THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 285, NO. 53, pp. 41348–41355, December 31, 2010

Biosynthetic Origin of the Galactosamine Substituent of Arabinogalactan in *Mycobacterium tuberculosis*[*] [S]

Received for publication, September 23, 2010, and in revised form, October 26, 2010 Published, JBC Papers in Press, October 28, 2010, DOI 10.1074/jbc.M110.188110

Henrieta Škovierová[‡1,2], Gérald Larrouy-Maumus[§¶1], Ha Pham[‡], Martina Belanová[‖], Nathalie Barilone[**], Arunava DasGupta[3], Katarina Mikušová[‖], Brigitte Gicquel[**], Martine Gilleron[§¶], Patrick. J. Brennan[‡], Germain Puzo[§¶], Jérôme Nigou[§¶4,5], and Mary Jackson[‡4,6]**

From the [‡]Mycobacteria Research Laboratories, Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, Colorado 80523-1682, the [§]Département Mécanismes Moléculaires des Infections Mycobactériennes, Institut de Pharmacologie et de Biologie Structurale (IPBS), CNRS, 205 route de Narbonne, F-31077 Toulouse, France, the [¶]Université de Toulouse, Université Paul Sabatier, IPBS, F-31077 Toulouse, France, the [‖]Department of Biochemistry, Faculty of Natural Sciences, Comenius University, Mlynska dolina CH-1, 84215 Bratislava, Slovak Republic, and the [**]Unité de Génétique Mycobactérienne, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France

The arabinogalactan (AG) of slow growing pathogenic *Mycobacterium* spp. is characterized by the presence of galactosamine (GalN) modifying some of the interior branched arabinosyl residues. The biosynthetic origin of this substituent and its role(s) in the physiology and/or pathogenicity of mycobacteria are not known. We report on the discovery of a polyprenyl-phospho-*N*-acetylgalactosaminyl synthase (PpgS) and the glycosyltransferase Rv3779 from *Mycobacterium tuberculosis* required, respectively, for providing and transferring the GalN substrate for the modification of AG. Disruption of either *ppgS* (*Rv3631*) or *Rv3779* totally abolished the synthesis of the GalN substituent of AG in *M. tuberculosis* H37Rv. Conversely, expression of *ppgS* in *Mycobacterium smegmatis* conferred upon this species otherwise devoid of *ppgS* ortholog and any detectable polyprenyl-phospho-*N*-acetylgalactosaminyl synthase activity the ability to synthesize polyprenyl-phospho-*N*-acetylgalactosamine (polyprenyl-P-GalNAc) from polyprenyl-P and UDP-GalNAc. Interestingly, this catalytic activity was increased 40–50-fold by co-expressing *Rv3632*, the encoding gene of a small membrane protein apparently co-transcribed with *ppgS* in *M. tuberculosis* H37Rv. The discovery of this novel lipid-linked sugar donor and the involvement of a the glycosyltransferase C-type glycosyltransferase in its transfer onto its final acceptor suggest that pathogenic mycobacteria modify AG on the periplasmic side of the plasma membrane. The availability of a *ppgS* knock-out mutant of *M. tuberculosis* provides unique opportunities to investigate the physiological function of the GalN substituent and the potential impact it may have on host-pathogen interactions.

Galactosamine (D-GalN) was recognized as a minor covalently bound amino sugar component of the cell wall of slow growing mycobacteria (*Mycobacterium tuberculosis* H37Ra, *Mycobacterium lepraemurium*, *Mycobacterium microti*) as early as in the 1970s (1–3). Advances in mass spectrometry later coupled with the use of an arabinanase isolated from *Mycobacterium smegmatis* allowed the GalN substituent to be mapped to the arabinogalactan (AG)[7] component of the cell wall (3), and more specifically to the C2 position of a portion of the internal 3,5-branched D-Ara*f* residues in the AG of *M. tuberculosis* (4, 5). D-GalN was estimated to occur at the level of about one residue per entire AG molecule (3–5). Interestingly, a similar sugar residue was found to substitute the AG of *Mycobacterium avium*, *Mycobacterium kansasii*, *Mycobacterium bovis* BCG (3) and *Mycobacterium leprae*,[8] but not that of *M. smegmatis*, *M. neoaurum*, and *M. phlei* (3–5), suggesting that fast growing *Mycobacterium* spp. are devoid of GalN substituent.

The function and biosynthetic origin of this substituent are not known. Non-*N*-acetylated hexosamine residues, particularly GalN residues, are rather uncommon in prokaryotes. Thus far, they essentially have been described as components of the lipopolysaccharide (LPS) of some Gram-negative bacteria (6–9). In *Francisella tularensis* and subsp. *novicida* for instance, GalN covalently modifies the lipid A moiety of LPS neutralizing its negative charge and enhancing the virulence of the bacterium in mice (8, 9). In Gram-positive bacteria, GalN and glucosamine (GlcN) have been reported in the glycosylation motif of a glycoprotein toxin from *Bacillus thuringiensis* subsp. *israelensis* and proposed to be important to the

* This work was supported, in whole or in part, by National Institutes of Health Grants AI064798 and AI018357 through the NIAID. This work was also supported by AIDS-Fogarty International Research Collaboration Award Grant TW 006487 and its parent National Institutes of Health Grant R37 AI18357 through the NIAID.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1–S8.

[1] Both authors contributed equally to this work.

[2] Present address: Dept. of Biochemistry, Faculty of Natural Sciences, Comenius University, Mlynska dolina CH-1, 84215 Bratislava, Slovak Republic.

[3] Present address: Dept. of Immunology, Max Planck Institute for Infection Biology, Berlin, Germany.

[4] Co-senior authors.

[5] To whom correspondence may be addressed: CNRS-IPBS, Département Mécanismes Moléculaires des Infections Mycobactériennes, 205 route de Narbonne, F-31077 Toulouse, France. Tel.: 33 5 61 17 55 54; Fax: 33 5 61 17 59 94; E-mail: Jerome.Nigou@ipbs.fr.

[6] To whom correspondence may be addressed. Tel.: 970-491-3582; Fax: 970-491-1815; E-mail: Mary.Jackson@colostate.edu.

[7] The abbreviations used are: AG, arabinogalactan; APTS, 8-aminopyrene-1,3,6-trisulfonate; CE, capillary electrophoresis; CE-LIF, CE monitored by laser-induced fluorescence; GT, glycosyltransferase; kan, kanamycin; PpgS, polyprenyl-phospho-*N*-acetylgalactosaminyl synthase.

[8] S. Bhamidi, unpublished observation.

biological activities of the toxin (10). GalN was also identified as a possible component of a galactose-containing polysaccharide from *Streptococcus mutans* (11).

Recent studies have partially elucidated the biosynthetic origin of the D-GalN unit modifying the lipid A of *Francisella* (8, 9, 12). It was shown to originate in the lipid-linked sugar donor, undecaprenyl phosphate-β-D-*N*-acetylgalactosamine which, upon deacetylation by a specialized deacetylase, yields undecaprenyl phosphate-β-D-galactosamine. This is then used as the D-GalN donor in a glycosyl transfer reaction onto lipid A catalyzed by the membrane-associated glycosyltransferase FlmK (8, 9, 12). This process is reminiscent of that involved in the addition of aminoarabinose to lipid A in enteric bacteria where an enzyme (termed ArnC in *Escherichia coli*) catalyzes the synthesis of undecaprenyl-phospho-*N*-formyl-aminoarabinose from UDP-*N*-formyl-aminoarabinose, and a glycosyltransferase (ArnT in *E. coli*) catalyzes the transfer of the deformylated aminoarabinose residue from this lipid donor onto lipid A (13).

