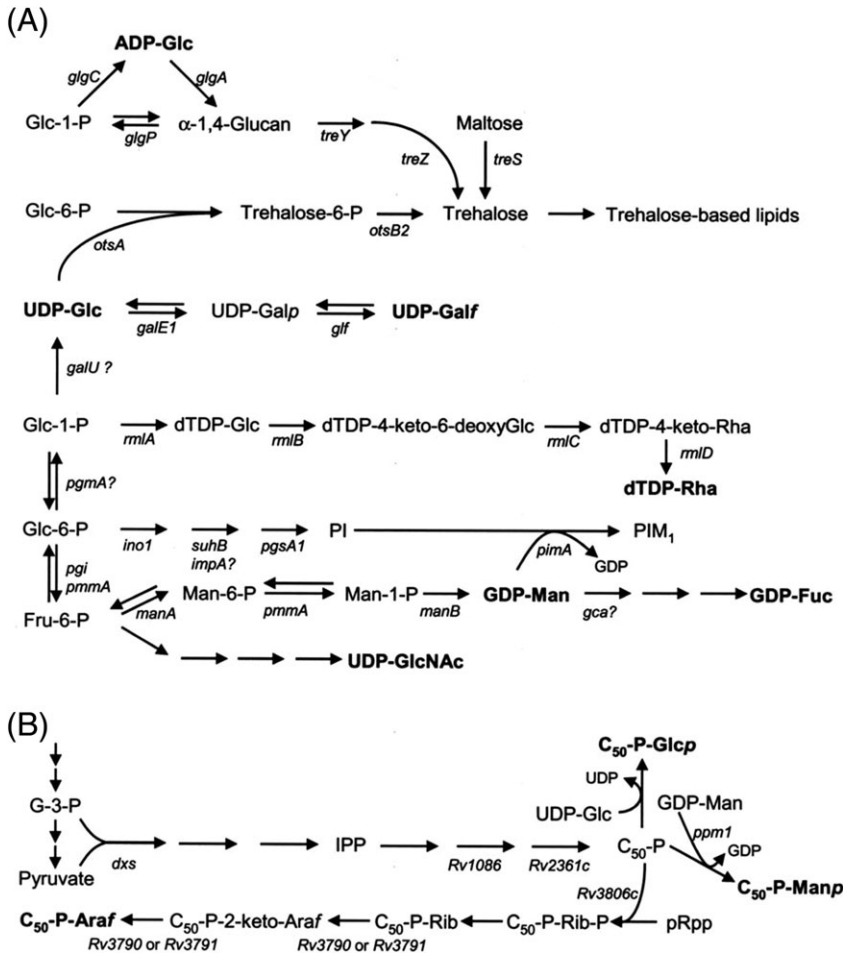

Fig. 4. Metabolic pathways for biosynthesis of the NDP-sugars (**A**) and C_{50}-P-linked sugars (**B**) in *M. tuberculosis* H_{37}Rv. All named genes are of the Sanger nomenclature and their functions have been characterized, except those labeled with question mark, which instead are annotated based on strong sequence similarities to genes with known functions. *pmmA* have dual functions by carrying both phosphogluco- and phosphomanno-mutase activities (McCarthy et al. 2005), and *impA* is a second candidate for the function of *suhB*. IPP, and isopentenyl diphosphate.

1990; Chatterjee et al. 1992), revealed their relationships in being based on PI, leading to the suggestion and evidence of shared biosynthesis (Figure 2) (Besra et al. 1997). In recent times the ManTs responsible for the initial reactions have been described. The synthesis of PIM_1 and PIM_2 are catalyzed by PimA (Rv2610c) and PimB (Rv0557), respectively, two enzymes which require the sugar donor GDP-Man (Schaeffer et al. 1999; Kordulakova et al. 2002). Such is also the case for PimC, a ManT which synthesizes PIM_3 in *M. tuberculosis* CDC1551; however, no strong homolog of PimC has been annotated in the genome of H_{37}Rv (Kremer et al. 2002). Further elongation of PIM_3 should lead to formation of the linear α1,6-linked mannan backbone of LM and LAM. However, such a pathway should include a branch point at PIM_4, one direction leading to LM/LAM, the other, through addition of two consecutive α1,2-linked Man*p* residues, resulting in PIM_6, an apparent dead-end product, not involved in LM and subsequent LAM synthesis (Figure 2) (Morita et al. 2004).

Early work had proposed that mannosylation of the more polar/mannosylated PIMs and LM involve both GDP-Man and C_{50}-P-Man (Yokoyama and Ballou 1989). However, inhibition studies with amphomycin were shown to inhibit the synthesis of PIM_4, PIM_5, and PIM_6, suggesting that these enzymatic steps utilize C_{50}-P-Man as a donor substrate (Morita et al. 2004). Indeed, PimE (Rv1159) was recently identified as a probable C_{50}-P-Man-dependent ManT responsible for the formation of PIM_5 from PIM_4 (Morita et al. 2006). Whether PimE also transfers the sixth mannose to form both PIM_5 and PIM_6 remains to be determined. Our laboratory recently created the *M. smegmatis* mutant Δ*MSMEG4250* (ortholog to *Rv2181* of *M. tuberculosis*), lacking α1,2-linked Man*p* on the mannan backbone of LAM, strongly suggesting this protein to be an α1,2-ManT in the synthesis of mature LAM (Kaur et al. 2006). The phenotype of mutant Δ*MSMEG4250*, which completely lacked LM but still produced a truncated form of LAM, has raised new speculations about the biosynthesis of mannan of LM/LAM. An earlier hypothesis of a "straight" pathway, in which linear LM served as the substrate for a branching enzyme leading to the formation of mature LM, subsequently used in LAM biosynthesis (Besra et al. 1997), has now been complemented with alternative routes (Kaur et al. 2006). The most innovative one suggests that smaller C_{50}-P-linked mannooligosaccharides

are being synthesized by Rv2181 and (an)other GT(s) and then used for chain extension at the nonreducing end of PIM_4 to form mature LM (Kaur et al. 2006). Future experiments should clarify these new ideas. Another ManT implicated in LAM biosynthesis of *M. tuberculosis* is Rv1635c, which recently was shown to be responsible for transferring the first Man*p* residue in Man-capping of ManLAM (Dinadayala et al. 2006). Both Rv2181 and Rv1635c have been proposed as polyprenyl-dependent GTs (see also section *Polyprenyl-dependent GTs of* M. tuberculosis), and characterization as such would be in accordance with the hypothesis stated earlier.

Additional GTs involved in synthesis of other glycoconjugates of Mycobacterium *spp.*

Trehalose is a precursor for the synthesis of the trehalose-containing LOSs, TMM, TDM, and several methyl-branched fatty acid-containing glycolipids such as sulfatides, and di-, tri-, and poly-acyltrehaloses. Mycobacteria have three alternative routes for trehalose synthesis (Figure 4) (De Smet et al. 2000), including the classical condensation of UDP-Glc and Glc-6-phosphate leading to an α,α-1,1-glycosidic linkage in the product, trehalose-6-phosphate. The glucosyltransferase involved, Rv3490 (OtsA), has been identified in *M. tuberculosis* (Pan et al. 2002), as has the phosphatase Rv3372 (OtsB2) which is responsible for dephosphorylation of the trehalose-6-phosphate (Murphy et al. 2005). In addition to this pathway, trehalose can be interconverted from maltose by trehalose synthase (TreS) (Nishimoto et al. 1996), or from glycogen involving the two enzymes TreY and TreZ (Maruta et al. 1996). Of these three pathways, the OtsAB-dependent route was recently shown to be predominant in *M. tuberculosis*, and OtsB2 shown to be essential for growth (Murphy et al. 2005).

It has been proposed that mycolic acids can be transferred via a mycolyl-mannosylphosphoheptaprenol (Besra et al. 1994) to trehalose-6-phosphate, arising from the OtsAB pathway, to yield phosphorylated TMM, which then can be dephosphorylated to yield TMM (Takayama et al. 2005). The screening of a transposon mutant library of *Corynebacterium matruchotii* for mutants with defects in corynemycolic acid synthesis led Wang et al. to propose that a putative polyprenyl-dependent GT, orthologous to Rv1459c of *M. tuberculosis* (section *Polyprenyl-dependent GTs of* M. tuberculosis), and a neighboring ATP-binding cassette (ABC) transporter were somehow involved in this process (Wang et al. 2006). However, all of this has to be proven since the intermediate, mycolyl-mannosylphosphoheptaprenol, has so far only been found in *M. smegmatis* (Besra et al. 1994). The subsequent synthesis of TDM, from two TMM molecules, and the transfer of mycolates to the nonreducing ends of AG have been show to involve a protein complex composed of antigens 85A, 85B, and 85C (Wiker and Harboe 1992; Belisle et al. 1997; Jackson et al. 1999).

Other carbohydrate-containing components in the cell envelope of *M. tuberculosis* are the phthiocerol/PGLs and *p*-hydroxybenzoic acid derivatives (*p*-HBADs), of which the PGLs were shown to be important virulence factors (Reed et al. 2004; Tsenova et al. 2005). Their biosyntheses involve the sequential addition of the three basic sugars, Rha, Rha, and Fuc catalyzed by the three GTs Rv2957, Rv2958c, and Rv2962c and several methylation steps, two of which are catalyzed by the methyltransferases Rv2952 and Rv2959c (Figure 5) (Perez, Constant, Lemassu, et al. 2004; Perez, Constant, Laval, et al. 2004). The PGLs of *M. tuberculosis, M. leprae*, and other species (reviewed in Brennan 1988) should not be confused with other mycobacterial 6-*O*-methyl hexose-containing cell wall glycolipids such as the glycopeptidolipids (GPLs), found in *M. avium* and *M. smegmatis* (Brennan et al. 1981; Daffé et al. 1983). The Rha moiety of the GPLs of *M. avium* is transferred by the characterized rhamnosyltransferase RtfA (Eckstein et al. 1998), and its strong ortholog, Gtf3, was recently suggested to carry an equivalent GT activity in *M. smegmatis* (Deshayes et al. 2005).

Other specific mycobacterial glycoconjugates are the 6-*O*-methyl glucose-containing lipopolysaccharides (MGLPs) of *M. tuberculosis, M. bovis* BCG, *M. smegmatis*, *M. phlei*, *M. xenopi*, and *M. leprae* (Lee 1966; Keller and Ballou 1968; Lornitzo and Goldman 1968; Hunter et al. 1986; Tuffal et al. 1995; Tuffal et al. 1998) and the 3-*O*-methyl mannose polysaccharides (MMPs) of *M. smegmatis* and several other mycobacterial species (Maitra and Ballou 1976; Weisman and Ballou 1984) both facilitators of long-chain fatty acid synthesis in vitro (Bloch 1977). An α1,4-ManT activity, with specificity for GDP-Man, has been characterized in the biosynthesis of MMPs (Weisman and Ballou 1984) but no gene has been annotated as such.

The cell wall "capsule" of mycobacteria contains D-arabinomannan similar to the arabinomannan domain of LAM, α-D-mannan, and a branched α-D-glucan (Lemassu and Daffé 1994; Ortalo-Magne et al. 1995). Glucan consists of linear α1,4-linked glucosyl residues occasionally substituted at position 6 with mono-, di-, tri-, tetra-, penta-, or hexa-glucosyl residues and thereby shares structural features with the glycogen stored in the cytosol of mycobacteria (Dinadayala et al. 2004). This similarity in structure suggests that they may use the same biosynthetic pathway.

Classification and structural aspects of GTs from *M. tuberculosis*

Families and structural superfamilies of GTs

The catalytic mechanisms described for sugar transfer reactions leads to either inversion or retention of the anomeric sugar binding. The enzymatic formation of an α- or β-glycosidic bond is consequently determined by (i) the mechanism used by the enzyme and (ii) the anomeric configuration of the donor substrate (Sinnott 1990). The classification, Carbohydrate-Active enZymes (CAZy; http://afmb.cnrs-mrs.fr/CAZY), has grouped GTs into about 87 families based on sequence homology, and which consistently differentiates inverting from retaining enzymes (Campbell et al. 1997). The many resolved X-ray structures of GTs over the last few years have revealed the number of structural superfamilies to be much less, suggesting that, despite low sequence identity between GT families, common evolutionary origins are probable (Murzin et al. 1995). To date, only three large superfamilies have been proposed (Unligil and Rini 2000; Liu and Mushegian 2003); GT-A and GT-B contain mainly soluble and peripheral membrane proteins which use NDP-sugar as donor substrates, while the third superfamily, GT-C, comprise integral membrane proteins with dependency for polyprenyl-linked sugar donors (Figure 6). No crystal structure has been

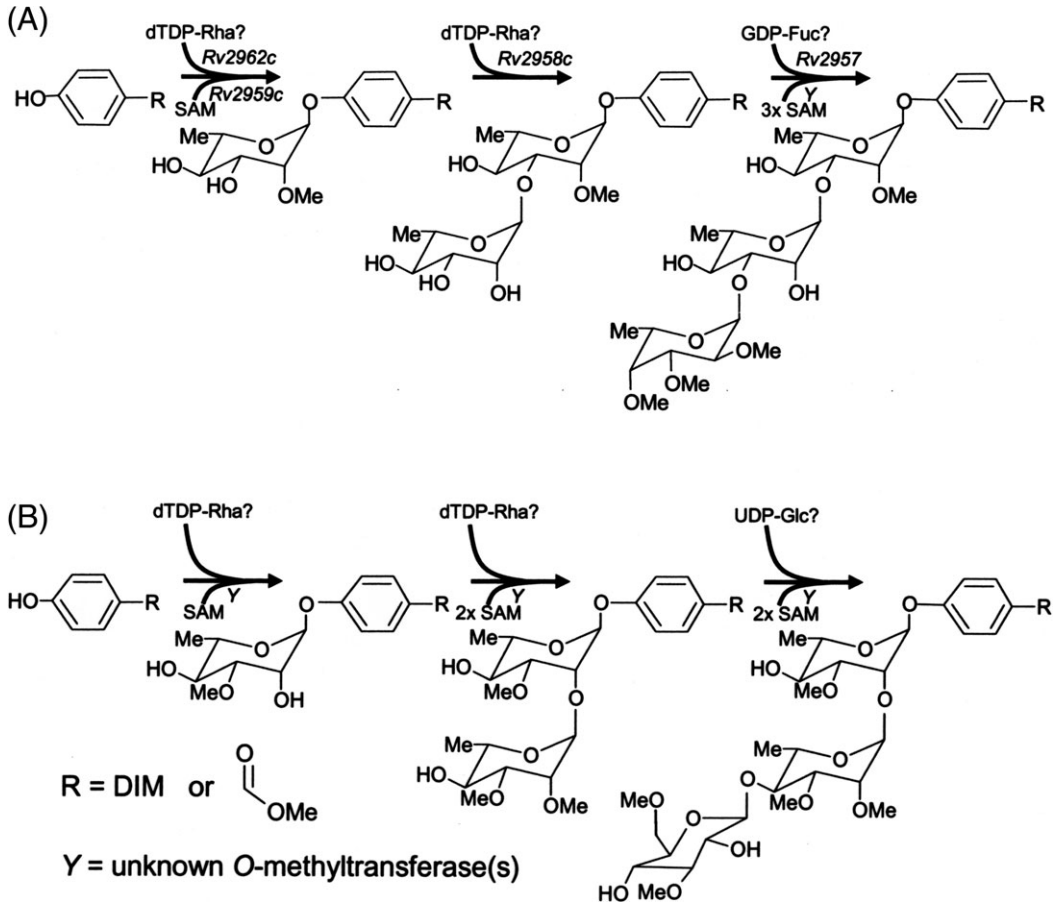

Fig. 5. Biosynthesis of PGLs and *p*-HBADs in *M. tuberculosis* (**A**) and *M. leprae* (**B**). The likely NDP-sugar donors utilized by the GTs are indicated, as well as *S*-adenosylmethionine (SAM), the suggested substrate for the methyltransferases. Genes identified in *M. tuberculosis* are shown (Perez, Constant, Laval, et al. 2004; Perez, Constant, Lemassu, et al. 2004). Notice the structural difference of the nonreducing sugar in these PGLs. DIM, dimycocerosate of phthiocerol.

solved for any members of the GT-C superfamily; instead, this group of GTs was discovered by iterative basic local alignment search tool (BLAST) searches and by structural comparisons (Oriol et al. 2002; Liu and Mushegian 2003).

