Genetics of Capsular Polysaccharides and Cell Envelope (Glyco)lipids

MAMADOU DAFFÉ,1 DEAN C. CRICK,2 and MARY JACKSON2
1CNRS, Institut de Pharmacologie et de Biologie Structurale, Département Mécanismes Moléculaires des Infections Mycobactériennes, and the Université de Toulouse Paul Sabatier, F-31077 Toulouse, France; 2Mycobacteria Research Laboratories, Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins CO 80523-1682

ABSTRACT This article summarizes what is currently known of the structures, physiological roles, involvement in pathogenicity, and biogenesis of a variety of noncovalently bound cell envelope lipids and glycoconjugates of Mycobacterium tuberculosis and other Mycobacterium species. Topics addressed in this article include phospholipids; phosphatidylinositol mannosides; triglycerides; isoprenoids and related compounds (polyprenyl phosphate, menaquinones, carotenoids, noncarotenoid cyclic isoprenoids); acyltrehaloses (lipooligosaccharides, trehalose mono- and di-mycoculates, sulfolipids, di- and poly-acyltrehaloses); mannosyl-beta-1-phosphomycoketides; glycopeptidolipids; phthiocerol dimycocerosates, para-hydroxybenzoic acids, and phenolic glycolipids; mycobactins; mycolactones; and capsular polysaccharides.

GLOBAL STRUCTURE AND COMPOSITION OF THE MYCOBACTERIAL CELL ENVELOPE

The compositional and architectural complexity of the mycobacterial cell envelope is probably the most distinctive feature of the Mycobacterium genus. It is the basis of many of the physiological and pathogenic features of these bacteria and the site of susceptibility and resistance to many antimycobacterial drugs (1, 2). In the context of the increasing incidence of multidrug-resistant strains of Mycobacterium tuberculosis, elucidating the complex pathways allowing mycobacteria to synthesize and assemble this complex structure represents a crucial area of research.

The mycobacterial cell envelope is made up of three major segments: the plasma membrane, the cell wall core, and the outermost layer. The cell wall core consists of peptidoglycan (PG) in covalent attachment via phosphoryl-N-acetylgalactosaminosyl-rhamnosyl linkage units with the heteropolysaccharide arabinogalactan, which is in turn esterified at its nonreducing ends to α-alkyl, β-hydroxy long-chain (C70 to C90) mycolic acids. The latter form the bulk of the inner leaflet of the outer membrane, with the outer layer consisting of a variety of noncovalently attached (glyco)lipids, polysaccharides, lipoglycans, and proteins (1, 3, 4) (Fig. 1). Only recently have developments in cryo-electron microscopy techniques allowed the different layers of the mycobacterial cell envelope to be visualized in their native state (3–5). These studies provided direct evidence of the existence of an outer bilayer and periplasmic space in M. tuberculosis, Mycobacterium bovis BCG, Mycobacterium marinum, Mycobacterium smegmatis, and the closely related Corynebacterium glutamicum (Fig. 1) (3–5). Together with classical subfractionation and biochemical
approaches, they also provided significant insights into the compositional diversity of the outermost layers of the cell envelopes of mycobacteria (5–9). All Mycobacterium species studied to date elaborate more or less abundant “capsule”-like structures both in vitro and during host infection that primarily consist of polysaccharides and proteins with generally minor amounts of lipids (7, 9). In some cases, however (e.g., Mycobacterium leprae, Mycobacterium avium), abundant quantities of species-specific glycolipids may be found (glycopeptidolipids [GPLs] and phenolic glycolipids [PGLs] in particular). Many of the proteins and lipids typically found in the capsules of mycobacteria also occur in the outer membrane and periplasm, and their relative distribution between these three compartments seems to be species-dependent.
This diversity in terms of surface composition most likely reflects differences in the cell envelope organization of mycobacteria and is likely to significantly impact the way that Mycobacterium species interact with the host (11, 12).

Developments in the genetic manipulation of mycobacteria in the 1990s and the publication of the complete genome sequence of M. tuberculosis in 1998, followed later by that of several other fast-growing and slow-growing mycobacteria, have provided a major impetus to the study of cell envelope biosynthesis in various Mycobacterium species, with the result that many of the enzymes involved in their synthesis have now been identified. The molecular genetics of the cell wall core proper (PG, arabinogalactan, mycolic acids) is reviewed in other articles. This article focuses on what is known of the biosynthesis and translocation of the major noncovalently bound (extractable) lipid and glycoconjugate constituents populating the inner and outer membranes and capsule-like structures of mycobacteria. For those constituents ubiquitously distributed in mycobacteria, the gene nomenclature used is that of M. tuberculosis H37Rv.

PHOSPHOLIPIDS, PHOSPHATIDYLINOSITOL MANNOSIDES, AND TRIGLYCERIDES

Phospholipids and Triacylglycerols of Mycobacteria

The mycobacterial phospholipids include phosphatidylglycerol, diphosphatidylglycerol (i.e., cardiolipin; CL), phosphatidylethanolamine (PE), phosphatidyl-my-o-inositol (PI), and mannosylated forms of PI known collectively as the phosphatidyl-my-o-inositol mannosides (PIMs) (Fig. 2). Phosphatidylserine also occurs in limited amounts (Fig. 2), but phosphatidylcholine is apparently not produced by mycobacteria (13). Phospholipids represent the main structural amphipathic polar lipids of the mycobacterial inner membrane and also populate the outer membrane (Fig. 1). PE and PIMs, in particular, were identified in the surface-exposed lipids of all Mycobacterium species investigated (M. tuberculosis, M. avium, M. kansasii, M. gastri, M. smegmatis, and M. aurum) (8). Palmitic (C16:0), oleic (C18:1), and tuberculostearic (C19) acids appear to be the major fatty acid substituents in the phospholipids of mycobacteria, with the unsaturated or branched C18:1 and C19 fatty acids principally esterifying position 1 of glycerol, and C16:0 preferentially occupying position 2.

FIGURE 2 Structures of mycobacterial phospholipids. doi:10.1128/microbiolspec.MGM2-0021-2013.f2
Triacylglycerols (TAGs), triglycerides, have similarly been isolated from all mycobacterial species examined and represent the major apolar lipids when glycerol is the major carbon source in the culture medium (14). Mycobacteria grown in vitro or recovered from human samples essentially accumulate TAG in the form of intracellular lipid droplets, but TAGs have also been identified in the surface-exposed lipids of M. smegmatis and M. avium (8). They are proposed to act as a source of energy for actively replicating bacteria as well as a means by which free fatty acids are detoxified. TAGs are also proposed to serve as an energy reserve for the long-term survival of M. tuberculosis during the persistence phase of infection (14, 15). In M. bovis BCG and M. smegmatis, position 1 of TAG is occupied principally by stearic (C18:0), C18:1, and C19 fatty acids; position 2 is mostly esterified with C16 fatty acids, whereas the third position predominantly bears fatty acids with more than 20 carbons (C20 to C33) (16). The fatty acids acylating phospholipids and triglycerides in axenically grown bacteria are thought to be synthesized by fatty acid synthase I (FAS-I) (Rv2524c) (17, 18).

Phosphatidic Acid Synthesis

Phosphatidic acid (Fig. 2) is a common intermediate in the biosynthesis of both TAG and phospholipids. The pathway begins with glycerol-3-phosphate, which is formed by reduction of dihydroxyacetone phosphate by the glycerol-3-phosphate synthase GpsA. Two gene candidates were annotated for this function in the genome of M. tuberculosis H37Rv, gpdA1 (Rv0564c) and gpdA2 (Rv2982c), but neither of them has been confirmed biochemically. Glycerol-3-phosphate is first acylated by acyl-coenzyme A (CoA), acyl-ACP, or acyl-phosphate to form lysophosphatidate and is then acylated again by acyl-CoA or acyl-ACP to yield phosphatidate (19). Again, based on sequence similarities, two putative glycerol-3-phosphate acyltransferase genes, plsB1 (Rv1551) and plsB2 (Rv2482c), and one putative lysophosphatidate acyltransferase gene, plsC (Rv2483c), have been proposed to be involved in those acyl transfer reactions, but they have not yet been biochemically validated (Table 1).

TAG Synthesis

In the synthesis of TAG, phosphatidate is hydrolyzed by a specific phosphatase to yield diacylglycerol (DAG). This intermediate is then acylated to TAG in a reaction catalyzed by diglyceride acyltransferases (or triglyceride synthases). Although no phosphatidic acid phosphatases have yet been identified in mycobacteria, two proteins displaying this activity were recently characterized in Streptomyces coelicolor (20), one of which (SCO1102) displays sequence similarity with Rv0308 of M. tuberculosis H37Rv (H. Gramajo, personal communication). In the genome of M. tuberculosis H37Rv 15 genes were identified whose protein products display triglyceride synthase activity in vitro, generating triolein from diolien and oleyl-CoA (15, 21). Interestingly, Ag85A (FbpA, Rv3804c) is also endowed with a similar acyltransferase activity, transferring long-chain acyl-CoA onto DAG (22) (Table 1).

Phospholipid Biosynthesis

CDP-DAG appears to be the common precursor for the biosynthesis of phospholipids in mycobacteria and is synthesized from phosphatidic acid and CTP by the CDP-DAG synthase (CTP:phosphatidate cytidylyltransferase). Such enzymatic activity was detected in M. smegmatis and found to be membrane-associated (23). The structural gene for CDP-DAG synthase in the genome of M. tuberculosis H37Rv is predicted to be cdaA (Rv2881c). Phosphatidyl-myoinositol (PI) is made de novo from CDP-DAG and myo-inositol (24) in a reaction catalyzed by the PI synthase, PgsA1 (Rv2612c) (25). However, an alternative pathway for PI synthesis has been suggested wherein myo-inositol is first phosphorylated to form myo-inositol 3-phosphate, which then reacts with CDP-DAG to form PI 3-phosphate (PI3P). It was proposed that pgsA1 encodes a PI3P synthase rather than a PI synthase and that PI3P is subsequently dephosphorylated (by an as yet unknown enzyme) to yield PI (26). Evidence based on sequence homology or changes in the phospholipid composition of M. smegmatis upon gene overexpression strongly suggests that the pgsA3 (Rv2746c) and pssA (Rv0436c) genes of M. tuberculosis encode the phosphatidyl-glycerophosphate synthase and phosphatidylserine synthase involved in the formation of, respectively, phosphatidylglycerol and phosphatidylserine (23). As in other bacteria, PE is likely to arise from the decarboxylation of phosphatidylserine in a reaction catalyzed by the product of pssA (Rv0437c). Cardiolipin may be formed from the condensation of two phosphatidylglycerol molecules by a cardiolipin synthase as in most prokaryotes or through the transfer of a phosphatidyl group from CDP-DAG onto phosphatidylglycerol as in yeast and as recently shown in S. coelicolor (27). M. tuberculosis H37Rv possesses a eukaryotic-type cardiolipin synthase bearing sequence similarity to the Streptomyces enzyme (PgsA2; Rv1822), whereas proteins displaying the characteristic phospholipase D-type features of classical prokaryotic cardiolipin synthases are missing, suggesting that the second pathway may
<table>
<thead>
<tr>
<th>Rv number</th>
<th>Gene name</th>
<th>Function</th>
<th>Evidence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv0221</td>
<td>Putative acyl-CoA:diacylglycerol acyltransferase</td>
<td>E</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Rv0308</td>
<td>Putative phosphatidic acid phosphatase</td>
<td>H</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Rv0436c</td>
<td>pssA</td>
<td>Putative phosphatidylserine synthase</td>
<td>P, H</td>
<td>25</td>
</tr>
<tr>
<td>Rv0437c</td>
<td>psd</td>
<td>Putative phosphatidylserine decarboxylase</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>Rv0534c</td>
<td>menA</td>
<td>Demethylmenaquinone synthase</td>
<td>E</td>
<td>114</td>
</tr>
<tr>
<td>Rv0542c</td>
<td>menE</td>
<td>o-Succinylbenzoyl-CoA synthase</td>
<td>E</td>
<td>111-113</td>
</tr>
<tr>
<td>Rv0548c</td>
<td>menB</td>
<td>1,4-Dihydroxy2-naphthoic acid synthase</td>
<td>E</td>
<td>108-110</td>
</tr>
<tr>
<td>Rv0562</td>
<td>ω,E,E,E-Geranylgeranyldiphosphate synthase</td>
<td>E</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>Rv0564c</td>
<td>gpdA1</td>
<td>Putative glycerol-3-phosphate synthase</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>Rv0654</td>
<td>Carotenoid oxygenase</td>
<td>E</td>
<td>132</td>
<td></td>
</tr>
<tr>
<td>Rv0895</td>
<td>GeranylCoA synthase</td>
<td>E</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Rv0989c</td>
<td>Geranyldiphosphate synthase</td>
<td>E</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>Rv1011</td>
<td>ispE</td>
<td>4-Diphosphocytidyl-2C-methyl-α-erythritol kinase</td>
<td>E</td>
<td>61, 62</td>
</tr>
<tr>
<td>Rv1086</td>
<td>ω,E,Z-Farnesyldiphosphate synthase</td>
<td>E</td>
<td>79, 80</td>
<td></td>
</tr>
<tr>
<td>Rv1159</td>
<td>pimE</td>
<td>Polyprenol phosphomannose-dependent α-1,2-mannosyltransferase</td>
<td>E, P</td>
<td>34</td>
</tr>
<tr>
<td>Rv1411c</td>
<td>lprG</td>
<td>Lipoprotein; putative PIM, LM, and LAM transporter</td>
<td>P</td>
<td>38</td>
</tr>
<tr>
<td>Rv1425</td>
<td>Putative acyl-CoA:diacylglycerol acyltransferase</td>
<td>E</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Rv1551</td>
<td>Putative acyl-CoA:diacylglycerol acyltransferase</td>
<td>E</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Rv1760</td>
<td>Putative acyl-CoA:diacylglycerol acyltransferase</td>
<td>E</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Rv1822</td>
<td>Putative cardiolipin synthase</td>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv2188c</td>
<td>pimB</td>
<td>GDP-Man-dependent α-1,6-phosphatidylglycerol transferase</td>
<td>E, P</td>
<td>31</td>
</tr>
<tr>
<td>Rv2267c</td>
<td>stf3</td>
<td>Putative sulfotransferase</td>
<td>P, H</td>
<td>117</td>
</tr>
<tr>
<td>Rv2285</td>
<td>Putative acyl-CoA:diacylglycerol acyltransferase</td>
<td>E</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Rv2361c</td>
<td>ω,E,poly-Z-Decaprenyl-CoA synthase</td>
<td>E</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>Rv2482c</td>
<td>pbsB2</td>
<td>Putative glycerol-3-phosphate acyltransferase</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>Rv2483c</td>
<td>pbsC</td>
<td>Putative lysophosphatidate acyltransferase</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>Rv2484c</td>
<td>Putative acyl-CoA:diacylglycerol acyltransferase</td>
<td>E</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Rv2524c</td>
<td>fas</td>
<td>Fatty acid synthase type I</td>
<td>E, P</td>
<td>17, 18</td>
</tr>
<tr>
<td>Rv2610c</td>
<td>pimA</td>
<td>GDP-Man-dependent α-1,2-phosphatidylinositol mannosyltransferase</td>
<td>E, P</td>
<td>28, 30</td>
</tr>
<tr>
<td>Rv2611c</td>
<td>Acyltransferase involved in the 6-O-acetylation of the Manp residue linked to the 2-position of myo-inositol in PIM1 and PIM2</td>
<td>E, P</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>Rv2612c</td>
<td>pgsA1</td>
<td>Phosphatidyl-myo-inositol synthase and/or phosphatidyl-myo-inositol phosphatase synthase</td>
<td>E, P</td>
<td>25, 26</td>
</tr>
<tr>
<td>Rv2682c</td>
<td>dks</td>
<td>1-Deoxy-α-xylulose-5-phosphate synthase</td>
<td>E</td>
<td>49</td>
</tr>
<tr>
<td>Rv2746c</td>
<td>pgsA3</td>
<td>Phosphatidylglycerol synthase synthase</td>
<td>P, H</td>
<td>25</td>
</tr>
<tr>
<td>Rv2868c</td>
<td>ispG</td>
<td>1-Hydroxy-2-methyl-2(E)-butenyl 4-diphosphate synthase</td>
<td>H</td>
<td>50</td>
</tr>
<tr>
<td>Rv2870c</td>
<td>ispC</td>
<td>1-Deoxy-α,xylulose-5-phosphate dehydratase synthase</td>
<td>E</td>
<td>54, 55</td>
</tr>
<tr>
<td>Rv2881c</td>
<td>cdsA</td>
<td>Putative CDP-diacylglycerol synthase</td>
<td>H</td>
<td>23</td>
</tr>
<tr>
<td>Rv2982c</td>
<td>gpdA2</td>
<td>Putative glycerol-3-phosphate synthase</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>Rv3087</td>
<td>Putative acyl-CoA:diacylglycerol acyltransferase</td>
<td>E</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Rv3088</td>
<td>tgs4</td>
<td>Putative acyl-CoA:diacylglycerol acyltransferase</td>
<td>E</td>
<td>15</td>
</tr>
<tr>
<td>Rv3130c</td>
<td>tgs1</td>
<td>Putative acyl-CoA:diacylglycerol acyltransferase</td>
<td>E, P</td>
<td>15, 21</td>
</tr>
<tr>
<td>Rv3233c</td>
<td>tgs2</td>
<td>Putative acyl-CoA:diacylglycerol acyltransferase</td>
<td>E</td>
<td>15</td>
</tr>
<tr>
<td>Rv3234c</td>
<td>tgs3</td>
<td>Putative acyl-CoA:diacylglycerol acyltransferase</td>
<td>E</td>
<td>15</td>
</tr>
<tr>
<td>Rv3371</td>
<td>Putative acyl-CoA:diacylglycerol acyltransferase</td>
<td>E</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Rv3377c</td>
<td>Tuberculossyldiphosphate synthase</td>
<td>E</td>
<td>138</td>
<td></td>
</tr>
<tr>
<td>Rv3378c</td>
<td>Isotuberculossin synthase</td>
<td>E</td>
<td>135, 139, 140</td>
<td></td>
</tr>
<tr>
<td>Rv3383c</td>
<td>ω,E,E,E-Geranylgeranyldiphosphate synthase</td>
<td>E</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>Rv3398c</td>
<td>ω,E,E-Farnesyldiphosphate synthase</td>
<td>E</td>
<td>86</td>
<td></td>
</tr>
</tbody>
</table>