Numerous glycosylation events in the biosynthesis of *M. tuberculosis* cell wall and other glycoconjugates are known to occur on the periplasmic face of the inner membrane, catalyzed by membrane-associated glycosyltransferases of the GT-C superfamily dependent on decaprenyl-phosphate-linked sugar donors rather than nucleotide-sugar donors (for a review, see Refs. 14–16). However, to date, the known polyprenyl-based sugar donors of *Mycobacterium* spp. only include β-D-glucosyl-1-monophosphoryl-decaprenol, β-D-mannosyl-1-monophosphoryl-decaprenol, β-D-ribosyl-1-monophosphoryl-decaprenol, and β-D-arabinofuranosyl-1-monophosphoryl-decaprenol (17–19). To the best of our knowledge, polyprenyl-linked galactosamine or *N*-acetylgalactosamine has never been reported in mycobacteria. This study reports on the discovery of the biosynthetic pathway for the synthesis of lipid-linked-D-GalNAc in mycobacteria and subsequent D-GalNAc or D-GalN transfer onto AG.

EXPERIMENTAL PROCEDURES

Bacterial Strains and Growth Conditions—*M. tuberculosis* H37Rv ATCC 25618 and H37Rv TMC102 were grown in Sauton's and Middlebrook 7H9 medium. *M. smegmatis* mc^2155 was grown in 7H9 or LB medium. Kanamycin (kan) and hygromycin were added to final concentrations of 20 μg ml^{-1} and 50 μg ml^{-1}, respectively.

M. tuberculosis Knock-out Mutants—The Ts*/sacB* method was used to achieve allelic replacement at the *Rv3631* (*ppgS*) locus of *M. tuberculosis* H37Rv (20). The *ppgS* gene and flanking regions was PCR-amplified from *M. tuberculosis* H37Rv genomic DNA with primers Rv3631.5 (5′-ttggtacctctagacccg-caccgtggtcg-3′) and Rv3631.6 (5′-ttgaattccagatcaacgcgtcgccc-3′) and a disrupted allele, *ppgS::kan*, was obtained by inserting the *Tn903* kanamycin resistance cassette at the FseI restriction site of *ppgS*. *ppgS::kan* was then cloned into the XbaI-cut and blunt-ended pPR27-*xylE* (20), yielding pPR27*ppgS*KX. The *Rv3779* knock-out mutant of *M. tuberculosis* H37Rv TMC102 and complemented mutant strains were described elsewhere (21). The *mshB* knock-out mutant of *M. tuberculosis* Erdman (22) was kindly provided by Drs. Fahey and Newton (University of California San Diego), and the *mca* transposon mutant of *M. tuberculosis* CDC1551 was obtained from CSU TB Vaccine Testing and Research Materials Contract (National Institutes of Health,NIAID N01-AI-40091).

Expression of ppgS and Rv3632 in M. smegmatis—The entire coding sequences of *ppgS* and *Rv3632* were PCR-amplified from *M. tuberculosis* H37Rv genomic DNA and cloned into the expression vector pVV16 (23), yielding pVV*ppgS* and pVV*Rv3632*, respectively. pVV*ppgS-Rv3632* was generated by amplifying the *ppgS* and *Rv3632* ORFs together and cloning the resulting PCR fragment into the same expression plasmid. These constructs allow the constitutive expression of *ppgS*, *Rv3632*, or both genes under control of the *hsp60* promoter. Primer sequences are available upon request.

Enzyme Assays—Membranes from *M. smegmatis* mc^2155 were prepared as described (24) and resuspended in 50 mM MOPS buffer (pH 7.9) (buffer A). The typical reaction mixture for the production of polyprenyl-P-D-[^{3}H]GalNAc from UDP-α-D-[^{3}H]GalNAc by *M. smegmatis* membranes contained 0.5 μCi of UDP-α-D-[^{3}H]GalNAc (specific activity 20 Ci $mmol^{-1}$; American Radiolabeled Chemicals, Inc., St. Louis, MO), 5 mM 2-mercaptoethanol, 10 mM $MgCl_2$, 0.1 mM ATP, 0.1 mM NADH, membrane fraction (0.5 mg of proteins), and buffer A in a final volume of 80 μl; thus, the membranes were the source of the endogenous polyprenyl-P acceptors. After incubation at 37 °C, the reactions were stopped by addition of 1.2 ml of $CHCl_3/CH_3OH/NH_4OH$ (2:1:0.05, v/v). The upper aqueous phase was discarded, and the bottom organic phase was backwashed with $CHCl_3/CH_3OH/H_2O$ (3:47:48). The dried organic phase was dissolved in 50 μl of $CHCl_3/CH_3OH/NH_4OH/H_2O$ (65:25:0.5:3.6) prior to TLC analysis on aluminum-backed silica gel 60-precoated plates F_{254} (Merck, Darmstadt, Germany).

Capillary Electrophoresis—Purified AG (10 μg) was partially degraded by controlled acid hydrolysis (0.1 M HCl for 20 min at 110 °C), and the oligosaccharides liberated were then tagged with the fluorescent probe 8-aminopyrene-1,3,6-trisulfonate (APTS) as described earlier (25). Labeled oligosaccharides were then analyzed by capillary electrophoresis monitored by laser-induced fluorescence (CE-LIF) on a P/ACE 5000 instrument (Beckman Instruments, Inc.). Separations were performed using an uncoated fused-silica capillary column (Sigma) of 50-μm internal diameter and 40-cm effective length (47-cm total length). Analyses were carried out at a temperature of 25 °C, in the reverse mode, with an applied voltage of 20 kV using 1% acetic acid (w/v)-30 mM triethylamine, pH 3.5, as a running electrolyte (25). Picopreparative CE was performed as described previously (26).

MALDI-TOF/MS and MS/MS—Analyses were performed on a 4700 Proteomics Analyzer (with TOF/TOF optics, Applied Biosystems, Voyager DE-STR) in the reflectron mode. The matrix used was 2,5-dihydroxybenzoic acid (10 μg μl^{-1} in a mixture of ethanol/water (1:1, v/v)). 0.3 μl of the CE-collected oligosaccharides resuspended in 10 μl of water or TLC-extracted lipids resuspended in 10 μl $CHCl_3/CH_3OH$ (2:1) was mixed with 0.3 μl of the matrix solution. Mass spectra were recorded in the negative mode. The collision-induced

dissociation gas type for MS/MS experiments was atmosphere, and the gas pressure was set to medium.