Despite the fact that the catalytic mechanism used by enzymes within a GT family is consistent, families within a superfamily can use different mechanisms. It has been suggested that similar structural elements are employed in families having the same fold, irrespective of the stereochemistry of the glycosylation reaction (Persson et al. 2001). Thereby, the three-dimensional structure held by a GT does not dictate the catalytic mechanism. Although the catalytic mechanisms of both inverting and retaining GTs have been proposed, only the mechanism of inverting enzymes is to some extent understood. However, in both cases the involvement of an acidic amino acid, such as a glutamate or an aspartate residue, employed as a general base, is suggested to initiate the catalytic transfer (Unligil and Rini 2000; Davies 2001; Ramakrishnan et al. 2002). For further information on the structural and functional aspects of GTs, at least two recent reviews are relevant (Davies et al. 2005; Breton et al. 2006).

Classification of putative GTs of M. tuberculosis

It is difficult to estimate a rational number of GTs needed for biosynthesis of the complex cell wall structure of mycobacteria. However, GTs seem to be the largest group of enzymes involved in synthesis of the mycobacterial cell envelope. According to CAZy, among the approximately 3900 ORFs found in the genome of *M. tuberculosis* $H_{37}Rv$ (Cole et al. 1998), about 41 ORFs (approximately 1%) encode putative GTs. The majority of these classified GTs of *M. tuberculosis* is proposed to have a requirement for NDP-sugar donors, and they belong to families that have representatives in all kingdoms of life. In contrast, the polyprenyl-dependent GTs of *M. tuberculosis* are more confined to the order of *Actinomycetales* and mostly form their own GT families. Besides the CAZy classification, Wimmerova et al. used fold recognition analysis to study the genome of *M. tuberculosis* and found another 15 proteins with predicted similarity to the structural fold of GT-A and GT-B (Wimmerova et al. 2003). However, none of them has yet been shown to be a GT by biochemical means.

Most of the characterized and uncharacterized GT genes listed in Tables I and II are evenly distributed on the *M. tuberculosis* $H_{37}Rv$ chromosome. However, there are at least two obvious GT-containing gene clusters, each holding nine proposed GT genes; one is located in the region of *Rv1500* to *Rv1526c* and the other spans from *Rv3779* to *Rv3809c* (Figure 7). The former cluster contains mostly GTs proposed to utilize NDP-sugars but their functions are still unidentified,

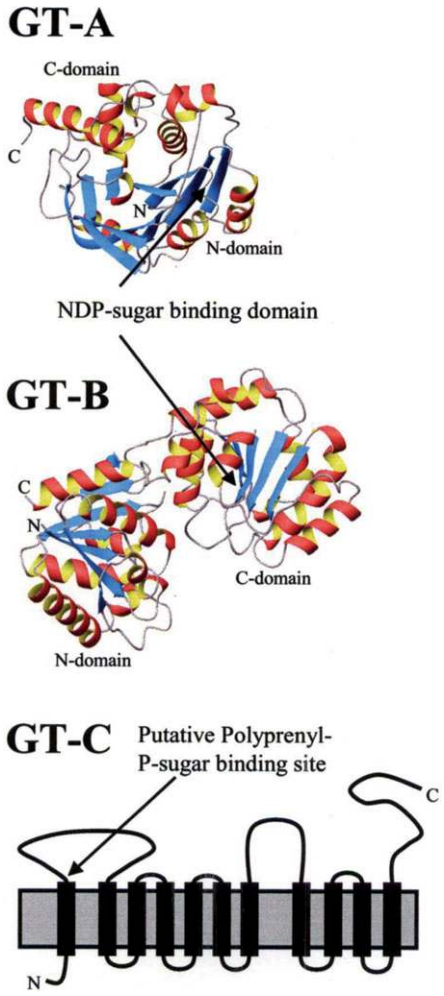

Fig. 6. The resolved structures of SpsA (PDB: 1QGS) from *B. subtilis* and MurG (PDB: 1F0K) from *E. coli* represent superfamily GT-A and GT-B, respectively. Representative structures of GT-A consist of a mixed α/β fold, organized into two tightly associated domains; a N-terminal NDP-sugar binding domain and a C-terminal acceptor binding domain (Tarbouriech et al. 2001). This formation should be compared with the typical structure of GT-B proteins, which comprises two distinct domains of Rossmann fold (α-β-α sandwiches) connected by a flexible hinge giving rise to a cleft between the domains (Ha et al. 2000; Mulichak et al. 2001). The binding domains for acceptor and donor substrates are reversed in GT-B enzymes, consequently leading to a NDP-sugar binding site in the C-terminal domain (Coutinho et al. 2003). In general, the most conserved domain of GTs in those two superfamilies is that for NDP-sugar binding (indicated). No protein structure of the GT-C superfamily has yet been solved. Here we suggest a generalized topology model for these integral membrane proteins of *M. tuberculosis* H_{37}Rv (Table II). The putative polyprenyl-P-sugar binding site, commonly located in the first or the second extracytoplasmic loop, is indicated.

except for LosA (Rv1500) which recently was annotated as a GT involved in LOS biosynthesis [(Burguiere et al. 2005); see section *Sugar-nucleotide dependent GTs of* M. tuberculosis]. The latter GT-containing gene cluster, on the other hand, has been described earlier as "the cell wall biosynthetic cluster" (Cole et al. 1998; Belanger and Inamine 2000) and includes several now characterized proteins implicated in AG, LAM, and mycolic acid biosynthesis (Mikušová et al. 2000). Interestingly, as many as five ORFs in this cluster are possibly polyprenyl-P-sugar dependent GTs (Table II) which is in agreement with the hypothesized use of C_{50}-P-Ara*f* and C_{50}-P-Man in the pathways of AG and LAM (Figure 3). Noteworthy, a third smaller cluster with three proposed GT genes (*Rv2174*, *Rv2181*, and *Rv2188c*) is located downstream of the *fts*/*mur*-gene cluster (*Rv2150c-2158c*), which carries a large number of genes involved in PG biosynthesis and cell division. Rv2181 was recently shown to be a ManT in LM biosynthesis (Kaur et al. 2006).

Sugar-nucleotide dependent GTs of M. tuberculosis

Comparisons of GTs of mycobacterial origins to GTs of known X-ray structures at the level of amino acid sequence and predicted secondary structure can help us towards structural and functional understanding of these enzymes. To explore such relationships, we investigated the amino acid sequences of all of the ORFs of *M. tuberculosis* H_{37}Rv (listed in Tables I and II) that had been proposed to belong to any of the three superfamilies, GT-A, GT-B, or GT-C. Most of the classified GTs of *M. tuberculosis* are proposed as NDP-sugar dependent, and they belong to the families GT-1, GT-2, GT-4, GT-20, GT-28, and GT-35, listed in Table I. Of these six families, GT-2 is the only one with GT-A fold; the remaining five belong to superfamily GT-B. Herein follows a closer examination of the properties of these GT families.

GTs of M. tuberculosis *with a proposed GT-A fold*. The characteristic fold of superfamily GT-A is here represented by SpsA (Figure 6), a GT-2 protein involved in spore formation of *Bacillus subtilis*. GT-2 enzymes use inverting mechanism, generally leading to a glycosidic bond in β-configuration, and carry a DxD motif found to bind a divalent cation as part of immobilization of the sugar-nucleotide donor (Charnock and Davies 1999). As many as 16 ORFs of *M. tuberculosis* H_{37}Rv (Table I) are members of GT-2 and they are consequently proposed to share the GT-A fold. So far, five of them (LosA, Ppm1, WbbL, Rv3782, and GlfT) have been biochemically characterized. The four that are highly conserved in the sequenced genomes of *M. leprae*, *M. smegmatis*, *C. glutamicum*, and *Nocardia farcinica*, Ppm1, WbbL, Rv3782, and GlfT, have been characterized as key enzymes in AG and LM synthesis. LosA (the ortholog of Rv1500 in *M. marinum*) was shown to be involved in LOS biosynthesis (see further below) (Burguiere et al. 2005).

The resolved crystal structure of SpsA showed that at least three invariant residues, Asp^{39}, Asp^{98}, and Asp^{99} in the N-terminal domain (Figure 6; Table I), are involved in binding of the donor substrate. Furthermore, Asp^{191} in the C-terminal domain (not shown here) may play a role in the catalytic event (Charnock and Davies 1999). The selected alignment of SpsA with the 16 GT candidates of *M. tuberculosis* shows that conserved Asp residues can be identified at comparable positions, and always within the two loops following Nβ2 and Nβ4 of SpsA (Table I). Also, predicted secondary structures in this region (not shown) suggest high similarity to the X-ray structure determined for SpsA, implying that these 16 proteins of H_{37}Rv carry a NDP-sugar binding site, with a structure similar to that of SpsA.

Rv3786c, which has not yet been classified by CAZy but shares the characteristic features of GT-2 proteins, is located in the cell wall biosynthetic cluster discussed earlier (Figure 7). We therefore suggest that it may be functioning as an additional GalT in the galactan synthesis of AG, in

Table I. CAZy classification of characterized and putative GTs of *M. tuberculosis* H_{37}Rv from the two superfamilies GT-A and GT-B. See Table II for footnotes

GT-A

Protein[a]	[b]	[c]	Conserved donor substrate binding motif[d]	Size[e]	TMD[f]	Donor[g]	Ml[h]	Ms[h]	Cg[h]	Nf[h]
GT-2 (Inverting)			Nβ2 Nα2 Nβ4 Nβ5 Cα5							
1QGS	bsSpsA		30-DFELFIMD**D**NSNEETLNVI-37-MAEGEYITYAT**DD**NIYMPDRLLK	256		UDP-Glc				
Rv0539			20-GYRALVV**D**NNSTDDTATVA-24-AATTPIVAVI**DAD**GSMDAGDLPK	210			*22*	66	23	25
Rv0696			105-GLRVIVVD**D**GS--ACPVES-30-ACTTDFVAFL**DSD**V---TPRRGW	470	1		*16*	68	-	52
Rv1208			77-VLDSGSTD**D**TEIRAIASGA-27-ATSGDIVVFI**DSD**LINPHPLFVP	324			80	71	37	61
Rv1500	LosA		33-DVEIIFVD**D**GSPDAALQQA-30-HATGDLVFLI**DSD**L---EEDPAL	342	2		*16*	15	20	23
Rv1514c			34-RIEHIVI**D**GGSGDDVVAYL-28-ASGDLLWFLHSA**D**--RFSGPDVV	262			-	15	18	16
Rv1516c			37-PVEIIVAD**D**ASTDATPAII-32-ARGEYVALCEA**DD**---YWIDPLK	336			-	*13*	22	-
Rv1518			45-EFEIVVVD**D**DSSDDCAAIA-29-ARGGLIQFVHG**DD**RLLPGALQTL	319			-	16	18	-
Rv1520			29-PVEVIIAD**D**ASTDATPRII-32-ARGEYLALCEG**DD**---YWTDPLK	346			-	*16*	21	-
Rv1525			31-GADYLIV**D**NRGDYPRIGTE-23-AEGYSHAMTLNN**D**TRVSKGFVAA	261			-	*18*	-	-
Rv2051c	Ppm1		637-AVHVLVVD**D**SSPDGTGQLA-34-SREYSVLVEM**DAD**GSHAPEQLQR	874	7	GDP-Man	67	56	40	45
Rv2957			51-DFELVLV**D**GGSTDETLDIA-31-ATGTWLLFLGA**DD**--SLYEADTL	275	1		-	18	20	-
Rv3265c	WbbL		35-PVSVLLA**D**NGSTDGTPQAA-29-EMAGDAGEPWV**DD**WVIVANPDVQ	301		dTDP-Rha	83	68	50	57
Rv3631			39-FDHVVCVD**D**GSTDGTGDIA-29-QPGAQVFATF**D**G**D**G---QHRVKD	241			80	*23*	57	23
Rv3782			32-DHLIVV**D**NDGCGDSPVREL-31-AQGADWVWLA**DD**GHAQDARVLA	304		UDP-Gal*f*	75	75	60	65
Rv3786c[i]			26-SDDIYVID**D**RSTDDTAEIL-37-FCRPDWVMMV**DAD**WLVETDIDLR	407			*16*	18	15	18
Rv3808c	GlfT		190-VIGAVIVP**D**QGERKVRDHP-36-NTDCQQILFM**DD**IRLEPDSILR	637		UDP-Gal*f*	85	73	53	66
Consensus			e vvDdgstd a a ddD							

GT-B

Protein[a]	[b]	[c]	Conserved donor substrate binding motif[d]	Size[e]	TMD[f]	Donor[g]	Ml[h]	Ms[h]	Cg[h]	Nf[h]
GT-1 (Inverting)			Cβ4 Cα4 Cβ5 Cα5							
1RRV	aoGtfD		297-ALFRRVAAVIH**H**G**S**AG**T**EHVATRAGVPQLVIPRNT**DQ**PYFAGRVAA	408		TDP-L-vancosamine				
Rv1524			306-AVFPTCRAVVH**H**GGAG**T**TAAGLRAGIPTLILWVTS**DQ**PIWAAQIKQ	414			61	58	-	-
Rv1526c			308-TILPKCRAVVH**H**GGAG**T**TAAGLRAGMPTLILWDVA**DQ**PIWAGAVQR	420			*53*	46	-	-
Rv2739c			285-ELLTRADLVICGGGHGMVAKTLLAGVPMVVVPGGG**DQ**WEIANRVVR	388			80	77	-	61
Rv2958c	RhaT		322-AAAARSAVVLCNGGSP**T**TQQALAAGVPVIGLPSNM**DQ**HLNMEALER	428			70	*23*	18	19
Rv2962c	RhaT		333-AAAARSAVVVCNGGSL**T**TQQALVAGVPVIGVAGNL**DQ**HLNMEAVER	449			72	*23*	17	*20*
Consensus			r aV hhGGagtta l AGvP p DQp a v r							
GT-4 (Retaining)										
Q44571	axAceA		275-EASFFGCLSAH**E**GFGLAAV**E**AMSAGLVPILS-NITPFARLMQQGAA	393		GDP-Man				
Rv0225			281-PSRK-------**E**GWGLAVI**E**AAQHGVPTIGYRSSGGLADSIVDGVT	384			83	74	51	63
Rv0486	MshA		348-PSYS-------**E**SFGLVAV**E**AQACGTPVVAA-AVGGLPVAVRDGIT	480		UDP-GlcNAc	78	69	44	60
Rv0557	PimB		277-SGEH-------**E**TFCQVVQ**E**ALASGLPVIAP-DAGGPRDLITPHRT	378		GDP-Man	pg	74	45	61
Rv1212c			284-PSVY-------**E**PLGIVNL**E**AMACATAVVAS-DVGGIPEVVADGIT	387			pg	77	59	61
Rv2188c			293-PCRTRGAGMDV**E**GLGIVFL**E**ASAAGVPVIAG-NSGGAPETVQHNKT	399			78	68	48	62
Rv2610c	PimA		268-PNTGG------**E**SFGIVLV**E**AMAAGTAVVAS-DLDAFRRVLRDGEV	378		GDP-Man	82	79	49	69
Rv3032			298-PSHY-------**E**PFGLVAL**E**AAAAGTPLVTS-NIGGLGEAVINGQT	414			83	69	24	69
Consensus			P E fg v EA p va gg v dg t							
GT-20 (Retaining)			Cβ4 Cα4 Cβ5							
1GZ5	ecOtsA		349-YSDVGLVTPLR**D**GM**NLV**AK**E**YVAAQDPANPGVLVLSQFAGAANELT	473		UDP-Glc				
Rv3490	OtsA		391-ASDVMLVTPLR**D**GMNLVAK**E**YVACRSDLGGALVLSEFTGAAAELRH	500		UDP-Glc	80	76	50	70
GT-28 (Inverting)			Cβ4 Cα4 Cβ5 Cα5							
1F0K	ecMurG		249-AAYAWADVVVC**R**SG**ALT**VS**E**IAAAGLPALFVPFQHKDR**QQ**YW**N**ALP	359		UDP-GlcNAc				
Rv2153c			294-LAYAAADLVIC**R**AGAMTVA**E**VSAVGLPAIYVPLPIGNGEQRLNALP	410			74	69	42	60
GT-35 (Retaining)			Cβ4 Cα4 Cβ5 Cα5							
1E4O	ecMalP		625-DISEQISTAGK**E**ASGTGNMKLALNGALTVGTLDGANVEIAEKVGEE	796		Glc-1-P				
Rv1328			599-DVWLNNPLRPL**E**ACGTSGMKSALNGGLNLSIRDGWWDEWYDGENGW	863			pg	79	47	71

