Genetics of Mycobacterial Arabinogalactan and Lipoarabinomannan Assembly

MONIKA JANKUTE,† SHIPRA GROVER,† HELEN L. BIRCH, and GURDYAL S. BESRA
School of Biosciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom

ABSTRACT The cell wall of Mycobacterium tuberculosis is unique in that it differs significantly from those of both Gram-negative and Gram-positive bacteria. The thick, carbohydrate- and lipid-rich cell wall with distinct lipoglycans enables mycobacteria to survive under hostile conditions such as shortage of nutrients and antimicrobial exposure. The key features of this highly complex cell wall are the mycolyl-arabinogalactan-peptidoglycan (mAGP)–based and phosphatidyl-myoinositol–based macromolecular structures, with the latter possessing potent immunomodulatory properties. These structures are crucial for the growth, viability, and virulence of M. tuberculosis and therefore are often the targets of effective chemotherapeutic agents against tuberculosis. Over the past decade, sophisticated genomic and molecular tools have advanced our understanding of the primary structure and biosynthesis of these macromolecules. The availability of the full genome sequences of various mycobacterial species, including M. tuberculosis (3), Mycobacterium marinum (4), and Mycobacterium bovis BCG (5), have greatly facilitated the identification of large numbers of drug targets and antigens specific to tuberculosis. Techniques to manipulate mycobacteria have also improved extensively; the conditional expression-specialized transduction essentiality test (CESTET) is currently used to determine the essentiality of individual genes. Finally, various biosynthetic assays using either purified proteins or synthetic cell wall acceptors have been developed to study enzyme function. This article focuses on the recent advances in determining the structural details and biosynthesis of arabinogalactan, lipoarabinomannan, and related glycoconjugates.

The cell wall of Mycobacterium tuberculosis is unique in that it differs significantly from both Gram-negative and Gram-positive bacteria. The thick, carbohydrate- and lipid-rich cell wall with distinct lipoglycans enables mycobacteria to survive under hostile conditions such as shortage of nutrients and antimicrobial exposure. The key features of this highly complex cell wall are the mycolyl-arabinogalactan-peptidoglycan (mAGP)–based and phosphatidyl-myoinositol–based macromolecular structures, with the latter possessing potent immunomodulatory properties. These structures are crucial for the growth, viability, and virulence of M. tuberculosis and therefore are often the targets of effective chemotherapeutic agents against tuberculosis (TB). Over the past decade, sophisticated genomic and molecular tools have advanced our understanding of the primary structure and biosynthesis of these macromolecules (1, 2). The availability of the full-genome sequences of various mycobacterial species, including M. tuberculosis (3), Mycobacterium marinum (4), and Mycobacterium bovis BCG (5), have greatly facilitated the identification of large numbers of drug targets and antigens specific to TB. Techniques to manipulate mycobacteria have

Received: 18 April 2013, Accepted: 23 July 2013, Published: 8 August 2014
Editors: Graham F. Hatfull, University of Pittsburgh, Pittsburgh, PA, and William R. Jacobs, Jr., Howard Hughes Medical Institute, Albert Einstein College of Medicine, Bronx, NY
Correspondence: G. S. Besra, g.besra@bham.ac.uk
†Authors contributed equally to this work.
© 2014 American Society for Microbiology. All rights reserved.
also improved extensively; the conditional expression-specialized transduction essentiality test (CESTET) is currently used to determine the essentiality of individual genes (6). Finally, various biosynthetic assays using either purified proteins or synthetic cell wall acceptors have been developed to study enzyme function. This article focuses on the recent advances in determining the structural details and biosynthesis of arabinogalactan (AG), lipoarabinomannan (LAM), and related glycoconjugates.

**ARABINOGLACTAN**

**Structural Features of AG**

AG is a major cell wall heteropolysaccharide of mycobacteria. This highly branched structure is covalently attached to peptidoglycan (PG) via a phosphodiester bond to approximately 10 to 12% of the muramic acid residues (7). Collectively, PG and AG form a covalently linked network positioned between the plasma membrane and the mycolic acid layer, resulting in an exceptionally robust cell wall. AG is composed predominantly of arabinose and galactose residues, both in their furanose ring form, that are extremely rare in nature (8). Unlike most bacterial polysaccharides, AG lacks repeating units and is composed of a few distinct structural motifs (9–12).

The whole mycolyl-AG structure is attached to PG via a specific linkage unit, and thus it is believed to be the most vulnerable part of the complex. Detailed characterization of per-O-alkylated oligosaccharide alditols together with gas chromatography–mass spectrometry, fast atom bombardment–mass spectrometry, and nuclear magnetic resonance (NMR) analysis established the detailed structure of AG (Fig. 1).

The galactan domain is composed of approximately 30 alternating β(1→5) and β(1→6) galactofuranosyl (Gal f) residues connected in a linear fashion. At the reducing end of AG, the galactan chain is linked to the C-6 position of selected N-glycolylmuramic acid residues of PG via an α-l-Rha-p-(1→3)-α-d-GlcNAc-1-phosphate linkage unit (11). Three similar d-arabinan chains comprising roughly 30 arabinofuranosyl (Araf) residues each are attached to the C-5 of specific β(1→6) linked-Galf residues (9). Since the AG structure is essential to *M. tuberculosis*, many gene deletion studies investigating AG have been performed in the *Corynebacterium* genus, in which aspects of AG biosynthesis are nonessential. Knockout mutants in *Corynebacterium glutamicum* together with mass spectrometry determined that the arabinan chains of AG are attached distinctively to the 8th, 10th, and 12th residues of the linear galactan chain (13) (Fig. 1). Previous work demonstrated that the arabinan domain is present as a highly branched network built on a backbone of α(1→5) linked sugars with branching introduced by the presence of 3,5-α-d-Araf residues. Further α(1→5) linked Araf sugars are attached subsequent to this branch point, with the nonreducing ends terminating with β(1→2) Araf residues. The final structural motif is a distinct hexa-arabinofuranoside (14), present as [β-α-Araf-(1→2)-α-d-Araf]_3-3,5-α-d-Araf -(1→5)-α-d-Araf. Analysis of per-O-methylated mAGP and per-O-alkylated oligoglycosyl alditols determined that position 5 of both the terminal β-d-Araf and the penultimate 2-α-d-Araf are the attachment sites for the mycolic acids (12). Follow-up studies determined that the mycolyl residues are located in clusters of four on the terminal hexa-arabinofuranoside motifs, with only two-thirds of these being mycolated.

An endogenous arabinase, which can cleave the arabinan, has been partially purified from *Mycobacterium smegmatis* (15). The use of this enzyme together with matrix-assisted laser desorption/ionization–time of flight–mass spectrometry and NMR allowed the sequencing of very large fragments of arabinan chains released from the mycobacterial cell wall. Significantly, galactosamine (GalN) residues, previously detected as a minor covalently bound sugar residue of the cell envelope of slow-growing mycobacteria such as *M. tuberculosis* and *Mycobacterium avium* (16), were shown to be located on the C-2 position of some of the internal 3,5-α-d-branched Araf residues (17), and the stereochemistry of the GalN moiety was confirmed to be an α-anomer (18) (Fig. 1). In addition, the succinyl groups were found on the interior branched arabinosyl residues (19). Approximately one of the three arabinan chains linked to the linear galactan contains a GalN group, and one of three is also succinylated (19). In addition, the succinyl residues were also shown to be present only on the nonmycolated chains. It is speculated that the GalN residue of AG may serve a specific function during host infection (20).