(continued)
be the one used by mycobacteria (25, 27). However, whether PgsA2 is endowed with such enzymatic activity remains to be established (Table 1).

### PIMs

The PI dimannosides (PIM$_2$) are considered both metabolic end products and intermediates in the biosynthesis of polar PI (PIM$_5$, PIM$_6$), lipomannan (LM), and lipoarabinomannan (LAM) (for more details about these molecules and their biosynthetic pathways, see reference 36.3). We will only briefly describe the initial steps of PIM synthesis leading to the formation of PIM$_2$ and PIM$_6$, the two most abundant forms of PIM found in mycobacteria. The first step in PIM synthesis involves the transfer of a mannose residue from GDP-Manp to the 2-position of the myo-inositol ring of PI to form PI monomannoside, PIM$_1$. We have identified PimA (Rv2610c) as the α-mannosyltransferase responsible for this catalytic step and found it to be an essential enzyme (28–30). The second step involves the action of another essential α-mannosyltransferase, PimB’ (Rv2188c), which transfers a Manp residue from GDP-Manp to the 6-position of the myo-inositol ring of PIM$_1$ (31). Both PIM$_1$ and PIM$_2$ can be acylated with palmitate at position 6 of the Manp residue transferred by PimA by the acyltransferase Rv2611c to form Ac$_1$PIM$_1$ and Ac$_2$PIM$_2$, respectively (32). The acyltransferase responsible for the transfer of a fourth acyl group to position 3 of the myo-inositol ring has not yet been identified. Likewise, the identity of the enzymes involved in the mannosylation of the dimannosylated forms of PIM to form PIM$_3$ and PIM$_4$ is at present unclear (33). PimE (Rv1159) has been identified as the α-1,2-mannosyltransferase involved in the biosynthesis of PIM$_4$ from PIM$_3$ (34). PimE belongs to the GT-C superfamily of glycosyltransferases, which comprises integral membrane proteins that use polypropenyl-linked sugars as donors (33, 35). Whether PimE also catalyzes the transfer of the second α-1,2-linked Man residue onto PIM$_3$ to yield PIM$_6$ or whether the formation of PIM$_6$ results from the action of an independent mannosyltransferase is at present not known.

### Translocation of Phospholipids, PIM, and TAG to the Outer Membrane and Cell Surface

Phospholipids and TAG are synthesized in the cytoplasm or at the periphery of the inner leaflet of the plasma membrane. Likewise, the early steps of PIM biosynthesis take place on the cytosolic face of the plasma membrane until PIM intermediates, believed to be PIM$_2$ or PIM$_3$, are translocated across the plasma membrane by an as yet unknown flippase to serve as substrates for further mannosylation reactions catalyzed by PimE and other GT-C polypropenyl-phosphate mannosyl-dependent glycosyltransferases (33, 35, 36). Beyond their translocation across the plasma membrane, the further export of phospholipids, TAG, and PIM to the outer membrane and cell surface most likely requires dedicated translocation machineries. Thus far, none of the flippases and transporters involved have been formally identified. Evidence based on physical interactions and cocrystallography suggests that the lipoprotein LprG (Rv1411c), which shares structural resemblance to LppX, a lipoprotein thought to carry phthiocerol dimyocerosates (PDIM) across the periplasm (37), may participate in the transport of PIM, LM, and LAM to the cell surface (38). This exciting hypothesis awaits further genetic and biochemical validation.

### ISOPRENOIDs AND RELATED LIPIDS

#### Biosynthesis of Isoprenoid Precursors

A number of isoprenoids have been observed and characterized in *Mycobacterium* species including polypropenyl diphosphates, polypropenyl phosphates, lipid I and lipid II, carotenoids, menaquinones, sulfomenaquinones, and cyclic isoprenoids. These molecules have diverse and in

---

**TABLE 1** *M. tuberculosis* H37Rv genes involved or thought to be involved in the biogenesis of phospholipids, triglycerides, isoprenoids, and related lipids (continued)

<table>
<thead>
<tr>
<th>Rv number</th>
<th>Gene name</th>
<th>Function</th>
<th>Evidence$^a$</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv3480c</td>
<td>ispF</td>
<td>Putative acyl-CoA:diacylglycerol acyltransferase</td>
<td>E</td>
<td>15</td>
</tr>
<tr>
<td>Rv3581c</td>
<td>ispD</td>
<td>2C-Methyl-α-erythritol 2,4-cyclodiphosphate</td>
<td>E</td>
<td>63</td>
</tr>
<tr>
<td>Rv3582c</td>
<td>tgs2</td>
<td>4-Diphosphorytidyl-2C-methyl-α-erythritol</td>
<td>E</td>
<td>58</td>
</tr>
<tr>
<td>Rv3734c</td>
<td></td>
<td>Putative acyl-CoA:diacylglycerol acyltransferase</td>
<td>E</td>
<td>15</td>
</tr>
<tr>
<td>Rv3740c</td>
<td></td>
<td>Putative acyl-CoA:diacylglycerol acyltransferase</td>
<td>E</td>
<td>15</td>
</tr>
<tr>
<td>Rv3804c</td>
<td>fbpA</td>
<td>Acyl-CoA:diacylglycerol acyltransferase</td>
<td>E, P</td>
<td>22</td>
</tr>
</tbody>
</table>

$^a$The experimental evidence for the annotation of a gene may either be “enzymatic” (E) (i.e., an enzymatic activity was associated with the gene’s product in vitro) or “phenotypic” (P) (i.e., the annotation results from the biochemical analysis of mycobacterial recombinant strains—e.g., knockout/knockdown mutants, complemented mutant strains, overexpressors—or from the functional complementation of defined *E. coli* mutants). In some cases, the function of a gene is exclusively based on its homology to other known (myco)bacterial genes (H).
Genetics of Capsular Polysaccharides and Cell Envelope (Glyco)lipids

some cases multiple functions. For example, polyisoprenyl phosphate (Pol-P) is involved in the biosynthesis of the arabinan portion of arabinogalactan, arabinomannan, and LAM (39), and in that of the PG precursors lipid I and lipid II (40, 41), as a lipid carrier of the activated saccharide subunits. Pol-P is also involved in the biosynthesis of the “linker unit” between two essential cell wall components, arabinogalactan and PG (42).

All isoprenoids are derived from the repetitive condensation of isopentenyl diphosphate (IPP) and allylic diphosphates (43) catalyzed by enzymes known as prenyldiphosphate synthases or prenyltransferases. To date, two distinct pathways for the biosynthesis of the IPP and dimethylallyl diphosphate (DMAPP; the smallest allylic diphosphate) have been identified: the mevalonate (MVA) pathway and the 2C-methyl-D-erythritol 4-phosphate (MEP) pathway. In mycobacteria, IPP and DMAPP (Fig. 3) are biosynthesized exclusively via the MEP pathway.

The MEP pathway of M. tuberculosis

The initial enzyme in the MEP pathway, 1-deoxy-D-xylulose-5-phosphate synthase (DXS), catalyzes the condensation of glyceraldehyde-3-phosphate and pyruvate, forming 1-deoxy-D-xylulose-5-phosphate (DXP) (44). The product of DXS is used not only as a biosynthetic intermediate of IPP but also as the precursor of thiamin (vitamin B1) and of pyridoxol (vitamin B6) in E. coli (45–47); thus, DXS is not a committed step in the MEP pathway.

The dxs gene was first identified in E. coli (45, 46). Sequence alignment with E. coli DXS demonstrated that Rv2682c has approximately 38% identity with the conserved DRAG motif and a key amino acid (His49) required for catalytic activity (48) (Table 1). The function of Rv2682c was demonstrated empirically, because the purified recombinant enzyme is capable of producing DXS by condensation of pyruvate and glyceraldehyde-3-phosphate in the presence of thiamine pyrophosphate (49). Interestingly, M. tuberculosis contains a second ortholog of E. coli DXS, Rv3379c. However, an alignment with E. coli DXS indicated that Rv3379c, despite a relatively high level of identity (38%), was truncated due to the positioning of an insertion element (IS6110) and, more importantly, the His49 residue is missing and the recombinant protein showed no DXS activity (49). This and the fact that Rv2682c is essential for bacterial survival (50) suggest that it encodes the only functional M. tuberculosis DXS.

1-Deoxy-d-xylulose-5-phosphate reductoisomerase (IspC), the second enzyme in the MEP biosynthetic pathway, catalyzes the rearrangement and reduction of DXP in the presence of NADPH to generate MEP (51). As mentioned above, DXP is a precursor not only of IPP and DMAPP but also of thiamine and pyridoxol; therefore, IspC catalyzes the first committed step for biosynthesis of IPP and DMAPP (52).

Alignments with E. coli IspC indicated that the primary structure of Rv2870c of M. tuberculosis is 25% identical to that of the E. coli IspC with conserved amino acid residues (53, 54). Recombinant Rv2870c efficiently catalyzes the conversion of DXP to MEP in the presence of NADPH and the reverse reaction in the presence of NADP+ (54–56).

Incubation of MEP with crude, cell-free extracts of E. coli in the presence of CTP produces 4-diphosphocytidyl-2C-methyl-D-erythritol (CDP-ME), and the gene encoding the activity was identified as ygbP (57), which was later renamed ispD. The Rv3582c gene product has approximately 31% identity with E. coli IspD, and the recombinant Rv3582c protein was shown to be a functional IspD in M. tuberculosis (58).