Table II. CAZy classification of identified and proposed polyprenyl-dependent GTs of *M. tuberculosis* H_{37}Rv, all potential members of the GT-C superfamily. The two putative murein polymerases of family GT-51 are not shown

GT-C Protein [a]	[b]	[c]	GT-C Motif [j]	Size[e]	TMD[f]	Donor[g]	Ml[h]	Ms[h]	Cg[h]	Nf[h]
GT-39 (inverting)										
P33775	scPMT1	72-	SVVF**DE**VHFGGFASQYIRGTYFMDVH**PP**LAKMLYAGVASLGGFQGDFD	817	12	Dol-P-Man				
Rv1002		51-	TPIF**DE**KHYAPQAWQVLNNHGVEDNPGYGLVVH**PP**VGKQLIAIGEAIF	503	8		76	73	57	41
GT-53 (inverting)										
Rv3793	EmbC	289-	ANTS**DD**GYILTMARVSEHAGYMANYYRWFGTPEA**P**FGWYYDLLALWAH	1094	13		83	71	43	37
Rv3794	EmbA	269-	ATSS**DD**GYLLTVARVAPKAGYVANYYRYFGTTEA**P**FDWYTSVLAQLAA	1094	13		79	67	*40*	32
Rv3795	EmbB	295-	ANSS**DD**GYILGMARVADHAGYMSNYFRWFGSPED**P**FGWYYNLLALMTH	1098	13		78	67	*41*	34
GT-85 (inverting)										
Rv3792	AftA	124-	GISV**D**QQFRTEYLTRLTDTAALRDMTYIGL**PP**FY**PP**GWFWIGGRAAAL	643	13	C_{50}-P-Ara*f*	69	65	46	34
GT-86 (inverting)										
Rv1635c		81-	SLWF**DE**GATISASASRTLPELWSLLGHIDAVHGLYYLLMHGWFAIF**PP**	556	10		62	–	–	–
GT-87 (inverting)										
Rv1159	PimE	24-	GANFV**D**LHVYVSGAASLDHPGTLYGYVYADQTPDFPLPFTY**PP**FAAVV	431	10		73	64	50	33
Rv2181		41-	PYRI**D**I**D**IYQMGARAWLDGRPLYGGGVLFHTPIGLNLPFTY**PP**LAAVL	427	10		80	60	34	24
Rv2673		60-	GNIT**DD**FKPVYRAVLNFRRGWDIYNEHFDYVDPHYLY**PP**GGTLLMAPF	433	8		70	69	49	35
NC [i] (inverting?)										
Rv0051		130-	SSWI**E**T**D**SNGTPQLRYDGQIAVRYMEY**P**VLTGIYQYLSMAIAKTYTAL	560	8		74	69	36	56
Rv0236c		470-	GTSH**DE**PLQVLGDGPWGVRDSIPLT**PP**QTIRALDSVQRLFAAGRPSAG	1400	9		79	71	42	53
Rv0541c		117-	LTTR**DE**YLWQVPGIADIPATLRTFTSRILDFQPNSWVTHVSGH**PP**GAL	449	10		pg	68	–	–
Rv1459c		144-	IGR**D**GL**D**PYRVGPASGLGLGHVFTLSVPSLWRETPA**P**YG**P**LFLWIGRG	591	10		68	73	44	31
Rv1508c		77-	RLPG**D**RVGNHGALLPFRAEPRRIQMKR**PP**EVLRGAVTASRERLWAIGS	599	10		–	16	–	–
Rv2174		144-	LLR**D**GL**D**PYAVGPVGNPNALLDDVSPIWTITTA**P**YG**P**AFILVAKFVTV	516	12		71	68	46	32
Rv3779		82-	ARFR**D**L**D**AEALAVSRWPAVTVAAGVLLGALLIGWAAYRGI**P**HWQSIPS	666	14		56	–	–	23

[a]ORF from *M. tuberculosis* H_{37}Rv; Genebank accession number; or PDB ID number of crystal structure at the National Center for Biotechnology Information.
[b]Name of GT: bsSpsA, SpsA of *Bacillus subtilis*; aoGtfD, GtfD of *Amycolatopsis orientalis*; axAceA, AceA of *Acetobacter xylinum*; ecOtsA, OtsA of *E. coli*; ecMurG, MurG of *E. coli*; and ecMalP, MalP of *E. coli*; scPMT1, PMT1 of *Saccharomyces cerevisiae*.
[c]Number of amino acids preceding the first residue in alignment.
[d]Residue in bold are proposed as part of a NDP-sugar binding motif conserved within a GT family. Underlined residue has been shown to affect NDP-sugar binding or catalysis.
[e]Total number of residues in protein.
[f]TMD, number of transmembrane domains predicted by TMHMM 2.0.
[g]Characterized sugar donor.
[h]Percentage of identity with the strongest ortholog found by BLAST search in the genomes of Ml, *M. leprae* TN, Ms, *M. smegmatis* mc^2155, Cg, *Corynebacterium glutamicum* ATCC 13032, and Nf, *Nocardia farcinica* IFM 10152. Ortholog with higher identity to another H_{37}Rv-protein of the same GT-family is shown in italic. The identity cut-off was set to 15%; pg, pseudogene.
[i]NC, protein not classified by CAZy.
[j]Residue in bold is part of the suggested GT-C motif (section *Polyprenyl-dependent GTs*)

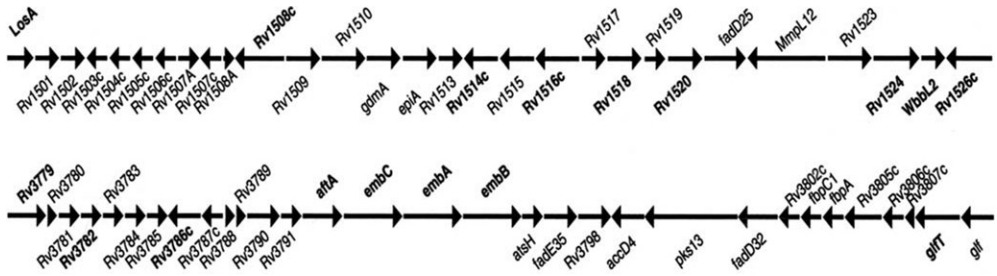

Fig. 7. Two major GT gene clusters are contained in the genome of *M. tuberculosis*: the Rv1500-cluster holds eight putative GTs classified as inverting enzymes of the GT-1 and the GT-2 families, and one proposed GT-C protein (Rv1508c). Only LosA (Rv1500) has been characterized and is implicated in LOS biosynthesis. The cell wall biosynthetic cluster (Rv3779-Rv3809c) contains many genes directly or indirectly involved in the synthesis of AG, mycolic acid, and LAM, and of which seven are identified or putative GTs (*Classification of putative GTs of* M. tuberculosis). Genes annotated or suggested as GTs are marked in bold and those having predicted TM domains are marked above its arrow.

conjunction with the two characterized GalTs, Rv3782 and GlfT (Rv3808c). The latter contains over 600 residues and is thereby significantly larger than other members of the GT-2 family, which commonly have 200–400 amino acids (Table I). However, its bifunctional role in galactan synthesis (Figure 3) (Mikušová et al. 2000) is a plausible reason for its extended size.

Unlike most of the GT-2 enzymes, WbbL is known to generate a product with the glycosidic bond in α-configuration. However, it is in fact an inverting enzyme since it utilizes dTDP-β-Rha in concert with an inverting mechanism. Rv2957 is another GT of this family that may function similarly. A recent study and the restricted distribution of Rv2957 to mycobacterial species producing fucosylated forms of PGLs and *p*-HBADs support the idea that this enzyme catalyzes the transfer of Fuc in the synthesis of these molecules (Perez, Constant, Lemassu, et al. 2004). The transferred sugar is in α-configuration (Figure 5A). Thus, with Rv2957 classified as an inverting enzyme, the most likely substrate is GDP-β-Fuc. One of the most conserved GTs in the GT-2 family is Ppm1, a polyprenyl-P-Man synthase (Gurcha et al. 2002). However, this protein probably has dual functionality in that it carries two domains. Its N-terminal domain contains seven predicted TM segments and shows high sequence and topology similarities to an apolipoprotein *N*-acyltransferase recently characterized in *E. coli* (Robichon et al. 2005), while its soluble C-terminal domain carries the ManT activity (Baulard et al. 2003). Interestingly, these two domains are expressed as two different proteins in other mycobacteria (Gurcha et al. 2002). Rv0539 has 25% amino acid identity to the soluble C-terminal domain of Ppm1 and may also function as a polyprenyl-P-Man synthase. If that is the case, these two genes may complement for each other since neither of them have been classified as essential (Sassetti et al. 2003).

As many as six ORFs of the large GT cluster, Rv1500-Rv1526c (Figure 7), are in accord with the properties of the GT-2 family. Interestingly, none of those $H_{37}Rv$ proteins are highly conserved among *Actinomycetales* spp., suggesting involvement in glycosylations of either components specific to *M. tuberculosis* or glycoconjugates with high structural variability among mycobacterial spp., such as the trehalose-based LOS components (Brennan 1988). A strong homolog to Rv1500, LosA of *M. marinum* (76% amino acid identity), was recently suggested as a GT involved in LOS biosynthesis (Burguiere et al. 2005) but the sugar donor used by LosA was not identified. Both Rv1500 and LosA diverge from most GTs of the GT-2 family by being predicted to have TM domains, located in the C-terminus of the protein. A very interesting comparison can be drawn to GtrB, a polyprenyl-P-Glc synthase expressed by *Shigella flexneri* bacteriophage SfX. The lipid-linked sugar synthesized by GtrB is utilized in *O*-antigen glucosylation leading to a serotype specific *S. flexneri* (Guan et al. 1999). GtrB shares 31% amino acid identity with Rv1500 and a topology study has shown that GtrB actually contains two TM domains in the C-terminus (Korres et al. 2005). Analogous functions are thereby likely for Rv1500 and LosA, and they may catalyze the formation of polyprenyl-P-hexose (from polyprenyl-P and NDP-hexose) then used in a coupled reaction in LOS biosynthesis.

GTs of M. tuberculosis *with a proposed GT-B fold*. The five families GT-1, GT-4, GT-20, GT-28, and GT-35 contain some 15 *M. tuberculosis* ORFs (Table I). Resolved crystal structures of at least one representative from each family have shown that all belong to the GT-B superfamily; GtfD from *Amycolatopsis orientalis* (aoGtfD; GT-1), OtsA from *E. coli* (ecOtsA; GT-20), MurG from *E. coli* (ecMurG; GT-28), and MalP from *E. coli* (ecMalP; GT-35) display the characteristic features of a GT-B structure (Watson et al. 1999; Ha et al. 2000; Gibson et al. 2002; Mulichak et al. 2004). In the absence of a published structure for GTs of the GT-4 family, a distant relationship between GT-4 and GT-28 [exemplified in Table I by AceA from *Acetobacter xylinum* (axAceA)] allowed for structural modeling of GT-4 enzymes into the GT-B fold (Abdian et al. 2000; Edman et al. 2003). This proposed classification of the GT-4 family has now been confirmed with the recent determination of the three-dimensional structure of PimA from *M. smegmatis* (msPimA; GT-4). MsPimA has all of the expected features of GTs from the GT-B superfamily being organized into two $\beta-\alpha-\beta$ Rossmann fold domains, with a deep fissure at the interface that includes the catalytic center (Guerin et al. Forthcoming).

A comparison between the NDP-sugar binding regions in these GT-B families shows that the structural elements are very similar in spite of the different catalytic mechanisms they employ. The two inverting enzymes, aoGtfD and ecMurG of the GT-1 and GT-28 families, respectively, contain a loop between β-strand Cβ4 and α-helix Cα4, which is particularly noteworthy (Table I). His^{309} of aoGtfD and Arg^{261} of ecMurG belong to this loop and may play a role in catalysis (Ha et al. 2001; Mulichak et al. 2004). Arg^{261} is part of the motif $[K/R]X_7E$ which is strictly

conserved among members of the GT-28 family. In comparison, the related motif [D/E]X_7E is conserved among retaining enzymes of the GT-4 and GT-20 families in an analogous structural element, as found in ecOtsA and predicted for axAceA (Table I) (Abdian et al. 2000; Gibson et al. 2002). The importance of this motif among GT-4 enzymes has been shown by site-directed mutagenesis. Exchange of the first glutamic acid in this motif in axAceA and msPimA (E287A and E274A, respectively), resulted in a complete loss of activity of these α-ManTs (Abdian et al. 2000; Guerin et al. Forthcoming). These results, and studies of others (Yep et al. 2004), suggest that an acidic residue in the first position of this motif is essential and may play a catalytic role in enzymes of the GT-4 and GT-20 families.

M. tuberculosis has five ORFs classified into the GT-1, a family characterized by enzymes using the inverting mechanism. As most of them use α-linked donors, it leads to sugar transfers in β-configuration. However, this family also includes many GTs with specificity for β-linked donors, such as dTDP-β-Rha and TDP-β-vancosamine, and consequently gives rise to α-linked products. The five proteins from *M. tuberculosis* within this GT-1 classification seem also to be using β-linked donors. Two of them, Rv2958c and Rv2962c, are functioning in the biosynthesis of PGLs and *p*-HBADs (Perez, Constant, Lemassu, et al. 2004) and are likely to use dTDP-β-Rha as a donor substrate (Figure 5). The remaining three (Rv1524, Rv1526c, and Rv2739) may also utilize β-linked sugar donors as they share 25–60% amino acid identity with the rhamnosyltransferase RtfA of *M. avium*. Since RtfA acts in the synthesis of GPLs (Eckstein et al. 1998), which so far have not been found in *M. tuberculosis*, other but similar functions should be designated for these RtfA homologs. In particular, Rv1524 and Rv1526c share significant over-all similarity (approximately 25% amino acid identity) and several motifs with aoGtfD, the latter responsible for vancomycin synthesis. The sequence stretch of aoGtfD shown in Table I contains the motif H^{308}HGSAGT, including the proposed catalytic residue His^{309} (Mulichak et al. 2004). This motif is conserved in Rv1524 and Rv1526c, suggesting a strong functional relationship. Further investigations may clarify the connection between aoGtfD and these putative GTs from *M. tuberculosis*.