**Precursor Formation**

The biosynthesis of the linkage unit employs two high-energy substrates, UDP-GlcNAc and dTDP-Rha (Fig. 2). UDP-GlcNAc, a sugar donor for both the AG linkage unit and the biosynthesis of PG, is formed via a four-step reaction. Three enzymes (glutamine fructose-6-phosphate transferase, GlmS; phosphoglucomutase, GlmM; and glucosamine-1-phosphate acetyl transferase/
N-acetylglucosamine-1-phosphate uridylyl transferase, GlmU) catalyze the conversion of fructose-6-phosphate to UDP-GlcNAc in *Escherichia coli* (21–24). Analysis of the genome sequence of *M. tuberculosis* determined that the proteins encoded by *Rv3436c*, *Rv3441c*, and *Rv1018c* are homologous to the *E. coli* GlmS, GlmM, and GlmU enzymes, respectively (25). GlmS is responsible for conversion of fructose-6-phosphate to glucosamine-6-phosphate, which is then converted to glucosamine-1-phosphate by GlmM. Recent gene deletion studies demonstrated that *MSMEG_1556*, an *M. smegmatis* gene encoding the homologue of *glmM* from *E. coli*, is essential for survival (26). Furthermore, it was shown that *M. tuberculosis Rv3441c* possesses phosphogluco-6-phosphate mutase activity and was able to compensate for the loss of *MSMEG_1556* in the conditional mutant (26).

FIGURE 1 Schematic structure of mycobacterial arabinogalactan (AG). A linkage unit, composed of rhamnose and N-acetyl-glucosamine residues, anchors the whole AG structure to peptidoglycan. The galactan domain is composed of alternating β(1→5) and β(1→6) galactofuranose residues with three chains of arabinan attached to each linear galactan chain at positions 8, 10, and 12. The highly branched nonreducing end of AG terminates with a hexa-arabinofuranoside motif, two-thirds of which is substituted with mycolic acids. doi:10.1128/microbiolspec.MGM2-0013-2013.f1
thus demonstrating that they share the same function. Mycobacterial GlmU is a bifunctional enzyme involved in the last two sequential steps of UDP-GlcNAc synthesis (Fig. 2). Disruption of glmU in M. smegmatis resulted in gross morphological changes and loss of viability (25). Biochemical characterization as well as the structure of M. tuberculosis GlmU has recently been established (27–29).

The second nucleotide donor utilized in the biosynthesis of the linkage unit is dTDP-Rha (Fig. 2). The presence of l-rhamnose, a sugar absent in humans, makes the biosynthetic machinery of the mycobacterial linkage unit an attractive drug target. As a result, the rhamnosyl biosynthetic pathway has come under close scrutiny, and hence a number of inhibitors targeting this pathway have been described (30–32). Synthesis of dTDP-Rha occurs via a linear four-step reaction. Recognition of the genes involved in this pathway was revealed by comparison to known polysaccharide biosynthetic enzymes found in other bacteria, namely E. coli (33). RmlA (Rv0334) sets in motion a sequence of reactions, converting dTTP and α-D-glucose-1-P into dTDP-glucose. A strain of E. coli lacking four rhamnose biosynthetic genes was complemented with rmlA from M. tuberculosis. Analysis of cellular extracts revealed an abundance of α-D-Glc-P thymidylyltransferase activity, thus confirming its proposed function (33). The product of RmlA activity is then shuttled through three sequential reactions catalyzed by dTDP-D-glucose-4,6-dehydratase RmlB (Rv3464), dTDP-4-oxo-6-deoxyglucose-3,5-epimerase RmlC (Rv3465), and dTDP-6-deoxy-L-lyxo-4-hexulose reductase RmlD (Rv3266) to finally form the nucleotide donor dTDP-Rha. Gene deletion studies in the presence of a rescue plasmid with a temperature-sensitive origin of replication determined that the rmlA (34), rmlB (35), rmlC (35), and rmlD (36) genes were all essential for M. smegmatis. Hence, dTDP-Rha is an essential sugar donor for mycobacterial growth, and enzymes involved in its synthesis are potential chemotherapeutic targets. Finally, enzyme assays employing RmlA-D from M. tuberculosis to screen inhibitors for developing novel anti-TB therapeutics have been established (30, 37).

FIGURE 2 The biosynthesis of sugar donors required for mycobacterial AG biosynthesis. Both UDP-GlcNAc and dTDP-Rha are utilized in the formation of the linkage unit. UDP-Galf is the sugar donor of the galactofuranosyl residues used in the galactan chain formation. Decaprenylphosphoryl-d-arabinofuranose (DPA) is the only known high-energy nucleotide providing arabinofuranosyl residues to the arabinan domain of AG. doi:10.1128/microbiolspec.MGM2-0013-2013.f2
The high-energy nucleotide substrate UDP-Galf is formed via a three-step reaction (Fig. 2). GalU (Rv0993), a glucose-1-phosphate uridylyltransferase, catalyzes the formation of UDP-GlcP from UTP and glucose-1-P (38). Recently, galU from *M. tuberculosis* was successfully expressed, purified, and biochemically characterized (38). The second enzyme, GalE, is responsible for the epimerization reaction, which forms UDP-Galp from UDP-GlcP in *E. coli* (39). Studies in *M. smegmatis* examined the reverse reaction using radiolabeled UDP-Galp and observed UDP-glucose-4-epimerase activity. Sequentially, the *M. smegmatis* protein was purified, and its N-terminal sequence was shown to be similar to that of the *M. tuberculosis* product GalE1 (Rv3634c) (40). The further conversion of UDP-Galp to the furanose form occurs via a ring contraction catalyzed by the enzyme UDP-galactopyranose mutase Glf, which was identified initially in *E. coli* (41) and subsequently in *M. smegmatis* (MSMEG_6404) and *M. tuberculosis* (Rv3809c) (40). Allelic exchange experiments highlighted the essentiality of *glf* to *M. smegmatis* (42). In addition, the crystal structures of Glf mutases from *M. tuberculosis*, *Escherichia coli*, and *Klebsiella pneumoniae* have been solved (43, 44).

Arabinan biosynthesis utilizes β-Δ-arabinofuranosyl-1-monophosphodecaprenol (DPA), the only known donor of Ara residues in mycobacteria and corynebacteria (45). Recently, its membrane-linked synthesis was investigated in detail (13, 46) (Fig. 2). The initial reaction involves activation of ribose-5-phosphate by a phosphoribosyl-1-pyrophosphate synthetase, PrsA (Rv1017c), to yield 5-phosphoribosyl-1-pyrophosphate (pRpp) (46). UbiA (Rv3806c) then transfers pRpp to a decaprenylmonophosphate, producing decaprenylphosphoryl-5-phosphoribosyl (DPPR) (46). Disruption of *ubiA* (NCgl2781) in *C. glutamicum* resulted in a complete loss of cell wall arabinan, demonstrating that DPA is the only Ara sugar donor used in AG biosynthesis (13). Remarkably, the mutant still generated a modified LAM version, which was arabinosylated even in the absence of DPA. An alternative source and mechanism by which these Ara residues are added to this glycolipid is yet to be resolved (47). DPPR is then dephosphorylated to decaprenyl-5-phosphoribose (DPR) by the putative phospholipid phosphatase encoded by *Rv3807c*. Its homologue in *M. smegmatis* (MSMEG_6402) was shown to be a nonessential gene (48). The DprE1 (Rv3790) and DprE2 (Rv3791) heterodimer catalyzes the epimerization of DPR to DPA, which occurs via an oxidation-reduction mechanism. DPR is initially oxidized at the C2-OH group to form the keto-sugar intermediate decaprenol-1-monophosphoryl-

