A Glycosyltransferase Involved in Biosynthesis of Triglycosylated Glycopeptidolipids in *Mycobacterium smegmatis*: Impact on Surface Properties

Caroline Deshayes,¹ Françoise Laval,² Henri Montrozier,² Mamadou Daffé,²
Gilles Etienne,² and Jean-Marc Reyrat^{1*}

Inserm-UMR 570, Unité de Pathogénie des Infections Systémiques, Groupe Avenir,
Paris Cedex 15, F-75730, France¹; and IPBS-UMR 5089, Toulouse Cedex,
F-31062, France²

Received 24 June 2005/Accepted 28 July 2005

The cell envelope of mycobacteria is a complex structure that plays an important role in the interactions of the cell with its environment and in the protection against the antimicrobial activity of the immune system. Glycopeptidolipids (GPLs) are species- or type species-specific glycolipids that are present at the surface of a number of mycobacteria and that are characterized by a high variability in glycosylation patterns. These GPLs possess various biological activities that depend mostly on the sugars capping the core molecule. In *Mycobacterium smegmatis*, the GPL core can be substituted by either two or three deoxyhexoses. In this study, we show that Gtf3 is a glycosyltransferase responsible for the synthesis of the triglycosylated GPLs. Biochemical analysis of these molecules, with a combination of mass spectrometry and chemical degradation methods, has shown that they contain three deoxyhexose moieties. The presence of the triglycosylated GPLs is associated with cell surface modifications that lead to a decrease in sliding motility as well as a modification in cellular aggregation and colony appearance on Congo red. Phylogenetic analysis indicated that Gtf3 is a member of a yet-uncharacterized glycosyltransferase family conserved among the mycobacteria.

The mycobacterial envelope confers to mycobacteria a high impermeability to chemical disinfectants and to some antibiotics and contributes also, in the case of the pathogenic species, to the ability to survive in macrophages. This envelope is composed of a plasma membrane surrounded by a complex cell wall, which in turn is covered by a superficial layer, also called a capsule in the case of pathogenic species. The cell wall consists of a monolayer of mycoloyl residues covalently linked to the peptidoglycan-arabinogalactan complex and includes other lipids which are probably arranged to form a bilayer with the mycoloyl residues. The outermost structure, composed of proteins, carbohydrates, and (to a lesser extent) lipids, represents a privileged interface between bacilli and their environment. Both the outer lipid layer of the cell wall and the outermost capsule contain species-specific glycolipids or phospholipids (8, 12, 16). Glycopeptidolipids (GPLs) are the predominant glycolipids in members of the *Mycobacterium avium* complex, a group of subspecies involved in zoonotic infection and in the infection of immunocompromised patients. GPLs are also present at the surface of *M. smegmatis*, a saprophytic species (17). Purified GPLs are able to disturb macrophage membrane ultrastructure (42) and to insert into phospholipid monolayers (47) or to inhibit nonopsonic phagocytosis of mycobacteria by human macrophages (48), thus suggesting a potential role in the virulence of mycobacteria. It has been shown that GPLs can decrease the phosphorylation efficiency of isolated mitochondria without modifying the active respiration

(30). GPLs also play a role in sliding motility and in biofilm formation (33), probably through the interaction between the support and the bacterial surface.

GPLs are made of a tripeptide-amino alcohol core (D-Phe-D-*allo*-Thr-D-Ala-L-alaninol) linked to a fatty acyl residue. This lipopeptide core is invariably substituted with 6-deoxytalose (dTal) linked to the *allo*-Thr residue and an *O*-methylated rhamnosyl (Rha) unit linked to the terminal alaninol residue to generate the nonspecific apolar GPLs (Fig. 1A) (3). In *M. smegmatis*, dTal can be acetylated, while the fatty acid and the rhamnosyl residue can be modified with one and three methyl groups on position 2, 3, or 4, respectively (26, 44). In addition, polar GPLs have been recently identified for *M. smegmatis*: triglycosylated GPLs (Fig. 1A, GPL IIb and GPL IIIb) are synthesized in glucose-limited culture (37). The additional sugar was later identified as rhamnose on position 2 of the first rhamnosyl residue (36, 48). Polar GPLs with a succinyl residue acylating the terminal rhamnosyl unit (Fig. 1A, GPL IIIa and GPL IIIb) have also been characterized for *M. smegmatis* (48).

In a recent study, the polar GPLs of *M. smegmatis* were shown to be more active than the apolar forms in the inhibition of nonopsonic phagocytosis of mycobacteria by human macrophages (48), indicating the major role of sugars in the biological activities of GPLs.

The GPL biosynthetic gene cluster of *M. smegmatis* was characterized in 1999 through the identification of a morphological transpositional mutation in a gene encoding a nonribosomal peptide synthase (*mpe*) involved in the biosynthesis of the peptide core of GPLs (5). Five genes responsible for the addition of various substitutions on the lipopeptide core have since been identified: *atf1* encodes the acetyltransferase (44) and *fnt*, *rmt2*, *rmt3*, and *rmt4* encode the four methyltrans-

* Corresponding author. Mailing address: Inserm-UMR570, Faculté de Médecine Necker-Enfants Malades, 156 rue de Vaugirard, Paris Cedex 15, F-75730, France. Phone: 33 (0)1 40 61 53 79. Fax: 33 (0)1 40 61 56 77. E-mail: jmreyrat@necker.fr.

FIG. 1. (A) General structures of *M. smegmatis* mc²155 GPLs (adapted from references 40 and 48). Ac, acetyl. (B) Part of the *M. smegmatis* glycopeptidolipid biosynthetic gene cluster, corresponding to genes adding substitutions on the lipopeptide core of the GPLs. The three *gtf* open reading frames are represented by hatched arrows. Genes which have been functionally characterized are marked by asterisks (adapted from reference 25).

ferases (25, 26, 40) (Fig. 1B). The *rmlA* and *rmlB* genes may be involved in rhamnose biosynthesis, according to their sequence homology with *rmlA* of *M. tuberculosis* and *gepiA* of *M. avium*, respectively (13, 25, 31). Rv1174c is uncharacterized, and nothing is known about the three putative glycosyltransferases (*gtf1*, *gtf2*, and *gtf3* genes) present in the GPL locus. The only glycosyltransferase involved in GPL biosynthesis characterized so far for mycobacteria is RtfA of *M. avium* (4). Heterologous expression of the *rtfA* gene in *M. smegmatis* as well as allelic exchange in *M. avium* have proven that RtfA transfers a rhamnosyl residue onto dTal in the *M. avium* polar GPLs (20, 34).