RESULTS

Disruption of ppgS in M. tuberculosis H37Rv—Rv3631 (thereafter renamed *ppgS*) is predicted to encode a GT-A-fold glycosyltransferase (GT) (14, 15) and, according to the Carbohydrate-Active enZymes classification of glycosyltransferases (27), to belong to the GT-2 family of inverting GTs dependent on NDP-sugars as donor substrates. Orthologs of *ppgS* are found in the genomes of slow growing mycobacteria including *M. bovis*, *M. bovis* BCG, *M. leprae*, *Mycobacterium ulcerans*, *Mycobacterium marinum*, *M. avium*, and *Mycobacterium paratuberculosis*, as well as in *Mycobacterium abscessus*, but not in the genomes of other rapidly growing *Mycobacterium* species such as *M. smegmatis* (supplemental Fig. S1). Interestingly, in all of the genomes analyzed, *ppgS* and orthologs map in close vicinity to *galE1*, a gene thought to encode the UDP-glucose 4-epimerase responsible for the conversion of UDP-D-Glc*p* to UDP-D-Gal*p* in AG synthesis (16). *ppgS* and orthologs appear to be co-transcribed with a gene potentially encoding a small membrane protein (13 kDa) of unknown function (Rv3632 in *M. tuberculosis* H37Rv) (supplemental Fig. S1). PpgS shares significant sequence similarities with polyprenyl-monophosphomannose synthases from various prokaryotic sources, including Ppm1 (Rv2051c) from *M. tuberculosis* H37Rv (24% identity, 39% similarity on a 228-amino acid overlap).

To investigate the putative involvement of this GT in AG synthesis, the *ppgS* gene from *M. tuberculosis* H37Rv was disrupted by homologous recombination using the Ts*/sacB* method (20). Allelic replacement at the *ppgS* locus was confirmed by Southern hybridization (supplemental Fig. S2*A*). The mutant grew similarly to its wild-type (WT) H37Rv parent in 7H9-ADC-Tween 80 broth at 37 °C (supplemental Fig. S2*B*).

Phenotypic Characterization of the ppgS Mutant Strain—Soluble AG was prepared from the WT and mutant strains as described previously (5), and the Ara/Gal ratios were determined by CE or GC upon total acid hydrolysis. No significant differences between WT and mutant strains were observed, indicating that PpgS was probably not involved in the polymerization of the arabinan or galactan domains. Because the distribution of *ppgS* within the *Mycobacterium* genus seemed to correlate with that of D-GalN on AG, we decided to probe the mutant for the presence of this substituent. To this end, a fast and sensitive CE-based methodology was first developed to facilitate the detection of this rare substituent. By analogy to the procedure that we previously developed to analyze the mannose caps of lipoarabinomannan (25, 28), we postulated that a controlled mild acid hydrolysis of AG would cleave the arabinofuranosidic bonds of the arabinan domain while preserving the pyranosidic linkage of D-GalN onto the relevant Ara*f* unit, thereby releasing an oligosaccharide that could be detected by CE-LIF upon tagging with an APTS fluorescent probe. A typical electrophoretogram of *M. tuberculosis* H37Rv AG resulting from this analysis is shown in Fig. 1*A*. The monosaccharide region is dominated by two main peaks

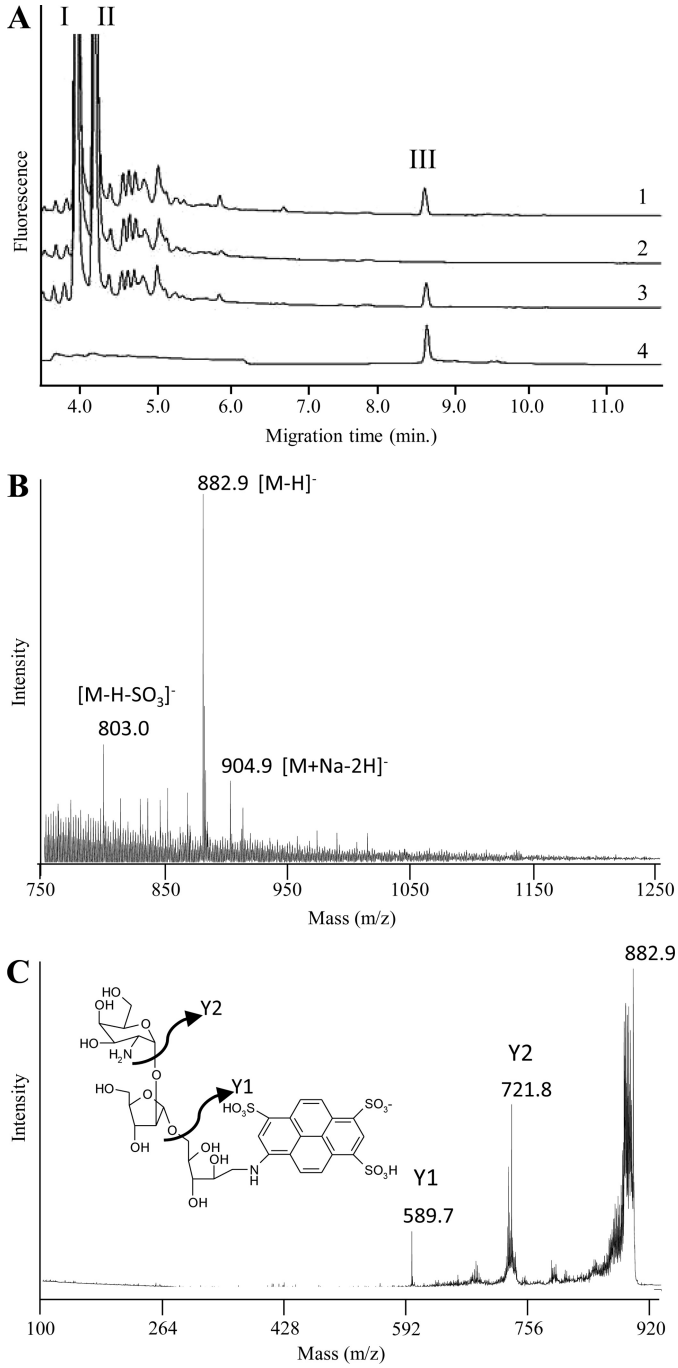

FIGURE 1. **Arabinogalactan composition of the *ppgS* knock-out mutant of *M. tuberculosis* H37Rv.** *A*, CE-LIF analysis of AG after mild acid hydrolysis and APTS derivatization. Shown are the partial electrophoretograms obtained with AG prepared from *M. tuberculosis* H37Rv wild-type (*trace 1*), the *ppgS* knock-out mutant (*trace 2*), and the complemented mutant strain, H37RvΔ*ppgS*/pVV*ppgS* (*trace 3*). An electrophoretogram of CE-collected peak III is shown in *trace 4*. *I*, D-Ara*f*-APTS; *II*, D-Gal*p*-APTS; *III*, D-GalN*p*-D-Ara*f*-D-Ara*f*-APTS. *B*, negative MALDI-TOF mass spectrum of compound III. *C*, MALDI-MS/MS collision-induced dissociation mass spectrum of the precursor ions 882.9 assigned to the deprotonated molecular ions of compound III.

attributed to D-Ara*f*-APTS (I) and D-Gal*p*-APTS (II) arising from the hydrolysis of the arabinan and galactan domains. A third peak (III) of lower intensity was detected with a migration time matching that of a maltoheptaose-APTS, tentatively attributed to an oligosaccharide-APTS containing D-Gal*p*N

and D-Ara*f* units. This assumption was confirmed by collecting peak III in a picopreparative CE step (26) and subjecting it to MALDI/MS analysis (Fig. 1*B*). The negative MALDI mass spectrum was dominated by a peak at *m/z* 882.9 attributed to deprotonated molecular ions $[M-H]^-$ that could be assigned to an APTS derivative containing two pentosyl and one hexosaminyl units. The identity of these units as two arabinosyl and one galactosaminyl residues was revealed by CE analysis upon acid hydrolysis and labeling with APTS or fluorescein-succinimidyl ester. The exact trisaccharide sequence, D-GalN-(1→2)-α-D-Ara*f*-(1→5)-α-D-Ara*f*, was revealed by MS/MS analysis (Fig. 1*C*) showing fragment ions at *m/z* 721.8 (y2) and *m/z* 589.7 (y1) arising from the loss of anhydro-D-GalN*p* and anhydro-D-GalN*p*-Ara*f*, and based on the described structure of AG (5) (see also Fig. 6). Peak III was thus further used as a reporter peak for the presence of D-GalN on AG.