The retaining GT-4 family is the largest family of GTs responsible for the formation of anomeric bindings in α-configurations. This family contains many bacterial GTs involved in synthesis of cell envelope structures, such as lipopolysaccharides and capsular polysaccharides (http://afmb.cnrs-mrs.fr/CAZY). *M. tuberculosis* contains seven representatives of this class, including the two ManTs PimA (Rv2610c) and PimB (Rv0557) involved in PIM biosynthesis (Schaeffer et al. 1999; Kordulakova et al. 2002). A third identified protein of the GT-4 type is MshA (Rv0486), a GlcNAc-inositol-phosphate synthase which catalyzes the first step in mycothiol biosynthesis, important for cellular detoxification of thiol-reactive agents (Newton and Fahey 2002; Newton et al. 2006). MshA is so far the only identified GT in *M. tuberculosis* not associated with cell wall biosynthesis.

It is likely that the remaining enzymes of the GT-4 type (Rv0225, Rv1212c, Rv2188c, Rv3032) are involved in biosynthesis of LAM, glycogen, α-glucan, and methylated (lipo)-polysaccharides of mycobacteria, since these molecules solely or essentially contain glycosidic bonds in α-configurations. A study on the ortholog of Rv1212c in *C. glutamicum* (Tzvetkov et al. 2003) and our preliminary data on Rv1212c and Rv3032 from *M. tuberculosis* actually suggest that these GTs act as α1,4-glucosyltransferases in the synthesis of glucan/glycogen and MGLPs, respectively (Jackson et al. unpublished data). All glycogen synthases from archaea, prokaryotic and eukaryotic origin classified to date in CAZy belong to either the GT-3 or the GT-5 family, both suggested to share the GT-B structural fold. Surprisingly, *M. tuberculosis* has no GT belonging to either of these two families. Instead the putative glycogen synthase, Rv1212c, is a GT-4 protein. This unexpected finding further emphasizes the structural similarities between these three families of GTs. Apart from *Rv0557* and *Rv1212c*, which lack orthologs in *M. leprae*, and *Rv3032* which lacks strong orthologs in corynebacteria, all other GT-4 proteins of *M. tuberculosis* H_{37}Rv are well conserved in the four genomes evaluated in this study (Table I).

Three other families of GTs, namely GT-20, GT-28, and GT-35, are represented by only one ORF each in *M. tuberculosis* (Table I). The enzymes that catalyze synthesis of trehalose-6-phosphate (OtsA) constitute their own family of GT-20 enzymes. Rv3490 belongs to this family and has already been characterized as a trehalose-phosphate synthase (Pan et al. 2002). The GT-28 family includes many glycolipid-synthesizing GTs such as MurG, an essential bacterial GT that catalyzes the synthesis of Lipid II. In *M. tuberculosis*, Rv2153c is a very strong candidate as MurG, since it shares the typical motifs, including RX_7E, and a sequence identity of 33% to ecMurG (Ha et al. 2000). Enzymes of the GT-35 family function as glycogen or starch phosphorylases and do not utilize NDP-sugar but instead degrade α1,4-linked glucans to Glc-1-phosphate. Rv1328 (GlgP) shares high homology to this family and has been proposed to be involved in metabolism of either glycogen or α-glucan, or both (Schneider et al. 2000). Homologs of MurG, OtsA, and GlgP are well conserved among the four species of *Actinomycetales* compared here, except for the latter (GlgP), which again lacks an ortholog in *M. leprae* (Table I). Incidentally, the two proposed glycogen/glucan biosynthesizing enzymes, Rv1212c and Rv1328, are pseudo genes in *M. leprae* in which no glycogen or glucan have so far been reported.

GTs of GT-B fold with membrane binding properties. Many of the GTs implicated in cell wall biosynthesis must be membrane-associated, particularly those with their substrates located in the cytoplasmic membrane. One such enzyme is MurG (GT-28) of *E. coli*, and its resolved crystal structure (Figure 6) suggested that the binding site of its acceptor substrate (Lipid I) is located in the N-terminal domain (Ha et al. 2000). This domain was shown to contain an amphipathic helix surrounded by basic residues that can create both hydrophobic and electrostatic interactions with the lipid bilayer, and thereby facilitate binding of the lipid acceptor. Interestingly, the surplus of basic amino acids in the N-terminal domain of MurG is reflected in the basic value of its calculated isoelectric point (pI), while the C-terminal domain has a neutral pI value (Edman et al. 2003). These features are also applicable to the MurG homolog of *M. tuberculosis* (Table III). Because enzymes of the GT-4 family share the structural characteristics of MurG, some of these enzymes may also share the

Table III. Theoretical pI calculated for selected GTs of superfamily GT-B

Protein	CAZy family	pI[a]	pI_N^b	pI_C^c
ecMurG	28	9.7	10.6	6.4
Rv2153c	28	10.0	10.8	7.8
Rv0225	4	10.1	10.9	6.3
Rv0486 (MshA)	4	7.8	6.0	9.8
Rv0557 (PimB)	4	10.6	11.1	9.9
Rv1212c	4	6.4	6.7	5.2
Rv2188c	4	9.8	10.9	6.0
Rv2610c (PimA)	4	7.8	9.5	5.3
Rv3032	4	7.4	6.0	9.9
Rv3490 (OtsA)	20	6.3	5.5	8.1

[a]pI calculated on the entire protein sequence.
[b]pI calculated on the N-terminal half of the protein sequence.
[c]pI calculated on the C-terminal half of the protein sequence.

membrane binding properties of MurG. Edman et al. used this proposed similarity to predict the surface charge distribution of lipid synthesizing enzymes of GT-4 (Edman et al. 2003). In *M. tuberculosis*, PimA (Rv2610c) and PimB (Rv0557) have been shown to carry ManT activities in the synthesis of glycolipids PIM_1 and PIM_2 (Schaeffer et al. 1999; Kordulakova et al. 2002), and these enzymes are therefore most likely associated to the cytoplasmic membrane. The basic pI of the N-terminal domains of PimA and PimB is consistent with this hypothesis (Table III). In contrast, MshA and OtsA are, due to their role in mycothiol and trehalose biosynthesis, proposed as soluble proteins without membrane association. That is in line with their low or neutral pI in the N-terminal domain. Likewise for Rv1212c and Rv3032 that are proposed to participate in the synthesis of glycogen/α-glucan and MGLPs. Rv0225 and Rv2188c, however, share the property of high pI in the N-terminal domain with PimA and PimB and can thus be proposed as membrane-associated GTs (Table III).

Polyprenyl-dependent GTs

General features. It has long been established that the classical protein glycosylation pathways, leading to the assemblies of asparagine-linked glycans (ALGs or *N*-glycan) and phosphatidylinositol glycan (PIG) anchors at the endoplasmatic reticulum (ER) of eukaryotes, utilize not only nucleotide-sugar donors but also polyprenyl-P-linked sugars (Lennarz 1975). In more recent years have the implicated dolichyl-dependent glucosyltransferases (ALG6, ALG8, and ALG10) and ManTs (ALG3, ALG9, ALG12, PIG-B, PIG-M, PIG-V, and SMP3) of these pathways been characterized (Burda and Aebi 1999; Kang, Hong, et al. 2005, and references therein). They constitute the basis for the CAZy classification of these type of enzymes. The CAZy has grouped them into six families, GT-22, GT-50, GT-57, GT-58, GT-59, and GT-76, which so far only contain members from the eukaryotic kingdom. All of these enzymes have been proposed to have a common evolutionary path (Oriol et al. 2002). Polyprenyl-dependent GTs associated with protein *O*-mannosylation, on the other hand, are organized into the GT-39 family. Both eukaryotic and prokaryotic proteins are found within this family, suggesting that their functions are evolutionary conserved. Besides these seven GT families with functions in protein glycosylation, CAZy has classified polyprenyl-dependent GTs into another four families (GT-53, GT-85, GT-86, and GT-87). Interestingly, these families consist of proteins restricted to mycobacteria and related species of the order *Actinomycetales*, consistent with the fact that they have recently been characterized as AraTs and ManTs involved in AG and LAM biosynthesis (see further discussion in sections *Polyprenyl-dependent GTs of* M. tuberculosis and *The Emb proteins and their relationship to other GTs*).

All 11 families of polyprenyl-dependent GTs consist of integral membrane proteins having 8–13 predicted TM domains. The sequence homology between them is in general very low, but conserved amino acid motifs have been found (Oriol et al. 2002). In common is a modified DxD motif (e.g., DxE, ExD, DDx, DEx, or EEx), typically located in the first or the second predicted extracytoplasmic loop. The position of this motif and a similar topology pattern among these polyprenyl-dependent GTs has suggested that they are structurally related, and therefore they have been organized into a superfamily named GT-C (Figure 6) (Liu and Mushegian 2003). The importance of the modified DxD motif has been investigated by site-directed mutagenesis for some of these GT-C proteins; substitution of an aspartic acid in this motif in the human PIG-M (ManT in PIG biosynthesis; Maeda et al. 2001), in PimE of *M. smegmatis* (ManT in PIM_5 biosynthesis; Morita et al. 2006), and in EmbC of *M. smegmatis* (AraT in LAM biosynthesis; Berg et al. 2005), resulted in all cases in loss or reduction of the enzyme activity. However, the exact function of this motif has not been elucidated. Nevertheless, a comparison can be made to NDP-sugar dependent GTs, many of which carry a DxD motif involved in binding of the donor substrate via a divalent cation (Unligil and Rini 2000). Thus, the acidic motifs of PIG-M, PimE and EmbC, and the corresponding motifs of other GT-C proteins may be part of a binding site for polyprenyl-P-sugar donors. Interestingly, about 20–40 amino acids downstream of the modified DxD motif is an aromatic residue commonly clustered together with one or several prolines (Table II). Furthermore, these residues are next to an additional partially conserved acidic residue (Liu and Mushegian 2003). These conserved residues, that are part of the same predicted loop in the GT-C proteins, will here be referred to as "the GT-C motif" and they may constitute elements important for binding of a lipid-linked sugar donor and/or for catalytic activity.

Polyprenyl-dependent GTs of M. tuberculosis. The recognition in the past of the lipid-linked sugar donors C_{50}-P-Ara*f*, C_{50}-P-Man, and C_{50}-P-Glc (section *Biosynthesis of sugar donors*) has revealed that mycobacteria and related species should be endowed with GT-activities dependent on these substrates. Early on, the three Emb proteins were suggested to have such activity (Belanger et al. 1996) and they are now classified by CAZy into their own family, GT-53 (section *The Emb proteins and their relationship to other GTs*). Later, sequencing of the complete genome of *M. tuberculosis* (Cole et al. 1998) led initially to the finding of only one ORF with similarity to known polyprenyl-dependent GTs, namely, Rv1002c (GT-39), the now characterized ManT involved in a Sec-dependent pathway for protein *O*-mannosylation (VanderVen et al.

2005). More sophisticated genome comparisons and advanced BLAST searches have then led to the identification of over 10 ORFs from *M. tuberculosis* H_{37}Rv as potential members of the GT-C superfamily (Table II) (Liu and Mushegian 2003; Kaur et al. 2006; Morita et al. 2006). Recent progress has allowed four of these candidates (Rv3792, Rv1635c, Rv1159, and Rv2181) to be characterized as such. The identification of AftA (Rv3792) as an AraT involved in AG biosynthesis is encouraging in particular due to its potential as a drug target in *M. tuberculosis* (Alderwick et al. 2006). The CAZy database has classified AftA and its homologs into an individual GT family, GT-85 (Table II). Another recently identified GT is Rv1635c of *M. tuberculosis* that catalyzes the addition of the first mannose residue in Man-capping of ManLAM (Dinadayala et al. 2006), most likely from the lipid-linked donor C_{50}-P-Man, conferring α1,5 linkages to the β-Ara*f* moieties at the nonreducing ends of LAM-arabinan (Figure 1) (Chatterjee et al. 1993). The availability of a *Rv1635c* mutant of *M. tuberculosis* provides a unique opportunity to measure the contribution of ManLAM to pathogenicity when carried by the whole tubercle bacillus, including its contribution to the modulation of the host immune response, and to the entry and intracellular survival of *M. tuberculosis* inside phagocytic cells (For a review, Briken et al. 2004; Kang, Azad, et al. 2005). As expected, BLAST searches in sequenced genomes of *Actinomycetales* spp. have identified orthologs to *Rv1635c* only in species known to have ManLAM (Table II; GT-86). The two newly characterized α1,2-ManTs, Rv1159 (PimE) and Rv2181, are responsible for biosynthesis of PIM_5 (Morita et al. 2006) and branching of the mannan backbone of LM/LAM (Kaur et al. 2006), respectively, and their suggested donor substrate is C_{50}-P-Man. The related functions of PimE and Rv2181 are reflected in similar amino acid sequences (25% identity). Interestingly, these two proteins also share sequence similarities with the as yet uncharacterized Rv2673, suggesting that this putative GT-C protein may carry an analogous function e.g., as a ManT involved in PIM_6 biosynthesis. The CAZy has grouped these three proteins of *M. tuberculosis* into the GT-87 family, a family that contains no members from species beyond *Actinomycetales*.

In addition to the now classified GTs with supposed dependency for lipid-linked sugar donors, Table II lists another seven ORFs with characteristics of the GT-C superfamily. Designated functions for these proteins are most probably found in the still incomplete pathways of AG, LAM, and LOS biosynthesis, and, not to be forgotten, in protein mannosylation. *M. tuberculosis* contains two fully characterized glycoproteins (Dobos et al. 1996; Michell et al. 2003), glycosylated with linear α1,2- and α1,3-oligomannosides. It is likely that the mannosylation of these proteins entirely takes place on the extracytoplasmic side of the plasma membrane, initiated by Rv1002c (VanderVen et al. 2005) with subsequent glycosylations being catalyzed by other GT-C proteins. One interesting candidate for such ManT activity is Rv0541c, based on the conservation of the amino acid patterns DEx and HPP of its GT-C motif with Rv1002c and other GT-39 proteins (Table II). Furthermore, *Rv0541c* is clustered on the chromosome of *M. tuberculosis* with *Rv0539*, a proposed polyprenyl-P-Man synthase gene homologous to *ppm1* (section *GTs of* M. tuberculosis *with a proposed GT-A fold*). Altogether, these observations suggest that Rv0541c may utilize the C_{50}-P-Man substrate generated by Rv0539 for protein mannosylation.

A topology study of the putative mycobacterial GT-C proteins was performed (TMHMM 2.0) and we found interesting structural patterns. The N-terminus of these proteins is most often predicted as cytosolic and the C-terminus is often a larger soluble domain. In general, three sections of clustered TM domains were observed, and these clusters are connected with two longer loops of which the GT-C motif is part of the first of these loops (Figure 6). This generalized picture of their topology is similar to what was predicted for GT-C proteins of eukaryotic origin (Oriol et al. 2002), indicating that they may have developed from a common ancestor of which only the GT-C motif has remained fairly conserved.