The fourth step in the MEP pathway involves the conversion of CDP-ME to 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate (CDP-ME2P) in the presence of ATP catalyzed by IspE, which was initially identified in E. coli and tomatoes (59, 60). Alignment of E. coli IspE with genes of the M. tuberculosis genome indicated that Rv1011 encodes a protein that harbors around 22% identity with conserved amino acids involved in forming the CDP-ME and ATP binding and crucial active sites and catalyzes CDP-ME phosphorylation in an ATP-dependent manner (61, 62).

The fifth step of the MEP pathway involves the formation of a metabolite containing a cyclodiphosphate moiety. The product of IspE, CDP-ME2P, is converted into

FIGURE 3 Structures of IPP and DMAPP. These molecules are precursors of all isoprenoid compounds. doi:10.1128/microbiolspec.MGM2-0021-2013.f3
2C-methyl-D-erythritol 2,4-cyclodiphosphate (MECDP) with corresponding release of CMP by the ispF gene product (59). Rv3581c encodes M. tuberculosis IspF (63) and is essential for bacterial survival (64). The crystal structure of M. smegmatis IspF, harboring around 73% amino acid sequence identity with M. tuberculosis IspF, has been solved (64).

Recombinant E. coli ispC, ispD, ispF, and ispG were shown to catalyze the conversion of 1-deoxy-D-xylulose (DX) into 1-hydroxy-2-methyl-2(E)-butenyl-4-prenyldiphosphate (HMBPP) (65), and the ispH gene product is responsible for the conversion of HMBPP into IPP and DMAPP (66, 67). Recombinant IspG catalyzes the reduction of MECDP, resulting in opening of the cyclodiphosphate ring structure using a photoreduced deazaflavin derivative as an artificial electron donor (68, 69). Interestingly, in vivo experiments using an E. coli strain overexpressing ispH resulted in the formation of IPP and DMAPP from HMBPP in a molar ratio of 5:1 (70). Blast searches of E. coli IspG or IspH with the M. tuberculosis genome indicate that Rv2868c, an essential gene (50), is the likely M. tuberculosis IspG, and either Rv1110 or Rv3382c is the candidate gene encoding M. tuberculosis IspH.

IPP isomerase

Upon biosynthesis of IPP and DMAPP by IspH, IPP isomerase (Idi) catalyzes the interconversion of the two isoforms (71), but the equilibrium favors the forward reaction, from IPP to DMAPP (72). In organisms capable of synthesizing isoprenoid units by the MVA pathway, Idi is reported to be essential (73), as pyrophosphomevalonate decarboxylase in the MVA pathway produces only IPP and both DMAPP and IPP (Fig. 4, Table 1) are required for further biosyntheses of isoprenoids. Orthologs of idi are also found in many organisms that utilize the MEP pathway, most of which are reported to encode nonessential enzymes (74), presumably because IspH of the MEP pathway produces both IPP and DMAPP. Two forms of bacterial Idi have been discovered to date: type I, which includes Idi from E. coli, and type II, which was identified in Streptomyces species strain CL190 (75); M. tuberculosis has an ortholog of a type I Idi, while M. smegmatis has an ortholog of the type II Idi.

Prenyldiphosphate synthases

As mentioned above, the universal precursors of all isoprenoid compounds are synthesized from IPP, DMAPP, or linear IPPs that are synthesized by sequential 1’4 condensations of IPP with DMAPP. The enzymes catalyzing this sequential process are known as prenyltransferases or prenyldiphosphate synthases. These enzymes can be divided into two families depending on the stereochemistry of the double bonds formed during product formation and the chain length of the final product. Thus, prenyldiphosphate synthases can be categorized as E-prenyldiphosphate synthases or Z-prenyldiphosphate synthases, and there is no similarity between the two in terms of amino acid sequence. The E-prenyldiphosphate synthases can be further characterized as short-chain, with a product containing 10 to 25 carbons, medium-chain, 30 to 35 carbons, and long-chain, 40 to 50 carbons (76). Similarly, the Z-prenyldiphosphate synthases can be characterized as short-chain, medium-chain, and long-chain (77). Both the E- and Z-prenyldiphosphate synthase families generate products with the correct chain lengths via a molecular ruler mechanism, where one or two bulky amino acids occupy the bottom of each of the enzyme active sites to block extra chain elongation of the products, thereby determining the ultimate chain lengths (78). Both E- and Z-prenyldiphosphate synthase have been identified and characterized in Mycobacterium species.

Mycobacterium species are unusual in that they harbor two or three Z-prenyldiphosphate synthases, whereas most bacteria have only one of these enzymes. In M. tuberculosis, Rv1086 encodes a short-chain Z-prenyldiphosphate synthase that generates ω,E,Z-farnesylphosphate (Fig. 4, Table 1). This gene has been cloned and expressed and the enzyme activity characterized (79, 80), and it was the first representative of this class of enzyme described. The crystal structure and mechanism of chain length determination have been solved (81, 82). Rv2361c has been identified as a long-chain Z-prenyldiphosphate synthase that synthesizes ω,E,poly-Z-decaprenylphosphate (79, 83) (Fig. 4). In Mycobacterium vanbaalenii, three Z-prenyldiphosphate synthase were identified and characterized (84). Mvan_4662 accepts only geranyldiphosphate as the allylic primer, producing only ω,E,Z-farnesylphosphate, indicating a function similar to Rv1086. Mvan_1705 accepts only ω,E,Z-farnesylphosphate, synthesizing ω,Z,E,E-geranylgeranyldiphosphate whereas Mvan_3822 is a bifunctional Z-prenyldiphosphate synthase that preferentially synthesizes C35 or C50 products, depending on the allylic reaction primer.

A number of E-prenyldiphosphate synthases have also been identified in mycobacteria, which synthesize E-prenyldiphosphates of various chain lengths (Fig. 4, Table 1). These include Rv0989c, which is reported to synthesize geranyldiphosphate (85); Rv3398c, which encodes an ω,E,E-farnesylphosphate synthase (86);
and Rv0562 and Rv3383c, both of which are reported to encode ω,E,E,E-geranylgeranyldiphosphate synthases (87). It should be noted that stereochemistry of the products of the E-prenyldiphosphate synthases is assumed based on the amino acid sequence of the enzyme, not on empirical observation.

Polyprenyl Phosphate
 Structures of mycobacterial polyprenyl phosphates

The most common structures of polyisoprenol (and therefore Pol-P) found in nature tend to be confined to four main groups: (i) ω,E-polyisoprenol, (ii) ω,di-E, poly-Z-polyprenol, (iii) ω,tri-E,poly-Z-polyprenol, and (iv) ω,Z-polyprenol (88). Most bacteria utilize undecaprenylphosphate (or bactoprenylphosphate), a ω,di-E, octa-Z-prenylphosphate, as a carrier of activated sugars primarily for synthesis of oligo- and polysaccharides on the outside of the plasma membrane as is seen in PG synthesis. However, mycobacteria synthesize and utilize at least two and perhaps three forms of Pol-P. In M. smegmatis two forms of Pol-P have been reported (Fig. 4): (i) decaprenyl phosphate (Dec-P) containing one ω, one E-, and eight Z-isoprene units (ω,E,poly-Z) (39) and (ii) a heptaprenyl phosphate (89) containing four saturated isoprene units on the omega end of the molecule and two E- and one Z-isoprene units (90) or four saturated and three Z-isoprene units (91). M. tuberculosis, however, appears to utilize a single predominant Pol-P (Dec-P). To date, the stereochemistry of the individual isoprene units of Dec-P from M. tuberculosis have not been determined (92), but it is likely that they are the same as those of the M. smegmatis Dec-P. In all three cases described above, the mycobacterial Pol-P molecules are structurally unusual.

Pol-P biosynthesis

In general, all Pol-P molecules are synthesized via sequential condensation of IPP with allylic diphosphates catalyzed by the prenyldiphosphate synthases described above, forming polyisoprenyldiphosphates (Pol-PP) that are subsequently dephosphorylated. In mycobacteria, Rv1086 and Rv2361c (Table 1) can catalyze the addition of IPP to ω,E-GPP; however, kinetic analyses (80, 83) suggest that Rv1086 and Rv2361c act sequentially in the synthesis of ω,E,poly-Z-decaprenyl diphosphate (Dec-PP), the precursor of the ω,E,poly-Z-Dec-P found in mycobacteria (39, 89, 91–93), with Rv2361c adding seven isoprene units to the ω,E,Z-FPP synthesized by Rv1086. Thus, it seems likely that Rv0989c, Rv1086, and Rv2361c act in concert to generate decaprenyl-diphosphate (Dec-PP), with isoprene units of the required stereochemistry. Once the Dec-PP has been

|FIGURE 4| Structures of representative short-chain IPPs synthesized by mycobacteria. The stereochemical conformation is shown. [doi:10.1128/microbiolspec.MGM2-0021-2013.f4](http://www.asmscience.org/microbiolspectrum)

synthesized, it must be dephosphorylated to form Dec-P (Fig. 5). Currently, there is little information regarding this biosynthetic transformation in mycobacteria; however, an ortholog of BacA, a phosphatase reported to be involved in dephosphorylation of Pol-PP in \textit{E. coli} (94), may be involved.

**Menaquinones**

**Structure of mycobacterial lipoquinones**

The lipoquinones involved in the respiratory chains of bacteria consist of menaquinones and ubiquinones (95), while mammals have only ubiquinone. Menaquinones (2-methyl-3-polyprenyl-1,4-naphthoquinones) are the predominant isoprenoid lipoquinones of mycobacteria and many Gram-positive bacteria, whereas Gram-negative bacteria typically utilize both menaquinone and ubiquinone or only ubiquinone (which has a benzoquinone ring rather than a naphthoquinone ring) (96–100).

Menaquinones are identified by the variable portions of the molecules. Generally, the only variation seen in the naphthyl ring structure is whether or not the C-2 position is methylated (Fig. 6). The most variant portion of the molecule is the polyprenyl side chain found at the C-3 position. Menaquinones (and ubiquinones) are identified by the length and chemical structure of this side chain. For example, a menaquinone with a side chain of eight isoprene units as seen in \textit{E. coli} is identified as MK-8. The predominant form of menaquinone in mycobacteria has nine isoprene units, with the \(\beta\) position being saturated (96) (Fig. 6). Hence, this menaquinone is identified as MK-9 (II-H\(_2\)).

**Functional significance of the menaquinone structure**

Historically, respiratory quinones have been utilized for taxonomic purposes because the length and degree of saturation of the isoprenoid chain often reflect the phylogenetic affiliation of bacteria (101). The taxonomic distribution of structural features suggests that they are both functional and evolutionarily conserved. A great deal of effort was put into defining the significance of the various structural variations in the 1960s, but this area of research has been largely ignored since then. In 1970, Brodie et al. summarized the state of the knowledge (102). Thus, it is known that the substitution at C-2 of the naphthyl ring is required for both oxidation and phosphorylation and must be a methyl group because conversion to a hydroxyl permits oxidation but not phosphorylation. The C-3 position must be substituted with an isoprenoid chain to function as a membrane-bound electron transporter. The double bond in the \(\alpha\)-isoprene unit must be in the \(E\)-configuration; the \(Z\)-isomer does not allow phosphorylation. Thus, it appears that menaquinone in the electron transport is more than a simple electron transporter, because structural modifications allow uncoupling of oxidation and phosphorylation, suggesting that the menaquinone structure may regulate ATP synthesis. The single bond in the \(\beta\)-isoprene unit is conserved in many Gram-positive bacteria. However, the function of this modification is unknown. Recently, a novel sulfated menaquinone was isolated from \textit{M. tuberculosis}, which appears to regulate virulence in mouse infection studies (103), but the precise function of this molecule is also unknown.

**FIGURE 6** Structures of the predominant menaquinone and menaquinone sulfate reported from \textit{M. tuberculosis}. Carbon positions 2 and 3 and the \(\beta\)-isoprene unit are indicated by the arrows and call-out. doi:10.1128/microbiolspec.MGM2-0021-2013.f6
Menaquinone biosynthesis

The biosynthesis of menaquinone takes place via the intersection of two separate pathways. 1,4-Dihydroxy-2-naphthoate is synthesized via the shikimate pathway. The naphthoate ring is then prenylated with a prenyldiphosphate, derived from a series of prenyl transferase reactions, to form demethylmenaquinone and, subsequently, the C-2 position of the ring structure is methylated. The details of the biosynthesis of menaquinone studied in species other than *Mycobacterium* species have been reviewed (96, 104–106). In mycobacteria, the β-isoprene unit of the prenyl group is reduced to form MK-9 (II-H2) after the formation of menaquinone (107).

In *E. coli*, the synthesis of menaquinone is accomplished by seven enzymes (*menA* to *menG*). These enzymes are encoded by two clusters of genes: the *men* cluster consisting of *menB,C,D,E,F* and a separate cluster containing *menA* and *menG*. Menaquinone synthesis in Gram-positive bacteria in general has largely been ignored; however, the general pathway in *M. tuberculosis* appears to be similar. In *M. tuberculosis*, the *menA-E* genes appear to be found in a single cluster, whereas, the gene with the most homology to *menF* in *E. coli* is *Rv3215*, annotated as *entC* (isochorismate synthase).

Although menaquinone synthesis has been relatively extensively studied in *E. coli* (due in part to the availability of the *men* mutants, which can easily be generated in this organism, because it can utilize ubiquinone as an electron carrier in aerobic conditions), the synthesis of this compound in other organisms has received relatively little attention; however, MenB (1,4-dihydroxy-2-naphthoic acid synthase, *Rv0548c*) (108–110) (**Table 1**), MenE (o-succinylbenzoyl-CoA synthase, *Rv0542c*) (111–113), and MenA (Rv0534c) (114, 115) from mycobacteria have been studied as potential drug targets.