Suggestive of an involvement of PpgS in the formation of the GalN substituent of AG, peak III was noticeably absent from the electrophoretogram of H37RvΔ*ppgS::kan* but restored in the complemented mutant strain H37RvΔ*ppgS::kan*/pVV*ppgS*. Corroborating this observation, peak III was also detected in the AG of several *ppgS*-expressing *Mycobacterium* species (*M. bovis* BCG, *M. marinum*, and *M. abscessus*) but, as expected, was missing from that of *M. smegmatis* (supplemental Fig. S3). Although PpgS shares significant sequence similarities with polyprenol-monophosphomannose synthases, no significant differences were found between the mannolipid and lipoglycan (lipomannan and lipoarabinomannan) contents of the WT and mutant strains (supplemental Fig. S4).

Effect of Overexpressing ppgS on the Synthesis of Lipid-linked Sugars and AG in M. smegmatis—The absence of a *ppgS* ortholog in *M. smegmatis* provided a convenient mycobacterial host for the heterologous expression and functional characterization of this gene. Transformation of mc^2155 with pVV*ppgS*, the plasmid used for complementation studies in *M. tuberculosis*, resulted in the production of a C-terminal His_6-tagged recombinant protein of the expected size (~26 kDa) (Fig. 2, *A* and *B*). The recombinant protein was almost exclusively found in the membrane and P60 fractions of the bacterium, suggestive of its association with membranes (supplemental Fig. S5). To optimize the chances of producing an active form of PpgS in *M. smegmatis*, a recombinant strain of *M. smegmatis* expressing *ppgS* in combination with *Rv3632* and a control strain expressing *Rv3632* alone were also constructed. In all cases, recombinant strains grew similarly to the control in liquid broth and the production of recombinant proteins of the expected sizes was confirmed (Fig. 2*B*).

Analysis of the AG of the various recombinant strains as described above failed to reveal the presence of a D-GalN*p* (or *N*-acetyl-D-GalN*p*) substituent on the AG of mc^2155/pVV16, mc^2155/pVV*ppgS*, mc^2155/pVV*Rv3632*, and mc^2155/pVV*ppgS-Rv3632* (supplemental Fig. S6). It thus appeared that the expression of *ppgS* alone or in combination with *Rv3632* was not sufficient for the formation of this substituent in *M. smegmatis*.

Cell-free assays using membranes from the recombinant *M. smegmatis* strains were then conducted to determine whether

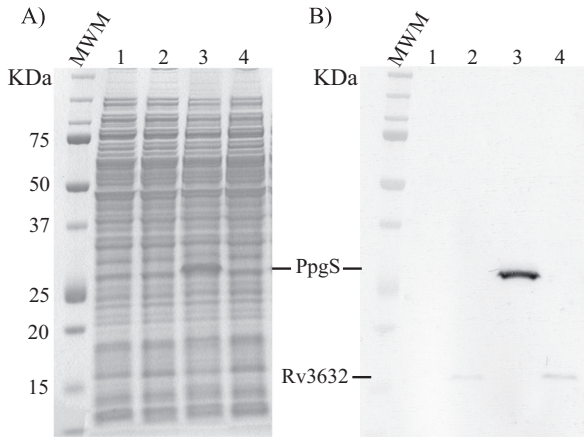

FIGURE 2. **Production of recombinant forms of PpgS and Rv3632 in *M. smegmatis* mc^2155.** The production of a recombinant C-terminal His_6-tagged form of PpgS from *M. tuberculosis* was detected by Coomassie Blue staining (*A*) and Western blot using a monoclonal anti-His tag antibody (*B*) in the mc^2155/pVV*ppgS* strain (*lanes 3*) but not in the control strain, mc^2155/pVV16 (*lanes 1*). The production of a recombinant C-terminal His_6-tagged form of Rv3632 from *M. tuberculosis* was detected by Western blotting (*B*) in the mc^2155/pVV*Rv3632* strain (*lanes 2*) and mc^2155/pVV*ppgS-Rv3632* strain (*lanes 4*) but not in the control strain, mc^2155/pVV16 (*lanes 1*).

PpgS is involved in some aspects of lipid-linked sugar biosynthesis. TLC autoradiographs clearly revealed the incorporation of D-[^{3}H]GalNAc from UDP-α-D-[^{3}H]GalNAc into two major glycolipids (A and B) by mc^2155/pVV*ppgS* membranes that were not detected in the control strain mc^2155/pVV16 or in the strain expressing the only *Rv3632* gene, mc^2155/pVV*Rv3632* (Fig. 3*A*). Strikingly, this enzymatic activity was increased 40–50-fold in the strain co-expressing *ppgS* and *Rv3632* (Fig. 3*A*, *lane 4*). The accumulated D-[^{3}H]GalNAc- (or D-[^{3}H]GalN)-containing glycolipids were mild alkali-stable and mild acid-labile (supplemental Fig. 7*SA*) with TLC mobility properties suggestive of lipid-linked sugars of the mycobacterial polyisoprenyl-P class (19, 21, 29). Further assays aimed at testing the effects of time and protein or acceptor substrate concentration on D-[^{3}H]GalNAc transfer by mc^2155/pVV*ppgS-Rv3632* membranes showed clear enzyme and substrate concentration-dependent increases in activity (Fig. 3, *B* and C). Increases in activity were similar whether decaprenyl (C_{50})-P or the shorter heptaprenyl (C_{35})-P were used as acceptor substrates (Fig. 3*D*) but inexistent when C_{10}-P served as the acceptor (supplemental Fig. S7*B*) indicative of PpgS preference for longer lipid carriers. Suggestive of PpgS specificity for the sugar transferred, assays performed under similar conditions using other sugar donors including GDP-D-[^{14}C]Man, UDP-D-[^{14}C]GlcNAc, ADP-D-[^{14}C]Glc, and UDP- D-[^{14}C]Gal failed to reveal any effect of overexpressing *ppgS* on the incorporation of these radiolabeled sugars into glycolipids (data not shown).