In this study, we show by using a combination of mass spectrometry and chemical degradation methods that Gtf3 in *M. smegmatis* is responsible for the synthesis of the triglycosylated forms of GPLs, which contain three deoxyhexose moieties. A phylogenetic analysis shows that Gtf3 belongs to a family of glycosyltransferases which includes RtfA and which is conserved across the mycobacterial genus.

MATERIALS AND METHODS

Bacterial strains, medium, and growth. *Escherichia coli* DH5 α was used to propagate plasmids and was grown in LB medium. *M. smegmatis* strain mc²155 (46) was grown in LB or M9 medium (Difco). When required, antibiotics were added to the medium at the following concentrations: kanamycin, 25 μ g/ml, or hygromycin, 50 μ g/ml (200 μ g/ml for *E. coli*).

Construction of the glycosyltransferase expression plasmids. The wild-type *gtf1* gene (GenBank accession number AY138899) coding sequence was amplified by PCR with *Pfu* Turbo DNA polymerase (Stratagene), with *M. smegmatis* mc²155 genomic DNA as the template and primers *gtf1*_trans.5 and *gtf1*_trans.3 (Table 1). After purification with a QIAGEN PCR purification kit, PCR products were digested with XbaI and cloned into the dephosphorylated integrative expression vector pNIP40b at the unique XbaI site to generate pNIP*gtf1* (18). Enzymatic digestions were used to select clones having the *gtf1* gene inserted in the opposite orientation of the hygromycin resistance gene. One clone was selected and sequenced. A similar strategy was applied to clones *gtf2* (GenBank

accession number AY138899) (using the primers *gtf2*_trans.5 and *gtf2*_trans.3) and *gtf3* (GenBank accession number AY138899) (using the primers *gtf3*_trans.5 and *gtf3*_trans.3) in pNIP40b, yielding pNIP*gtf2* and pNIP*gtf3*. These plasmids were electroporated into *M. smegmatis* mc²155, and the transformants were selected on hygromycin and named mc²*gtf1*, mc²*gtf2*, and mc²*gtf3*.

RNA isolation and RT-PCR assay. Total RNA was extracted from 5 ml of log phase cultures of *M. smegmatis* mc²155 and *gtf*-overexpressing strains grown in M9 medium. Bacterial cells were pelleted by centrifugation for 15 min at 4,000 $\times$ g, resuspended in 350 μ l of RLT buffer (RNeasy mini kit; QIAGEN) containing 0.1% of β -mercaptoethanol, and transferred to lysing matrix B tubes (Polylabo) containing 0.1-mm silica beads. The mixtures were homogenized in a FastPrep FP120 instrument for 60 s at speed 6.5. After a short centrifugation, nucleic acids in the supernatants were precipitated with 0.7 volume of ethanol and the RNA were purified using an RNeasy mini spin column. Contaminating DNA was

TABLE 1. List of oligonucleotides used in this study

Primer	Sequence (5' to 3') ^a
<i>gtf1</i> _trans.5	TCAGTCTAGACTGTACGAGCGGTTGTC ACCCGGG
<i>gtf1</i> _trans.3	CTGATCTAGAGCCTCAGTGAGGCGCGC GCCGACG
<i>gtf2</i> _trans.5	TCGATCTAGAAAAGTCGACGCAGAAGCC GTGGCTT
<i>gtf2</i> _trans.3	CTGATCTAGATCTTCTGAAGTATCTCGC ACAATCT
<i>gtf3</i> _trans.5	TCAGTCTAGAGGAATTCTTCGAGGAAG AAGCGCT
<i>gtf3</i> _trans.3	CTGATCTAGAGACAAGAGTCGACGCCT TCCGTTG
<i>gtf1</i> _RT_reverse	GCGCGCCAGTACCGGCGGCAGGAA
<i>gtf1</i> _RT_direct	GTCCTGATGGCTGTGCCGCTGAC
<i>gtf2</i> _RT_reverse	TTCGGCGATCCTCCGCGGGGACGG
<i>gtf2</i> _RT_direct	TCTGCGCGACGATTTCTTCGCAA
<i>gtf3</i> _RT_reverse	GGCGCCATACCGTGCCCACTCGTC
<i>gtf3</i> _RT_direct	GCGTGACGATCTCTCAGTGGGAGG

^a Engineered XbaI sites are printed in bold.

FIG. 2. Results from RT-PCR assays. RT-PCR was done for each *gtf* gene, with cDNA of the *mc*²155 and *gtf*-overexpressing strains (+) as templates and respective primers listed in Table 1. Negative controls without reverse transcriptase were made for each PCR (–). MW, molecular weight ruler.

removed by digestion with DNase I according to the manufacturer's instructions (Roche Molecular Biochemicals). The DNase I enzyme was removed with two phenol-chloroform-isoamylalcohol extractions, followed by ethanol precipitation. Reverse transcriptase PCRs (RT-PCR) were carried out as described previously (10), and the PCR products were then resolved by horizontal electrophoresis on a 1.5% agarose gel.

Aggregation assay. Bacteria were grown to stationary phase. Unicellular mycobacteria were separated from aggregates by sedimentation (1 g) for 10 min. The optical density at 600 nm (OD_{600}) of the supernatant was measured and compared to the OD_{600} of resuspended cultures where the aggregates were broken up by vortexing with glass beads. The aggregative index was calculated as the ratio between the two OD_{600} (adapted from reference 22). The aggregation assay was performed in triplicate.