2-keto-β-erythro-pentofuranose (DPX), which is subsequently reduced to DPA (49). Deletion studies in *C. glutamicum* showed that *dprE1* (NCgl0187) is essential to bacterial growth, whereas *dprE2* (NCgl0186) is not (50). In the absence of *dprE2*, a different enzyme encoded by NCgl1429 was proposed to carry out the function of DprE2 since NCgl1429 showed a similar function in vivo and appeared to be essential in the NCgl0186-inactivated mutant (50). Further investigation demonstrated that *dprE1* (MSMEG_6382) is also an essential gene in *M. smegmatis* (51). These results highlighted DprE1 as a novel drug target. Indeed, recent studies led to the discovery of two classes of potent compounds with specific activities against mycobacteria: dinitrobenzamide derivatives (DNBs) and nitrocompounds related to DNBs—nitro-benzothiazinones (BTZs)—both of which were revealed to target the decaprenylphosphoryl-β-D-ribose 2'-epimerase encoded by *dprE1* (52, 53). The structural complex of DprE1-BTZ has been determined, revealing the mode of inhibitor binding (52, 54).

**Biosynthesis of AG**

The biosynthesis of AG begins with the formation of the linkage unit synthesized on a decaprenyl phosphate (C50-P) lipid carrier. WecA (Rv1302) catalyzes the first reaction by transferring GlcNac-1-P from the sugar donor UDP-GlcNac to the lipid carrier (55, 56). Lipopolysaccharide analysis of a *wecA*-defective strain of *E. coli* complemented with either *M. tuberculosis* (Rv1302) or *M. smegmatis* (MSMEG_4947) homologue showed restoration of lipopolysaccharide biosynthesis, thus providing evidence that it has the same function as the WecA protein from *E. coli* (55). In addition, inactivation of *wecA* from *M. smegmatis* using a homologous recombination strategy resulted in drastic morphological changes and loss of viability (55). Rhamnosyltransferase WbbL (Rv3265c) is responsible for the transfer of the rhamnose residue from the dTDP-Rha substrate to the 3-position of the GlcNac of C50-P-GlcNac, thus yielding the full linkage unit C50-P-P-GlcNac-Rha of AG. The key to the discovery of mycobacterial WbbL was the successful complementation of an *E. coli* mutant lacking WbbL activity with the *Rv3265c* gene from *M. tuberculosis* (57). *M. tuberculosis wbbL* was expressed in *E. coli* and was used together with bioinformatics analysis to establish its preliminary structure and characteristics (58). Moreover, it was demonstrated that *wbbL* (MSMEG_1826) is crucial to the growth and viability of *M. smegmatis* (57).
The previously synthesized linkage unit serves as an acceptor for the addition of Galf residues from the sugar donor UDP-Galf. GlfT1 (Rv3782) recognizes the linkage unit and transfers the initial two Galf residues to C₅₀-P-P-GlcNAc-Rha, resulting in C₅₀-P-P-GlcNAc-Rha-Galf₂ (59–61) (Fig. 3). Further galactan polymerization is carried out by the second transferase GlfT2 (Rv3808c), identified through the use of a neoglycolipid acceptor assay together with UDP-Galf and isolated E. coli membranes expressing glfT2. It was demonstrated that the enzyme had dual functionality, acting as both a UDP-Galf:β-D-(1→5) galactofuranosyltransferase (GalfT) and the UDP-Galf:β-D-(1→6) GalfT (62–64). Structural data together with site-directed mutagenesis and kinetic studies provided evidence for a mechanism that explains the unique ability of GlfT2 to generate β(1→5) and β(1→6) linkages using a single active site (64).

An arabinofuranosyltransferase (ArafT) from the emb locus, AftA (Rv3792), is responsible for addition of the first key Araf residue to the 8th, 10th, and 12th Galf residues, thus “priming” the galactan chain for further attachment of α(1→5)-linked Araf units (13, 65) (Fig. 3). A homologue of aftA in M. smegmatis, MSMEG_6386, was shown to be essential for survival of mycobacteria. However, deletion of aftA in C. glutamicum resulted in a slow-growing but viable mutant. Cell wall analysis revealed the complete loss of arabinose leading to a truncated cell wall structure.

**FIGURE 3** Schematic representation of mycobacterial arabinogalactan biosynthesis. WecA catalyzes the transfer of GlcNAc to decaprenyl phosphate, which is then used as an acceptor for addition of rhamnosyl residue by WbbL, thereby forming the full linkage unit. The first two galactofuranosyl (Galf) residues are added to the linkage unit via GlfT1. The bifunctional GlfT2 adds the remaining Galf residues forming a linear galactan chain. Before the polymerization with arabinofuranosyl (Araf) residues, the galactan domain is thought to be translocated across the plasma membrane by the unknown flippase. AftA initiates the transfer of Araf residues from the sugar donor DPA to the 8th, 10th and 12th β(1→6)-linked Galf residues of the galactan chain. EmbA and EmbB proteins act as α-1,5-arabinosyltransferases utilizing the same nucleotide donor DPA. The 3,5-linked Araf branching is introduced by AftC and AftD enzymes. Finally, the terminal Araf residues are added to the arabinan domain by a “capping” enzyme AftB.

containing only a galactan chain and greatly diminished cell wall-bound mycolic acids (65). EmbA (Rv3794) and EmbB (Rv3795) catalyze further polymerization of arabinan. Emb proteins were first discovered as a target for ethambutol (EMB), a first-line TB drug. Individual inactivation of embA and embB in M. smegmatis resulted in diminished incorporation of arabinin into AG, specifically the terminal disaccharide β-D-Araf-(1→2)-α-D-Araf, normally situated on the 3-OH of the 3,5-linked Araf residue (66). Efforts to generate a viable embA/embB mutant in M. tuberculosis and an embAB double mutant in M. smegmatis have so far proven unsuccessful, highlighting their essentiality to mycobacteria. However, a singular emb gene (NCgl0184) was successfully disrupted in C. glutamicum (13). Corynebacterium is deemed the archetype of Corynebacteriaceae since it maintains a low frequency of gene duplications and modifications, and thus it is reasonable that C. glutamicum possesses only one emb gene. Surprisingly, NCgl0184 exhibited higher identity to embC, which encodes ArafT, involved exclusively in LAM biosynthesis, even though C. glutamicum lacks an elaborately arabinosylated lipomannan (LM) product (47, 67).

Chemical analyses of the tolerable emb deletion mutant from C. glutamicum revealed an almost total loss of cell wall arabinan, except for terminal t-Araf residues decorating the galactan backbone (13). Moreover, EMB treatment of wild-type C. glutamicum produced a profile identical to that of the mutant, illustrating that emb is indeed the target of EMB. Recent deletion studies in M. smegmatis identified a branching enzyme, AftC (Rv2673), that is responsible for the transfer of Araf residues from DPA to the arabamin domain to form α(1→3)-linked Araf residues of the internal arabamin domain at the nonreducing end of AG and LAM (68, 69). Yet another functional ArafT encoded by aftD (Rv0236c) has been shown to have α,1,3-branaching activity on linear α,1,5-linked synthetic acceptors in vitro. Inactivation of the aftD homolog in M. smegmatis, MSMEG_0339, was shown to be lethal to mycobacteria, while overexpression of aftD in M. smegmatis resulted in an overall increase of Araf residues (70). Finally, AftB (Rv3805c) catalyzes the transfer of Araf residues to the arabamin domain to form the terminal β(1→2)-linked Araf residues (Fig. 3). Disruption of aftB in C. glutamicum resulted in a viable mutant with complete absence of terminal β(1→2)-linked arabinofuranosyl residues and decreased abundance of cell-wall-bound mycolic acids, consistent with a partial loss of mycolylation sites (71).