The isoprenoid tail of the menaquinone must be generated by a prenyldiphosphate synthase as described above, and together with 1,4-dihydroxy-2-naphthoic acid is the substrate for MenA (Rv0534c). However, the specific prenyldiphosphate synthase generating this prenyldiphosphate has yet to be identified. As noted above, other functions have been assigned to the potential candidates, suggesting that additional study is required.

In addition, the saturation of the second isoprene unit from the head group of menaquinone in mycobacteria (**Fig. 6**) is not seen in *E. coli* or *Bacillus subtilis*. However, this modification is seen in many Gram-positive bacteria (100, 116). Based on the chemical mechanism of the prenyl diphosphate synthases, it is likely that this modification is introduced after the mature prenyldiphosphate is synthesized and potentially after the formation of demethylmenaquinone or menaquinone.

There is a single report that cell free extracts of *Mycobacterium phlei* are capable of reducing MK-9 to MK-9 (II-H2) (107). The reduction required either NADH or NADPH, but nothing further has been reported regarding the nature of this enzyme, and it is, as yet, unknown whether this modification is required for function in mycobacteria.

Sulfated menaquinone

Sulfated menaquinone, where the sulfate is found on the ω-end of the isoprenoid tail (**Fig. 6**), has been isolated from *M. tuberculosis* (103). The function of this unique lipid is, as yet, unknown. However, it has been reported that sulfated menaquinone, previously known as S881, negatively regulates the virulence of the organism in mouse infection models (117). It has been postulated that this molecule is synthesized from MK-9 (II-H2) in at least two steps: (i) oxidation of the terminal position of the isoprenoid tail and (ii) sulfation of the resulting hydroxyl residue. It has been shown that the putative sulftotransferase encoded by *stf3* (*Rv2267c*) (**Table 1**), is required for the production of S881 (117) and has been hypothesized that Cyp128, encoded by *Rv2268c*, hydroxylates the MK-9 (II-H2). However, this remains to be definitively demonstrated, and Cyp124, encoded by *Rv2266*, has been shown to have appropriate ω-hydroxylase activity and a marked preference for lipids containing methyl branching such as isoprenoid compounds (118).

**Carotenoids**

The carotenoids of mycobacteria

Carotenoids are a diverse family of isoprenoids that typically have six to eight isoprene units. These molecules are structurally diverse but are similar in general structure with a long chain of conjugated double bonds. More than 700 carotenoids have been identified and are widespread among bacteria, including mycobacteria (**Fig. 7**). These often pigmented compounds play significant roles in protecting the organisms from oxidative damage and modify membrane fluidity (119, 120). The carotenoids can be divided into two classes based on the presence or absence of oxygen atoms. Carotenoids without oxygen atoms in the molecule are known as carotenes, whereas those with oxygen atoms in their structure are known as xanthophylls.
Many *Mycobacterium* species produce yellow, orange, or pink pigments in the dark (scotochromogens) or in the light (photochromogens), although these pigments may not be visible in culture. Very early on, mycobacteria were shown to contain carotenoid pigments (see reference 121 for a review). Chargaff reported the presence of carotenoid pigments in *M. phlei* in 1930, and subsequent analysis showed that the major carotenoid in *M. phlei* was leprotene (or iso-neriatene), a carotene that was first isolated from an organism mistakenly identified as *M. leprae* (96). In addition, many bacteria, including mycobacteria, produce carotenoid glycosides, which act as surfactants, stabilize membranes, and possibly contribute to regulating the permeability of membranes to oxygen (122–125). The first complete structure of glycosylated carotenoids, phleixanthophyll and 4-keto-phleixanthophyll isolated from *M. phlei*, was determined in 1967 (126).

**Carotenoid biosynthesis**

Carotenoid synthesis is well understood in many microorganisms (reviewed in reference 127) but has received limited attention in mycobacteria; however, the generally accepted pathway for carotenoid synthesis in mycobacteria, reviewed by Minnikin (96), appears to be similar to that of most nonphotosynthetic microbes (127). That is, the pathway consists of a geranylgeranyldiphosphate synthase, phytoene synthase, phytoene dehydrogenase, and lycopene cyclase. In the carotenoid literature these enzymes are designated CrtE, CrtB, CrtI, and CrtY, respectively. It should be noted that in nonphotosynthetic bacteria, CrtI catalyzes multiple dehydrogenations (usually two to four) that generate the conjugated double bond system and that there are multiple CrtY type cyclases with multiple designations (127). Once lycopene has been generated in mycobacteria, the pathway splits to form α- and β-carotene (96), one of which is presumably the precursor of leprotene.

As described above, orthologs of Rv0562 and Rv3383c, both of which are reported to encode E,E,E-geranylgeranyldiphosphate synthases (87), have the potential to provide the CrtE functionality in mycobacteria. Studies, aimed primarily at the development of genetic tools for manipulating mycobacteria, have provided information about other genes and enzymes involved in carotenoid synthesis in mycobacteria as well. Thus, orthologs of CrtB, CrtI, and CrtY have been identified in *M. marinum* (128, 129) and *M. aurum* (130, 131). In addition, an ortholog of CrtU, a β-carotene desaturase, has been reported in *M. aurum* (131), and a carotenoid oxygenase, Rv0654, has been identified in *M. tuberculosis* (132). In terms of regulation of carotenoid synthesis in mycobacteria, orthologs of *crtR* and *crtP* encode a putative repressor and a positive regulator, respectively, in *M. marinum* and *M. tuberculosis* (128), and SigF controls carotenoid production in *M. smegmatis* (133). Details regarding carotenoid synthesis in *M. tuberculosis*...
are not clear. The *M. tuberculosis* H37Rv genome encodes an ortholog of CrtB (PhyA), which may be nonfunctional (129).

**Noncarotenoid Cyclic Isoprenoids**
A novel class of cyclic C\textsubscript{35} terpenes isolated from nonpathogenic *Mycobacterium aichiense*, *Mycobacterium chlorophenolicum*, *Mycobacterium parafortuitum*, *M. smegmatis*, *Mycobacterium thermoresistible*, and *Mycobacterium vanbaalenii* has been described (84, 134). These compounds, designated heptaprenylcyclines (Fig. 8), are synthesized via the cyclization of \(\omega,E\), polyZ-heptaprenyldiphosphate or \(\omega,E,E\),polyZ-heptaprenyldiphosphate; thus, the prenyldiphosphate synthases described in these species are likely involved in the production of these molecules, but little else is currently known about their synthesis or function.

A labdane-related diterpenoid compound, isotubercolusinol (Fig. 8), is produced by *M. tuberculosis*. This molecule appears to be immunomodulatory because it has been shown to block phagosome maturation in macrophages (135, 136). This role was first suggested when genes encoding enzymes involved in isotubercolusinol synthesis, *Rv3377c* and *Rv3378c*, were identified in a screen for mutants defective in arresting phagosome maturation (137). *Rv3377c* (Table 1) was demonstrated to be a class II diterpene cyclase, catalyzing bicyclization and rearrangement of geranylgeranyldiphosphate to form halimadienyl/tubercolusinylidiphosphate (138). It was then shown that halimadienyl/tubercolusinylidiphosphate was hydrolyzed to tuberculosinol and isotubercolusinol by *Rv3378c* (135, 136, 139, 140).

**ACYLTREHALOSES**
The outer membrane of mycobacteria contains a number of trehalose esters. Among them, trehalose monomycolates (TMMs) and trehalose dimycolates (TDMs; cord factor) are ubiquitously found across the *Mycobacterium* genus. Species-specific trehalose esters include di-, tri-, and poly-acyltrehaloses (DATs, TATs, and PATs); sulfolipids (SLs); and lipooligosaccharides (LOSs). Species-specific trehalose esters are found in the outermost capsule in addition to the outer membrane (8). TMM and TDM, in contrast, were identified in the surface-exposed capsular materials of *M. avium* and *M. smegmatis* but not in those of *M. tuberculosis*, *M. kansasii*, and *M. gastri*, indicating that they may be more deeply buried in the cell envelope of some *Mycobacterium* species (8). Interest in trehalose esters stems from their demonstrated or postulated roles in host-pathogen interactions and from their potential as diagnostic tools (for reviews see references 1, 141).

The presence and abundance of species-specific acyltrehaloses (SL, DAT, TAT, and PAT) and phthiocerol dimycocerosates (PDIMs; see “PDIMs, PGLs, and Related Compounds,” below) in the cell envelope of *M. tuberculosis* impact the ability of the bacilli to stain with the cationic dye neutral red (142, 143), a property known since Dubos and Middlebrook’s early studies in the 1940s to correlate with virulence (144).

Here, we focus on steps in the formation of acyltrehaloses, including the biosynthesis of the fatty acyl substituents, their transfer onto trehalose, and what is known of the translocation of biosynthetic intermediates and end products across the cell envelope.

**FIGURE 8** Structures of representative noncarotenoid cyclic isoprenoids found in mycobacteria. doi:10.1128/microbiolspec.MGM2-0021-2013.f8
TMMs and TDM (Cord Factor)

In TMM and TDM, trehalose is esterified with long-chain α-branched β-hydroxy fatty acids known as the mycolic acids. The structure and biosynthesis of mycolic acids is reviewed in reference 364. Any structural type of mycolic acid may esterify positions 6 and 6′ of TDM and position 6 of TMM (Fig. 9). The biosynthesis of mycolic acids occurs in the cytoplasm, and so does that of trehalose. We recently identified MmpL3 (Rv0206c) (Table 2) as an inner membrane transporter required for the translocation of TMM to the periplasm, where TMM can then serve as a mycolic acid donor for the mycolylation of arabinogalactan and the formation of TDM (Fig. 1) (145, 146). This finding indicates that TMM is most likely the form under which mycolic acids are exported to the cell wall and outer membrane and, therefore, that TMM is probably made on the cytosolic side of the plasma membrane. The catalytic process underlying the cytoplasmic formation of TMM from fully elongated and functionalized mycolic acid chains and trehalose has not yet been elucidated. The subsequent synthesis of TDM from two TMM molecules and the transfer of mycolates to the nonreducing ends of arabinogalactan have been shown to involve antigens 85A (Rv3804c; FbpA), 85B (Rv1886c; FbpB), and 85C (Rv0129c; FbpC) (Table 2) (147–149). In vitro, these three mycolyltransferases display apparent redundant catalytic activities (147). Consistent with this finding, none of the fbpA, B, or C genes are individually required for the growth of M. tuberculosis. Their combined inactivation or chemical inhibition, however, leads to cell death (147, 150) (our unpublished data). Although the phenotypic characterization of fbpA, B, or C null mutants of M. tuberculosis and M. smegmatis indicates that the function of these genes may in fact only partially overlap in whole cells, to date, the precise contribution of each of the three paralogs to the transfer of mycolic acids to their cell wall and outer membrane glycolipid acceptors remains unclear. FbpC appears to be essentially involved in the transfer of mycolic acids to arabinogalactan, and FbpA, in the formation of TDM (148, 149, 151–153).

Numerous biological activities have been associated with the TDM from tuberculous and nontuberculous mycobacteria both in vitro and in vivo (for reviews see references 1, 154–157). In fact, TDM seems to be a major contributor to the inflammation seen in mycobacterial infections. TDM contributes to protecting M. tuberculosis from killing by macrophages, is a potent modulator of the activation of macrophages, stimulates the formation of lung granulomas, and enhances the resistance of mycobacteria to antibiotics (152, 154, 156, 158, 159). The binding of TDM from M. tuberculosis to the C-type lectin Mincle is required for activation of macrophages and granuloma formation (158, 160). It is important to note that the biological activities of TDM are very dependent on the fine structure of their mycolyl substituents (156, 161).

Sulfolipids

Sulfolipids (SLs), also known as sulfatides and sulfoglycolipids, are sulfated trehalose esters that are acylated with three or four acyl groups consisting of one middle-chain saturated fatty acid (palmitic or stearic acid) at the 2-position and different combinations of the heptadecyl and octa-methyl-branched phthioceranic and hydroxyphthioceranic acids (C31 to C46) at the 3′-, 6'-, and 6′-positions. Monomethyl-branched unsaturated C16 to C20 fatty acids have also been found as minor constituents of SL (162). Sulfolipid-1 (SL-1), whose structure is shown in Fig. 10, is the most abundant form of sulfolipid produced by M. tuberculosis (163). This family of lipids is exclusively found in the human pathogen M. tuberculosis.
The genes involved in the biogenesis of SL-1 have for the most part been identified and, with the exception of the sulfotransferase Sft0, found to cluster on the chromosome of *M. tuberculosis* (Table 2). The sulfotransferase Sft0 (Rv0295c) catalyzes the first committed step in the pathway by sulfating trehalose to form trehalose-2-sulfate (164). The acyltransferase PapA2 (Rv3820c) then catalyzes the esterification of trehalose-2-sulfate with a straight-chain saturated fatty acid (e.g., palmitic acid) at the 2'-position to generate a monoacyl intermediate, SL659 (165). The polyketide synthase Pks2 (Rv3825c) synthesizes the methyl-branched phthioceranic and hydroxyphthioceranic acids (166), most likely using an activated long-chain fatty acid starter unit (an acyl-adenylate) provided by the fatty acid AMP ligase FadD23 (Rv3826) (167). The polyketide-associated

### Table 2

*M. tuberculosis* H37Rv genes involved in the biogenesis of trehalose mono- and dimycolates, sulfolipids, di- and poly-acyltrehaloses, and mannosyl-β-1-phosphomycoketides