Structural Characterization of Glycolipids A and B—To establish the nature of the products formed in the assay using mc^2155/pVV*ppgS-Rv3632* membranes and UDP-α-D-[^{3}H]GalNAc, a nonradioactive assay was up-scaled, and the products of the reaction, purified by preparative TLC, were submitted to MALDI/MS analyses. The negative MALDI-TOF mass spectra of purified products A and B were dominated by peaks at *m/z* 980.5 and 784.4, respectively, attributed

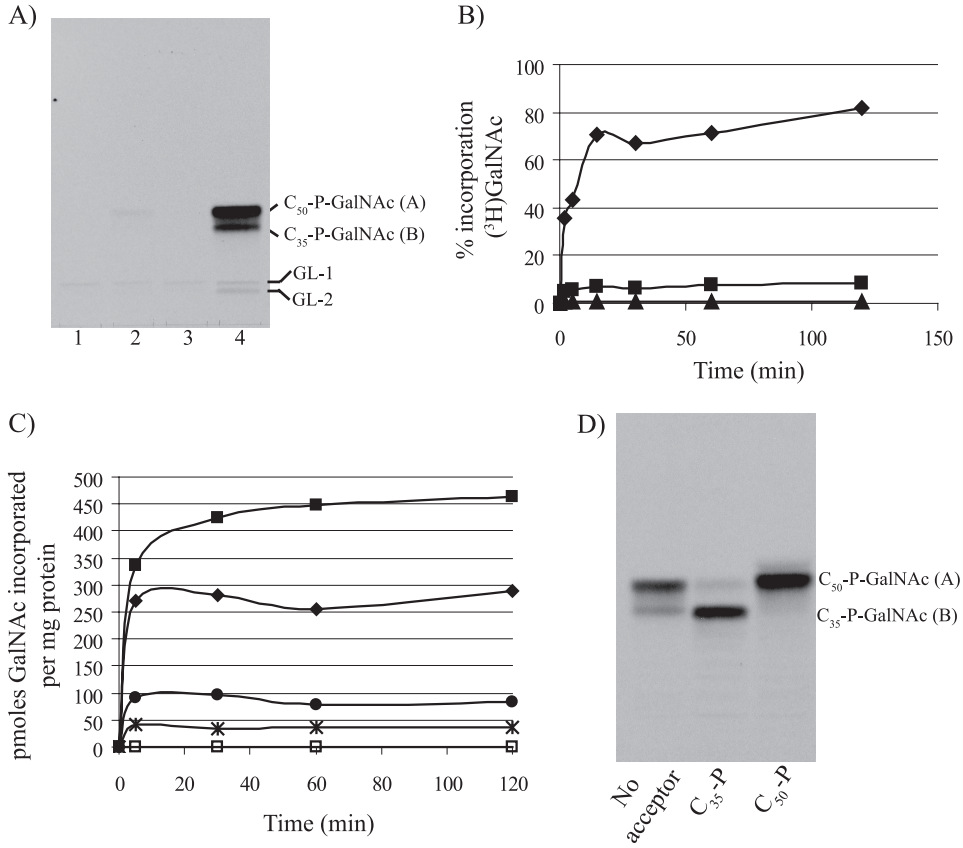

FIGURE 3. **Effect of overexpressing *ppgS* alone and in combination with *Rv3632* on the incorporation of D-[^{3}H]GalNAc from UDP-α-D-[^{3}H]GalNAc into lipid-linked sugars by membrane extracts from *M. smegmatis*.** *A*, TLC analysis of an *in vitro* cell-free assay using UDP-[^{3}H]-α-D-GalNAc and membranes from mc^{2}155/pVV16 (*lane 1*), mc^{2}155/pVV*ppgS* (*lane 2*), mc^{2}155/pVV*Rv3632* (*lane 3*), and mc^{2}155/pVV*ppgS-Rv3632* (*lane 4*). The synthesized lipid-linked sugars were extracted with $CHCl_3/CH_3OH/NH_4OH$ (2:1:0.05, v/v), and a 10% aliquot of each sample was analyzed by TLC in the solvent system $CHCl_3/CH_3OH/NH_4OH/H_2O$ (65:25:0.5:3.6, v/v) followed by autoradiography. The radiolabeled products migrating close to the origin in all assays were identified by co-migration as C_{50}-P-P-D-GlcNAc (GL-1) and C_{50}-P-P-D-GlcNAc-Rha (GL-2) and probably arose from the randomization of UDP-α-D-[^{3}H]GalNAc into UDP-α-D-[^{3}H]GlcNAc which is known to serve as a donor substrate in the initial steps of the biosynthesis of AG (15). *B*, percentage incorporation of D-[^{3}H]GalNAc into polyprenyl-P-D-GalNAc over time by membranes of mc^{2}155/pVV*ppgS-Rv3632* as a function of the amount of membrane proteins used in the assay. *Triangles*, 0.5 μg of protein; *squares*, 5 μg; *diamonds*, 50 μg. *C*, incorporation of D-[^{3}H]GalNAc into polyprenyl-P-D-GalNAc over time by membranes of mc^{2}155/pVV16 (*open symbols*) and mc^{2}155/pVV*ppgS-Rv3632* (*filled symbols*) as a function of the concentration of C_{50}-P present in the assay mixture. *Squares*, 250 μM C_{50}-P; *diamonds*, 25 μM C_{50}-P; *circles*, 2.5 μM C_{50}-P; *stars*, no added acceptor substrate. *D*, TLC analysis of *in vitro* cell-free assays using UDP-α-D-[^{3}H]GalNAc and membranes from mc^{2}155/pVV*ppgS-Rv3632*, in the presence or absence of added acceptor substrates (C_{35}-P and C_{50}-P). Assays were run and analyzed as in *A*.

to deprotonated molecular ions $[M-H]^-$ that could be assigned to C_{50}-P-D-GalNAc and C_{35}-P-D-GalNAc (Fig. 4, *A* and *B*). This assignment was confirmed by MS/MS analyses showing for compound A fragment ions at *m/z* 300.5 and 778.2, arising from the loss of the decaprenyl and *N*-acetylgalactosaminyl moieties, respectively (Fig. 4*C*). The negative MALDI-TOF/TOF mass spectrum of purified compound B exhibited the same reporter fragment ions at *m/z* 300.5 and 581.1 (Fig. 4*D*). Altogether, our results thus indicate that *ppgS* encodes a polyprenyl-phospho-*N*-acetylgalactosaminyl synthase required for the synthesis of the GalN substituent of AG.

Identification of the Glycosyltransferase Responsible for GalN Transfer onto AG—Because the involvement of a lipid-linked D-GalN(Ac) donor suggested that the transfer of the amino sugar onto AG occurred on the periplasmic side of the plasma membrane, we thought to search for the responsible GT among the membrane-associated GT-C enzymes of *M. tuberculosis* (14). Of the 17 proposed GT-Cs of strain H37Rv, only 2 are missing an ortholog in *M. smegmatis* mc^{2}155 but have orthologs in *M. leprae*, *M. abscessus*, and other species harboring a D-GalN*p* substituent on AG, namely Rv1635c and Rv3779. Rv1635c was demonstrated earlier to catalyze the addition of the first mannose residue in the Man-capping of mannosylated lipoarabinomannan (30). *Rv3779* clusters on the chromosome of *M. tuberculosis* with genes involved in the biosynthesis of the cell wall core and was recently found to participate in the synthesis of polyprenylphosphomannose (21). Analysis of the AG of mutants deficient in the expression of each of these genes (21, 30) revealed that although the *M. tuberculosis* CDC1551 *Rv1635c* knock-out strain still carried the D-GalN*p* substituent (supplemental Fig. S8*A*), the AG of *M. tuberculosis* H37RvΔ*Rv3779* was totally devoid of it (Fig. 5). Production of galactosaminylated AG was restored in the *Rv3779* mutant upon complementation with a WT copy of *Rv3779* expressed from a replicative plasmid (Fig. 5).