Some of the proposed GT-C proteins of *M. tuberculosis* might be falsely predicted as GTs. The significantly larger Rv0236c has a different membrane topology and its proposed GT-C motif shown in Table II is not located at the beginning of the protein. Divergent are also Rv1459c and Rv2174 in their modified DxD motif; two amino acids are separating the acidic residues in a DGLD motif (Table II). The latter motif has also been reported in the well-characterized polyprenyl-P:GlcNAc-P-transferase WecA of *E. coli*, in which the two aspartic acids of a DGID motif have been proposed to bind polyprenyl-P (Amer and Valvano 2002). The presence of this motif in Rv1459c and Rv2174 may indicate that these proteins bind a similar substrate. The partial characterization of the ortholog of *Rv1459c* in *C. matruchotii*, however, suggests a function in lipid transport rather than in enzymatic sugar-transfer (Wang et al. 2006). The high over-all sequence identity between Rv2174 and Rv1459c (27%) supports the idea that Rv2174 may have a function similar to that of Rv1459c. Future experimental studies may clarify these listed differences.

The Emb proteins and their relationship to other GTs. The Emb proteins have been subject of research for more than a decade and early on they were suggested to carry AraT activities (Belanger et al. 1996). However, recent development of a cell-free AraT assay (Khasnobis et al. 2006) and the observation of the conserved GT-C motif (section *General features*) (Berg et al. 2005) are the strongest evidence so far to classify the Emb proteins as AraTs. These are large membrane proteins composed of approximately 1100 amino acids. A topology study of mycobacterial Emb proteins proposed that the first approximately 700 residues hold 13 TM spanning segments (Berg et al. 2005), and that prediction follows the same topology pattern as in the model of other GT-C proteins presented here (Figure 6).

Sequencing of entire genomes of *Mycobacterium* and *Nocardia* spp. have revealed three paralogous *emb* genes commonly clustered together in an operon, whereas genomes of *Corynebacterium* species contain only one ortholog (Table II; GT-53). In *Mycobacterium*, the corresponding proteins, EmbA, EmbB, and EmbC, have been partially characterized by variable approaches. Initial experiments demonstrated that overexpression of *embA* and *embB* produced low levels of resistance to the drug EMB, a first-line drug in the treatment of TB, in an otherwise susceptible *M. smegmatis* host. In addition, these strains showed elevated AraT activities implying that the Emb proteins are AraTs (Belanger et al. 1996). Later on, results gained from knock-out strains Δ*embA* and Δ*embB* of *M. smegmatis* pointed in the same direction with both strains showing similar defects in the characteristic branched Ara_6 motif ([β-

Ara*f*-1,2-α-Ara*f*]$_2$-3,5-α-Ara*f*-1,5-α-Ara*f*) at the nonreducing ends of the arabinan structure of AG (Figure 1A); AG-arabinan of the mutants carried instead the linear Ara_4 motif (β-Ara*f*-1,2-α-Ara*f*-1,5-α-Ara*f*-1,5-α-Ara*f*) (Escuyer et al. 2001). A recently developed cell-free assay indicated that simultaneous expression of EmbA and EmbB is essential for formation of Ara_6 motifs in AG-arabinan (Khasnobis et al. 2006). A knock-out of *embC* of *M. smegmatis* led, on the other hand, to complete absence of LAM synthesis, whereas the structure of AG remained relatively unaffected (Escuyer et al. 2001; Zhang et al. 2003). These phenotypes can be compared with studies of arabinan biosynthesis in corynebacteria. The fact that *C. glutamicum* contains a LM-like lipoglycan (Gibson et al. 2003) but no LAM-like structure, and furthermore, no branched terminal arabinan AG motifs (Dover et al. 2004), may all together explain the single *emb* copy in *Corynebacterium* species. Inactivation of the *emb* ortholog of *C. glutamicum* resulted in a knock-out strain devoid of the arabinan domain of AG, carrying only singular arabinose residues linked to the galactan chain (Alderwick et al. 2005).

Despite the fact that the Emb proteins seem to have no overall sequence similarity to other protein families, advanced BLAST searches have resulted in the finding of two functional amino acid motifs in the Emb proteins, which previously have been described for proteins with known function. The first was the identification of the GT-C motif (Liu and Mushegian 2003), an important link towards categorization of the Emb proteins as AraTs. The result gained from site-directed mutagenesis in the GT-C motif of EmbC of *M. smegmatis* showed its essentiality; exchange of the first Asp^{279} residue in the conserved DDx motif (Table II) resulted in complete abrogation of LAM-arabinan biosynthesis (Berg et al. 2005), the same phenotype as was achieved in the *embC* knock-out strain.

The second conserved motif that was found in the Emb proteins, a proline-rich motif (not shown here; Berg et al. 2005), has previously been characterized in polysaccharide co-polymerases, a protein family that is involved in polymerization or chain length determination of polysaccharides in both Gram-negative and Gram-positive bacteria (Paulsen et al. 1997; Becker and Puhler 1998; Morona et al. 2000). For example, Wzz of *S. flexneri* is functioning as such in the Wzy-dependent pathway for *O*-antigenic polysaccharides (Daniels and Morona 1999). Site-directed mutagenesis of several residues in the proline-rich motif of EmbC of *M. smegmatis* affected the LAM-arabinan biosynthesis negatively, leading to smaller versions of LAM. In short, the interpretation of the results was that smaller arabinan units, containing the array of linkages found in native LAM, are precursors in a polymerization process, in which the proline-rich motif of EmbC is implicated (Berg et al. 2005). Thus, in contrast to other characterized GT-C proteins, the Emb proteins are most likely multifunctional proteins involved in both sugar transfer and in processes governing polymerization or chain length regulation.

Postulated pathways for biosynthesis of AG and LAM

The cytoplasmic membrane plays a crucial role in the biosynthesis of mycobacterial cell wall components. Many of the involved proteins are either embedded in the lipid bilayer as integral membrane proteins, or associated with the membrane surface by hydrophobic and/or electrostatic interactions. It is well established that the lipid composition of biomembranes is important for events such as protein translocation, membrane protein folding, and for protein binding and function (Karlsson et al. 1994; Simons and Ikonen 1997; Van Klompenburg et al. 1997; Bogdanov and Dowhan 1999). The phospholipids in the plasma membrane of mycobacteria are mainly derivatives of phosphatidic acid, namely PI, phosphatidylglycerol, cardiolipin, phosphatidylethanolamine, and PIMs (Brennan and Nikaido 1995; Haites et al. 2005). With the exception of PIMs which are restricted to *Actinomycetales*, this lipid composition is similar to that of other prokaryotes, suggesting that the general properties of lipid bilayers, such as a certain fraction of negatively charged lipids, required to fulfill biomembrane functions (Reviewed in Dowhan 1997), are valid also for mycobacteria.

About 20 of the nearly 50 herein proposed GTs of *M. tuberculosis* have been functionally characterized and, for the most part, found to be associated with cell wall biosynthesis. However, the precise subcellular location of the majority of these enzymes is still uncertain. Here, we present a topology model for the biosynthesis of the major mycobacterial heteropolysaccharides: the PG-AG cell wall core, the apolar and polar PIMs, and the lipoglycans LM and LAM (Figure 8). An important assumption made in this model is that GTs utilizing the soluble NDP-sugar donors are catalytically active on the cytosolic side of the plasma membrane, whereas GTs dependent on lipid-linked donors (i.e. polyprenyl-dependent GTs) are active mainly on the extracytoplasmic side.

Biosynthesis of the lipid-linked donors, C_{50}-P-Ara*f* and C_{50}-P-Man, are dependent on pRpp and GDP-Man, respectively (Figure 4B) (Baulard et al. 2003; Huang et al. 2005; Mikušová et al. 2005), and following our statement above, should occur on the cytoplasmic side of the plasma membrane. C_{50}-P-5-phosphoribose synthase (DPPRS; Rv3806c) utilizes pRpp and C_{50}-P to form C_{50}-P-5-phosphoribose in an early step towards biosynthesis of C_{50}-P-Ara*f*, and Ppm1 catalyzes the formation of C_{50}-P-Man (Figure 8). Since we hypothesized that these lipid-linked sugar donors are mainly utilized on the extracytoplasmic side, translocation across the plasma membrane would be required, processes suggested to be mediated by flippases (Bugg and Brandish 1994; Rush and Waechter 2005).

Both the biochemistry and the genetics of galactan biosynthesis in mycobacteria (Figure 3) share many features with the biosynthesis of a polymannan *O*-polysaccharide of *E. coli*. The latter is built in the cytosol as an undecaprenyl-P-P-GlcNAc-[Man_2]-[polymannan], by WecA and processive ManTs (Clarke et al. 2004). The nascent polymer is then exported across the plasma membrane where it is ligated to the Lipid A-core. This translocation is proposed to be dependent on an ABC transporter consisting of Wzm, a transmembrane (TM) protein with six TM domains, and Wzt, which carries the ABC (Bronner et al. 1994; Cuthbertson et al. 2005). Figure 8A depicts a similar model for the synthesis of the analogous polymer C_{50}-P-P-GlcNAc-[Rha-Gal-Gal]-[polygalactan] in mycobacteria, catalyzed in the cytosol by WecA, WbbL (Rv3265c), Rv3782, and the processive GalT Rv3808c. In a classification of ABC transporters of *M. tuberculosis*, Rv3781 and Rv3783 were proposed to be

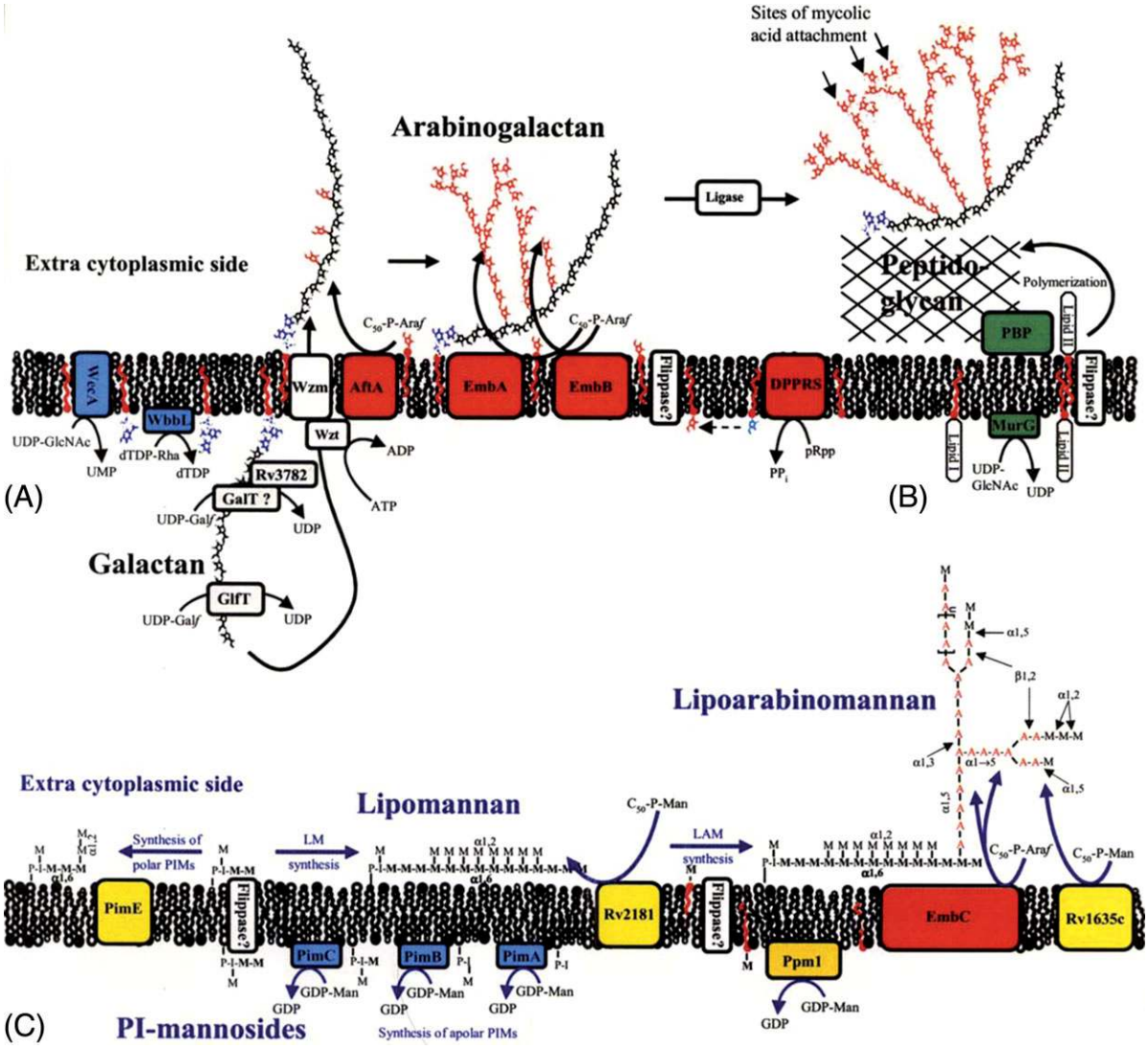

Fig. 8. Topology model of mycobacterial cell wall biosynthesis. Characterized and putative proteins implicated in the pathways for biosynthesis of arabinogalactan (**A**), PG (**B**), and PIMs and LAM (**C**), are shown. A compartmentalization is plausible with NDP-sugar and polyprenyl-P-linked sugar donors utilized by GTs on the cytosolic and extracytoplasmic side of the plasma membrane, respectively. See section *Postulated pathways for biosynthesis of arabinogalactan and lipoarabinomannan* for further details.

related to polysaccharide export (Braibant et al. 2000). These two ORFs are located in an operon together with Rv3782 and share significant sequence identity to Wzt and Wzm, respectively. Thus, Rv3781 and Rv3783 most likely form an ABC transporter responsible for the translocation of lipid-linked galactan chains across the mycobacterial plasma membrane. Like the *E. coli* genes required for polymannan *O*-polysaccharide biosynthesis and export (Guan et al. 2001), the majority of the mycobacterial genes implicated in the formation of the galactan core of AG are clustered on the chromosome, in the so-called mycobacterial cell-wall biosynthetic cluster (Figure 7).

The biosynthesis of the arabinan of AG is initiated by AftA, a priming enzyme that adds single Ara*f* residues to the galactan chain (Alderwick et al. 2006). In our model, it is possible that AftA is acting in concert with Rv3781/Rv3783 and transfers Ara*f* residues to the galactan polymer while it is entering the extracytoplasmic side of the membrane (Figure 8A). An analogous priming enzyme that would add single Ara*f* residues to the mannan backbone of LAM has not been characterized. However, the LM phenotype in the *embC* knock-out strain of *M. smegmatis* did contain two to three Ara*f* residues attached to the mannan backbone (Zhang et al. 2003), suggesting that such AraT activity should occur. An understanding of the subsequent arabinosylation steps of AG and LAM is still vague. Based on the proposed existence of a common Ara_{18} arabinan of both AG and LAM (as shown in Figure 1) (Shi et al. 2006), the overall sequence similarity of the Emb paralogs [39–43% identity in *M. tuberculosis* (ClustalW)], and the presence of conserved motifs within these three proteins, it is likely that similar strategies are used for assembling the arabinans of AG and LAM. On the other hand, different terminal motifs at the nonreducing ends of the arabinans of AG and LAM and the occurrence in these heteropolysaccharides of differently linked Ara*f* units (Figure 1) imply that several AraT activities are necessary for synthesis, but, of which, the Emb proteins

are probably the most significant. It has been proposed that arabinan biosynthesis may have similarities to the Wzy-dependent pathway of Gram-negative bacteria (Berg et al. 2005) (see section *The Emb proteins and their relationship to other GTs*), which involves an intermediate unit built on a polyprenyl-diphosphoryl-lipid followed by polymerization/ligation events. Berg et al. (2005) suggested that the GT-C motif of the Emb proteins carries a processive AraT activity, utilizing C_{50}-P-Ara*f* to synthesize a nonbranched α1,5-linked arabinan backbone core linked to a C_{50}-P anchor. The extension of such a lipid-linked oligoarabinoside chain may either occur at its reducing or its nonreducing end. The proline-rich motif of the Emb proteins [with homology to a functionally important segment in polysaccharide co-polymerases (section *The Emb proteins and their relationship to other GTs*)] may be involved in determining the chain length of this linear oligoarabinoside-intermediate and/or be implicated in polymerization of several oligosaccharides to form branched arabinan units subsequently transferred to an acceptor in the AG or LAM pathways. Since the arabinose of C_{50}-P-Ara*f* is linked in β-configuration, transfer of the majority of the Ara*f* residues in the assembly of arabinan demands an inverting mechanism, as has been suggested for the Emb proteins. However, formation of the terminal β1,2-linked Ara*f* would need to be catalyzed by a different AraT using the retaining mechanism.