<table>
<thead>
<tr>
<th>Gene number</th>
<th>Gene name</th>
<th>Function</th>
<th>Evidence*</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv0129c</td>
<td>fbpC</td>
<td>Mycolyltransferase (antigen 85C)</td>
<td>E, P</td>
<td>147-149</td>
</tr>
<tr>
<td>Rv0206c</td>
<td>mmpL3</td>
<td>Inner membrane transporter of the RND superfamily involved in the translocation of TMM</td>
<td>P</td>
<td>145, 146</td>
</tr>
<tr>
<td>Rv0295c</td>
<td>Sft0</td>
<td>Sulfotransferase responsible for the formation of the trehalose-2-sulfate moiety of sulfolipids</td>
<td>E, P</td>
<td>164</td>
</tr>
<tr>
<td>Rv0757-Rv0758</td>
<td>phoP-phoR</td>
<td>Two-component transcriptional regulator involved in the regulation of SL, DAT, and PAT</td>
<td>P</td>
<td>143, 176</td>
</tr>
<tr>
<td>Rv1180-Rv1181</td>
<td>pks3-pks4</td>
<td>Polypeptide synthase responsible for the elongation of the methyl-branched mycosanoic and mycolipenic acids found in DAT, TAT, and PAT</td>
<td>P</td>
<td>196, 197</td>
</tr>
<tr>
<td>Rv1182</td>
<td>papA3</td>
<td>Acyltransferase catalyzing the sequential transfer of the first straight-chain saturated fatty acyl chain followed by the first mycolipenoyl group onto the 2- and 3-positions of trehalose, respectively, in the biosynthesis of DAT and PAT</td>
<td>E, P</td>
<td>199</td>
</tr>
<tr>
<td>Rv1183</td>
<td>mmpL10</td>
<td>Inner membrane transporter of the RND superfamily thought to be involved in the translocation of DAT and PAT</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>Rv1184c</td>
<td>chp2</td>
<td>Acyltransferase thought to catalyze the last three acylations leading to the formation of PAT from DAT</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>Rv1185c</td>
<td>fadD21</td>
<td>Putative fatty acid AMP ligase providing Pks3/4 with activated long-chain fatty acid starter units</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>Rv1662-Rv1663</td>
<td>pks8-pks17</td>
<td>Polypeptide synthase responsible for the elongation of the monomethyl-branched unsaturated C_{16} to C_{20} fatty acids found in DAT and PAT</td>
<td>P</td>
<td>162</td>
</tr>
<tr>
<td>Rv1886c</td>
<td>fbpB</td>
<td>Mycolyltransferase (antigen 85B)</td>
<td>E</td>
<td>147, 149</td>
</tr>
<tr>
<td>Rv2048c</td>
<td>pks12</td>
<td>Polypeptide synthase involved in the elongation of the alkyl backbone of mycocetides</td>
<td>E, P</td>
<td>218, 220</td>
</tr>
<tr>
<td>Rv3416</td>
<td>whiB3</td>
<td>Regulator of SL, DAT, and PAT synthesis</td>
<td>P</td>
<td>174</td>
</tr>
<tr>
<td>Rv3804c</td>
<td>fbpA</td>
<td>Mycolyltransferase (antigen 85A)</td>
<td>E, P</td>
<td>147, 149, 152</td>
</tr>
<tr>
<td>Rv3820c</td>
<td>papA2</td>
<td>Acyltransferase catalyzing the transfer of the first straight-chain saturated fatty acyl chain onto trehalose-2-sulfate in the biosynthesis of sulfolipids</td>
<td>E, P</td>
<td>165</td>
</tr>
<tr>
<td>Rv3821</td>
<td>sap</td>
<td>Integral membrane protein thought to facilitate the translocation of SL-1 to the cell surface</td>
<td>P</td>
<td>168</td>
</tr>
<tr>
<td>Rv3822</td>
<td>chp1</td>
<td>Acyltransferase catalyzing the acylation at the 6- and 6'-positions of sulfolipids</td>
<td>E, P</td>
<td>168</td>
</tr>
<tr>
<td>Rv3823c</td>
<td>mmpL8</td>
<td>Inner membrane transporter of the RND superfamily involved in the translocation of sulfolipids</td>
<td>P</td>
<td>168, 170, 171</td>
</tr>
<tr>
<td>Rv3824c</td>
<td>papA1</td>
<td>Acyltransferase catalyzing the transfer of the first (hydroxy) phthioceranoyl group at the 3'-position of the product of PapA2</td>
<td>E</td>
<td>165</td>
</tr>
<tr>
<td>Rv3825c</td>
<td>pks2</td>
<td>Polypeptide synthase responsible for the elongation of the methyl-branched phthioceranic and hydroxyphthioceranic acids found in sulfolipids</td>
<td>P</td>
<td>166</td>
</tr>
<tr>
<td>Rv3826</td>
<td>fadD23</td>
<td>Putative fatty acid AMP ligase providing Pks2 with activated long-chain fatty acid starter units</td>
<td>H</td>
<td></td>
</tr>
</tbody>
</table>

*E, P, H: see Table 1 footnote.
FIGURE 10 Structures of SLs, DATs, and PATs and biosynthetic gene clusters. The major sulfolipid, SL-I (2,3,6,6′-tetraacyl α-α′-trehalose-2′-sulfate), is represented. In SL-I, trehalose is sulfated at the 2′ position and esterified with palmitic acid and the multimethyl-branched phthioceranic and hydroxyphthioceranic acids. In DAT (2,3-di-O-acyltrehalose), trehalose is esterified with palmitic acid and the multimethyl-branched mycosanoic acid. In PAT, trehalose is esterified with palmitic acid and the multimethyl-branched mycolipenic acids. doi:10.1128/microbiolspec.MGM2-0021-2013.f10
protein-1 (PapA1; Rv3824c) catalyzes the transfer of the first (hydroxy)phthioceranoyl group at the 3′-position of the product of PapA2, yielding a diacylated form of SL known as SL-1278 (165). The additional two acylations at the 6′- and 6′-positions of SL-1278 are catalyzed by the acyltransferase Chp1 (Rv3822) (168). PapA1 and PapA2 are related to the acyltransferase PapA5, which esterifies phthiocerol with mycocerosic acids in the biosynthesis of PDIM (see “PDIMs, PGLs, and Related Compounds,” below). Chp1 (cutinase-like hydrolase protein-1), in contrast, more closely resembles cutinase-like proteins (168). All three acyltransferases are essential for the synthesis of SL-1 as demonstrated by the absence of fully elaborated SL-1 from the corresponding knockout mutants (165, 168, 169).

Evidence of the involvement of MmpL8 (Rv3823c), an inner membrane transporter of the RND (resistance, nodulation, and division) superfamily, in the translocation of SL-1 to the cell surface was provided in 2003–2004 by two independent research groups (170, 171). M. tuberculosis mmpL8 knockout mutants failed to produce SL-1 and instead accumulated the diacylated SL-1278 intracellularly. A possible finding of this finding was that the first two acylation steps catalyzed by PapA2 and PapA1 occurred on the cytoplasmic side of the plasma membrane, whereas the two subsequent acylations catalyzed by Chp1 and yielding SL-1 required the prior MmpL8-mediated translocation of the diacylated SL-1278 precursor across the plasma membrane. This model was, however, recently revised in light of the finding that the catalytic domain of the membrane-associated acyltransferase Chp1 is cytosolic and that its activity is potentiayed by MmpL8 (168). These observations are consistent with a model similar to that proposed for PDIM (see “PDIMs, PGLs, and Related Compounds,” below) wherein the biosynthesis and transport of SL-1 is coupled and MmpL8 acts as a scaffold for a cytoplasmically oriented macromolecular complex consisting of the SL biosynthetic machinery. Further support for this assumption was recently obtained by Zheng et al. (172) in identifying MmpL8 among the components of a membrane-associated protein complex containing Pks2, PapA1, and FadD23 in M. bovis BCG. Sap (sulfolipid-1-addressing protein) (Rv3821) is an integral membrane protein that appears to facilitate the translocation of SL-1 to the cell surface. Its disruption in M. tuberculosis causes the intracellular build-up of SL-1278 similar to that observed in mmpL8 knockouts, although the sap mutant retains the ability to synthesize small amounts of SL-1 (168). Beyond MmpL8 and Sap, it is likely that the translocation of SL-1 to the cell surface requires additional periplasmic and/or outer membrane transporters, but their identity is at present not known.

SL production appears to be regulated in M. tuberculosis, but the environmental factors governing the synthesis of these glycolipids are still poorly understood. Supporting a role for SL during host infection, the expression of the pks2 gene was found to be strongly upregulated upon phagocytosis of M. tuberculosis by human primary macrophages (173). It appears that one of the roles of methyl-branched fatty acid–containing lipids such as PDIM, SL, DAT, and PAT during infection is to alleviate the propionate-mediated stress undergone by M. tuberculosis when the bacterium switches to host cholesterol as a major carbon source (174, 175). The propionyl-CoA generated upon β-oxidation of cholesterol is converted to methylmalonyl-CoA by the propionyl-CoA carboxylase, which is then used by dedicated polyketide syntheses such as Pks2, Mas, and Pks3/4 (see below) in the elongation of the methyl-branched fatty acids found in PDIM, SL, DAT, and PAT. The regulator facilitating this metabolic switching to fatty acids was identified as WhiB3 (Rv3416), which binds the promoter region of pks2 (174). Another important regulator of SL production is the two-component transcriptional regulator PhoP-PhoR (Rv0757-Rv0758). PhoP-PhoR positively regulates the synthesis of SL, and M. tuberculosis mutants deficient in the expression of this regulator are totally deficient in SL-1 production (143, 176). It was shown that a mutation in the phoP gene of M. tuberculosis H37Ra accounts for the inability of this avirulent strain to produce SL-1 (177). PhoP binds the promoter region of pks2 in vitro (178, 179).

The restriction of SL-1 to the human pathogen M. tuberculosis together with the observation some 50 years ago of a positive correlation between the levels of SL-1 produced by M. tuberculosis clinical isolates and their virulence in animal models has prompted extensive research aimed at elucidating the biological functions of sulfolipids during host infection (for reviews see references 141, 163, 180, 181). Numerous and sometimes controversial activities were associated with purified SL-1 molecules. Among these, the ability of SL-1 to potentiate the toxicity of TDM in mice, to inhibit mitochondrial oxidative phosphorylation, to prevent phagosome-lysosome fusion in cultured macrophages, and to modulate the oxidative and cytokine responses of human monocytes and neutrophils has probably received the most attention. In more recent years, the diacylated SL biosynthetic precursor SL-1278 was
shown to stimulate CD1b-restricted T cells through mechanisms dependent on the number of C-methyl substituents on the fatty acyl chains, the configuration of the chiral centers, and the length and respective localization of the two acyl chains on the sugar moiety (182, 183). In the last decade, the elucidation of the biosynthetic pathway of SL finally allowed the generation of isogenic mutants of \textit{M. tuberculosis} specifically deficient in their synthesis and an evaluation of the roles of these glycolipids during infection when carried by whole bacilli.

Unexpectedly, \textit{pks2}, \textit{papA1}, and \textit{papA2} knockout mutants, which all lack fully elaborated SL-1 while in some cases retaining the ability to synthesize sulfated trehalose and mono- and/or di-acylated forms of SL, were found to be undistinguishable from their wild-type parents in their ability to replicate and persist in mice or guinea pigs (165, 184). In contrast, three independent studies indicated that \textit{mmplL8} knockout mutants that accumulate diacylated SL-1278 at the periphery of the plasma membrane display some level of attenuation in mice, although the attenuation phenotypes considerably differed between studies, possibly as a result of the different \textit{M. tuberculosis} strains and models of infection that were used (170, 185, 186). Recently, Gilmore et al. (187) provided evidence that a \textit{sft0} null mutant of \textit{M. tuberculosis} survives better than its wild-type parent in human but not in murine macrophages, possibly as a result of the increased resistance of this strain to human antimicrobial peptides. These results suggest that SL may only have a detectable impact on infection in the human host.

\textbf{DATs and PATs}

The 2,3-di-O-acyltrehaloses (DATs) consist of trehalose acylated at the 2-position with one middle-chain saturated fatty acid (C16 to C19) and at the 3-position with the di-methyl-branched mycosanoic acids (C21 to C25) (Fig. 10). In other less common forms of DAT, the tri-methyl-branched C25 to C27 mycolipenic (phtthienoic) or mono-hydroxyacetyl tri-methyl-branched C24 to C28 mycolipanolic acids replace the mycosanoic acids (188–190). 2,3,6-Triacyltrehaloses (TATs) harboring stearic, palmitic, and mycolipenic acyl substituents have also been reported in \textit{M. tuberculosis} (191). Polyacyltrehaloses (PATs) are trehalose esters acylated with five acyl groups consisting of one middle-chain saturated fatty acid (C16 to C19) at the 2-position and different combinations of the tri-methyl-branched C27-mycolipenic and C27-mycolipanolic acids at the 2’, 3’, 4, and 6’-positions (Fig. 10) (188, 192). Monomethyl-branched unsaturated C16 to C20 fatty acids have also been found as minor constituents esterifying PAT and DAT (162). So far, the mycolipenic acyl substituents found in DAT, TAT, and PAT have only been isolated from virulent isolates of the \textit{M. tuberculosis} complex species \textit{M. tuberculosis}, \textit{M. bovis}, and \textit{M. africanum} but were not found in the avirulent laboratory strain \textit{M. tuberculosis} H37Ra or in the vaccine strain \textit{M. bovis} BCG. While 2,3-diacyltrehaloses and 2,3,4- and 2,3,6-triacyltrehaloses may be found in nonpathogenic species of mycobacteria such as \textit{M. fortuitum}, the fatty acyl substituents identified in this species consist of straightchain (C14 to C18) and mono-methyl-branched unsaturated C16 to C20 fatty acids (193, 194).