DISCUSSION

The presence of a covalently bound GalN residue typifies the AG of slow growing mycobacteria. We have identified the

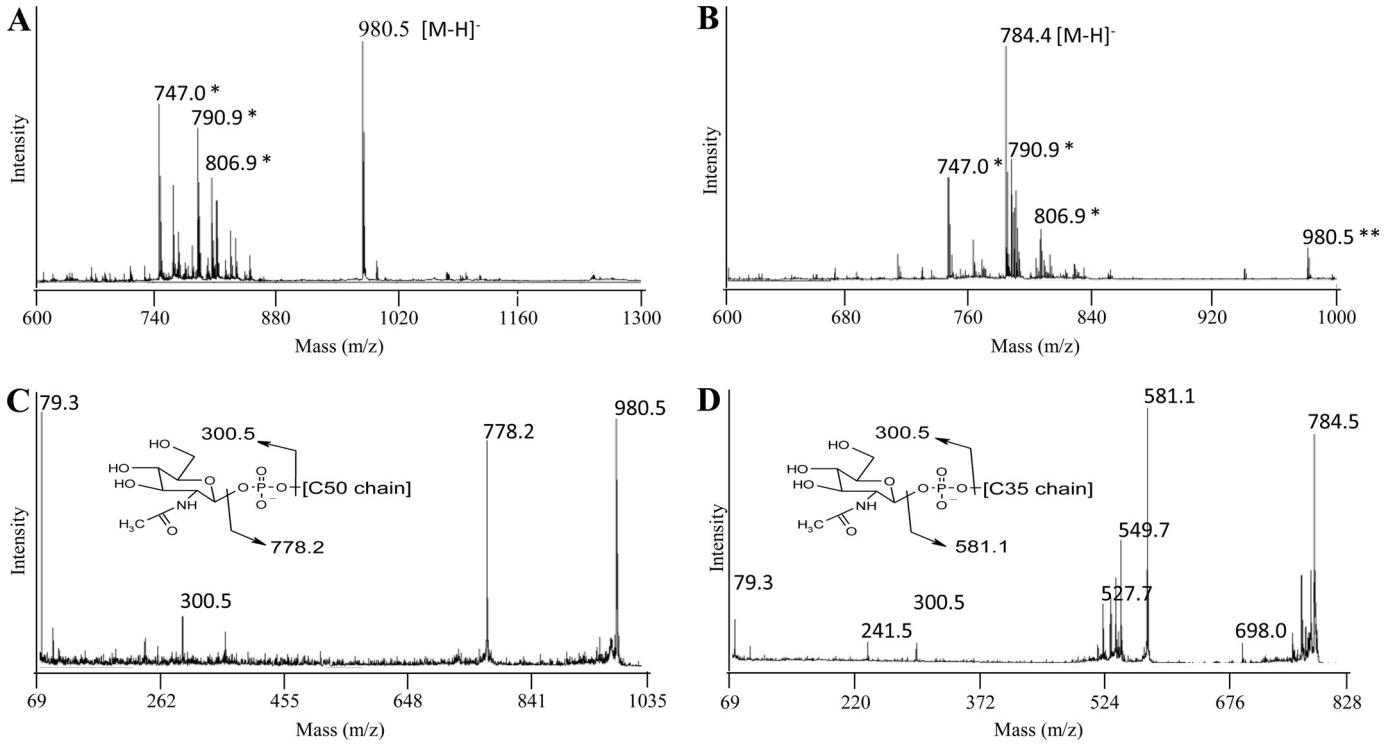

FIGURE 4. **Structural characterization of products A and B by MALDI/MS.** *A* and *B*, negative MALDI-TOF mass spectra of product A (*A*) and product B (*B*). *C* and *D*, negative MALDI-TOF/TOF mass spectra of product A (*C*) and product B (*D*). *, matrix ions; **, $[M-H]^-$ ions from contaminating C_{50}-P-D-GalNAc.

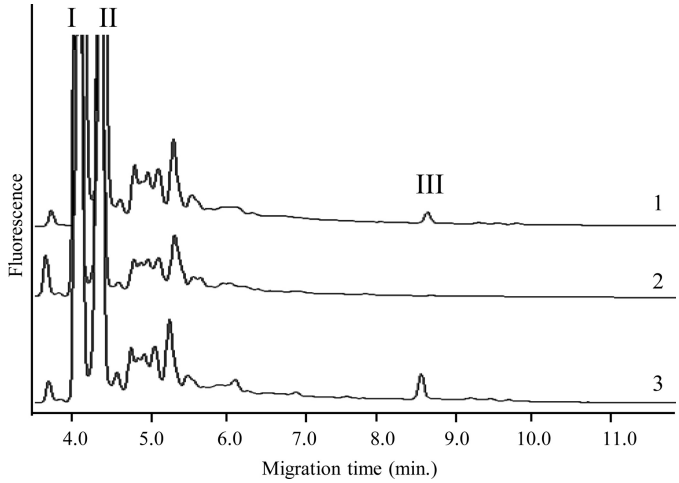

FIGURE 5. **Arabinogalactan composition of a *Rv3779* knock-out mutant of *M. tuberculosis* H37Rv.** CE-LIF analysis of the AG prepared from *M. tuberculosis* H37Rv wild-type (*trace 1*), the *Rv3779* knock-out mutant (*trace 2*), and the complemented mutant strain, H37RvΔ*Rv3779*/pVV*Rv3779* (*trace 3*). *I*, D-Ara*f*-APTS; *II*, D-Gal*p*-APTS; *III*, D-GalN*p*-D-Ara*f*-D-Ara*f*-APTS.

key components of the biosynthetic machinery required for the synthesis and transfer of this amino sugar in *M. tuberculosis* and provided evidence that a previously unknown mycobacterial lipid-linked sugar, polyprenyl-P-(*N*-acetyl)-galactosamine served as the sugar donor in this process. PpgS, a sugar nucleotide-utilizing GT, catalyzes the formation of polyprenyl (C_{35}/C_{50})-P-D-GalNAc from UDP-α-D-GalNAc and polyprenyl-P on the cytosolic side of the plasma membrane. This sugar donor is then presumably deacylated by an as yet unknown deacetylase before or after being translocated to the outer leaflet of the plasma membrane where the membrane-associated GT-C enzyme Rv3779 transfers the D-GalN*p* (or D-GalNAc) residue from polyprenyl-P-D-GalN(Ac) to the C2 position of a portion of the internal 3,5-branched D-Ara*f* residues of AG (Fig. 6). The identity of the enzymes involved and deduced compartmentalization of this biosynthetic pathway are in line with what is currently known of the biogenesis of AG, whose early biosynthetic steps take place on the cytosolic side of the plasma membrane whereas later stages leading to the elongation and branching of the arabinan domain occur extracytoplasmically (16). The distribution of the *ppgS*, *Rv3632*, and *Rv3779* genes within the *Mycobacterium* genus matches the reported presence or absence of a D-GalN*p* substituent on the AG of the mycobacterial species analyzed to date.