The PG synthesis of *M. tuberculosis* (Figure 8B) has been assumed to resemble the pathway established for *E. coli*. MurG is the last enzyme in the intracellular phase of PG synthesis, leading to the formation of Lipid II (Reviewed in Bugg and Walsh 1992). This key intermediate is translocated by a flippase to the extracytoplasmic face of the plasma membrane, and then used in transglycosylation and transpeptidation events, performed by penicillin-binding proteins (PBP), to form PG. The ligation of AG to PG has been described (Yagi et al. 2003) but the enzyme(s) responsible for this oligosaccharyl-transfer has not yet been characterized.

The three first mannosylations of PI in biosynthesis of PIMs involve the GDP-Man dependent PimA, PimB, and PimC (Figure 8C) and they are suggested to occur on the cytoplasmic face of the plasma membrane (Morita et al. 2004). Characterization of PimE and other GT-C proteins now shed light over the subsequent steps in PIMs and LM/LAM biosynthesis. Following the hypothesis that polyprenyl-dependent GTs are functioning on the extracytoplasmic side of the plasma membrane, we suggest that PIM_3 or PIM_4 is translocated across the bilayer, possibly by a flippase (Figure 8C). As being an intermediate at the branch point at which polar PIMs and LM/LAM biosynthesis diverges (Morita et al. 2006), PIM_4 then acts as a substrate for PimE towards biosynthesis of the polar PIM_6 and as a precursor in the formation of LM (Besra et al. 1997; Morita et al. 2004; Haites et al. 2005). The ManT(s) responsible for synthesis of the mannan backbone of LM is still unknown, but it is plausible that such activity is C_{50}-P-Man dependent, as was suggested for Rv2181, the α1,2-ManT responsible for the LM branching (Kaur et al. 2006). In the context of this branch point is the identification of lipoprotein LpqW of *M. smegmatis*, a very interesting contribution (Kovacevic et al. 2006). LpqW, which structure has now been solved (Marland et al. 2006), seems to control the relative amount of polar PIMs and LM/LAM in the cell wall, and accordingly, may regulate the implicated ManTs as they are competing for PIM_4. Future results may clarify this exciting development.

Conclusions

The envelope of *Mycobacterium* spp. is a source of unique carbohydrates, many of which play important roles in the physiology and virulence of these bacteria. With this in mind, considerable effort has been placed during the last decade or so towards the elucidation of the pathways leading to their synthesis. Aided by recent advances in the genetic manipulation of mycobacteria, the availability of a growing number of bacterial genome sequences and a better understanding of the mechanistic and structural aspects of GTs, considerable progress has been made towards these goals, as evidenced by the some 20 *M. tuberculosis* GTs characterized to date. Two of the most exciting aspects of the elucidation of part of the biosynthetic pathways of AG and LM/LAM have certainly been the findings that pathways reminiscent of those involved in the biosynthesis and export of polysaccharides in Gram-negative bacteria operate in mycobacteria and the characterization of the Emb proteins as multifunctional enzymes. Much work, however, remains to be done to get a complete picture of the topologically complex assembly of the mannan backbone of LM/LAM and the arabinans of AG and LAM, including the paths for the addition of the terminal β-Ara*f* and the mannose residues in Man-capped LAM. Likewise, the precise contribution of the Emb proteins to the arabinosylation of AG and LAM remains to be defined and most of the GTs, flippases, and export systems required for the biosynthesis of glycogen, glucan, LOS, MGLPs, MMPs, acyltrehaloses, and glycoproteins remain to be identified.

In addition to providing fundamental knowledge about the biochemistry of *Mycobacterium* spp., the identification of the enzymes required for the synthesis of their carbohydrates are expected to provide novel attractive targets for chemotherapeutic agents. The availability of defined mutants of *M. tuberculosis* deficient in the biosynthesis of PIM/LM/LAM or other immunogenic glycoconjugates should also provide a unique opportunity to measure the contribution of individual carbohydrates to immunopathogenesis when carried by whole live tubercle bacilli, as is now the case with the *Rv1635c* mutant of *M. tuberculosis* which produces a cap-less LAM.

Acknowledgments

Many thanks go to Drs Delphi Chatterjee, Dean Crick, Premkumar Dinadayala, Sebabrata Mahapatra, Michael McNeil, and Vara Vissa, all at CSU, for valuable scientific discussions; Drs P. Alzari, A. Buschiazzo and M. Guerin, at Institut Pasteur, for sharing information about the crystal structure of msPimA; Drs M. Daffé, A. Lemassu, M. Rivière (IPBS-CNRS, Toulouse) for sharing information about Rv1212c and Rv3032; and Drs Åke Wieslander and Maria Rosén, both at Stockholm University, for their help with the GT structures in Figure 6. This work was supported by the National Institute of Allergy and Infectious Diseases/National Institutes of Health grant AI-18357.

Conflict of interest statement

None declared.

Abbreviations

ABC, ATP-binding cassette; AG, arabinogalactan; ALG, asparagine-linked glycan; AraT, arabinosyltransferase; BLAST, basic local alignment search tool; C_{50}-P, decaprenyl-phosphate; C_{50}-P-Ara*f*, β-D-arabinofuranosyl-monophosphoryl-decaprenol; C_{50}-P-Man, β-D-mannosyl-monophosphoryl-decaprenol; CAZy, Carbohydrate-Active enZyme; D-Araf, D-arabinofuranose; D-Gal*f*, D-galactofuranose; ER, endoplasmatic reticulum; GalT, galactosyltransferase; GPL, glycopeptidolipid; GT, glycosyltransferase; LAM, lipoarabinomannan; LM, lipomannan; LOS, lipooligosaccharide; mAGP, mycolate-arabinogalactan-peptidoglycan; ManT, mannosyltransferase; MGLP, 6-*O*-methyl glucose lipopolysaccharide; MMP, 3-*O*-methyl mannose polysaccharide; NDP, nucleotide-diphosphate; ORF, open reading frame; PG, peptidoglycan; PGL, phenolic glycolipid; *p*-HBAD, *p*-hydroxybenzoic acid derivatives; PI, phosphatidyl-*myo*-inositol; PIG, phosphatidyl-inositol glycan; PIM, phosphatidyl-inositol mannoside; pRpp, 5-phosphoribose-pyrophosphate; TB, tuberculosis; TDM, trehalose-dimycolate; TM, transmembrane; TMM, trehalose-monomycolate.

References

Abdian PL, Lellouch AC, Gautier C, Ielpi L, Geremia RA. 2000. Identification of essential amino acids in the bacterial α-mannosyltransferase AceA. J Biol Chem. 275:40568–40575.

Alderwick LJ, Radmacher E, Seidel M, Gande R, Hitchen PG, Morris HR, Dell A, Sahm H, Eggeling L, Besra GS. 2005. Deletion of Cg-*emb* in Corynebacterianeae leads to a novel truncated cell wall arabinogalactan, whereas inactivation of Cg-*ubiA* results in an arabinan-deficient mutant with a cell wall galactan core. J Biol Chem. 280:32362–32371.

Alderwick LJ, Seidel M, Sahm H, Besra GS, Eggeling L. 2006. Identification of a novel arabinofuranosyltransferase (AftA) involved in cell wall arabinan biosynthesis in *Mycobacterium tuberculosis*. J Biol Chem. 281: 15653–15661.

Amer AO, Valvano MA. 2002. Conserved aspartic acids are essential for the enzymic activity of the WecA protein initiating the biosynthesis of *O*-specific lipopolysaccharide and enterobacterial common antigen in *Escherichia coli*. Microbiology. 148:571–582.

Baulard AR, Gurcha SS, Engohang-Ndong J, Gouffi K, Locht C, et al. 2003. In vivo interaction between the polyprenol phosphate mannose synthase Ppm1 and the integral membrane protein Ppm2 from *Mycobacterium smegmatis* revealed by a bacterial two-hybrid system. J Biol Chem. 278: 2242–2248.

Becker A, Puhler A. 1998. Specific amino acid substitutions in the proline-rich motif of the *Rhizobium meliloti* ExoP protein result in enhanced production of low-molecular-weight succinoglycan at the expense of high-molecular-weight succinoglycan. J Bacteriol. 180:395–399.

Belanger AE, Besra GS, Ford ME, Mikušová K, Belisle JT, Brennan PJ, Inamine JM. 1996. The *embAB* genes of *Mycobacterium avium* encode an arabinosyl transferase involved in cell wall arabinan biosynthesis that is the target for the antimycobacterial drug ethambutol. Proc Natl Acad Sci USA. 93:11919–11924.

Belanger AE, Inamine JM. 2000. Genetics of cell wall biosynthesis. In Hatfull GF, Jacobs WR Jr, editors. Molecular genetics of mycobacteria. Washington, DC: ASM Press, p 191–202.

Belisle JT, Vissa VD, Sievert T, Takayama K, Brennan PJ, Besra GS. 1997. Role of the major antigen of *Mycobacterium tuberculosis* in cell wall biogenesis. Science. 276:1420–1422.

Berg S, Starbuck J, Torrelles JB, Vissa VD, Crick DC, Chatterjee D, Brennan PJ. 2005. Roles of conserved proline and glycosyltransferase motifs of EmbC in biosynthesis of lipoarabinomannan. J Biol Chem. 280: 5651–5663.

Besra GS, Khoo KH, McNeil MR, Dell A, Morris HR, Brennan PJ. 1995. A new interpretation of the structure of the mycolyl-arabinogalactan complex of *Mycobacterium tuberculosis* as revealed through characterization of oligoglycosylalditol fragments by fast-atom bombardment mass spectrometry and 1H nuclear magnetic resonance spectroscopy. Biochemistry. 34:4257–4266.

Besra GS, Morehouse CB, Rittner CM, Waechter CJ, Brennan PJ. 1997. Biosynthesis of mycobacterial lipoarabinomannan. J Biol Chem. 272: 18460–18466.

Besra GS, Sievert T, Lee RE, Slayden RA, Brennan PJ, Takayama K. 1994. Identification of the apparent carrier in mycolic acid synthesis. Proc Natl Acad Sci USA. 91:12735–12739.

Bhakta S, Basu J. 2002. Overexpression, purification and biochemical characterization of a class A high-molecular-mass penicillin-binding protein (PBP), PBP1* and its soluble derivative from *Mycobacterium tuberculosis*. Biochem J. 361:635–639.

Bloch K. 1977. Control mechanisms for fatty acid synthesis in *Mycobacterium smegmatis*. Adv Enzymol Relat Areas Mol Biol. 45:1–84.

Bogdanov M, Dowhan W. 1999. Lipid-assisted protein folding. J Biol Chem. 274:36827–36830.

Braibant M, Gilot P, Content J. 2000. The ATP binding cassette (ABC) transport systems of *Mycobacterium tuberculosis*. FEMS Microbiol Rev. 24:449–467.

Brennan PJ. 1988. Mycobacterium and other actinomycetes. In: Ratledge C, Wilkinson SG (Eds). Microbial Lipids. London: Academic Press Limited. p. 203–298.

Brennan PJ, Mayer H, Aspinall GO, Nam Shin JE. 1981. Structures of the glycopeptidolipid antigens from serovars in the *Mycobacterium avium*/*Mycobacterium intracellulare*/*Mycobacterium scrofulaceum serocomplex*. Eur J Biochem. 115:7–15.

Brennan PJ, Nikaido H. 1995. The envelope of mycobacteria. Annu Rev Biochem. 64:29–63.

Breton C, Snajdrova L, Jeanneau C, Koca J, Imberty A. 2006. Structures and mechanisms of glycosyltransferases. Glycobiology. 16:29R–37R.

Briken V, Porcelli SA, Besra GS, Kremer L. 2004. Mycobacterial lipoarabinomannan and related lipoglycans: from biogenesis to modulation of the immune response. Mol Microbiol. 53:391–403.

Bronner D, Clarke BR, Whitfield C. 1994. Identification of an ATP-binding cassette transport system required for translocation of lipopolysaccharide *O*-antigen side-chains across the cytoplasmic membrane of *Klebsiella pneumoniae* serotype O1. Mol Microbiol. 14:505–519.

Bugg TD, Brandish PE. 1994. From peptidoglycan to glycoproteins: common features of lipid-linked oligosaccharide biosynthesis. FEMS Microbiol Lett. 119:255–262.

Bugg TD, Walsh CT. 1992. Intracellular steps of bacterial cell wall peptidoglycan biosynthesis: enzymology, antibiotics, and antibiotic resistance. Nat Prod Rep. 9:199–215.

Burda P, Aebi M. 1999. The dolichol pathway of *N*-linked glycosylation. Biochim Biophys Acta. 1426:239–257.

Burguiere A, Hitchen PG, Dover LG, Kremer L, Ridell M, Alexander DC, Liu J, Morris HR, Minnikin DE, Dell A, Besra GS. 2005. LosA a key glycosyltransferase involved in the biosynthesis of a novel family of glycosylated acyltrehalose lipooligosaccharides from *Mycobacterium marinum*. J Biol Chem. 280:42124–42133.

Campbell JA, Davies GJ, Bulone V, Henrissat B. 1997. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem J. 326:929–939.

Charnock SJ, Davies GJ. 1999. Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. Biochemistry. 38:6380–6385.

Chatterjee D, Bozic CM, McNeil M, Brennan PJ. 1991. Structural features of the arabinan component of the lipoarabinomannan of *Mycobacterium tuberculosis*. J Biol Chem. 266:9652–9660.

Chatterjee D, Hunter SW, McNeil M, Brennan PJ. 1992. Lipoarabinomannan. Multiglycosylated form of the mycobacterial mannosylphosphatidylinositols. J Biol Chem. 267:6228–6233.

Chatterjee D, Khoo KH, McNeil MR, Dell A, Morris HR, Brennan PJ. 1993. Structural definition of the non-reducing termini of mannose-capped LAM from Mycobacterium tuberculosis through selective enzymatic

degradation and fast atom bombardment-mass spectrometry. Glycobiology. 3:497–506.

Clarke BR, Cuthbertson L, Whitfield C. 2004. Nonreducing terminal modifications determine the chain length of polymannose *O* antigens of *Escherichia coli* and couple chain termination to polymer export via an ATP-binding cassette transporter. J Biol Chem. 279:35709–35718.

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE III, et al. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. 393:537–544.

Coutinho PM, Deleury E, Davies GJ, Henrissat B. 2003. An evolving hierarchical family classification for glycosyltransferases. J Mol Biol. 328:307–317.

Crick DC, Mahapatra S, Brennan PJ. 2001. Biosynthesis of the arabinogalactan-peptidoglycan complex of *Mycobacterium tuberculosis*. Glycobiology. 11:107R–118R.