**Decoration of AG and Its Attachment to PG and Mycolic Acids**

Two decorating structures, succinyl and GalN residues, have been identified in the interior AG-arabinan domain of M. tuberculosis, thus concluding a model of the complete primary structure of mycobacterial AG. Enzymes involved in succinylation of arabinan chains toward the nonreducing end are yet to be determined (19), but the key components of the biosynthetic pathway of GalN have recently been elucidated (20). Glycosyltransferase PpgS (Rv3631) catalyzes the transfer of GalNAc from UDP-GalNAc to polypropenyl-P, yielding a sugar donor polypropenyl-P-GalNAc. This high-energy substrate is then presumably deacylated by an as yet unknown decacylase before or after being translocated to the extracellular space where the membrane-associated enzyme Rv3779 transfers the GalNp (or GalNAc) residue to the C2 position of a portion of the internal 3,5-branched β-Araf residues of AG. The synthesis of GalN was demolished in both ppgS and Rv3779 deletion mutants in M. tuberculosis. It is worth noting that the GalN residue is only found in slow-growing mycobacteria. Hence, expression of ppgS in the fast-growing M. smegmatis species, otherwise devoid of the ppgS orthologue and any detectable polypropenyl-P-GalNAc synthase activity, allowed mycobacteria to synthesize polypropenyl-P-GalNAc in vivo. The physiological role and pathogenesis of both succinyl and GalN residues as well as the biosynthetic origin of succinylation remain to be elucidated.

Very little evidence has been obtained that shows how AG is ligated to PG to generate the complete cell wall core. In vitro assay in M. smegmatis utilizing cell-free extracts and radiolabeled substrates demonstrated the formation of simpler polypropenyl-P-P-GlcNAc-Rha-(Gal)_n intermediates, followed by addition of AG and, finally, ligation to PG (72). However, there are no enzymes reported to show the attachment of AG to PG. In vitro enzymatic assays have identified members of the antigen 85 complex, FbpA, FbpB, and FbpC (Rv3804c, Rv1886c and Rv0129c, respectively), that are responsible for the transfer of mycolic acids onto trehalose that leads to the formation of trehalose monomycolate (TMM) and trehalose dimycolate (TDM) (73). Inactivation of antigen 85 by transposon mutagenesis resulted in a mutant with reduced capacity to transfer mycolic acids to the mycobacterial cell wall (74). Similar mycolyltransferases from C. glutamicum, encoded by cmnytA and cmnytB genes, were deleted, leading to a viable double mutant with significantly impaired ability to transfer corynomycolates to AG (75).
PHOSPHATIDYL-MYO-INOSITOL BASED GLYCOLIPIDS

The plasma membrane of mycobacteria is rich in free lipids that play a crucial role in its pathogenesis. The phospholipids—phosphatidyl ethanolamine, phosphatidyl-myo-inositol (PI), phosphatidyl inositol mannosides (PIMs), LM, LAM, and cardiolipid (CL)—are among the major structural components of the mycobacterial plasma membrane, with PI/PIMs alone representing 28% of the total phospholipids in M. smegmatis (76). PIM, LM, and LAM are glycolipids built on a PI backbone by a series of modifications including glycosylation and acylation. It is still unclear whether these PI-based glycolipids are embedded in the plasma membrane or found in the outer membrane of mycobacteria. However, using surface labeling experiments it has recently been shown that the lipoglycans are exposed at the surface of mycobacteria, indicating their presence in the outer leaflet of the outer membrane (27). The surface exposed lipoglycans, PIMs, LM, and LAM possess significant immunomodulatory effects in macrophages, including cytokine production, inhibition of phagosome maturation, and apoptosis and also account for cross-protective immunity of mycobacteria (2).

Recent biochemical and genetic studies have indicated that the modifications of the PI anchor follow the order PI → PIM → LM → LAM (2), wherein mannosylation of the PI anchor with up to six mannose residues produces PIMs that are further modified with mannose and arabinose residues to generate LM and LAM. The enzymes for biogenesis of early PIM species are encoded by a gene cluster consisting of an operon of six open reading frames (ORFs), conserved in the Corynebacterineae family (3). These include Rv2614c, which encodes a probable threonyl-tRNA synthase. The second gene, Rv2613c, encodes a protein of unknown function that is predicted to be involved in nucleotide biosynthesis. The third gene of the gene cluster is Rv2612c, which encodes PgsA, a protein involved in PI synthesis, and the fourth gene, Rv2611c, encodes an acyltransferase (78) that acylates both PIM1 and PIM2. The fifth gene, Rv2610c, codes for PimA, an α-mannopyranosyltransferase of the GT-B superfamily and the first enzyme of the PIM biosynthetic pathway responsible for PIM1 production (79). Rv2609c encodes a putative guanosine diphosphate (GDP)-Manp hydrolase with a conserved MutT domain (80). The genes encoding proteins involved in biogenesis of higher PIMs, LM, and LAM are all found scattered in the genome. Interestingly, AraFTs such as EmbA, EmbB, and EmbC are found in the embCAB operon (81, 82) in mycobacteria; of these, only EmbC is employed in biosynthesis of LAM (83), while EmbA and EmbB are specific for AG biosynthesis (66). The core framework of the mycobacterial cell wall provides a template for the insertion of mannosylated molecules such as LAM and its structurally related glycolipids, LM, and PIMs (84). These highly complex immunomodulatory lipoglycans are found ubiquitously in the envelopes of all mycobacterial species, noncovalently associated to the plasma membrane and/or the mycolic acid layer via a conserved mannosyl-phosphate inositol (MPI) anchor (85), which extends to the exterior of the cell wall (84, 86).

Structural Features of PIMs

The PI unit is based on sn-glycero-3-phosphate-(1-α-myo-inositol), whereby the glycerol phosphate component is attached to the L-1-position of myo-inositol (87, 88). The PI unit is sequentially substituted with mannose residues at positions C-2 and C-6 of the inositol ring, thus forming the MPI anchor (89, 90). This anchor is highly heterogenous with respect to the number, location, and nature of the fatty acids attached. There are a total of four potential acylation sites on the MPI anchor, of which two are present on the glycerol unit, one is on the Manp unit linked to C-2 of myo-inositol, and the fourth is at the C-3 position of myo-inositol (91, 92). Of the initial PIMs, the PI in PIM1 is glycosylated with α-α-mannopyranosyl (Manp) at position O-2, whereas PIM2 is glycosylated with Manp at both O-2 and O-6. The terminal PIM species PIM6 consists of Manp-α(1→2)-Manp-α(1→2)-Manp-α(1→6)-Manp-α(1→6)-Manp-α(1→) attached to the MPI anchor (49) (Fig. 4). Acylated forms of the PIMs, such as Ac1PIM2 and Ac1PIM6, are the major PIM species, with acylation occurring at the Manp residue attached to C-2 of myo-inositol. The diacyl forms of PIMs, such as Ac2PIM2 and Ac2PIM6, also exist, but Ac1PIM2 is believed to be the preferred precursor for building higher PIMs, such as PIM6, LM, and LAM (93).