The identification of the PPM synthase Rv3779 (21) as the galactosaminyltransferase (or *N*-acetylgalactosaminyltransferase) of the system indicates that this GT-C enzyme is endowed with at least two distinct enzymatic activities and highlights the complexity of the GT-C enzymes responsible for much of the biosynthesis of mycobacterial cell wall polysaccharides. Consistent with this finding, in addition to the cytoplasmic GT-C signature motif (DLD at position 86) which we proposed earlier to be involved in the transfer of a mannose residue from GDP-α-D-Man*p* to polyprenyl phosphate (14, 21), a second putative GT-C motif predicted to be located in the last extracytoplasmic loop of the protein (TMHMM 2.0) and susceptible of catalyzing the transfer of D-GalN(Ac) was identified (DLD at position 536). This motif and associated downstream acidic, aromatic, and proline residues (YDYP at position 578) (14) are conserved among the *Mycobacterium* species reported to display a GalN-substituted AG. The pre-

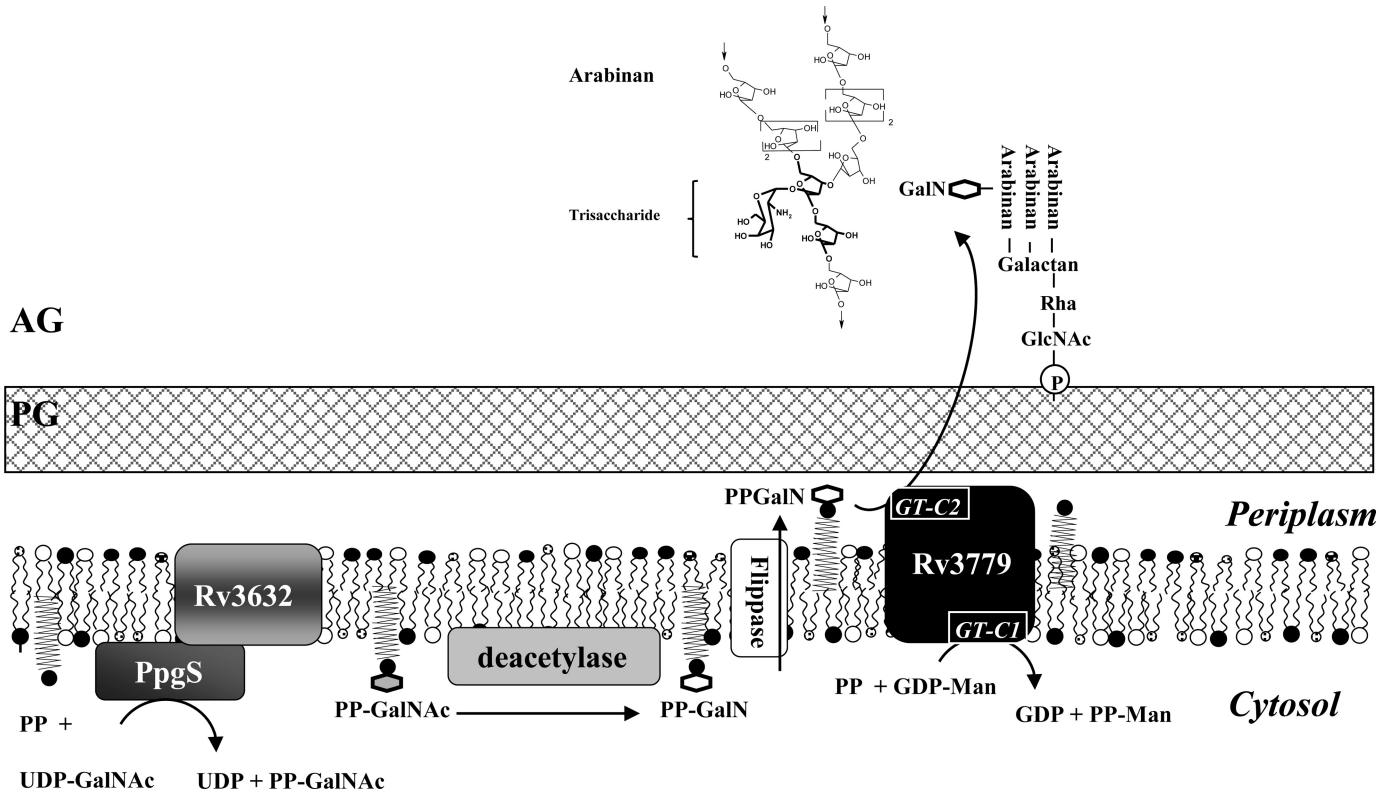

FIGURE 6. **Proposed model for the biosynthesis of the GalN substituent of AG.** *PG*, peptidoglycan; *AG*, arabinogalactan; *PP*, polyprenol phosphate; *PP-GalNAc*, polyprenyl-monophosphoryl-*N*-acetylgalactosaminyl; *PP-GalN*, polyprenyl-monophosphoryl-galactosaminyl; *PP-Man*, polyprenyl-phosphomannose. The detailed structure of a D-GalN residue modifying an arabinan chain of AG is shown. The trisaccharide moiety of compound III is in *bold font*. Also shown are the two predicted GT-C signature motifs of Rv3779. GT-C1 (DLD at position 86) located on the cytosolic side of the plasma membrane is involved in the transfer of a mannose residue from GDP-α-D-Man*p* to polyprenyl phosphate (21); GT-C2 (DLD at position 536) is predicted to be in the last extracytoplasmic loop of the protein and to catalyze the transfer of D-GalN(Ac) onto AG.

dicted spatial distribution of each GT-C motif, which is in line with that of their respective donor and acceptor substrates (cytosolic GDP-Man and periplasmic arabinan chains of AG in particular), would thus govern the dual catalytic activities of Rv3779 (Fig. 6). Further experiments are under way to validate this hypothesis.

Missing biosynthetic components required for producing the GalN substituent of AG include the deacetylase responsible for deacetylating polyprenyl-P-D-GalNAc to the corresponding D-GalN derivative and the translocase required for translocating polyprenyl-P-D-GalN/polyprenyl-P-D-GalNAc from the inner to the outer leaflet of the plasma membrane (Fig. 6). A search for the missing deacetylase of the system in the genome of *M. tuberculosis* H37Rv led to our interest in *Rv1170* (*mshB*) and *Rv1082* (*mca*). MshB participates in the deacetylation of 1-D-*myo*-inosityl-2-acetamido-2-deoxy-α-D-glucopyranoside in the biosynthesis of mycothiols and, *in vitro*, displays deacetylase activity on various other acetylated substrates. The mycothiol *S*-conjugate amidase Mca shares with MshB sequence similarities and deacetylase activity *in vitro*, although its primary function in the bacterium is the cleavage of various *S*-conjugates of mycothiol (22, 31, 32). Analysis of AG in the *M. tuberculosis mshB* and *mca* knock-out strains, however, revealed WT profiles in both mutants (supplemental Fig. S8*B*), suggesting either that these enzymes are not involved in the formation of the D-GalN*p* substituent or that compensatory enzymatic activities exist in the cell.