Cuthbertson L, Powers J, Whitfield C. 2005. The C-terminal domain of the nucleotide-binding domain protein Wzt determines substrate specificity in the ATP-binding cassette transporter for the lipopolysaccharide *O*-antigens in *Escherichia coli* serotypes O8 and O9a. J Biol Chem. 280: 30310–30319.

Daffé M, Brennan PJ, McNeil M. 1990. Predominant structural features of the cell wall arabinogalactan of *Mycobacterium tuberculosis* as revealed through characterization of oligoglycosyl alditol fragments by gas chromatography/mass spectrometry and by 1H and 13C NMR analyses. J Biol Chem. 265:6734–6743.

Daffé M, Draper P. 1998. The envelope layers of mycobacteria with reference to their pathogenicity. Adv Microb Physiol. 39:131–203.

Daffé M, Laneelle MA, Puzo G. 1983. Structural elucidation by field desorption and electron-impact mass spectrometry of the C-mycosides isolated from *Mycobacterium smegmatis*. Biochim Biophys Acta. 751: 439–443.

Daniels C, Morona R. 1999. Analysis of *Shigella flexneri* Wzz (Rol) function by mutagenesis and cross-linking: Wzz is able to oligomerize. Mol Microbiol. 34:181–194.

Davies GJ. 2001. Sweet secrets of synthesis. Nat Struct Biol. 8:98–100.

Davies GJ, Gloster TM, Henrissat B. 2005. Recent structural insights into the expanding world of carbohydrate-active enzymes. Curr Opin Struct Biol. 15:637–645.

De Smet KA, Weston A, Brown IN, Young DB, Robertson BD. 2000. Three pathways for trehalose biosynthesis in mycobacteria. Microbiology. 146: 199–208.

Deshayes C, Laval F, Montrozier H, Daffé M, Etienne G, Reyrat JM. 2005. A glycosyltransferase involved in biosynthesis of triglycosylated glycopeptidolipids in *Mycobacterium smegmatis*: impact on surface properties. J Bacteriol. 187:7283–7291.

Dinadayala P, Kaur D, Berg S, Amin AG, Vissa VD, Chatterjee D, Brennan PJ, Crick DC. 2006. Genetic basis for the synthesis of the immunomodulatory mannose caps of lipoarabinomannan in *Mycobacterium tuberculosis*. J Biol Chem. 281:20027–20035.

Dinadayala P, Lemassu A, Granovski P, Cerantola S, Winter N, Daffé M. 2004. Revisiting the structure of the anti-neoplastic glucans of *Mycobacterium bovis* Bacille Calmette-Guerin. Structural analysis of the extracellular and boiling water extract-derived glucans of the vaccine substrains. J Biol Chem. 279:12369–12378.

Dobos KM, Khoo KH, Swiderek KM, Brennan PJ, Belisle JT. 1996. Definition of the full extent of glycosylation of the 45-kilodalton glycoprotein of Mycobacterium tuberculosis. J Bacteriol. 178:2498–2506.

Dover LG, Cerdeno-Tarraga AM, Pallen MJ, Parkhill J, Besra GS. 2004. Comparative cell wall core biosynthesis in the mycolated pathogens, *Mycobacterium tuberculosis* and *Corynebacterium diphtheriae*. FEMS Microbiol Rev. 28:225–250.

Dowhan W. 1997. Molecular basis for membrane phospholipid diversity: Why are there so many lipids? Annu Rev Biochem. 66:199–232.

Eckstein TM, Silbaq FS, Chatterjee D, Kelly NJ, Brennan PJ, Belisle JT. 1998. Identification and recombinant expression of a *Mycobacterium avium* rhamnosyltransferase gene (rtfA) involved in glycopeptidolipid biosynthesis. J Bacteriol. 180:5567–5573.

Edman M, Berg S, Storm P, Wikström M, Vikström S, Öhman A, Wieslander Å. 2003. Structural features of glycosyltransferases synthesizing major bilayer and nonbilayer-prone membrane lipids in *Acholeplasma laidlawii* and *Streptococcus pneumoniae*. J Biol Chem. 278:8420–8428.

Escuyer VE, Lety MA, Torrelles JB, Khoo KH, Tang JB, Rithner CD, Frehel C, McNeil MR, Brennan PJ, Chatterjee D. 2001. The role of the *embA* and *embB* gene products in the biosynthesis of the terminal hexaarabinofuranosyl motif of *Mycobacterium smegmatis* arabinogalactan. J Biol Chem. 276: 48854–48862.

Geijtenbeek TB, van Vliet SJ, Koppel EA, Sanchez-Hernandez M, Vandenbroucke-Grauls CM, Appelmelk B, van Kooyk Y. 2003. Mycobacteria target DC-SIGN to suppress dendritic cell function. J Exp Med. 197:7–17.

Gibson KJ, Eggeling L, Maughan WN, Krumbach K, Gurcha SS, Nigou J, Puzo G, Sahm H, Besra GS. 2003. Disruption of Cg-Ppm1, a polyprenyl monophosphomannose synthase, and the generation of lipoglycan-less mutants in *Corynebacterium glutamicum*. J Biol Chem. 278:40842–40850.

Gibson RP, Turkenburg JP, Charnock SJ, Lloyd R, Davies GJ. 2002. Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA. Chem Biol. 9:1337–1346.

Guan S, Bastin DA, Verma NK. 1999. Functional analysis of the *O* antigen glucosylation gene cluster of *Shigella flexneri* bacteriophage SfX. Microbiology. 145:1263–1273.

Guan S, Clarke AJ, Whitfield C. 2001. Functional analysis of the galactosyltransferases required for biosynthesis of D-galactan I, a component of the lipopolysaccharide O1 antigen of *Klebsiella pneumoniae*. J Bacteriol. 183:3318–3327.

Gurcha SS, Baulard AR, Kremer L, Locht C, Moody DB, Muhlecker W, Costello CE, Crick DC, Brennan PJ, Besra GS. 2002. Ppm1, a novel polyprenol monophosphomannose synthase from *Mycobacterium tuberculosis*. Biochem J. 365:441–450.

Ha S, Gross B, Walker S. 2001. *E. coli* MurG: a paradigm for a superfamily of glycosyltransferases. Curr Drug Targets Infect Disord. 1:201–213.

Ha S, Walker D, Shi Y, Walker S. 2000. The 1.9 Å crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. Protein Sci. 9:1045–1052.

Haites RE, Morita YS, McConville MJ, Billman-Jacobe H. 2005. Function of phosphatidylinositol in mycobacteria. J Biol Chem. 280:10981–10987.

Hancock IC, Carman S, Besra GS, Brennan PJ, Waite E. 2002. Ligation of arabinogalactan to peptidoglycan in the cell wall of *Mycobacterium smegmatis* requires concomitant synthesis of the two wall polymers. Microbiology. 148:3059–3067.

Huang H, Scherman MS, D'Haeze W, Vereecke D, Holsters M, Crick DC, McNeil MR. 2005. Identification and active expression of the *Mycobacterium tuberculosis* gene encoding 5-phospho-α-D-ribose-1-diphosphate: decaprenyl-phosphate 5-phosphoribosyltransferase, the first enzyme committed to decaprenylphosphoryl-D-arabinose synthesis. J Biol Chem. 280:24539–24543.

Hunter SW, Brennan PJ. 1990. Evidence for the presence of a phosphatidylinositol anchor on the lipoarabinomannan and lipomannan of *Mycobacterium tuberculosis*. J Biol Chem. 265:9272–9279.

Hunter SW, Gaylord H, Brennan PJ. 1986. Structure and antigenicity of the phosphorylated lipopolysaccharide antigens from the leprosy and tubercle bacilli. J Biol Chem. 261:12345–12351.

Ikeda M, Wachi M, Jung HK, Ishino F, Matsuhashi M. 1991. The *Escherichia coli mraY* gene encoding UDP-*N*-acetylmuramoyl-pentapeptide: undecaprenyl-phosphate phospho-*N*-acetylmuramoyl-pentapeptide transferase. J Bacteriol. 173:1021–1026.

Jackson M, Raynaud C, Laneelle MA, Guilhot C, Laurent-Winter C, Ensergueix D, Gicquel B, Daffé M. 1999. Inactivation of the antigen 85C gene profoundly affects the mycolate content and alters the permeability of the *Mycobacterium tuberculosis* cell envelope. Mol Microbiol. 31:1573–1587.

Kang JY, Hong Y, Ashida H, Shishioh N, Murakami Y, Morita YS, Maeda Y, Kinoshita T. 2005. PIG-V involved in transferring the second mannose in glycosylphosphatidylinositol. J Biol Chem. 280:9489–9497.

Kang PB, Azad AK, Torrelles JB, Kaufman TM, Beharka A, Tibesar E, DesJardin LE, Schlesinger LS. 2005. The human macrophage mannose receptor directs *Mycobacterium tuberculosis* lipoarabinomannan-mediated phagosome biogenesis. J Exp Med. 202:987–999.

Karlsson OP, Dahlqvist A, Wieslander A. 1994. Activation of the membrane glucolipid synthesis in *Acholeplasma laidlawii* by phosphatidylglycerol and other anionic lipids. J Biol Chem. 269:23484–23490.

Kaur D, Berg S, Dinadayala P, Gicquel B, Chatterjee D, McNeil MR, Vissa VD, Crick DC, Jackson M, Brennan PJ. 2006. Biosynthesis of

mycobacterial lipoarabinomannan: role of a branching mannosyltransferase. Proc Natl Acad Sci USA. 103:13664–13669

Keller J, Ballou CE. 1968. The 6-*O*-methylglucose-containing lipopolysaccharide of *Mycobacterium phlei*. Identification of the lipid components. J Biol Chem. 243:2905–2910.

Khasnobis S, Zhang J, Angala SK, Amin AG, McNeil MR, Crick DC, Chatterjee D. 2006. Characterization of a specific arabinosyltransferase activity involved in mycobacterial arabinan biosynthesis. Chem Biol. 13: 787–795.

Khoo KH, Dell A, Morris HR, Brennan PJ, Chatterjee D. 1995. Structural definition of acylated phosphatidylinositol mannosides from *Mycobacterium tuberculosis*: definition of a common anchor for lipomannan and lipoarabinomannan. Glycobiology. 5:117–127.

Khoo KH, Douglas E, Azadi P, Inamine JM, Besra GS, Mikušová, K Brennan PJ, Chatterjee D. 1996. Truncated structural variants of lipoarabinomannan in ethambutol drug-resistant strains of *Mycobacterium smegmatis*. Inhibition of arabinan biosynthesis by ethambutol. J Biol Chem. 271: 28682–28690.

Koppel EA, Ludwig IS, Hernandez MS, Lowary TL, Gadikota RR, Tuzikov AB, Vandenbroucke-Grauls CM, van Kooyk Y, Appelmelk BJ, Geijtenbeek TB. 2004. Identification of the mycobacterial carbohydrate structure that binds the C-type lectins DC-SIGN, L-SIGN and SIGNR1. Immunobiology. 209:117–127.

Kordulakova J, Gilleron M, Mikušová K, Puzo G, Brennan PJ, Gicquel B, Jackson M. 2002. Definition of the first mannosylation step in phosphatidylinositol mannoside synthesis. PimA is essential for growth of mycobacteria. J Biol Chem. 277:31335–31344.

Kordulakova J, Gilleron M, Puzo G, Brennan PJ, Gicquel B, Mikušová K, Jackson M. 2003. Identification of the required acyltransferase step in the biosynthesis of the phosphatidylinositol mannosides of *Mycobacterium* species. J Biol Chem. 278:36285–36295.

Korres H, Mavris M, Morona R, Manning PA, Verma NK. 2005. Topological analysis of GtrA and GtrB proteins encoded by the serotype-converting cassette of *Shigella flexneri*. Biochem Biophys Res Commun. 328: 1252–1260.

Kovacevic S, Anderson D, Morita YS, Patterson J, Haites R, McMillan BN, Coppel R, McConville MJ, Billman-Jacobe H. 2006. Identification of a novel protein with a role in lipoarabinomannan biosynthesis in mycobacteria. J Biol Chem. 281:9011–9017.

Kremer L, Dover LG, Morehouse C, Hitchin P, Everett M, Morris HR, Dell A, Brennan PJ, McNeil MR, Flaherty C, et al. 2001. Galactan biosynthesis in *Mycobacterium tuberculosis*. Identification of a bifunctional UDP-galactofuranosyltransferase. J Biol Chem. 276:26430–26440.

Kremer L, Gurcha SS, Bifani P, Hitchen PG, Baulard A, Morris HR, Dell A, Brennan PJ, Besra GS. 2002. Characterization of a putative α-mannosyltransferase involved in phosphatidylinositol trimannoside biosynthesis in *Mycobacterium tuberculosis*. Biochem J. 363:437–447.

Lee YC. 1966. Isolation and characterization of lipopolysaccharides containing 6-*O*-methyl-D-glucose from *Mycobacterium species*. J Biol Chem. 241: 1899–1908.

Lee YC, Ballou CE. 1965. Complete structures of the glycophospholipids of mycobacteria. Biochemistry. 4:1395–1404.

Lemassu A, Daffé M. 1994. Structural features of the exocellular polysaccharides of *Mycobacterium tuberculosis*. Biochem J. 297:351–357.

Lennarz WJ. 1975. Lipid linked sugars in glycoprotein synthesis. Science. 188: 986–991.

Liu J, Mushegian A. 2003. Three monophyletic superfamilies account for the majority of the known glycosyltransferases. Protein Sci. 12:1418–1431.

Lornitzo FA, Goldman DS. 1968. Intracellular localization of a 6-*O*-methyl-D-glucose containing soluble polysaccharide from *Mycobacterium tuberculosis*. Biochim Biophys Acta. 158:329–335.

Ma Y, Mills JA, Belisle JT, Vissa V, Howell M, Bowlin K, Scherman MS, McNeil M. 1997. Determination of the pathway for rhamnose biosynthesis in mycobacteria: cloning, sequencing and expression of the *Mycobacterium tuberculosis* gene encoding α-D-glucose-1-phosphate thymidylyltransferase. Microbiology. 143:937–945.

Maeda N, Nigou J, Herrmann JL, Jackson M Amara A, Lagrange PH, Puzo G, Gicquel B, Neyrolles O. 2003. The cell surface receptor DC-SIGN discriminates between *Mycobacterium* species through selective recognition of the mannose caps on lipoarabinomannan. J Biol Chem. 278:5513–5516.

Maeda Y, Watanabe R, Harris CL, Hong Y, Ohishi K, Kinoshita K, Kinoshita T. 2001. PIG-M transfers the first mannose to glycosylphosphatidylinositol on the lumenal side of the ER. EMBO J. 20:250–261.

Mahapatra S, Yagi T, Belisle JT, Espinosa BJ, Hill PJ, McNeil MR, Brennan PJ, Crick DC. 2005. Mycobacterial lipid II is composed of a complex mixture of modified muramyl and peptide moieties linked to decaprenyl phosphate. J Bacteriol. 187:2747–2757.

Maitra SK, Ballou CE. 1976. Characterization of a mannan-like oligosaccharide from *Mycobacterium smegmatis*. Biochem Biophys Res Commun. 73: 1101–1108.

Marland Z, Beddoe T, Zaker-Tabrizi L, Lucet IS, Brammananth R, Whisstock JC, Wilce MC, Coppel RL, Crellin PK, Rossjohn J. 2006. Hijacking of a substrate-binding protein scaffold for use in mycobacterial cell wall biosynthesis. J Mol Biol. 359:983–997.