Structural Features of LM and LAM

The mannan of LM and LAM is an extension of PIMs, containing on average 20 to 30 residues, emanating from the C-6 position of the inositol ring (92, 94) (Fig. 4). The α(1→6)-mannose residues are linked linearly to form a linear backbone with branching introduced by 5 to 10 units of single α(1→2) Manp residues at C-2 of the occasional α(1→6)-linked mannose in all mycobacterial species. However, in the case of Mycobacterium chelonae, the branching is introduced by single α(1→2) Manp residues at the C-3 position (93, 95). The arabinian domain of LAM consists of 55 to 70 Araf residues (92) arranged as a
linear α(1→5)-linked arabinosylfuranosyl backbone with branching introduced at the third position on some residues (69, 96) (Fig. 4). The branches in the arabinan domain of LAM consist of two distinct motifs: a linear tetra-arabinoside comprising β-D-Araf(1→2)-α-D-Araf(1→5)-α-D-Araf(1→5)-α-D-Araf and a hexa-arabinoside comprising [β-D-Araf(1→2)-α-D-Araf]2-3,5-α-D-Araf(1→5)-α-D-Araf (94, 97). Both tetra- and hexa-arabinoside moieties are decorated with a disaccharide unit, Araf-β(1→2)-Araf at the nonreducing end (14, 94, 97). Mycobacterial species differ in the nature and extent of the capping motifs modifying the nonreducing termini of the arabinan chains, specifically the β(1→2)-linked terminal Araf units. To date, three structural families have been recognized: mannose-capped LAM (Man-LAM), PI-capped LAM (PI-LAM), and noncapped LAM (Ara-LAM), of which “Man-caps” are an important feature of pathogenic species of slow-growing mycobacteria such as M. tuberculosis, M. bovis, M. bovis BCG, M. leprae, M. avium, M. xenopi, M. marinum, and M. kansasii (97).

The caps may be present as single Manp capping residues —di- or tri-α(1→2)-mannosides, of which dimannosides predominate (93, 97). Fast-growing mycobacteria such as M. smegmatis and M. fortuitum possess PI caps (98), and M. chelonae is the only known example of Ara-LAM (95). In addition to the caps, NMR studies have demonstrated the presence of succinyl residues on the C-2 of the 3,5-di-α-D-Araf units in Man-LAM isolated from M. bovis BCG (Pasteur, Glaxo, Copenhagen, and Japanese strains). The average number of succinyl residues varies from one to four per LAM molecule (99).

**Biosynthesis of PIMs, LM, and LAM**

**Precursor formation**

In mycobacteria, mannose is the key component of glycolipids, lipoglycans, and a number of glycosylated proteins, which makes it essential for growth and viability. The polymethylated polysaccharides are the components that contain mannose in mycobacteria (100).
Mycobacteria obtains mannose from two sources: first, the extracellular environment, from which mannose is obtained via sugar transporters and is phosphorylated with the aid of hexokinase (Rv2702) (101) to yield mannose-1-phosphate; and second, via the glycolytic pathway, where fructose-6-phosphate is converted to mannose-6-phosphate by a phosphomannose isomerase, ManA (Rv3255c) (102) (Fig. 5). Subsequently, mannose-1-phosphate is converted to mannose-6-phosphate in a reaction catalyzed by a phosphomannomutase, ManB (Rv3257c) (103). This mannose-1-phosphate generated from mannose obtained from both sources is loaded onto GDP to generate the mannose donor, GDP-Manp, by ManC (30, 104). GDP-Manp serves as an intracellular nucleotide-derived mannose donor for the synthesis of several glycolipids and mannosylated proteins by the GT-A/B superfamily of glycosyltransferases (105).

The GT-C superfamily of glycosyltransferases uses polyprenyl-phosphate-based mannose donors for peri-plasmic synthesis of lipoglycans, such as higher PIMs, LM, and LAM. The presence of a C50-polyprenol-based mannolipid, C50-decaprenol-phospho-mannose (C50-P-Manp, polyrenol monophosphomannose [PPM]), in M. tuberculosis was first reported by Takayama and Goldman in 1970 (106). However, in M. smegmatis, a C35-octahydroheptaprenyl-phospho-mannose, C35-P-Manp, was later identified by Wolucka and Hoffmann in 1998 (107).

In mycobacteria, PPM is synthesized using GDP-Man and C35/C50-P in a reaction catalyzed by a PPM synthase, Ppm1 (Rv2051c) (108) (Fig. 5). However, more recently, a transmembrane glycosyltransferase, Rv3779, was identified and suggested to be involved in the synthesis of C35/50-P-Manp as a second PPM synthase (109). However, in a study conducted by Skovierova et al. (20), Rv3779 was found to have glycosyltransferase activity, and it was suggested to be involved in transferring GalN from a polyprenyl-phospho-N-acetylgalactosamine to AG in M. tuberculosis. The roles of Rv2051c and Rv3779 were recently reinvestigated by Rana et al. (110) using genetic and biochemical methods in M. smegmatis. This study revealed that a conditional mutant of ppm1 generated using CESTET in M. smegmatis could only be rescued in the presence of plasmid encoding Rv2051c, while the second PPM synthase (Rv3779) failed to substitute for...
the loss of gene function \textit{in vivo}. Therefore, \textit{ppm1} (\textit{Rv2051}) is the sole gene responsible for generating the mannose donor PPM in \textit{M. tuberculosis}.

**Biosynthesis of PI**

The PI moiety plays a dual role, because it forms the backbone for synthesis of PIMs, LM, and LAM and also anchors the high-molecular-weight lipoglycans LM and LAM to the inner and outer membranes in the cell envelope. Synthesis of PI is a three-step process that begins with cyclization of glucose-6-phosphate by inositol phosphate synthase (InO1), encoded by \textit{Rv0046c}, to generate myo-inositol-1-phosphate, which is subsequently dephosphorylated by inositol mono-phosphatase (IMP) to yield myo-inositol (111, 112) (Fig. 5). In the last step, myo-inositol is esterified to diacylglycerol (DAG), which is transferred from cytidine diphosphatadiacylglycerol (CDP-DAG) by a PI synthase, \textit{PgsA} (\textit{Rv2612c}). Recently, gene mutational studies in \textit{M. tuberculosis} and \textit{M. smegmatis} have revealed that the inositol phosphate synthase (Ino1) and the PI synthetase (PgsA) are essential for the viability of mycobacteria (111, 113). Interestingly, ImpC (\textit{Rv3137}), an inositol monophosphatase, has been demonstrated to be essential for the growth of \textit{M. tuberculosis} and \textit{M. smegmatis} (114).

**Biosynthesis of PIMs**

PIM biogenesis follows a linear pathway that consists of a series of mannosylations of the PI anchor in the order PI → PIM$_2$ → PIM$_4$ → PIM$_6$ (Fig. 6). The higher PIM species have prominent structural and physiological roles; PIM$_4$ forms the structural basis for LM and LAM, while PIM$_6$ is required to maintain the integrity of the plasma membrane. The early steps of PIM biogenesis occur on the cytoplasmic side of the plasma membrane and utilize GDP-Manp as the sugar donor. The pathway for PIM biogenesis is initiated by transfer of Manp from GDP-Man to position O-2 of the inositol ring of PI to yield PIM$_1$. This reaction is catalyzed by PimA (\textit{Rv2601}), a GDP-Man-dependent $\alpha$-mannosyltransferase of the GT-B superfamily of glycosyltransferases (78–80). Recent biochemical and genetic studies conducted in \textit{M. smegmatis} have demonstrated PimA to be an essential enzyme required for growth of mycobacteria.