With regard to the translocase of the system, studies are in progress in our laboratories to investigate whether the small integral membrane protein Rv3632 may assume this function, thereby explaining its tight functional association with PpgS. Alternatively, the role of Rv3632 may be to target PpgS to the plasma membrane as shown for GTs involved in *O*-polysaccharide synthesis in Gram-negative bacteria (33) or to stabilize PpgS in the membrane as suggested for *M. smegmatis* polyprenyl phosphomannose synthase Ppm1 and associated transmembrane protein Ppm2 (fused as one single polypeptide, Rv2051c, in *M. tuberculosis* H37Rv) (34).

The availability of a *ppgS* knock-out mutant of *M. tuberculosis* now provides a unique opportunity to elucidate the function(s) of the D-GalN*p* substituent of AG. Preliminary studies indicate that the mutant does not display any defects in growth or susceptibility to drugs under optimal laboratory growth conditions. Thus, it is unlikely that D-GalN*p* serves a critical role in stabilizing the structure of the envelope as suggested earlier (3). Instead, the relative restriction of this substituent to pathogenic slow growing mycobacterial species (and emerging nosocomial pathogen, *M. abscessus*) together with the reported role of the D-GalN substituent of lipid A in the pathogenicity of *F. tularensis* (8) make it tempting to speculate that the D-GalN*p* substituent of AG serves a function during host infection. *In vivo* studies have been undertaken to verify this hypothesis.

Acknowledgments—We thank A. Noguera (IPBS) for technical assistance and Dr. M. McNeil (Colorado State University) for helpful discussions.

REFERENCES

1. Acharya, P. V., and Goldman, D. S. (1970) *J. Bacteriol.* **102,** 733–739
2. Draper, P. (1971) *J. Gen. Microbiol.* **69,** 313–324
3. Draper, P., Khoo, K. H., Chatterjee, D., Dell, A., and Morris, H. R. (1997) *Biochem. J.* **327,** 519–525
4. Lee, A., Wu, S. W., Scherman, M. S., Torrelles, J. B., Chatterjee, D., McNeil, M. R., and Khoo, K. H. (2006) *Biochemistry* **45,** 15817–15828
5. Bhamidi, S., Scherman, M. S., Rithner, C. D., Prenni, J. E., Chatterjee, D., Khoo, K. H., and McNeil, M. R. (2008) *J. Biol. Chem.* **283,** 12992–13000
6. Beer, W., Adam, M., and Seltmann, G. (1986) *J. Basic Microbiol.* **26,** 201–204
7. Sonesson, A., Jantzen, E., Tangen, T., and Zähringer, U. (1994) *Microbiology* **140,** 2663–2671
8. Kanistanon, D., Hajjar, A. M., Pelletier, M. R., Gallagher, L. A., Kalhorn, T., Shaffer, S. A., Goodlett, D. R., Rohmer, L., Brittnacher, M. J., Skerrett, S. J., and Ernst, R. K. (2008) *PLoS Pathog.* **4,** e24
9. Wang, X., Ribeiro, A. A., Guan, Z., and Raetz, C. R. (2009) *Biochemistry* **48,** 1162–1172
10. Pfannenstiel, M. A., Muthukumar, G., Couche, G. A., and Nickerson, K. W. (1987) *J. Bacteriol.* **169,** 796–801
11. Chiu, T. H. (1988) *Biochim. Biophys. Acta* **963,** 359–366
12. Song, F., Guan, Z., and Raetz, C. R. H. (2009) *Biochemistry* **48,** 1173–1182
13. Raetz, C. R., Reynolds, C. M., Trent, M. S., and Bishop, R. E. (2007) *Annu. Rev. Biochem.* **76,** 295–329
14. Berg, S., Kaur, D., Jackson, M., and Brennan, P. J. (2007) *Glycobiology* **17,** 35R–56R
15. Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008) *Annu. Rev. Biochem.* **77,** 521–555
16. Kaur, D., Guerin, M., Škovierová, H., Brennan, P. J., and Jackson, M. (2009) *Adv. Appl. Microbiol.* **69,** 23–78
17. Schultz, J., and Elbein, A. D. (1974) *Arch. Biochem. Biophys.* **160,** 311–322
18. Yokoyama, K., and Ballou, C. E. (1989) *J. Biol. Chem.* **264,** 21621–21628
19. Wolucka, B. A., McNeil, M. R., de Hoffmann, E., Chojnacki, T., and Brennan, P. J. (1994) *J. Biol. Chem.* **269,** 23328–23335
20. Pelicic, V., Jackson, M., Reyrat, J. M., Jacobs, W. R., Jr., Gicquel, B., and Guilhot, C. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94,** 10955–10960
21. Scherman, H., Kaur, D., Pham, H., Škovierová, H., Jackson, M., and Brennan, P. J. (2009) *J. Bacteriol.* **191,** 6769–6772
22. Buchmeier, N. A., Newton, G. L., Koledin, T., and Fahey, R. C. (2003) *Mol. Microbiol.* **47,** 1723–1732
23. Korduláková, J., Gilleron, M., Mikušová, K., Puzo, G., Brennan, P. J., Gicquel, B., and Jackson, M. (2002) *J. Biol. Chem.* **277,** 31335–31344
24. Mikušová, K., Yagi, T., Stern, R., McNeil, M. R., Besra, G. S., Crick, D. C., and Brennan, P. J. (2000) *J. Biol. Chem.* **275,** 33890–33897
25. Nigou, J., Vercellone, A., and Puzo, G. (2000) *J. Mol. Biol.* **299,** 1353–1362
26. Ludwiczak, P., Brando, T., Monsarrat, B., and Puzo, G. (2001) *Anal. Chem.* **73,** 2323–2330
27. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) *Nucleic Acids Res.* **37,** D233–D238
28. Nigou, J., Gilleron, M., Cahuzac, B., Bounéry, J. D., Herold, M., Thurnher, M., and Puzo, G. (1997) *J. Biol. Chem.* **272,** 23094–23103
29. Gurcha, S. S., Baulard, A. R., Kremer, L., Locht, C., Moody, D. B., Muhlecker, W., Costello, C. E., Crick, D. C., Brennan, P. J., and Besra, G. S. (2002) *Biochem. J.* **365,** 441–450
30. Dinadayala, P., Kaur, D., Berg, S., Amin, A. G., Vissa, V. D., Chatterjee, D., Brennan, P. J., and Crick, D. C. (2006) *J. Biol. Chem.* **281,** 20027–20035
31. Newton, G. L., Av-Gay, Y., and Fahey, R. C. (2000) *J. Bacteriol.* **182,** 6958–6963
32. Newton, G. L., Ko, M., Ta, P., Av-Gay, Y., and Fahey, R. C. (2006) *Protein Expr. Purif.* **47,** 542–550
33. Clarke, B. R., Greenfield, L. K., Bouwman, C., and Whitfield, C. (2009) *J. Biol. Chem.* **284,** 30662–30672
34. Baulard, A. R., Gurcha, S. S., Engohang-Ndong, J., Gouffi, K., Locht, C., and Besra, G. S. (2003) *J. Biol. Chem.* **278,** 2242–2248