Maruta K, Mitsuzumi H, Nakada T, Kubota M, Chaen H, Fukuda S, Sugimoto T, Kurimoto M. 1996. Cloning and sequencing of a cluster of genes encoding novel enzymes of trehalose biosynthesis from thermophilic archaebacterium *Sulfolobus acidocaldarius*. Biochim Biophys Acta. 1291:177–181.

McCarthy TR, Torrelles JB, MacFarlane AS, Katawczik M, Kutzbach B, DesJardin LE, Clegg S, Goldberg JB, Schlesinger LS. 2005. Overexpression of *Mycobacterium tuberculosis manB*, a phosphomannomutase that increases phosphatidylinositol mannoside biosynthesis in *Mycobacterium smegmatis* and mycobacterial association with human macrophages. Mol Microbiol. 58:774–790.

McNeil M, Daffé M, Brennan PJ. 1990. Evidence for the nature of the link between the arabinogalactan and peptidoglycan of mycobacterial cell walls. J Biol Chem. 265:18200–18206.

McNeil M, Wallner SJ, Hunter SW, Brennan PJ. 1987. Demonstration that the galactosyl and arabinosyl residues in the cell-wall arabinogalactan of *Mycobacterium leprae* and *Mycobacterium tuberculosis* are furanoid. Carbohydr Res. 166:299–308.

McNeil MR, Robuck KG, Harter M, Brennan PJ. 1994. Enzymatic evidence for the presence of a critical terminal hexa-arabinoside in the cell walls of *Mycobacterium tuberculosis*. Glycobiology. 4:165–173.

Michell SL, Whelan AO, Wheeler PR, Panico M, Easton RL, Etienne AT, Haslam SM, Dell A, Morris HR, Reason AJ, et al. 2003. The MPB83 antigen from *Mycobacterium bovis* contains *O*-linked mannose and 1,3-mannobiose moieties. J Biol Chem. 278:16423–16432.

Mikušová K, Belanova M, Kordulakova J, Honda K, McNeil MR, Mahapatra S, Crick DC, Brennan PJ. 2006. Identification of a novel galactosyl transferase involved in biosynthesis of the mycobacterial cell wall. J Bacteriol. 188:6592–6598.

Mikušová K, Huang H, Yagi T, Holsters M, Vereecke D, D'Haeze W, Scherman MS, Brennan PJ, McNeil MR, Crick DC. 2005. Decaprenylphosphoryl arabinofuranose, the donor of the D-arabinofuranosyl residues of mycobacterial arabinan, is formed via a two-step epimerization of decaprenylphosphoryl ribose. J Bacteriol. 187:8020–8025.

Mikušová K, Slayden RA, Besra GS, Brennan PJ. 1995. Biogenesis of the mycobacterial cell wall and the site of action of ethambutol. Antimicrob Agents Chemother. 39:2484–2489.

Mikušová K, Yagi T, Stern R, McNeil MR, Besra GS, Crick DC, Brennan PJ. 2000. Biosynthesis of the galactan component of the mycobacterial cell wall. J Biol Chem. 275:33890–33897.

Mills JA, Motichka K, Jucker M, Wu HP, Uhlik BC, Stern RJ, Scherman MS, Vissa VD, Pan F, Kundu M, et al. 2004. Inactivation of the mycobacterial rhamnosyltransferase, which is needed for the formation of the arabinogalactan-peptidoglycan linker, leads to irreversible loss of viability. J Biol Chem. 279:43540–43546.

Morita YS, Patterson JH, Billman-Jacobe H, McConville MJ. 2004. Biosynthesis of mycobacterial phosphatidylinositol mannosides. Biochem J. 378:589–597.

Morita YS, Sena CB, Waller RF, Kurokawa K, Sernee MF, Nakatani F, Haites RE, Billman-Jacobe H, McConville MJ, Maeda Y, et al. 2006. PimE is a polyprenol-phosphate-mannose-dependent mannosyltransferase that transfers the fifth mannose of phosphatidylinositol mannoside in mycobacteria. J Biol Chem. 281:25143–25155.

Morona R, Van den BL, Daniels C. 2000. Evaluation of Wzz/MPA1/MPA2 proteins based on the presence of coiled-coil regions. Microbiology. 146: 1–4.

Mulichak AM, Losey HC, Walsh CT, Garavito RM. 2001. Structure of the UDP-glucosyltransferase GtfB that modifies the heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics. Structure. 9:547–557.

Mulichak AM, Lu W, Losey HC, Walsh CT, Garavito RM. 2004. Crystal structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic pathway: interactions with acceptor and nucleotide ligands. Biochemistry. 43:5170–5180.

Murphy HN, Stewart GR, Mischenko VV, Apt AS, Harris R, McAlister MS, Driscoll PC, Young DB, Robertson BD. 2005. The OtsAB pathway is essential for trehalose biosynthesis in *Mycobacterium tuberculosis*. J Biol Chem. 280:14524–14529.

Murzin AG, Brenner SE, Hubbard T, Chothia C. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol. 247:536–540.

Newton GL, Fahey RC. 2002. Mycothiol biochemistry. Arch Microbiol. 178: 388–394.

Newton GL, Ta P, Bzymek KP, Fahey RC. 2006. Biochemistry of the initial steps of mycothiol biosynthesis. J Biol Chem. 281:33910–33920.

Nigou J, Gilleron M, Puzo G. 2003. Lipoarabinomannans: from structure to biosynthesis. Biochimie. 85:153–166.

Nishimoto T, Nakano M, Nakada T, Chaen H, Fukuda S, Sugimoto T, Kurimoto M, Tsujisaka Y. 1996. Purification and properties of a novel enzyme, trehalose synthase, from *Pimelobacter* sp. R48. Biosci Biotechnol Biochem. 60:640–644.

Oriol R, Martinez-Duncker I, Chantret I, Mollicone R, Codogno P. 2002. Common origin and evolution of glycosyltransferases using Dol-P-monosaccharides as donor substrate. Mol Biol Evol. 19:1451–1463.

Ortalo-Magne A, Dupont MA, Lemassu A, Andersen AB, Gounon P, Daffé M. 1995. Molecular composition of the outermost capsular material of the tubercle bacillus. Microbiology. 141:1609–1620.

Pan YT, Carroll JD, Elbein AD. 2002. Trehalose-phosphate synthase of *Mycobacterium tuberculosis*. Cloning, expression and properties of the recombinant enzyme. Eur J Biochem. 269:6091–6100.

Paulsen IT, Beness AM, Saier MH Jr. 1997. Computer-based analyses of the protein constituents of transport systems catalysing export of complex carbohydrates in bacteria. Microbiology. 143:2685–2699.

Perez E, Constant P, Laval F, Lemassu A, Laneelle MA, Daffé M, Guilhot C. 2004. Molecular dissection of the role of two methyltransferases in the biosynthesis of phenolglycolipids and phthiocerol dimycoserosate in the *Mycobacterium tuberculosis* complex. J Biol Chem. 279: 42584–42592.

Perez E, Constant P, Lemassu A, Laval F, Daffé M, Guilhot C. 2004. Characterization of three glycosyltransferases involved in the biosynthesis of the phenolic glycolipid antigens from the *Mycobacterium tuberculosis* complex. J Biol Chem. 279:42574–42583.

Persson K, Ly H, Dieckelmann M, Wakarchuk W, Withers S, NC S. 2001. Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs. Nat Struct Biol. 8:166–175.

Ramakrishnan B, Boeggeman E, Qasba PK. 2002. β-1,4-galactosyltransferase and lactose synthase: molecular mechanical devices. Biochem Biophys Res Commun. 291:1113–1118.

Reed MB, Domenech P, Manca C, Su H, Barczak AK, Kreiswirth BN, Kaplan G, Barry CE III. 2004. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. Nature. 431:84–87.

Robichon C, Vidal-Ingigliardi D, Pugsley AP. 2005. Depletion of apolipoprotein *N*-acyltransferase causes mislocalization of outer membrane lipoproteins in *Escherichia coli*. J Biol Chem. 280:974–983.

Rose NL, Completo GC, Lin SJ, McNeil M, Palcic MM, Lowary TL. 2006. Expression, purification, and characterization of a galactofuranosyltransferase involved in *Mycobacterium tuberculosis* arabinogalactan biosynthesis. J Am Chem Soc. 128:6721–6729.

Rush JS, Waechter CJ. 2005. Assay for the transbilayer movement of polyisoprenoid-linked saccharides based on the transport of water-soluble analogues. Methods 35:316–322.

Sassetti CM, Boyd DH, Rubin EJ. 2003. Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol. 48:77–84.

Schaeffer ML, Khoo KH, Besra GS, Chatterjee D, Brennan PJ, Belisle JT, Inamine JM. 1999. The *pimB* gene of *Mycobacterium tuberculosis* encodes a mannosyltransferase involved in lipoarabinomannan biosynthesis. J Biol Chem. 274:31625–31631.

Scherman MS, Kalbe-Bournonville L, Bush D, Xin Y, Deng L, McNeil M. 1996. Polyprenylphosphate-pentoses in mycobacteria are synthesized from 5-phosphoribose pyrophosphate. J Biol Chem. 271:29652–29658.

Schlesinger LS, Hull SR, Kaufman TM. 1994. Binding of the terminal mannosyl units of lipoarabinomannan from a virulent strain of *Mycobacterium tuberculosis* to human macrophages. J Immunol. 152: 4070–4079.

Schneider D, Bruton CJ, Chater KF. 2000. Duplicated gene clusters suggest an interplay of glycogen and trehalose metabolism during sequential stages of aerial mycelium development in *Streptomyces coelicolor* A3(2). Mol Gen Genet. 263:543–553.

Schultz J, Elbein AD. 1974. Biosynthesis of mannosyl- and glucosyl-phosphoryl polyprenols in *Mycobacterium smegmatis*. Evidence for oligosaccharide-phosphoryl-polyprenols. Arch Biochem Biophys 160: 311–322.

Shi L, Berg S, Lee A, Spencer JS, Zhang J, Vissa V, McNeil MR, Khoo KH, Chatterjee D. 2006. The carboxy terminus of EmbC from *Mycobacterium smegmatis* mediates chain length extension of the arabinan in lipoarabinomannan. J Biol Chem. 281:19512–19526.

Simons K, Ikonen E. 1997. Functional rafts in cell membranes. Nature. 387: 569–572.

Singh S, Hogan SE. 1994. Isolation and characterization of sugar nucleotides from *Mycobacterium smegmatis*. Microbios. 77:217–222.

Sinnott ML. 1990. Catalytic mechanisms of enzymatic glycosyl transfer. Chem Rev. 90:1171–1202.

Takayama K, Kilburn JO. 1989. Inhibition of synthesis of arabinogalactan by ethambutol in *Mycobacterium smegmatis*. Antimicrob Agents Chemother. 33:1493–1499.

Takayama K, Wang C, Besra GS. 2005. Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis*. Clin Microbiol Rev. 18: 81–101.

Tarbouriech N, Charnock SJ, Davies GJ. 2001. Three-dimensional structures of the Mn and Mg dTDP complexes of the family GT-2 glycosyltransferase SpsA: a comparison with related NDP-sugar glycosyltransferases. J Mol Biol. 314:655–661.

Tsenova L, Ellison E, Harbacheuski R, Moreira AL, Kurepina N, Reed MB, Mathema B, Barry CE III, Kaplan G. 2005. Virulence of selected *Mycobacterium tuberculosis* clinical isolates in the rabbit model of meningitis is dependent on phenolic glycolipid produced by the bacilli. J Infect Dis. 192:98–106.

Tuffal G, Albigot R, Riviere M, Puzo G. 1998. Newly found 2-*N*-acetyl-2,6-dideoxy-β-glucopyranose containing methyl glucose polysaccharides in *M. bovis* BCG: revised structure of the mycobacterial methyl glucose lipopolysaccharides. Glycobiology. 8:675–684.

Tuffal G, Ponthus C, Picard C, Riviere M, Puzo G. 1995. Structural elucidation of novel methylglucose-containing polysaccharides from *Mycobacterium xenopi*. Eur J Biochem. 233:377–383.

Tzvetkov M, Klopprogge C, Zelder O, Liebl W. 2003. Genetic dissection of trehalose biosynthesis in *Corynebacterium glutamicum*: inactivation of trehalose production leads to impaired growth and an altered cell wall lipid composition. Microbiology. 149:1659–1673.

Unligil UM, Rini JM. 2000. Glycosyltransferase structure and mechanism. Curr Opin Struct Biol. 10:510–517.

van Heijenoort J. 2001. Formation of the glycan chains in the synthesis of bacterial peptidoglycan. Glycobiology. 11:25R–36R.

Van Klompenburg W, Nilsson I, von Heijne G, de Kruijff B. 1997. Anionic phospholipids are determinants of membrane protein topology. EMBO J. 16:4261–4266.

VanderVen BC, Harder JD, Crick DC, Belisle JT. 2005. Export-mediated assembly of mycobacterial glycoproteins parallels eukaryotic pathways. Science. 309:941–943.

Wang C, Hayes B, Vestling MM, Takayama K. 2006. Transposome mutagenesis of an integral membrane transporter in *Corynebacterium matruchotii*. Biochem Biophys Res Commun. 340:953–960.

Watson KA, McCleverty C, Geremia S, Cottaz S, Driguez H, Johnson LN. 1999. Phosphorylase recognition and phosphorolysis of its oligosaccharide substrate: answers to a long outstanding question. EMBO J. 18: 4619–4632.

Weisman LS, Ballou CE. 1984. Biosynthesis of the mycobacterial methylmannose polysaccharide. Identification of an α-1,4-mannosyltransferase. J Biol Chem. 259:3457–3463.

Weston A, Stern RJ, Lee RE, Nassau PM, Monsey D, Martin SL, Scherman MS, Besra GS, Duncan K, McNeil MR. 1997. Biosynthetic origin of mycobacterial cell wall galactofuranosyl residues. Tuber Lung Disease. 78: 123–131.

Wiker HG, Harboe M. 1992. The antigen 85 complex: a major secretion product of *Mycobacterium tuberculosis*. Microbiol Rev. 56:648–661.

Wimmerova M, Engelsen SB, Bettler E, Breton C, Imberty A. 2003. Combining fold recognition and exploratory data analysis for searching for glycosyltransferases in the genome of *Mycobacterium tuberculosis*. Biochimie. 85:691–700.

Wolucka BA, McNeil MR, de Hoffmann E, Chojnacki T, Brennan PJ. 1994. Recognition of the lipid intermediate for arabinogalactan/arabinomannan biosynthesis and its relation to the mode of action of ethambutol on mycobacteria. J Biol Chem. 269:23328–23335.

Yagi T, Mahapatra S, Mikušová K, Crick DC, Brennan PJ. 2003. Polymerization of mycobacterial arabinogalactan and ligation to peptidoglycan. J Biol Chem. 278:26497–26504.

Yep A, Ballicora MA, Sivak MN, Preiss J. 2004. Identification and characterization of a critical region in the glycogen synthase from *Escherichia coli*. J Biol Chem. 279:8359–8367.

Yokoyama K, Ballou CE. 1989. Synthesis of α-1,6-mannooligosaccharides in *Mycobacterium smegmatis*. Function of β-mannosylphosphoryldecaprenol as the mannosyl donor. J Biol Chem. 264:21621–21628.

Zhang N, Torrelles JB, McNeil MR, Escuyer VE, Khoo KH, Brennan PJ, Chatterjee D. 2003. The Emb proteins of mycobacteria direct arabinosylation of lipoarabinomannan and arabinogalactan via an N-terminal recognition region and a C-terminal synthetic region. Mol Microbiol. 50: 69–76.