As their relative distribution to pathogenic species of the \textit{M. tuberculosis} complex may suggest, DAT, TAT, and PAT are biologically active molecules capable of modulating a number of host immune responses \textit{in vitro} (141, 195). Their precise role during host infection remains, however, poorly understood. Phenotypic observations made on a mutant of \textit{M. tuberculosis} deficient in the biosynthesis of DAT and PAT indicated a role for these lipids in the retention of the capsular material at the cell surface (196, 197). The modification of the surface properties of the mutant affected its binding and uptake by phagocytic and nonphagocytic cells, but preliminary infection studies indicated that the mutant did not significantly differ from its wild-type parent in its ability to replicate and persist in cultured macrophages and in mice (197). Interestingly, increased binding to phagocytic cells was also reported in the case of an SL-deficient mutant of \textit{M. tuberculosis} (198). It is thus likely that the different families of acyltrehaloses produced by \textit{M. tuberculosis} have partially redundant activities in whole cells hampering the clear delineation of their individual contribution to virulence and other physiological functions. Independent from their binding or immunomodulatory properties and as noted above, methyl-branched fatty acid–containing lipids such as PDIM, SL, DAT, and PAT appear to play an important role in alleviating the propionate-mediated stress undergone by \textit{M. tuberculosis} when the bacterium utilizes host cholesterol as a major carbon source during infection (174, 175). Consistently, WhiB3 acts as a positive transcriptional regulator of \textit{pks3/4} in addition to \textit{pks2} (174).

Gene knockout studies indicated that the polyketide synthase encoded by \textit{pks3/4} (\textit{Rv1180/Rv1181}) is responsible for the elongation of mycosanoic and mycolipenic acids, while \textit{pks8} and \textit{pks17} (\textit{Rv1662} and \textit{Rv1663}) together encode the polyketide synthase...
producing monomethyl-branched unsaturated C_{16} to C_{20} fatty acids (162, 196, 197) (Table 2). An M. tuberculosis mutant deficient in the expression of \textit{pks3/4} failed to produce PAT and DAT (196, 197). In some \textit{M. tuberculosis} strains, an intervening stop codon in \textit{pks3/4} results in two separate open reading frames (ORFs) (annotated as \textit{pks3} and \textit{pks4}). Strains containing this mutation do not synthesize PAT (186). Striking resemblance in the genetic organization of the regions encompassing the polyketide synthase gene \textit{pks3/4} and that involved in SL (Fig. 10) and, to a lesser extent, PDIM biosynthesis (see “PDIMs, PGLs, and Related Compounds,” below) are suggestive of the involvement of \textit{fadD21} (\textit{Rv1185c}), \textit{mmpL10} (\textit{Rv1183}), and \textit{Rv1184c} (\textit{chp2}) in the assembly and export of DAT and PAT (Table 2). To date, however, only \textit{papA3} (\textit{Rv1182}) has been characterized (199). It encodes the acyltransferase necessary for the sequential transfer of a palmitoyl group at the 2-position of DAT/PAT followed by a mycolepenoyl group at the 3-position (Fig. 10). As is the case for SL, the two-component transcriptional regulator PhoP-PhoR (\textit{Rv0757-Rv0758}) positively regulates the synthesis of DAT and PAT, and \textit{M. tuberculosis} mutants deficient in the expression of this regulator are totally deficient in DAT and PAT production (143, 176). The same mutation in the \textit{phoP} gene of \textit{M. tuberculosis} H37Ra that accounts for the inability of this strain to produce SL also accounts for the absence of DAT and PAT from this avirulent \textit{M. tuberculosis} isolate (177). PhoP was shown to bind the promoter regions of \textit{pks3/4} and \textit{fadD21} (178, 179).

**Lipoooligosaccharides**

Lipoooligosaccharides (LOSs) are surface-exposed glycolipids (8) produced by a number of \textit{Mycobacterium} species (200). They were first found in \textit{M. kansasii} (201) and \textit{M. smegmatis} (202), then in nine other mycobacterial species (200), including “\textit{canetti}” and related strains of the \textit{M. tuberculosis} complex (203). LOSs are otherwise virtually absent from \textit{M. tuberculosis} strains \textit{sensu stricto} such as H37Rv (203). LOSs (Fig. 11A and 11B) share a poly-O-acetylated trehalose core further glycosylated by a mono- or, more frequently, an oligosaccharidyl unit (200). Similar to the situation in other trehalose-based mycobacterial glycolipids such as sulfolipids and di- or tri-acetyltrehaloses, the trehalose moiety of LOS is invariably acylated by polymethyl-branched fatty acids that can be either saturated, e.g., in “\textit{canetti},” or unsaturated, e.g., in \textit{M. smegmatis}.

The biosynthesis of LOS molecules is still poorly understood, with only a few genes experimentally demonstrated to be involved in their elongation and assembly (204–206). The synthesis of polymethyl-branched fatty acids invariably requires a polyketide synthase (Pks) that uses methylmalonyl-CoA instead of malonyl-CoA as the elongation unit, resulting in the formation of a polymethyl branched aliphatic chain. The MSMEG_4727 (\textit{pks5}) gene, whose sequence is 65.6% identical to that of the \textit{M. tuberculosis} Mas-like gene \textit{Rv1527c}, was involved in the biosynthesis of LOS in \textit{M. smegmatis} (207). The genomic surroundings of \textit{pks5} from \textit{M. smegmatis} resemble those described earlier for other acyltrehaloses (see subsections covering SLs, DATs, and PATs in “Acyltrehaloses,” above, and Fig. 9) in that \textit{pap}- and \textit{fadD}-like genes likely to be required for the activation and transfer of the acyl groups of LOS (206), and an \textit{mmpL} gene putatively involved in the translocation of these lipids, are found (Fig. 11C). In addition, genes whose products were tentatively annotated as polysaccharide pyruvyltransferases are found in the biosynthetic cluster of pyruvylated LOS-producing species such as \textit{M. smegmatis} (202, 208). Consistent with the finding of various methylated glucosyl residues in LOS (Fig. 11A), genes encoding putative glycosyltransferases and O-methyltransferases also map in the vicinity of \textit{pks5} (Fig. 11C). In \textit{M. marinum}, several of these have been characterized (204, 205, 209). It is noteworthy that orthologs of five of the \textit{M. smegmatis} LOS-related genes (\textit{pks, pap, fadD, mmpL, and gap}) are conserved in the corresponding biosynthetic gene clusters of \textit{M. marinum} and \textit{M. tuberculosis} (205). Interestingly, homologous genes are also found in the glycopeptidolipid (GPL) biosynthetic gene clusters of \textit{M. smegmatis} (210), \textit{M. abscessus} (211), \textit{M. chelonae} (211), and \textit{M. avium} (212) (see “Glycopeptidolipids,” below). This conserved set of genes may delineate the minimum biosynthetic machinery required for the synthesis and export of GPL- and LOS-type glycolipids in mycobacteria. The remaining ORFs identified in the confirmed or putative mycobacterial LOS biosynthetic clusters are less conserved, an observation consistent with the fact that LOSs differ from other mycobacterial glycolipids in terms of the number and nature of their sugar constituents (200). Recently, the regulatory protein WhiB4 from \textit{M. marinum} was associated with LOS biosynthesis, but its precise function is not known (213).

LOSs are highly antigenic molecules (203). Recent observations suggest that they play an important role in retaining proteins at the cell surface of some \textit{Mycobacterium} species such as \textit{M. marinum} (213). Their precise
FIGURE 11  Structures of (A) major LOS (LOS-A) of *M. smegmatis* ATCC 356 (R<sub>1</sub> and/or R<sub>2</sub> : octanoic acid and tetra- or hexa-decanoic acid) and (B) *M. tuberculosis* “canettii”; R = Ac. (C) LOS biosynthetic gene cluster of *M. smegmatis* mc<sup>2</sup>155. Shown is the 25.15-kb region spanning MSMEG<sub>4727</sub> (pks5) to MSMEG<sub>4741</sub> (mmpL). ORFs are depicted as arrows. Black arrows indicate genes encoding biosynthetic enzymes; gray arrows indicate putative transporter genes; white arrows show hypothetical genes of unknown function. Abbreviations: Pks5, Mas-like polyketide synthase; Pap, putative acyltransferase; MSMEG<sub>4729</sub> and MSMEG<sub>4730</sub>, putative acyltransferases; FadD, putative acyl-CoA synthase; Gtf (MSMEG<sub>4732</sub>), putative glycosyltransferase; Gap2, putative transmembrane protein involved in glycolipid translocation; MSMEG<sub>4734</sub>, hypothetical PE/PPE-like protein; Gtf (MSMEG<sub>4735</sub>), putative glycosyltransferase; MSMEG<sub>4736</sub> and MSMEG<sub>4737</sub>, putative pyruvyl transferases; MSMEG<sub>4738</sub>, hypothetical protein; Mtf, possible O-methyltransferase; Gtf (MSMEG<sub>4740</sub>), putative glycosyltransferase; MmpL, putative inner membrane transporter. doi:10.1128/microbiolspec.MGM2-0021-2013.f11
role in the colony morphology of mycobacteria is still a matter of debate and seems to be species-specific (205, 214, 215). In M. marinum, for instance, LOSs have clearly been associated with colony morphology, sliding motility, biofilm formation, and the ability of this Mycobacterium species to enter macrophages (205). The M. marinum LOSs are also endowed with immunomodulatory activities (216) and modulate virulence in the zebrafish embryo model of infection (213).

MANNOSYL-\(\beta\)-1-PHOSPHOMYCOKETIDES

Mannosyl-\(\beta\)-1-phosphomycoketides consist of a mannosyl-\(\beta\)-1-phosphate moiety reminiscent of polyprenol phosphomannose (the lipid-linked mannose donor) and an alkyl chain of varying length (C30 to C34) and devoid of fully saturated 4,8,12,16,20-pentamethylpentacosyl unit (Fig. 12). Mycoketides were first isolated from M. avium based on their ability to activate human CD1c-restricted T-cells (217). This family of lipids was later identified in the slow-growing pathogenic species M. tuberculosis and M. bovis BCG but not in the rapidly growing saprophytes, M. phele, Mycobacterium fallax, and M. smegmatis (218). Under standard liquid culture conditions, mycoketides are produced in minute amounts and are found both inside the cells and released in the culture medium (219). Their restricted distribution to pathogenic slow-growing Mycobacterium species are suggestive of an involvement in pathogenicity, and several studies aimed at comparing the virulence of mycoketide-deficient mutants of M. tuberculosis, M. avium, and M. marinum to that of their wild-type parents in animal models of infection have provided support for this assumption (219). In addition to their potential role in modulating the host immune response, mycoketides were proposed to be mycobacterial secondary metabolites acting as signaling factors to regulate cell division and virulence and to contribute to the suppression of phagosomal acidification (219).

The alkyl backbone of mycoketides is elongated by the polyketal synthase Pks12 (Rv2048c) (Table 2) (218, 220). Pks12 is the largest predicted protein of M. tuberculosis (430 KDa) and consists of two complete sets of fatty acid synthase--like catalytic domains capable together of using alternating C2 (malonyl-CoA) and C3 (methylmalonyl-CoA) units to elongate the alkyl backbone of mycoketides. After five cycles of C3 and C2 chain elongation, it is believed that the alkyl chain is released from the polyketal synthase upon hydrolysis, yielding mycoketidic acid, which is further reduced to the corresponding long-chain alcohol, mycoketide, and finally phosphorylated and mannosylated to generate mannosyl-\(\beta\)-1-phosphomycoketides (219). The enzymes catalyzing the hydrolysis, reduction, phosphorylation, and mannosylation steps have not yet been identified. The finding of orthologs of pks12 in M. marinum, M. ulcerans, M. avium paratuberculosis, and several species of the M. tuberculosis complex suggests that the production mannosyl-\(\beta\)-1-phosphomycoketides may be a common feature of slow-growing mycobacterial pathogens.

PDIMs, PGLs, AND RELATED COMPOUNDS

Phthiocerol Diesters and Related Compounds: Structures, Distribution, and Cell Localization

Phthiocerol dimycocerosates (PDIMs) and diphthioceranates are part of a family of long-chain C33-C41 \(\beta\)-diols (phthiocerols) esterified by two moles of polymethyl-branched (C27-C34) fatty acids. When the configuration of the asymmetric centers bearing the methyl branches belong to the D series, the fatty acids are called mycocerosic acids, whereas those of the L series are known as phthioceranic acids (14) (Fig. 13). The major \(\beta\)-diols (phthiocerol A) are usually accompanied by structural variants of these alcohols containing either a keto group in place of the methoxy group (phthiodiolone A) or a methyl group rather than

![FIGURE 12](https://doi.org/10.1128/microbiolspec.MGM2-0021-2013.f12)


ASMscience.org/MicrobiolSpectrum
an ethyl group at the terminus of the molecules and near the methoxyl group (phthiocerol B) (Fig. 13). To date, PDIMs have been found in *M. tuberculosis*, *M. bovis*, *M. leprae*, *M. microti*, *M. kansasi*, *M. gastri*, and *M. haemophilum*, whereas diphthioceranates have been found in *M. ulcerans* and *M. marinum* (221).

Glycosylated phenolic derivatives of PDIM and DIP, called phenolic glycolipids (PGLs), are found in the same species, although they may not be present in all strains; for instance, the PGL from *M. tuberculosis* (PGL-tb) has only been identified in the “canettii” strain (222, 223) and in some East-Asian/Beijing isolates (224, 225). In PGLs, the β-diols (phenolphthiocerols) are esterified by two moles of polymethyl-branched (C27–C34) fatty acids, except in Beijing strains, where a palmitic acid is found esterifying the additional hydroxyl group occurring in the aliphatic core of phenolphthiotriol (225). The glycosyl moiety of PGL is composed of one to four sugar residues depending on the species, most of which are O-methylated deoxysugars (14, 200). Identical PGL structures may be found in phylogenetically related mycobacterial species, for instance, in species of the *M. tuberculosis* complex (*M. bovis*, *M. microti*, *M. pinnipedii*, and *M. africanum*), *M. kansasi* and *M. gastri*, and *M. marinum* and *M. ulcerans* (200, 226). The glycosyl moiety of PGL was also found attached to *p*-hydroxybenzoic acid, i.e., as methyl esters, to form *p*-hydroxybenzoic acid derivatives (*p*-HBADs) in *M. tuberculous* and *M. bovis* BCG (Fig. 13) (223). In an attempt to correlate the lipid content with the virulence of *M. tuberculosis* isolates, Goren and collaborators characterized a methoxylated phenolphthiocerol, the so-called attenuation indicator lipid (227). The correlation between the occurrence of this lipid and reduced virulence remains, however, unclear. This lipid and its

**FIGURE 13** Structures of the PDIMs, PGLs, and *p*-hydroxybenzoic acid derivatives (*p*-HBADs) of *M. tuberculosis*. In *M. tuberculosis*, *p*, *p’* = 3–5; *n*, *n’* = 16–18; *m* = 15–17; *m*1 = 20–22; R = CH2–CH3 or CH3. doi:10.1128/microbiolspec.MGM2-0021-2013.f13
unmethylated form were detected in East-Asian/Beijing strains and accumulated in all of the Indo-Oceanic strains of *M. tuberculosis* examined (228).