Congo red assay. Bacteria were grown to stationary phase, and 2 μ l was dropped on LB medium (Difco) supplemented with 1.5% agar and 100 μ g/ml Congo red (Sigma). The plates were scored for colony morphology and Congo red staining after 3 weeks at 37°C (11).

Sliding motility. Liquid culture (10 μ l) was dropped on plates containing 7H9 medium without any added carbon source (0.3% agar) and incubated at 37°C for a week.

Extraction and purification of mycobacterial lipids. Since it has been shown that polar GPLs are produced in nutrient starvation (36), strains were cultured in M9 medium (0.4% glucose). Lipids were extracted from cell pellets with a mixture of chloroform and methanol as previously described (48). The extracts were dried under vacuum and partitioned between water and chloroform (1:1, vol/vol). The organic phases were extensively washed with distilled water and evaporated to dryness. The lipid extracts were dissolved in chloroform and analyzed by thin-layer chromatography (TLC) on silica gel Durasil 25-precoated plates (0.25-mm thickness; Macherey-Nagel). The GPLs were resolved by TLC run in chloroform-methanol (9:1, vol/vol) and visualized by spraying the plates with 0.2% anthrone in concentrated sulfuric acid, followed by heating at 110°C. The purification of GPLs was completed by deacylating the lipids with 0.1 M KOH, according to the method of Brennan and Goren (7), and then extracting the alkali-resistant GPLs with chloroform.

Analytical procedures. Four different chemical degradation methods were applied to the GPLs (48): (i) de-*O*-acylated and both de-*O*-acylated and β -eliminated GPLs were obtained from treatment of native GPLs with 0.5 M sodium methanolate for 2 h at 37°C; (ii) perdeuteriomethylation of GPLs was carried out according to the method described by Blakeney and Stone (6), with trideuteriomethyl iodide (ICD₃) as the methylating agent; (iii) the *N*-acetyl-phenylalanyl moiety of GPLs was produced as methyl ester from native GPLs by methanolysis with anhydrous 1.5 M $CH_3OH-HCl$ for 16 h at 80°C; and (iv) the partially *O*-methylated alditol acetate derivatives were obtained from perdeuteriomethylated GPLs after hydrolysis with 2 M trifluoroacetic acid (100°C, 2 h), reduction with $NaBH_4$, and acetylation with 1:1 acetic anhydride-pyridine (100°C, 1 h). The native GPLs as well as the products resulting from the chemical degradations were analyzed by matrix-assisted laser desorption ionization-time-of-flight (MALDI-TOF) mass spectrometry. MALDI-TOF mass spectrometry detection in reflectron mode was performed on an Applied Biosystems 4700 Analyzer mass spectrometer (Applied Biosystems, Framingham, MA) equipped with an Nd: YAG laser (355-nm wavelength, <500-ps pulse, and 200-Hz repetition rate). Shots (2,500) were accumulated in positive ion mode, and mass spectrometric

data were acquired using the instrument default calibration. GPL samples were dissolved in chloroform, at a concentration of 1 mM, and were directly spotted onto the target plate as 0.5- μ l droplets, followed by the addition of 0.5 μ l of matrix solution. Samples were allowed to crystallize at room temperature. The matrix used was 2,5-dihydroxybenzoic acid (10 mg/ml) in $CHCl_3-CH_3OH$ (1:1, vol/vol).

Phylogenetic analysis. The NCBI website (<http://www.ncbi.nlm.nih.gov/BLAST/>) was searched with the BLASTP program, using *M. smegmatis* Gtf3 as a query and a cutoff limit of 50% identity. Sequences of *M. avium* were downloaded from the NCBI website, and those of *M. tuberculosis* and *M. leprae* genomes were downloaded from the Pasteur GenoList website (<http://genolist.pasteur.fr/>). Protein sequences were aligned using the MultAlin program (14), and a phylogenetic tree was inferred.

RESULTS

Gtf3 overproduction alters cell surface properties. The genetically manipulable *mc*²155 strain of *M. smegmatis* has recently been shown to produce low levels of polar GPLs; particularly, the triglycosylated GPLs were not detected in this strain (21). Thus, we reasoned that the *mc*²155 strain could be used to identify and characterize the enzyme involved in the synthesis of these forms of GPLs. Consequently, we overproduced the three glycosyltransferases (Gtf1, Gtf2, and Gtf3) whose genes are found in the GPL locus by cloning the corresponding open reading frames into a mycobacterial integrative expression vector. Each integrative plasmid was introduced into *mc*²155, and the transformants were named *mc*²*gtf1*, *mc*²*gtf2*, and *mc*²*gtf3*. Overexpression of the *gtf1*, *gtf2*, and *gtf3* genes was confirmed by semiquantitative RT-PCR. In the three overexpressing strains, the transcription level of each *gtf* gene was strongly increased compared to levels in the wild-type strain (Fig. 2). The modification of the cell surface properties of these strains was monitored by measuring aggregative properties of the bacterial cells. Strain *mc*²*gtf3* aggregated approximately four times more than did the wild-type, *mc*²*gtf1*, and *mc*²*gtf2* strains, indicating a modification of the cell surface in strain *mc*²*gtf3* (Fig. 3A).

The phenotypes of these strains were also investigated on Congo red agar plates. Congo red is a vital dye that binds to lipoproteins or lipids present on the mycobacterial surface (39) and is used to characterize modifications affecting the cell wall (11, 22, 29). While the wild-type strain and strains *mc*²*gtf1* and *mc*²*gtf2* were red smooth, strain *mc*²*gtf3* was red rough, again

FIG. 3. Macroscopic phenotypes of the *gtf*-expressing strains. (A) Aggregation properties. WT, wild type. (B) Phenotypes on Congo red plates. (C) Sliding abilities of the mc²155 and mc²gtf3 strains spreading on the surfaces of motility plates.

supporting a modification of the cell envelope of this strain (Fig. 3B).