The next step in the pathway is generation of the metabolic intermediate PIM$_2$, which serves as a scaffold on which higher PIMs, LM, and LAM are synthesized (4, 29). PIM$_2$ can exist in mono- and diacylated forms in \textit{M. tuberculosis}, while only the monoacylated form is observed to accumulate in corynebacteria (78). Two separate hypotheses have been suggested for the

**FIGURE 6** Biosynthesis of PIMs. The PI anchor synthesized by \textit{PgsA} undergoes multiple mannosylations to produce PIMs. Of the PIM species synthesized, AcPIM$_2$ and AcPIM$_4$ are the most abundant. The production of AcPIM$_4$ serves as the branch point in the PIM biosynthesis, with one branch leading to formation of higher PIM species such as AcPIM$_6$ and the other leading to LM and LAM production. The AcPIM$_4$ is synthesized by the mannosyltransferases PimC and/or PimD, both of which remain unidentified. The flipase required for translocating the AcPIM$_4$ from the cytosolic to extracellular side is also unknown. doi:10.1128/microbiolspec.MGM2-0013-2013.f6

---

**ASMscience.org/MicrobiolSpectrum**
formation of AcPIM2, which is a two-step reaction catalyzed by a mannosyltransferase and an acyltransferase. According to the first hypothesis, PIM1 is first acylated by Rv2611 at the sixth position of the Manp residue to yield Ac1/2PIM1, which is then subsequently mannosylated by PimB’ at position O-6 of the inositol ring to form AcPIM2 (78, 80). This model has been supported by recent studies conducted in C. glutamicum, where the gene deletion mutant of pimB’ (Rv2188c), α-δ-mannose-α-(1→6) phosphatidylinositol-mannopyranosyltransferase, which adds Manp to inositol to yield Ac/2PIM2, was found to accumulate AcPIM1, implying precedence over the acylation of PIM1 and the second mannosylation step (78, 115, 116). The second hypothesis suggests that acylation of PIM1 and PIM2 can occur together. However, the cell-free experiments conducted in M. smegmatis favor generation of AcPIM2 after PIM2 has been formed by the action of PimB’ (117). The acyltransferase involved in the transfer of the fourth acyl group to the third position of myo-inositol is yet to be identified.

The gene Rv2611c is required for growth of mycobacteria, because its disruption causes severe growth defects and the mutant accumulates nonacylated PIM1 and PIM2 (78). The enzyme PimB’, encoded by Rv2188c, is also an essential enzyme in M. smegmatis (118). The function of Rv2188c was initially assigned to Rv0557, earlier designated as PimB. However, genetic experiments in M. tuberculosis identified no changes to PIM biosynthesis due to disruption of the gene. The results obtained suggested that the role of PimB could either be substituted by complementary genes or that the function of PimB is different from that of PimB’ (119).

In addition, recent investigations in C. glutamicum have identified Rv2188c as the mannopyranosyltransferase involved in the second mannosylation step for generation of AcPIM2 (115, 116, 120) and caused Rv0557 to be renamed as MgtA, due to the α-mannopyranosylglucuronic acid transferase activity responsible for production of Cg-LM-B, a glucuronic acid–based LM variant, and ManGlcAGroAc2, a glucuronic acid diacylglycerol–based glycolipid found in C. glutamicum (120).

The next major step in the pathway is synthesis of AcPIM4; however, intermediary PIM forms such as Cg/2PIM3 and Ac/2PIM3 can also exist in mycobacteria. PimC (RvD2-ORF1) is a mannosyltransferase identified in M. tuberculosis strain CDC1551 and is responsible for synthesis of trimannosylated PIMs, but no strong homologues of pimC have been found in M. tuberculosis H37Rv or M. smegmatis. Additionally, M. bovis BCG with a deleted chromosomal copy of pimC showed normal PIM, LM, and LAM levels, suggesting redundancy of the gene or the presence of compensatory pathways (121). The nonreducing end of Ac1/2PIM3 is mannosylated by an α(1→6) mannosyltransferase to form Ac/2PIM4. This reaction is catalyzed either by PimC or by the unidentified putative “PimD” protein (122). Formation of Ac/2PIM4 marks the “junction” point as the pathway now diverges into two branches, with one branch leading to the formation of polar PIM species, such as Ac1/2PIM3 and Ac1/2PIM6, by two consecutive additions of α(1→2) Manp, probably by PimE (Rv1159), and the second branch leading to the formation of LM and LAM (123, 124). The intermediate Ac1/2PIM4 is translocated across the membrane to the extracellular space using an unknown flippase for subsequent glycosylation. The glycosyltransferases involved in further modifications of Ac1/2PIM4 employ polyprenylphosphate-based sugar donors, PPM and DPA, and belong to the GT-C superfamily (105). PimE, an α(1→2)-mannopyranosyltransferase, is an example of a GT-C superfamily glycosyltransferase, which utilizes PPM for addition of one or more mannose residues to Ac1/2PIM4 (124). Although the steps for biosynthesis of the final product Ac1/2PIM6 have yet to be deduced, involvement of PimE cannot be ruled out.

Biosynthesis of LM and LAM
The lipoglycans LM and LAM are generated by extensive glycosylation of the PIMs, specifically, Ac1/2PIM4 (Fig. 7). Recent studies conducted in M. smegmatis indicated the involvement of LpqW (Rv1166) in LM and LAM biogenesis (125, 126). The protein LpqW has a regulatory function where it promotes the channeling of Ac1/2PIM4 into LM and LAM biosynthesis by interacting with MptB, a mannosyltransferase involved in synthesis of the mannan core of LM and LAM (125, 127). The mannan of LM/LAM is composed of approximately 25 to 30 mannose residues associated linearly to form the α(1→6) mannan backbone that is punctuated occasionally at C-2 with single α(1→2)-linked mannose units in M. tuberculosis, M. leprae, and M. smegmatis. However, in the case of M. chelonae the mannan core is substituted at C-3 with single Manp units. Addition of mannose residues to build the mannan core in LM and LAM is mediated by mannosyltransferases of the GT-C superfamily that utilize PPM as the mannose donor. The α(1→6) mannosyltransferases MptA (Rv2174) and MptB (Rv1459c) are the key enzymes, which synthesize the
characteristic mannan core of LM and LAM. Recent studies in C. glutamicum suggested the involvement of MptB in the synthesis of the proximal end through the addition of 12 to 15 Man residues to the backbone \((116)\) and designated MptA to be responsible for the synthesis of the distal end of the core \((118, 128)\). However, M. tuberculosis and M. smegmatis MptB proteins were unable to complement the C. glutamicum \(\Delta mptB\) mutant, indicating that either differences in substrate specificity or redundancy of the gene function was responsible for the contradictory observations.