PDIMs and PGLs are found in the capsules of *M. tuberculosis* and other pathogenic mycobacteria (8). PDIMs are otherwise abundant components of the outer membrane of *M. tuberculosis*, where they contribute to its well-known impermeability (229). p-HBADs, in contrast, are released in culture filtrates and tend not to remain associated with the cell envelope (223).

**Biosynthesis of PDIM and PGL**

**Biosynthesis of phthiocerol and related compounds**

Common enzymes participate in the biosynthesis of the lipid core of PDIM and PGL. (Fig. 13), where *n*-C22-C24 fatty-acyl chains and *p*-hydroxyphenylalkanoates, respectively, are elongated to form the long-chain β-diols, phthiocerol or phenolphthiocerol. Coupled genetic and biochemical strategies have allowed much of the biosynthetic pathways of PDIM and PGL to be elucidated. On the basis of mutant phenotypes, genes such as *pks11* (230), *pks12* (231), *pks10* (232), *mb0100* (233), and *pks7* (234) have been associated with the biosynthesis of PDIM; however, in the absence of genetic complementation, definitive proof of their involvement in the pathway is lacking, and no clear biosynthetic roles have yet been assigned to these genes. The genes unambiguously demonstrated to participate in PDIM and PGL biosynthesis are shown in Fig. 14, and their specific function in the pathway is detailed in Fig. 15 and Table 3. They are clustered on a 73-kb fragment of the *M. tuberculosis* chromosome (Fig. 14), and the organization of this locus is apparently highly conserved in all PDIM/PGL-producing mycobacteria, with the exception of *M. leprae*, in which this locus is split into two loci.

The *M. tuberculosis* genome encodes 36 FadD proteins with homology to acyl-CoA synthases. As noted in “Acyltrehaloses,” above, some of them map in the vicinity of *pks* genes. FadD26 is required for the synthesis of PDIM but not that of PGL in *M. tuberculosis* and *M. bovis* (229, 235, 236). The role of the FadD proteins in activating Pks substrates was demonstrated by Trivedi et al. (167), who established that FadD26 and FadD28 belong to a large family of fatty-acyl AMP ligases responsible for the activation of long-chain *n*-C22-C24 fatty acids as acyl-adenylates. They showed that FadD26 loads the activated substrates directly onto PpsA. These substrates are then elongated with malonyl-CoA and methylmalonyl-CoA by PpsA-PpsE to yield phthiocerol (Table 3; Figs. 14, 15).

The enzyme encoded by *Rv2949c* catalyzes the formation of *p*-hydroxybenzoic acid from chorismate (237). *p*-Hydroxybenzoic acid is activated by FadD22, which displays *p*-hydroxybenzoyl-AMP ligase activity (236) and is subsequently elongated by the type 1 polyketide synthase Pks15/1 to form *p*-hydroxyphenylalkanoates; the reaction may involve eight or nine elongation cycles using malonyl-CoA as the extender unit. A frameshift mutation within the

---

**FIGURE 14** Genetic organization of the PDIM and PGL locus of *M. tuberculosis* H37Rv. ORFs are depicted as arrows. Black arrows indicate genes encoding biosynthetic enzymes; gray arrows indicate putative transporter genes; white arrows indicate hypothetical genes of unknown function. More details about the function of each gene are provided in Table 3 and Fig. 15. Adapted from reference 260. doi:10.1128/microbiolspec.MGM2-0021-2013.f14
FIGURE 15 The PDIM biosynthetic pathway. See text for details. doi:10.1128/microbiolspec.MGM2-0021-2013.f15
Table 3: M. tuberculosis H37Rv genes involved in the biogenesis of phthiocerol dimycocerosates, phenolic glycolipids, and p-hydroxybenzoic acids

<table>
<thead>
<tr>
<th>Gene number</th>
<th>Gene name</th>
<th>Function</th>
<th>Evidence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv2928</td>
<td>tesA</td>
<td>Type II thioesterase thought to be involved in the release of phthiocerol and phenolphthiocerol from PpsE in the biosynthesis of PDIM and PGL</td>
<td>P</td>
<td>230, 238, 239</td>
</tr>
<tr>
<td>Rv2930</td>
<td>fadD26</td>
<td>Long-chain fatty acyl-AMP ligase providing PpsA with activated long-chain fatty acid starter units for the generation of phthiocerol in the biosynthesis of PDIM</td>
<td>P</td>
<td>167, 229, 236, 244</td>
</tr>
<tr>
<td>Rv2931</td>
<td>ppsA</td>
<td>Type 1 polyketide synthase responsible with PpsB-C-D-E for the elongation of C22 to C24 fatty acids and p-hydroxyphenylalkanoates with malonyl-CoA and malonylmalonyl-CoA to yield phthiocerol and phenolphthiocerol derivatives, respectively</td>
<td>E, P</td>
<td>235, 244</td>
</tr>
<tr>
<td>Rv2932</td>
<td>ppsB</td>
<td>Type 1 polyketide synthase responsible with PpsA-C-D-E for the elongation of C22 to C24 fatty acids and p-hydroxyphenylalkanoates with malonyl-CoA and malonylmalonyl-CoA to yield phthiocerol and phenolphthiocerol derivatives, respectively</td>
<td>E, P</td>
<td>235, 244, 349</td>
</tr>
<tr>
<td>Rv2933</td>
<td>ppsC</td>
<td>Type 1 polyketide synthase responsible with PpsB-B-D-E for the elongation of C22 to C24 fatty acids and p-hydroxyphenylalkanoates with malonyl-CoA and malonylmalonyl-CoA to yield phthiocerol and phenolphthiocerol derivatives, respectively</td>
<td>P</td>
<td>235, 349</td>
</tr>
<tr>
<td>Rv2934</td>
<td>ppsD</td>
<td>Type 1 polyketide synthase responsible with PpsA-B-C-E for the elongation of C22 to C24 fatty acids and p-hydroxyphenylalkanoates with malonyl-CoA and malonylmalonyl-CoA to yield phthiocerol and phenolphthiocerol derivatives, respectively</td>
<td>P</td>
<td>235</td>
</tr>
<tr>
<td>Rv2935</td>
<td>ppsE</td>
<td>Type 1 polyketide synthase responsible with PpsA-B-C-D for the elongation of C22 to C24 fatty acids and p-hydroxyphenylalkanoates with malonyl-CoA and malonylmalonyl-CoA to yield phthiocerol and phenolphthiocerol derivatives, respectively</td>
<td>E, P</td>
<td>235, 244</td>
</tr>
<tr>
<td>Rv2936</td>
<td>drrA</td>
<td>Component of the DrrABC transporter involved in the translocation of PDIM (and PGL?) across the plasma membrane; ATP-binding protein</td>
<td>P</td>
<td>350</td>
</tr>
<tr>
<td>Rv2937</td>
<td>drrB</td>
<td>Component of the DrrABC transporter involved in the translocation of PDIM (and PGL?) across the plasma membrane; integral membrane protein</td>
<td>P</td>
<td>230, 350</td>
</tr>
<tr>
<td>Rv2938</td>
<td>drrC</td>
<td>Component of the DrrABC transporter involved in the translocation of PDIM (and PGL?) across the plasma membrane; integral membrane protein</td>
<td>P</td>
<td>229</td>
</tr>
<tr>
<td>Rv2939</td>
<td>papA5</td>
<td>Acyltransferase responsible for the transfer of mycocerosic acids to phthiocerol to form PDIM</td>
<td>E, P</td>
<td>244, 351, 352</td>
</tr>
<tr>
<td>Rv2940c</td>
<td>mas</td>
<td>Mycocerosic acid synthase (type I polyketide synthase) involved in the biosynthesis of PDIM and PGL</td>
<td>E, P</td>
<td>240–244</td>
</tr>
<tr>
<td>Rv2941</td>
<td>fadD28</td>
<td>Long-chain fatty acyl-AMP ligase responsible for providing the long-chain fatty acid starter unit to Mas for the generation of mycocerosic acids</td>
<td>E, P</td>
<td>167, 229, 235, 245, 246</td>
</tr>
<tr>
<td>Rv2942</td>
<td>mmpL7</td>
<td>RND superfamily inner membrane transporter involved in PDIM translocation to the periplasm (and PGL?)</td>
<td>P</td>
<td>229, 235, 252</td>
</tr>
<tr>
<td>Rv2945</td>
<td>lppX</td>
<td>Lipoprotein involved in the transport of PDIM (and PGL?) to the cell surface</td>
<td>P</td>
<td>253</td>
</tr>
<tr>
<td>Rv2946c</td>
<td>pks1</td>
<td>Together with Pks15, type I polyketide synthase involved in the elongation of p-hydroxybenzoic acid derivatives with malonyl-CoA to form p-hydroxyphenylalkanoates (precursors of PGL)</td>
<td>P</td>
<td>223</td>
</tr>
<tr>
<td>Rv2947c</td>
<td>pks15</td>
<td>Together with Pks1, type I polyketide synthase involved in the elongation of p-hydroxybenzoic acid derivatives with malonyl-CoA to form p-hydroxyphenylalkanoates (precursors of PGL)</td>
<td>P</td>
<td>223</td>
</tr>
<tr>
<td>Rv2948c</td>
<td>fadD22</td>
<td>p-Hydroxybenzoyl-AMP ligase involved in the biosynthesis of PGL; catalyzes the activation of p-hydroxybenzoic acid and its subsequent transfer onto Pks15/1 for the production of p-hydroxyphenylalkanoates</td>
<td>E, P</td>
<td>236, 353, 354</td>
</tr>
<tr>
<td>Rv2949c</td>
<td>fadD22</td>
<td>p-Hydroxybenzoyl acid synthase</td>
<td>E, P</td>
<td>237</td>
</tr>
<tr>
<td>Rv2950c</td>
<td>fadD29</td>
<td>Fatty acyl-AMP ligase involved in the biosynthesis of PGL; catalyzes the activation of hydroxyphenylalkanoates that are then transferred onto PpsA to yield phenolphthiocerol</td>
<td>E, P</td>
<td>236</td>
</tr>
<tr>
<td>Rv2951c</td>
<td></td>
<td>Ketoreductase catalyzing the reduction of [phenol]phthiocerol to yield (phenol)phthiothiol in the biosynthesis of PDIM and PGL</td>
<td>P</td>
<td>355, 356</td>
</tr>
<tr>
<td>Rv2952c</td>
<td></td>
<td>SAM-dependent O-methyltransferase involved in the formation of (phenol)phthiocerol dimycocerosates from (phenol)phthiothiols dimycocerosates. Catalyzes the transfer of a methyl group to the third hydroxyl group of (phenol)phthioliol in PDIM and glycosylated phenolphthioliol dimycocerosates</td>
<td>P</td>
<td>247, 356</td>
</tr>
</tbody>
</table>

(continued)
TABLE 3  M. tuberculosis H37Rv genes involved in the biogenesis of phthiocerol dimycocerosates, phenolic glycolipids, and p-hydroxybenzoic acids (continued)

<table>
<thead>
<tr>
<th>Gene number</th>
<th>Gene name</th>
<th>Function</th>
<th>Evidence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv2953</td>
<td></td>
<td>Enoyl-reductase acting in concert with PpsD in the biosynthesis of the (phenol) phthiocerol moiety of PDIM and PGL</td>
<td>P</td>
<td>357</td>
</tr>
<tr>
<td>Rv2954c</td>
<td></td>
<td>SAM-dependent methyltransferase responsible for the O-methylation of the hydroxyl group at position 3 of the fucosyl residue of PGL (and possibly p-HBAD)</td>
<td>P</td>
<td>249</td>
</tr>
<tr>
<td>Rv2955c</td>
<td></td>
<td>SAM-dependent methyltransferase responsible for the O-methylation of the hydroxyl group at position 4 of the fucosyl residue of PGL (and possibly p-HBAD)</td>
<td>P</td>
<td>249</td>
</tr>
<tr>
<td>Rv2956</td>
<td></td>
<td>SAM-dependent methyltransferase responsible for the O-methylation of the hydroxyl group at position 2 of the fucosyl residue of PGL (and possibly p-HBAD)</td>
<td>P</td>
<td>249</td>
</tr>
<tr>
<td>Rv2957</td>
<td></td>
<td>Fucosyltransferase responsible for the transfer of the third glycosyl residue of the triglycosyl appendage of PGL and p-HBAD</td>
<td>P</td>
<td>248</td>
</tr>
<tr>
<td>Rv2958c</td>
<td></td>
<td>Rhamnosyltransferase responsible for the transfer of the second rhamnosyl residue of the triglycosyl appendage of PGL and p-HBAD</td>
<td>P</td>
<td>248</td>
</tr>
<tr>
<td>Rv2959c</td>
<td></td>
<td>SAM-dependent methyltransferase responsible for the O-methylation of the hydroxyl group at position 2 of the rhamnosyl residue linked to the phenolic group of PGL and p-HBAD</td>
<td>P</td>
<td>247</td>
</tr>
<tr>
<td>Rv2962c</td>
<td></td>
<td>Rhamnosyltransferase responsible for the transfer of a rhamnosyl residue onto p-hydroxybenzoic ethyl ester and/or phenolphthiocerol dimycocerosates</td>
<td>P</td>
<td>248</td>
</tr>
</tbody>
</table>

*aE, P, H: see Table 1 footnote.