In various bacterial species, cell surface molecules have been shown to play an active role in motility (24, 49). The four strains were therefore assayed for sliding motility. The sliding ability of strain mc²gtf3 was noticeably diminished compared to that of the wild-type strain (Fig. 3C), whereas strains mc²gtf1 and mc²gtf2 retained the ability to slide as the wild-type strain (data not shown).

In conclusion, the overexpression of the *gtf3* gene in *M. smegmatis* induces modifications in the phenotype on Congo red plate, as well as the ability to slide and to aggregate. This suggests that *gtf3* overexpression leads to cell surface modification. None of these traits was observed in the case of mc²gtf1 or mc²gtf2. However, it is possible that the overexpression of *gtf1* and *gtf2* leads to subtle cell wall modifications investigated in the next section that could not be detected using these macroscopic assays.

Gtf3 overproduction leads to the accumulation of triglycosylated GPLs. To check whether the overexpression of the three *gtf* genes alters the GPL profiles in these strains, the whole lipids were extracted and analyzed by TLC (Fig. 4A). As already described (21), strain mc²155 mainly produced apolar GPLs (GPL I). No difference between the TLC profiles of strains mc²155, mc²gtf1, and mc²gtf2 was observed. In contrast, the TLC profile of strain mc²gtf3 was shifted towards the polar

forms of GPLs. As a consequence, the apolar GPLs were almost absent, all presumably being converted into the polar forms. In control experiments, the TLC analysis of the whole lipids extracted from the various strains showed that the modifications of the GPL profiles were the only difference in the lipid contents of the wild type and the *gtf*-expressing strains (data not shown).

To further characterize the nature of the GPLs produced in strain mc²gtf3, MALDI-TOF mass spectrometry analysis was performed on the lipids produced by strain mc²gtf3 and by the wild-type strain mc²155 (Fig. 4B). The MALDI-TOF mass spectrum of the native mc²155 GPLs (Fig. 4B) showed a series of major pseudomolecular ion $[M + Na]^+$ peaks at m/z 1,257.7, corresponding to the already-described apolar GPLs of *M. smegmatis* mc²155 (21). In addition, a series of minor peaks corresponding to succinylated GPLs (48) was also detected at m/z 1,343 (Fig. 4B). On the other hand, the $[M + Na]^+$ peaks of the GPLs produced by mc²gtf3 were observed at m/z 1,403.7 and 1,417.7 for the major species and at m/z 1,503.7 and 1,517.7 for the minor ones (Fig. 4B). As described recently, the polar GPLs differ from the apolar GPLs by the addition of a rhamnosyl residue (triglycosylated GPLs) or a succinyl residue (succinylated GPLs) linked to the terminal rhamnosyl unit (35, 47) (Fig. 1A). Thus, the m/z values of the mc²gtf3 GPLs that were consistently 160 or 174 atomic mass units (amu) higher than those of mc²155 probably correspond to the addition of a deoxyhexose unit with various *O*-methyl substituents but not to the addition of a succinyl residue. The relative intensities of mass peaks were measured, and the results are summarized in Fig. 4C. Strain mc²155 synthesized a majority of GPL I and 20% of GPL IIIa, the diglycosylated form with a succinyl residue on the rhamnose, whereas the triglycosylated forms were only just above the detection threshold. According to the observed pseudomolecular mass, mc²gtf3 is likely to synthesize only GPL IIb and GPL IIIb, corresponding to the triglycosylated forms of GPLs with or without the succinyl residue, respectively (Fig. 4).

Characterization of the polar GPLs produced by strain mc²gtf3. The structures of the GPLs produced by strain mc²gtf3 were determined by the analysis of the chemical degradation products as reported by Villeneuve et al. (48). The alkali-resistant GPLs were purified by deacylating the lipids with 0.1 M KOH. The resulting de-*O*-acylated GPLs of mc²155 displayed m/z values of 84 (major peaks) or 184 (minor peaks) amu lower than those of the corresponding native GPLs (Table 2). These values are consistent with the presence of two acetyl groups substituting the deoxytalosyl moiety (major GPL I and minor GPL IIIa) plus the succinyl residue acylating the terminal rhamnosyl residue (minor GPL IIIa). The alkali treatment of the GPLs by Na methanolate led to β -elimination of the di-*O*-acetylated deoxytalosyl residue (9), resulting in a 248-amu decrease of the m/z values (Table 2). The m/z values of the β -eliminated GPLs purified from mc²gtf3 retained the 160- to 174-amu difference with the alkali-treated GPLs of mc²155 (Table 2), implying that the putative additional deoxyhexosyl unit was not substituting the 6-deoxytalosyl residue. Methanolysis of the GPLs eventually resulted in the release of the acyl-Phe or acyl-Phe-*allo*-Thr-Ala methyl ester moiety of the lipopeptidic core. Whatever the origin of the methanolysed GPLs, the mass spectra of the acyl methyl esters showed the

FIG. 4. Biochemical characterization of the *gtf*-expressing strains. (A) TLC analysis of GPLs. Lipids were extracted from stationary-phase cultures of *M. smegmatis* strains *mc*²155, *mc*²*gtf*1 (column 1), *mc*²*gtf*2 (column 2), and *mc*²*gtf*3 (column 3), and GPLs were chromatographed on silica gel and visualized with anthrone. Apolar and polar GPLs are indicated with open and filled boxes, respectively. (B) Mass spectrometry of the GPLs of strains *mc*²155 and *mc*²*gtf*3. (C) Quantification, in relative percentage, of mass spectrometry data of the major GPL forms produced by strains *mc*²155 and *mc*²*gtf*3. dTal, deoxytalose; Rha, rhamnose; Ac, acetyl; Suc, succinyl; -, not detected; nd, not detected; -, absent.

same pseudomolecular peaks (Table 2), implying that the major fatty acyl residue was C₂₈(OCH₃):0 (21) in both strains. These results demonstrated that the putative additional deoxyhexose was substituting the rhamnosyl residue linked to the alaninol end.