The \(\alpha(1\rightarrow 6)\)-mannan core is further branched by the addition of \(\alpha(1\rightarrow 2)\)-Manp residues, which are catalyzed by MptC \((Rv2181)\) \((2)\). Kaur et al. \((129)\) recently identified the putative integral membrane proteins, MSMEG_4250 in M. smegmatis and Rv2181 in M. tuberculosis, as potential polyprenol-dependent glycosyltransferases based on shared characteristics with previously identified enzymes. A knockout of MSMEG_4250 in M. smegmatis possessed a truncated version of LAM with a decrease in the number of \(\alpha(1\rightarrow 2)\)-Manp branching residues and altered growth with an inability to synthesize LM. Complementation of the mutant with the corresponding orthologue of M. tuberculosis \((Rv2181)\) restored normal LM/LAM synthesis. However, regulation of LM and LAM biosynthesis in M. smegmatis appears to differ somewhat with M. tuberculosis, as M. tuberculosis \(\Delta Rv2181\) produced truncated versions of LM and Man-LAM \((96)\).

LAM is generated via the further elaboration of LM with 55 to 70 Araf units forming an arabinan domain akin to that found in AG \((84)\). Addition of arabinofuranosyl residues to LM is mediated by the AraTAls: EmbC \((Rv3793)\), AftC \((Rv2673)\), AftD \((Rv0236c)\), and an uncharacterized AraT that is considered to prime LM for LAM synthesis \((65)\). The transition of LM to LAM occurs by the addition of 12 to 16 \(\alpha(1\rightarrow 5)\)-Araf residues to the primed LM, a reaction exclusively catalyzed by EmbC, an \(\alpha(1\rightarrow 5)\)Araf transferase \((69, 130, 131)\). Recently, the crystal structure of the truncated C-terminal domain of EmbC was solved \((130)\). The linear Ara polymer of LAM is branched.

**FIGURE 7** Biosynthesis of LM and LAM. Hyperglycosylation of AcPIM\(_4\) produces LM and LAM. The mannosyltransferases MptB, MptA, and MptC are involved in the synthesis of the mannan core, while the arabinosyltransferases EmbC, AftC, and AftD and an unknown transferase are responsible for the synthesis of the arabinan domain. The arabinan in LAM is capped with mannose residues in M. tuberculosis at the nonreducing termini referred to as Man-LAM. The enzymes, MptC and CapA, mediate this reaction. doi:10.1128/microbiolspec.MGM2-0013-2013.f7
similarly to AG by addition of α(1→3) Araf residues using AftC, resulting in 3,5 Araf branch points (68, 69). Recently, AftD was designated as the second branching enzyme with α(1→3) Araf transferase activity that extends the Ara branches (70). In vitro assays using artificial chemical acceptors and cell-free extracts from M. smegmatis and C. glutamicum showed that the enzyme was able to add α(1→3) Araf residues to the linear α-1,5-linked neoglycolipid acceptor, resulting in branching of the linear arabinan core. Therefore, its function is considered similar to AftC. The branched Ara motif is further extended to the tetra-arabinoside and hexa-arabinoside motifs. The terminal β(1→2) Araf residues present in both tetra and hexa motifs are likely to be added by AftB (Rv3805c) and considered to have a dual role similar to AftC in AG and LAM biosynthesis (69).

The arabinose domains, though similar in AG and LAM, differ in several respects. The arabinan domain in AG is more complicated and terminates in hexa-arabinoside motifs (10), with the terminal and penultimate Araf residues serving as attachment points for mycolic acids (12). In contrast, LAM consists of both tetra-arabinoside and hexa-arabinoside motifs, and the terminal Araf residue is modified with α-mannose residues in all pathogenic strains of mycobacteria, thus generating Man-LAM. Mannose capping of LAM consists of one, two, or three Manp residues, with di-mannose caps being the most abundant (93, 97). The addition of mannose caps is predicted to be a two-step process that requires two mannosyltransferases, one for identification of the arabinan domain and addition of the primary Araf residue following recognition and a second for elongation of the cap with a second α(1→2) mannose. Subtractive genomics of the M. tuberculosis genome, with those species of the genus that do not contain mannose-capped LAM, such as M. smegmatis, highlighted the GT-C enzyme CapA (Rv1635c) (132). The MT1671 (Rv1635c) M. tuberculosis CDC1551 transposon mutant produced LAM devoid of Manp capping. Further studies supported this finding, with Rv1635c mutants in M. marinum and M. bovis BCG showing that Rv1635c encoded a mannosyltransferase involved in the addition of the first Manp residue on the nonreducing arabinan termini of LAM (133). More recently, Kaur et al. (96) have shown that MptC (Rv2181c) possesses varied substrate specificity, capable of adding α(1→2)-Manp residues onto the mannan backbone, as well as the nonreducing end of LAM in combination with Rv1635c, thus generating Man-LAM.

**DRUGS TARGETING AG AND LAM**

**EMB and Its Mechanism of Action**

EMB [(S,S)-2′(ethylenediimino)di-1-butanol] is a bacteriostatic agent widely used as a frontline drug for the treatment of TB. It is a synthetic compound that was first recognized as an antimycobacterial agent in 1961 (134). Early work by Kilburn and Greenberg observed an unanticipated increase in viable cells during the initial 4 hours after addition of EMB to M. smegmatis cultures. It was postulated that large bacillary clusters disaggregated due to a possible reduction in lipid content, which would lead to the apparent increase in CFUs (135). This theory was supported by a series of studies of the effects of EMB on M. smegmatis, demonstrating that inhibition of mycolic acid transfer into the cell wall and the simultaneous accumulation of TMM, TDM, and free mycolic acids occurred within 15 minutes of drug administration (136, 137). This suggested that the target might be a mycolyltransferase responsible for the transfer of mycolic acids onto the arabinan polymer. However, it was later discovered that the earliest point of drug inhibition occurred during arabinan synthesis, demonstrated by the immediate inhibition of incorporated label from [14C]glucose into the cell wall α-arabinose (138), while synthesis of the galactan of AG remained unaffected (139). Furthermore, the arabinans of both AG and LAM were inhibited, although reduction of label into the latter was less pronounced and at a later stage of its biosynthesis (139, 140).

The generation of EMB-resistant M. smegmatis mutants greatly aided the discovery of the primary EMB target, implicating it as an arabinan-specific inhibitor. This was illustrated by a study in which an EMB-resistant strain subjected to subinhibitory levels of EMB possessed a normal cell wall AG structure but a truncated version of LAM due to arabinan inhibition (139). Extending this observation using higher concentrations of EMB, it was shown that the degree of truncation in LAM was dose dependent, and at higher concentrations, the arabinan of AG was also impaired (92). Collectively, these studies indicate that the effects of EMB on the synthesis of both arabinan moieties are uncoupled, and the time difference of inhibition implies that synthesis occurs via distinct pathways, involving multiple AraT targets with varying EMB sensitivities. A number of concurrent EMB studies provided evidence of an accumulation of DPA, the source of Araf residues in arabinan biosynthesis, confirming that EMB effects were due not to inhibition of Araf donor synthesis but rather to its utilization (141, 142). Taken together, all the evidence points to arabinan polymerization, specifically the arabinan of AG, as the primary target of EMB.
A major breakthrough in the discovery of the precise EMB cellular target arose through exploitation of a moderately resistant strain from the related *M. avium* species. The genomic library from the aforementioned strain was screened and overexpressed in an otherwise susceptible *M. smegmatis* host, leading to the identification of a resistance-conferring region encompassing three complete ORFs, *embR*, *embA*, and *embB* (81). Interestingly, neither *embA* nor *embB* alone was sufficient to confer multicopy resistance, thus supporting the supposition that they are translationally coupled, forming a multienzyme complex (81). EMB resistance was also used to identify the *embCAB* gene cluster from *M. smegmatis*, which was subsequently characterized in *M. tuberculosis* and *M. leprae*, all of which possess the same syntenic organization (82, 143) and encode homologues of *embA* and *embB* genes from *M. avium*. Individual genetic knockouts of *embC*, *embA*, and *embB* were generated in *M. smegmatis*, all of which were viable, with the most profound affects observed in the *embB* mutant (66). Chemical analysis revealed that there was a decrease in the arabinose content of AG in both *embA* and *embB* mutants, specifically, the terminal disaccharide β-D-Araf (1→2)-α-D-Araf. Thus, a substantial amount of the otherwise hexarabinofuranoside motifs were present as terminal linear tetra-arabinofuranoside structures akin to the terminal motif of LAM (66), also leading to a loss of cell wall-bound mycolates. Based on the above observations, it appears that the Emb proteins are involved in hexarabinofuranoside biosynthesis. Indeed, cell-free assays using wild-type *M. smegmatis* membranes were used to investigate the putative activity of EmbA/EmbB and demonstrated the formation of the nonreducing terminal disaccharide, which was absent in both *embA* and *embB* mutants (144). Moreover, the transferase activity was re-established upon mixing the membrane preparations from the disrupted strains with wild-type membranes (144).