Acid hydrolysis of the perdeuteromethylated GPLs, followed by reduction, acetylation, gas chromatography, and gas chromatography-mass spectrometry (GC-MS) analysis of the sugar derivatives (48), identified in the *mc*²155 strain (Fig. 5A) two alditol acetate peaks by use of authentic standards. Analysis of the fragmentation patterns of the eluted compounds (data not shown) demonstrated that the first peak corresponded to a 1,5-di-*O*-acetyl-2,3,4-*O*-CD₃-6-deoxytalosyl unit (terminal 6-deoxytalose). The second peak contained per-*O*-

methylated rhamnosyl residues (terminal rhamnose) composed mainly of 1,5-di-*O*-acetyl-2,3,4-tri-*O*-CH₃-rhamnosyl (33%) and 1,5-di-*O*-acetyl-2-*O*-CD₃-3,4-di-*O*-CH₃-rhamnosyl (53%) residues. The two peaks with retention times similar to those of the *mc*²155 strain observed in the mixture of alditol acetates obtained from the *mc*²*gtf*3 perdeuteromethylated GPLs (Fig. 5B) contained the same terminal hexose derivatives. The additional peak observed in the GC-MS spectrum of the *mc*²*gtf*3 perdeuteromethylated GPLs (Fig. 5B) was composed of monosubstituted rhamnosyl units consisting of 1,2,5-tri-*O*-acetyl-4-*O*-CD₃-3-*O*-CH₃-rhamnosyl (76%) and 1,2,5-tri-*O*-acetyl-3,4-di-*O*-CH₃-rhamnosyl (10%) residues. MALDI-TOF mass spectrometry and GC-MS analyses confirmed our previous observations (17, 21, 48) and showed that the major

TABLE 2. MALDI-TOF analysis of the chemical degradation products of the GPL species of the mc²155 and mc²gtf3 strains of *M. smegmatis*

GPL treatment	Degradation product	[M + Na] ⁺ m/z			
		mc ² 155		mc ² gtf3	
		GPL I	GPL IIIa	GPL IIb	GPL IIIb
Untreated	Native GPLs	1,257.7	1,343.7	1,417.7	1,517.7
Alkali treatment (Na methanolate)	De- <i>O</i> -acylated GPLs	1,173.5	1,159.5	1,333.5	1,333.5
	β -Eliminated GPLs	1,009.5	995.5	1,169.5	1,169.5
Methanolysis (CH ₃ OH-HCl)	Acyl-PheOCH ₃	638.5	638.5	638.5	638.5
	Acyl-Phe- <i>allo</i> -Thr-AlaOCH ₃	810.5	810.5	810.5	810.5

GPLs of the strain mc²155 were the diglycosylated apolar GPL I (Fig. 1A), which contained mainly a 6-deoxytalosyl residue linked to *allo*-threonyl and a 2,3,4-tri-*O*-CH₃-rhamnosyl residue attached to the alaninol residue. Likewise, the major GPLs of strain mc²gtf3 were the triglycosylated polar GPL IIb (Fig. 1A), in which a diglycosyl unit composed of one 3-*O*-CH₃-rhamnosyl and one 3,4-di-*O*-CH₃-rhamnosyl residue modified the alaninol end of the lipopeptidic core. All together, these results indicate that Gtf3 is responsible for the synthesis of the triglycosylated forms of the GPLs.

Phylogenetic analysis of related mycobacterial glycosyltransferases. To test whether Gtf3 homologs were present in other species, a BLASTP analysis was carried out. The bioinformatic search allowed us to identify eight homologs with a value above the cutoff limit. The identities between homologs ranged from 52 to 62%. These homologs were present exclu-

sively in the mycobacterial genus. Indeed, two other putative glycosyltransferases in *M. smegmatis* (Gtf1 and Gtf2, with 59 and 62% identity, respectively) we examined in this study are also homologs, and three homologs are present in *M. avium* (GtfA, GtfB, and RtfA, with 61, 61, and 59% identity, respectively). Gtf3 has two homologs in *M. tuberculosis* (Rv1524 and Rv1526c, with 56 and 52% identity, respectively) and one in *M. leprae* (ML2348, with 56% identity). These homologs are characterized by a predicted glycosyltransferase domain of 150 amino acids that is located in the amino-terminal region of the protein. This domain is predicted to contain the sugar binding site (15).

The Gtf3 homologs having the strongest identity to Gtf1 and Gtf2 of *M. smegmatis*, and GtfA, GtfB, and RtfA of *M. avium* are all located in a locus dedicated to GPL biosynthesis. Interestingly, RtfA has also been shown to transfer a rhamnosyl residue but on a deoxytalose moiety. A phylogenetic tree was inferred

FIG. 5. GC-MS analysis (in the chemical ionization mode) of the alditol acetate derivatives obtained from perdeuteromethylated deacylated lipids from mc²155 (A) and mc²gtf3 (B): terminal 6-deoxytalose (Rt, 9.70 min), terminal rhamnose (Rt, 9.90 min), and monosubstituted rhamnose (Rt, 11.27 min). Rt, retention time.

FIG. 6. (A) Phylogenetic tree of Gtf3 homologs from the genomes of *M. smegmatis*, *M. avium*, *M. leprae*, and *M. tuberculosis*. (B) Radial tree depicting the various Gtf3 homologs and the glycosyltransferases of *M. tuberculosis* which have been characterized experimentally. PAM, percent amino acid mutation.

from amino acid sequence alignment of the various Gtf3 homologs and is shown in Fig. 6A. Four mycobacterial glycosyltransferases that were functionally characterized were added to this tree (Fig. 6B). Indeed, Rv2962c, Rv2957, and Rv2958c are involved in the glycosylation of phenolglycolipids of *M. tuberculosis* (41) and Rv3265c is required for the attachment of the rhamnosyl unit of the arabinogalactan to the peptidoglycan (35).

Although the predicted glycosyltransferases of *M. tuberculosis* (Rv1524 and Rv1526c) and *M. leprae* (ML2348) are located on branches distinct from those of *M. smegmatis* and *M. avium* (Fig. 6A), they are also evolutionarily distinct from Rv2962c, Rv2957, Rv2958c, and Rv3265c (Fig. 6B).

DISCUSSION

Glycopeptidolipids are small molecules of the mycobacterial cell surface that present great diversity in their glycosylation patterns (16). The glycosyl substitution is variable from species to species but can also be variable from isolate to isolate (19, 27). These surface molecules, which are highly antigenic, have been used to produce monoclonal antibodies that are used in serodiagnostic tests (23). In this study, we have shown that Gtf3 is responsible for the production of the triglycosylated forms of GPLs in *M. smegmatis*. Although the predicted amino acid sequence and the biochemical experiments strongly suggest that Gtf3 is a rhamnosyltransferase, the final demonstration will involve the development of an in vitro-reconstituted assay. Indeed, one may argue that Gtf3 could be only an accessory protein required for the activity of the glycosyltransferase. From a mechanistic point of view, Gtf3 could either add one rhamnosyl unit on position 2 of the rhamnosyl residue attached to the alaninol end of the lipopeptide core or both rhamnosyl units on the apolar GPL core. The enzymology of

the transfer of rhamnose units remains to be characterized. The approach of expressing the *gtf3* gene from a heterologous promoter has been successful because *M. smegmatis* mc²155 does not produce detectable quantities of polar GPLs (21). This finding is fortunate as ATCC 607, which is the ancestor of *M. smegmatis* mc²155 and does produce significant amounts of polar GPLs (21), is not genetically easily tractable (46). The difference in distribution between apolar and polar forms of these two isolates remains to be understood. The production of these particular forms of GPLs may be highly regulated and may be controlled under special circumstances that remain to be determined.

In *M. smegmatis*, triglycosylated GPLs have been reported to be produced in late stationary phase and under conditions of low carbon source (37). One may hypothesize that these polar GPLs possess biochemical properties that protect the cell from the environment encountered in these conditions. It is very likely that the distinct physicochemical properties of the polar GPLs alter the nature of the interaction with the environment, leading to the various phenotypes observed.

M. smegmatis and *M. avium* contain, respectively, two (Gtf1 and Gtf2) and three (GtfA, GtfB, and RtfA) putative glycosyltransferases that are highly similar to Gtf3. In both species, the predicted glycosyltransferases are located in a locus dedicated to GPL biosynthesis. This observation suggests that in *M. smegmatis*, Gtf1 and Gtf2 are responsible for adding the talose and the rhamnose moieties on the tripeptide-alcohol core. Consequently, the function of Gtf1 and Gtf2 will only be amenable to study by the construction of the knockout mutants.

Phylogenetic analyses have shown that the genome of *M. tuberculosis* contains two homologs of Gtf3. These two members of the Gtf3 family, Rv1524 and Rv1526c, are located in a

region that has not yet been functionally characterized. Interestingly, Rv1525, which is located between the two Gtf3 homologs, is annotated as a putative glycosyltransferase but is from a phylogenetically distinct origin. Neither of these two predicted glycosyltransferases has been characterized so far. Interestingly, the corresponding genomic region is deleted in *M. bovis* BCG, the attenuated strain that is used as a vaccine in humans (9). It can therefore be hypothesized that *M. tuberculosis* produces a glycosylated product that is absent from *M. bovis* BCG. Recently, a number of glycosyltransferases of *M. tuberculosis* have been characterized. This is the case for Rv2962c, Rv2957, and Rv2958c, which are involved in the glycosylation of the phenolglycolipid, a molecule related to the virulence of the tubercle bacilli (41). The Rv3265c gene (WbbL), which has been shown to attach the rhamnosyl unit of the arabinogalactan to the peptidoglycan, is essential for both *M. smegmatis* and *M. tuberculosis* (35, 45). Interestingly, Rv2962c, Rv2957, Rv2958c, and Rv3265c are all rhamnosyltransferases but they are evolutionarily distinct from Gtf3. It is striking that Rv2957, a rhamnosyltransferase involved in phenolglycolipid modification, is evolutionarily closer to WbbL, required for the attachment of the rhamnosyl unit of the arabinogalactan to the peptidoglycan, than are the two other glycosyltransferases involved in the phenolglycolipid glycosylation. One may speculate that the evolutionary grouping is due either to the substrate specificity or to the mechanism of transfer. According to the CAZy classification, both Rv2958c and Rv2962c as well as the Gtf3 orthologs are part of the GT1 family (15). This GT1 family is widely distributed from bacteria to humans and is characterized by a hexose-transferring activity on a variety of small molecules.

The nature of the glycosylation pattern greatly influences the biological activity of the modified molecules. This is true for viruses for which glycosylation of the surface alters the entry into the host cell (1, 32). Similarly, glycosylation of nodulation factors has been shown to change rhizobium-host specificity (28). Glycosyltransferases have very high donor and acceptor substrate specificities and are in general limited to the establishment of one glycosidic linkage (43). The consequence is that glycosyltransferases show a limited similarity between them, and it is practically impossible to infer from the amino acid sequence the nature of the sugar that will be transferred, as well as the nature of the molecule that will accept the sugar. However, genomic analyses of the galactosyl- and fucosyltransferase families have shown that the glycosyltransferases tend to group according to the nature of the sugar that is transferred and the nature of the saccharidic linkage (2, 38). A collaborative effort between genetics, biochemistry, and structural biology will clearly help to unravel the specificity of these sugar-transferring enzymes.

ACKNOWLEDGMENTS

We thank C. Jeanneau and M. Bertili for bacterial medium preparation. We thank Patricia Martin and Eric Stewart for critical reading of the manuscript and useful suggestions.

C.D. is funded by a doctoral grant of Inserm—Région Ile de France. We gratefully acknowledge Inserm for funding this project under the Avenir program to J.M.R., Chargé de Recherches at Inserm.