Identification of the *emb* gene cluster has provided the opportunity to analyze the molecular basis of resistance of mycobacteria to EMB. Telenti et al. (82) demonstrated that high-level resistance to EMB in *M. smegmatis* could be related to either overproduction of the Emb protein(s), a structural mutation in a conserved region of EmbB, or both. A number of reports presented additional genetic evidence for a key role of the EmbB protein in cell wall biosynthesis, highlighting the fact it is the most EMB-sensitive protein in the gene cluster (145). Mutations in EmbB have been recorded in up to 65% of EMB-resistant clinical isolates of *M. tuberculosis*, with the majority of mutations present at codon 306 or in the immediate surrounding area (143, 146). This region is highly conserved among mycobacteria, and topological analysis of the Emb proteins (82) positioned this EMB resistance-determining region (ERDR) in the second intracellular loop of EmbB (147). Five distinct mutations have been recognized at codon 306, resulting in a substitution of the wild-type methionine with isoleucine, leucine, or valine (147). Other mutations have been identified in the second intracellular loop region and the large C-terminal globular region of EmbB (146). It should be noted that there are a number of EMB-resistant strains that do not possess ERDR mutations, so other genes may be involved in EMB resistance.

**SQ109: Potent EMB Analog**

SQ109 [N’-(2-adamantyl)-N-[2(E)-3,7-dimethyl octa-2, 6-dienyl]ethane-1,2-diamine] was identified through a high-throughput screen from a chemical library of EMB derivatives sharing the same ethylenediamine core (148). SQ109 has improved *in vivo* efficacy against *M. tuberculosis* as well as low cytotoxicity (148). More importantly, it is active against multidrug-resistant and extensively drug-resistant TB clinical strains (148). Substitution of EMB with SQ109 in drug combination regimens demonstrated additive effects with EMB and strong synergistic interactions when combined with isoniazid, rifampin, and the newly discovered TMC207 compound (149, 150). SQ109 has been reported to be a safe and well-tolerated agent that is currently in phase IIa clinical trials to evaluate its efficacy and safety in patients with pulmonary TB (151).

Recently, the molecular target of SQ109 was determined to be an essential *mmpL3* (*Rv0206c*) gene that encodes a conserved transmembrane transporter protein. Treatment of *M. tuberculosis* with SQ109 resulted in a rapid decrease in the attachment of mycolic acids to both the cell wall AG and TDM. The pool of mycolates remained unaffected, but the levels of TMM, the precursor of TDM, and mycolic acids accumulated in SQ109-treated cells (152), pointing to TMM transport as the SQ109 target. Efforts to generate a spontaneous SQ109-resistant mutant in *M. tuberculosis* were unsuccessful. However, similar ethylenediamine compounds related to SQ109 were then used to spontaneously generate resistant mutants that were shown to have cross-resistance to SQ109 (152). The whole-genome sequencing of these mutants identified mutations in the essential *mmpL3* gene from *M. tuberculosis*, suggesting that the target of SQ109 is likely to be MmpL3 (152).
BTZs and DNBs

BTZ and DNB are novel and highly potent antimycobacterial with high bactericidal activity against mycobacteria (53). Both DNBs and BTZs are nitro compounds that target decaprenylphosphoryl-β-d-ribose 2’ epimerase, which is encoded by dprE1 (Rv3790) and dprE2 (Rv3791) and is responsible for epimerization of DPR to DPA (52, 53). The lipid-based sugar donor DPA is a precursor molecule supplying Araf residues to the AG and LAM of mycobacteria (45, 49). The nitro group in BTZ and DNB is proposed to be reduced to a nitroso group by the action of DprE1. This nitroso group forms a semi-mercaptal linkage to the conserved cysteine residue (Cys387) in the active site of the enzyme, thus rendering it nonfunctional (153, 154). Therefore, BTZs are also classified as suicide substrates of the enzyme DprE1 (153). Recent studies have demonstrated that mis-sense mutations in Cys387 residue can confer resistance to both drugs (52, 53). BTZs and DNBs can also be inactivated by the nitroreductase NfnB in M. smegmatis (155). While orthologue of nfnB is absent in M. tuberculosis, its presence in eukaryotes raises questions about the efficacy of BTZ and DNB as antitubercular drugs (155). The lead compound of the BTZ series, BTZ043, is in the late stages of preclinical trials and has successfully demonstrated high bactericidal activity against the multidrug-resistant strains, thus paving the way for the clinical trials (156).

CONCLUDING REMARKS

Overall, the sequencing of several of mycobacterial genomes, progress toward genetic tools to manipulate mycobacteria, and the use of surrogate systems such as C. glutamicum and M. smegmatis have contributed significantly to the identification of enzymes involved in the biosynthesis of AG and LAM. However, there are still some missing pieces to be found in these complex biosynthetic pathways. For example, the biosynthetic origin and function of succinyl residues found in both AG and LAM are not clearly understood and have merely been hypothesized. Galactosamine residue of AG, found only in pathogenic slow-growing mycobacteria, is speculated to play a role during host infection, but further in vivo studies have to be undertaken to confirm this theory. Furthermore, very little is known about the transporters required for the translocation of lipid-linked sugar donors and oligosaccharide intermediates from the cytoplasm to the extracellular space of mycobacteria. Only one study so far proposed a small multidrug-resistance-like gene, Rv3789, to encode a transporter, which reorients sugar donor DPA to the extracellular space, thus allowing further polymerization of arabinan (157). Finally, our understanding of the regulatory mechanisms in cell wall biosynthesis is lacking and requires comprehensive research.

We have attempted in this article to describe the latest progress of biochemical and genetic studies of mycobacterial cell wall assembly. Deciphering the biosynthetic pathways of AG and LAM has increased our knowledge of the biology of pathogenic M. tuberculosis and set the stage for the next research step—high-throughput screening assays for identifying potent inhibitors against essential enzymes of mycobacteria, protein–protein complexes, and the structural elucidation of the transmembrane glycosyltransferases involved in AG and LAM biosynthesis.

REFERENCES

arabinogalactan complex of Mycobacterium tuberculosis as revealed through characterization of oligoglycosylalditol fragments by fast-atom bombardment mass spectrometry and $^1$H nuclear magnetic resonance spectroscopy. *Biochemistry* 34:4257–4266.


Genetics of Mycobacterial Arabinogalactan and Lipoarabinomannan Assembly