REFERENCES

- Abe, Y., E. Takashita, K. Sugawara, Y. Matsuzaki, Y. Muraki, and S. Hongo. 2004. Effect of the addition of oligosaccharides on the biological activities and antigenicity of influenza A/H3N2 virus hemagglutinin. *J. Virol.* **78**:9605–9611.
- Amado, M., R. Almeida, T. Schwientek, and H. Clausen. 1999. Identification and characterization of large galactosyltransferase gene families: galactosyltransferases for all functions. *Biochim. Biophys. Acta* **1473**:35–53.
- Aspinall, G. O., D. Chatterjee, and P. J. Brennan. 1995. The variable surface glycolipids of mycobacteria: structures, synthesis of epitopes, and biological properties. *Adv. Carbohydr. Chem. Biochem.* **51**:169–242.
- Belisle, J. T., K. Klaczekiewicz, P. J. Brennan, W. R. Jacobs, Jr., and J. M. Inamine. 1993. Rough morphological variants of *Mycobacterium avium*. Characterization of genomic deletions resulting in the loss of glycopeptidolipid expression. *J. Biol. Chem.* **268**:10517–10523.
- Billman-Jacobe, H., M. J. McConville, R. E. Haite, S. Kovacevic, and R. L. Coppel. 1999. Identification of a peptide synthetase involved in the biosynthesis of glycopeptidolipids of *Mycobacterium smegmatis*. *Mol. Microbiol.* **33**:1244–1253.
- Blakeney, A. B., and B. A. Stone. 1985. Methylation of carbohydrates with lithium methylsulphiny carbanion. *Carbohydr. Res.* **140**:319–324.
- Brennan, P. J., and M. B. Goren. 1979. Structural studies on the type-specific antigens and lipids of the *Mycobacterium avium* · *Mycobacterium intracellulare* · *Mycobacterium scrofulaceum* serocomplex. *Mycobacterium intracellulare* serotype 9. *J. Biol. Chem.* **254**:4205–4211.
- Brennan, P. J., and H. Nikaido. 1995. The envelope of mycobacteria. *Annu. Rev. Biochem.* **64**:29–63.
- Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeyer, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen, and S. T. Cole. 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci. USA* **99**:3684–3689.
- Camacho, L. R., P. Constant, C. Raynaud, M. A. Laneelle, J. A. Triccas, B. Gicquel, M. Daffe, and C. Guilhot. 2001. Analysis of the phthiocerol dimycocerosate locus of *Mycobacterium tuberculosis*. Evidence that this lipid is involved in the cell wall permeability barrier. *J. Biol. Chem.* **276**:19845–19854.
- Cangelosi, G. A., C. O. Palermo, J.-P. Laurent, A. M. Hamlin, and W. H. Brabant. 1999. Colony morphotypes on Congo red agar segregate along species and drug susceptibility lines in the *Mycobacterium avium-intracellulare* complex. *Microbiology* **145**:1317–1324.
- Chatterjee, D., and K. H. Khoo. 2001. The surface glycopeptidolipids of mycobacteria: structures and biological properties. *Cell. Mol. Life Sci.* **58**:2018–2042.
- Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeyer, S. Gas, C. E. Barry III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, and B. G. Barrell. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**:537–544.
- Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**:10881–10890.
- Coutinho, P. M., E. Deleury, G. J. Davies, and B. Henrissat. 2003. An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* **328**:307–317.
- Daffe, M., and A. Lemmassu. 2000. Glycobiology of the mycobacterial surface. Structures and biological activities of the cell envelope glycoconjugates. Kluwer Academic/Plenum Publishers, New York, N.Y.
- Daffe, M., M. A. Laneelle, and G. Puzo. 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.
- de Mendonça-Lima, L., M. Picardeau, C. Raynaud, J. Raubier, Y. O. de la Salmoniere, L. Barker, F. Bigi, A. Cataldi, B. Gicquel, and J. M. Reyat. 2001. Erp, an extracellular protein family specific to mycobacteria. *Microbiology* **147**:2315–2320.
- Eckstein, T. M., J. T. Belisle, and J. M. Inamine. 2003. Proposed pathway for the biosynthesis of serovar-specific glycopeptidolipids in *Mycobacterium avium* serovar 2. *Microbiology* **149**:2797–2807.
- Eckstein, T. M., F. S. Silbaq, D. Chatterjee, N. J. Kelly, P. J. Brennan, and J. T. Belisle. 1998. Identification and recombinant expression of a *Mycobacterium avium* rhamnosyltransferase gene (*rtfA*) involved in glycopeptidolipid biosynthesis. *J. Bacteriol.* **180**:5567–5573.
- Etienne, G., F. Laval, C. Villeneuve, P. Dinadayala, A. Abouwarda, D. Zerbib, A. Galamba, and M. Daffe. 2005. The cell envelope structure and properties of *Mycobacterium smegmatis* mc²155: is there a clue for the unique transformability of the strain? *Microbiology* **151**:2075–2086.
- Etienne, G., C. Villeneuve, H. Billman-Jacobe, C. Astarie-Dequeker, M. A. Dupont, and M. Daffe. 2002. The impact of the absence of glycopeptidolipids on the ultrastructure, cell surface and cell wall properties, and phagocytosis of *Mycobacterium smegmatis*. *Microbiology* **148**:3089–3100.
- Ikawa, H., S. Oka, H. Murakami, A. Hayashi, and I. Yano. 1989. Rapid identification of serotypes of *Mycobacterium avium-M. intracellulare* complex by using infected swine sera and reference antigenic glycolipids. *J. Clin. Microbiol.* **27**:2552–2558.
- Jackson, K. D., M. Starkey, S. Kremer, M. R. Parsek, and D. J. Wozniak.

2004. Identification of *psl*, a locus encoding a potential exopolysaccharide that is essential for *Pseudomonas aeruginosa* PAO1 biofilm formation. *J. Bacteriol.* **186**:4466–4475.
25. Jeevarajah, D., J. H. Patterson, M. J. McConville, and H. Billman-Jacobe. 2002. Modification of glycopeptidolipids by an *O*-methyltransferase of *Mycobacterium smegmatis*. *Microbiology* **148**:3079–3087.
26. Jeevarajah, D., J. H. Patterson, E. Taig, T. Sargeant, M. J. McConville, and H. Billman-Jacobe. 2004. Methylation of GPLs in *Mycobacterium smegmatis* and *Mycobacterium avium*. *J. Bacteriol.* **186**:6792–6799.
27. Krzywinska, E., and J. S. Schorey. 2003. Characterization of genetic differences between *Mycobacterium avium* subsp. *avium* strains of diverse virulence with a focus on the glycopeptidolipid biosynthesis cluster. *Vet. Microbiol.* **91**:249–264.
28. Laeremans, T., C. Snoeck, J. Marien, C. Verreth, E. Martinez-Romero, J. C. Prome, and J. Vanderleyden. 1999. Phaseolus vulgaris recognizes *Azorhizobium caulinodans* Nod factors with a variety of chemical substituents. *Mol. Plant-Microbe Interact.* **12**:820–824.
29. Laurent, J.-P., K. Hauge, K. Burnside, and G. Cangelosi. 2003. Mutational analysis of cell wall biosynthesis in *Mycobacterium avium*. *J. Bacteriol.* **185**:5003–5006.
30. Lopez-Marin, L. M., D. Quesada, F. Lakhdar-Ghazal, J. F. Toccanne, and G. Laneelle. 1994. Interactions of mycobacterial glycopeptidolipids with membranes: influence of carbohydrate on induced alterations. *Biochemistry* **33**:7056–7061.
31. Ma, Y., J. A. Mills, J. T. Belisle, V. Vissa, M. Howell, K. Bowlin, M. S. Scherman, and M. McNeil. 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 thymidyltransferase. *Microbiology* **143**:937–945.
32. Mardberg, K., K. Nystrom, M. A. Tarp, E. Trybala, H. Clausen, T. Bergstrom, and S. Olofsson. 2004. Basic amino acids as modulators of an O-linked glycosylation signal of the herpes simplex virus type 1 glycoprotein gC: functional roles in viral infectivity. *Glycobiology* **14**:571–581.
33. Martinez, A., S. Torello, and R. Kolter. 1999. Sliding motility in mycobacteria. *J. Bacteriol.* **181**:7331–7338.
34. Maslow, J. N., V. R. Irani, S. H. Lee, T. M. Eckstein, J. M. Inamine, and J. T. Belisle. 2003. Biosynthetic specificity of the rhamnosyltransferase gene of *Mycobacterium avium* serovar 2 as determined by allelic exchange mutagenesis. *Microbiology* **149**:3193–3202.
35. Mills, J. A., K. Motichka, M. Jucker, H. P. Wu, B. C. Uhlik, R. J. Stern, M. S. Scherman, V. D. Vissa, F. Pan, M. Kundu, Y. F. Ma, and M. McNeil. 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.
36. Mukherjee, R., M. Gomez, N. Jayaraman, I. Smith, and D. Chatterji. 2005. Hyperglycosylation of glycopeptidolipid of *Mycobacterium smegmatis* under nutrient starvation: structural studies. *Microbiology* **151**:2385–2392.
37. Ojha, A. K., S. Varma, and D. Chatterji. 2002. Synthesis of an unusual polar glycopeptidolipid in glucose-limited culture of *Mycobacterium smegmatis*. *Microbiology* **148**:3039–3048.
38. Oriol, R., R. Mollicone, A. Cailleau, L. Balanzino, and C. Breton. 1999. Divergent evolution of fucosyltransferase genes from vertebrates, invertebrates, and bacteria. *Glycobiology* **9**:323–334.
39. Ortalo-Magne, A., A. Lemassu, M. A. Laneelle, F. Bardou, G. Silve, P. Gounon, G. Marchal, and M. Daffe. 1996. Identification of the surface-exposed lipids on the cell envelopes of *Mycobacterium tuberculosis* and other mycobacterial species. *J. Bacteriol.* **178**:456–461.
40. Patterson, J. H., M. J. McConville, R. E. Haite, R. L. Coppel, and H. Billman-Jacobe. 2000. Identification of a methyltransferase from *Mycobacterium smegmatis* involved in glycopeptidolipid synthesis. *J. Biol. Chem.* **275**:24900–24906.
41. Perez, E., P. Constant, A. Lemassu, F. Laval, M. Daffe, and C. Guilhot. 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.
42. Pourshafie, M., Q. Ayub, and W. W. Barrow. 1993. Comparative effects of *Mycobacterium avium* glycopeptidolipid and lipopeptide fragment on the function and ultrastructure of mononuclear cells. *Clin. Exp. Immunol.* **93**:72–79.
43. Qasba, P. K., B. Ramakrishnan, and E. Boeggeman. 2005. Substrate-induced conformational changes in glycosyltransferases. *Trends Biochem. Sci.* **30**:53–62.
44. Recht, J., and R. Kolter. 2001. Glycopeptidolipid acetylation affects sliding motility and biofilm formation in *Mycobacterium smegmatis*. *J. Bacteriol.* **183**:5718–5724.
45. Sassetti, C. M., D. H. Boyd, and E. J. Rubin. 2003. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.* **48**:77–84.
46. Snapper, S. B., R. E. Melton, S. Mustafa, T. Kieser, and W. R. Jacobs, Jr. 1990. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol. Microbiol.* **4**:1911–1919.
47. Vergne, I., M. Prats, J. F. Toccanne, and G. Laneelle. 1995. Mycobacterial glycopeptidolipid interactions with membranes: a monolayer study. *FEBS Lett.* **375**:254–258.
48. Villeneuve, C., G. Etienne, V. Abadie, H. Montrozier, C. Bordier, F. Laval, M. Daffe, I. Maridonneau-Parini, and C. Astarie-Dequeker. 2003. Surface-exposed glycopeptidolipids of *Mycobacterium smegmatis* specifically inhibit the phagocytosis of mycobacteria by human macrophages. Identification of a novel family of glycopeptidolipids. *J. Biol. Chem.* **278**:51291–51300.
49. Wang, X., J. F. Preston III, and T. Romeo. 2004. The *pgaABCD* locus of *Escherichia coli* promotes the synthesis of a polysaccharide adhesin required for biofilm formation. *J. Bacteriol.* **186**:2724–2734.