*pks15/1* gene accounts for the lack of production of PGL by the *M. tuberculosis* reference strains H37Rv, Erdman, and CDC1551 (223). The fatty acyl-AMP ligase FadD29 activates p-hydroxyphenylalkanoates that are then transferred onto PpsA and finally elongated with malonyl-CoA and methylmalonyl-CoA by PpsA-PpsE to yield phenolphthiocerol (236) (Table 3). The type II thioesterase TesA is thought to be involved in the release of the phthiocerol and phenolphthiocerol moieties of PDIM and PGL, respectively, from the polyketide synthase PpsE (230, 238). The demonstrated interaction of TesA with the C-terminal half of PpsE tends to support this assumption (239).

**Biosynthesis and transfer of mycocerosates**

The mycocerosic acids that esterify the β-diols of phthiocerol and phenolphthiocerol are elongated from C$_{16}$ and C$_{20}$ fatty acids with three or four propionate units by a dedicated type I polyketide synthase known as Mas (for mycocerosic acid synthase) (240–244) (Fig. 15; Table 3). Mas preferentially uses methyl-malonyl-CoA instead of malonyl-CoA for fatty acid elongation, thereby introducing methyl branches into the mycocerosic acid chain. In addition, the ketoreductase and enoylreductase activities of this enzyme require NADPH as a cofactor (241). FadD28 is the fatty-acyl AMP ligase responsible for the activation of the C$_{16}$ and C$_{20}$ fatty acid starter units of Mas as acyl-AMP and their transfer onto the polyketide synthase (167, 229, 235, 245, 246) (Fig. 15; Table 3). The synthesized mycocerosates are not released from Mas by a conventional thioesterase but rather are directly transferred by PapA5 onto their phthiocerol or phenolphthiocerol acceptors through an interaction with Mas to catalyze the final esterification step (244) (Figs. 14, 15; Table 3).

**Synthesis of the saccharide moiety of PGL-tb and p-HBADs**

Consistent with their conserved structures (Fig. 13), the biosynthesis of the glycosyl moiety of PGL-tb and p-HBADs involves the same set of enzymes (Table 3). In the case of *M. tuberculosis* “canetti,” four genes (Rv2962c, Rv2957, Rv2958c, and Rv2959c) encoding three glycosyltransferases and one methyltransferase are involved in the formation of this structure (247, 248) (Fig. 14; Table 3). The glycosyltransferase encoded by Rv2962c is involved in the transfer of the first rhamnosyl residue onto p-hydroxy-phenolphthiocerol dimycocerosates and p-hydroxy-phenolmethylster. A single nucleotide polymorphism (SNP) at position 880 of Rv2962c in the Indo-Oceanic isolates of *M. tuberculosis* results in a truncated ORF accounting for the accumulation of phenolphthiocerol dimycocerosates and the related “attenuation lipid” observed in this lineage (228). Rv2958c encodes the rhamnosyltransferase responsible for the transfer of the second rhamnosyl residue onto the mono-rhamnosylated PGL or p-hydroxy-phenolmethylster (248), and a frameshift
mutation within this gene explains the lack of production of triglycosylated PGL by \textit{M. bovis}, \textit{M. microti}, \textit{M. pinnipedii}, and \textit{M. africanum} (226). \textit{Rv2957} encodes the fucosyltransferase responsible for the transfer of the third glycosyl residue of the triglycosyl appendage of PGL and \textit{p-HBAD-II} (Fig. 13) (248). \textit{Rv2959c} encodes a methyltransferase involved in the methylation of position 2 of the first rhamnosyl residue of PGL-tb and \textit{p-HBADs} (247). \textit{Rv2954c}, \textit{Rv2955c}, and \textit{Rv2956} encode the methyltransferases that catalyze the \textit{O}-methylation of the hydroxyl groups located, respectively, at positions 3, 4, and 2 of the terminal fucosyl residue of PGL-tb in a sequential process, starting with methylation at position 2, followed by positions 4 and 3 (249). The genes involved in the production of the glycosyl moiety of the PGL of \textit{M. leprae} were identified through genetic complementation of \textit{M. bovis} BCG, leading to the synthesis of \textit{M. leprae}–specific PGL-1 by the vaccine strain (250).

**Translocation of PDIM and Related Molecules**

PDIM and PGL-tb are found in the outermost layers of the \textit{M. tuberculosis} cell envelope (8), and \textit{p-HBADs} are secreted in the culture medium (223). Since at least some of the enzymes involved in the biosynthesis of PDIM and PGL are cytosolic (e.g., polyketide synthases and FadD enzymes), the presence of these two lipids at the surface implies the existence of a translocation machinery. All the published work on this topic to date has focused on PDIM because the \textit{M. tuberculosis} strains used to generate knockout mutants were naturally deficient in the production of PGL-tb.

The \textit{mmpL7} or \textit{draC} genes, both in the PDIM and PGL locus (Fig. 14), have been involved in the translocation of PDIM. \textit{draC} and \textit{mmpL7} null mutants synthesized PDIMs structurally identical to those of the wild-type strain but failed to translocate these compounds to the cell surface (229, 235). PDIMs in these mutants were apparently retained in deeper layers of the cell envelope. \textit{DraC} is an integral membrane protein belonging to an ABC transporter involving two other subunits encoded by \textit{draA} and \textit{draB} (Fig. 14; Table 3). The \textit{MmpL7} protein belongs to the RND superfamily of transporters (251). Like other members of this family, \textit{MmpL7} is predicted to consist of 12 transmembrane domains and two large soluble periplasmic loops. Using a yeast two-hybrid system, Jain and Cox (252) showed that the loop between the seventh and eighth transmembrane domains interacts with the polyketide synthase \textit{PpsE} involved in PDIM and PGL synthesis. Based on this finding, a model was proposed wherein the synthesis and transport of PDIM are coupled (252). Another gene, \textit{lpplX}, which encodes a lipoprotein, has been found to be required for PDIM to reach the cell surface (253). Interestingly, \textit{LppX} shares a similar fold with the periplasmic chaperone LolA and the outer membrane lipoprotein LolB, which in Gram-negative bacteria are involved in the localization of lipoproteins to the outer membrane. The crystal structure of \textit{LppX} revealed a large hydrophobic cavity suitable to accommodate a single PDIM molecule (253). It is possible that \textit{LppX} acts downstream from \textit{MmpL7} and \textit{DrrABC}, carrying PDIM across the periplasm to the outer membrane once the two membrane transporters have translocated the fully synthesized lipid products across the plasma membrane. The exact role of each of these transporters in the translocation process remains, however, to be determined. Given the nature of the enzymes involved in the biosynthesis of PGL and \textit{p-HBADs}, it is likely that most if not all of their biosynthesis takes place in the cytoplasm or at the periphery of the plasma membrane. The same transporters as those involved in the export of PDIM may be involved in their translocation to the cell surface.

**Roles of PDIM, \textit{p-HBADs}, and PGL-tb in the Organization of the Cell Envelope and Virulence**

As glycosylated capsular or secreted components (8, 223), \textit{p-HBADs} and PGLs are serologically active. Accordingly, several studies have explored the potential of PGLs as serodiagnostic tools for the detection of tuberculosiss and leprosy (14, 254–258). Because most clinical isolates of \textit{M. tuberculosis} do not produce PGL-tb (221–223), it is likely that the antibodies detected in patients were in fact directed against \textit{p-HBADs} (223).

PDIMs have been found in all \textit{M. tuberculosis} clinical isolates tested (221, 227). Their nonamphipathic character and abundance in the cell envelope have long suggested that they play a structural role, providing a hydrophobic barrier around \textit{M. tuberculosis} cells and possibly a platform for anchoring other components of the cell envelope (259). The roles of PDIM in the permeability barrier, intracellular survival, and virulence of \textit{M. tuberculosis} have been extensively discussed in previous reviews and will therefore not be detailed here (141, 260). Likewise, the reader is referred to earlier reviews for details on the roles of PGL-tb and \textit{p-HBADs} in the modulation of the host immune response and pathogenicity of \textit{M. tuberculosis} (260).

Numerous biological activities have also been associated with the PGLs of nontuberculous mycobacteria in general. The PGLs from \textit{M. leprae} and \textit{M. kansasii}, like
those of *M. bovis* BCG, seem to nonspecifically inhibit lymphoproliferative responses to various stimuli, including several antigens and mitogens (261). Other biological activities are dependent on the nature of the carbohydrate moiety. For instance, specific suppression of cell-mediated immunity is a feature of lepromatous leprosy, and the PGL-1 from *M. leprae* has been involved in many aspects of this process. In contrast to the PGLs from *M. microti* and *M. kansasii*, PGL-1 from *M. leprae* has the ability to suppress the “oxidative response” of human macrophages (262, 263), probably explaining why this response is abnormally low in leprosy patients. PGL-1, but not the PGLs from *M. bovis* and *M. kansasii*, also has been reported to be active in an indirect test of specific immunosuppression, inhibiting the concanavalin-A stimulation of lymphocytes from patients with lepromatous leprosy (264). PGL-1 also can neutralize hydroxyl and superoxide radicals *in vitro*, a property shared by deacylated PGL-1 and to some extent by the carbohydrate moiety of the molecule (265).

**GLYCOPEPTIDOLIPIDS**

**Structure and Subcellular Location**

Mycobacteria synthesize type- or species-specific GPLs that differ from one another by their sensitivity to alkali. Alkali-stable GPLs, also known as C-mycosides (266), are produced by a number of both fast- and slow-growing mycobacterial species, including *M. avium*, *M. abscessus*, and *M. smegmatis* (for detailed reviews, see references 200, 267, 268). Alkali-labile GPLs have thus far only been described in *Mycobacterium xenopi* (269, 270). Their structures greatly differ from those of the C-type GPLs. The C-type GPLs share a common lipopeptidyl core that consists of a mixture of 3-hydroxylated and 3-methoxylated long-chain (C26 to C34) fatty acids (271), amidated by a tripeptide (d-Phe-d-alloThr-d-Ala) and terminated by an aminoalcohol (l-alaninol) (Fig. 16A). The position of hydroxy/methoxyl group A on the fatty acyl chain has been recently questioned, and position 5 has been proposed (272). C-type GPLs differ by the number and the nature of the glycosyl residues that substitute the lipopeptidyl core. In the most abundant molecular species, the apolar C-type GPLs (Fig. 16A), also known as nonspecific GPLs (nsGPLs), the alaninol is glycosylated by a mono- or di-O-methylated rhamnosyl residue, while a di-O-acetylated 6-deoxytalosyl unit is attached to the alloThr residue. In the polar C-type GPLs, also known as serospecific GPLs (ssGPLs), additional sugar units are attached to the 6-deoxytalosyl residue; at least 14 out of the 28 described ssGPLs from the *M. avium-intracellularare* complex have been structurally...
characterized (200, 268). The serological variance among the members of the *M. avium-intracellulare* complex is due to subtle differences in the structure of the oligosaccharide chain that substitutes the communal C-type nsGPL core. The oligosaccharide happens from several polar C-type GPLs contain unusual sugars: glucuronic acid and variants, acetamido-dideoxy-hexosyl residues, and other branched sugars.

The correlation between the presence of C-type GPLs, smooth colony morphotype, and staining of the outer membrane with Ruthenium Red (273–275), and the fact that polar GPLs correspond to “Schaefer typing antigens” used in the identification of isolates of the *M. avium-intracellulare* complex (276) suggested that GPLs were present at the cell surface. Consistently, the capsular materials of *M. lepraemurium* (277) and *M. avium* (278) have been shown to consist of C-mycosides.

**Biological Properties of GPLs**

The antigenic properties of ssGPLs and the relationships between GPL production and colonial morphotype or drug resistance have been abundantly reviewed (268, 279). Other properties associated with GPL production are as follows. Although mycobacteria are nonflagellated, *M. smegmatis* and the slow-growing *M. avium* can spread on the surface of solid media by a sliding mechanism (280). Rough strains lacking GPLs appear to be devoid of such sliding motility (210, 275). Consistently, all of the nonsliding mutants isolated and analyzed by Recht et al. (281) had a rough morphotype and showed no detectable levels of GPLs. These nonsliding mutants were also defective for attachment and biofilm formation on PVC plastic (281). This observation emphasizes the importance of GPLs in determining the cell surface properties of *M. smegmatis*. Suggestive of the important role played by GPLs in the permeability of the cell envelope, the absence of nsGPLs from the cell envelope of a defined knockout mutant of *M. smegmatis* was shown to have a profound effect on the uptake of chenodeoxycholate (273), a hydrophobic molecule that diffuses through lipid domains of the mycobacterial cell envelope.

The species *M. avium-intracellulare* has received attention as a major opportunistic pathogen in AIDS patients. Although the specific mechanisms that define its pathogenicity have not been entirely clarified, it is becoming apparent that GPL antigens have a variety of biological activities that could influence host responses. During infection, *M. avium* synthesizes GPLs that accumulate in macrophages (278, 282–284). Early studies have shown that *M. avium* ssGPLs suppress the mitogen-induced proliferative responses of murine splenic cells (285–287) but not those of human peripheral blood mononuclear cells (288). More recently, GPLs from *M. avium* serovar 4 have been proposed to participate in the ability of *M. avium* to invade human macrophages and escape bactericidal responses (289). GPLs are also thought to impact adaptive immunity. Pretreatment of human blood mononuclear cells with serovar-specific GPLs, for instance, suppresses the production of Th1 cytokines including interleukin 2 (IL-2) and interferon γ (IFNγ) (290). In contrast, ssGPLs induce the production of two important immunomodulatory substances, tumor necrosis factor α (TNF-α) and prostaglandin E2 (PGE2) (288, 291). A group of polar GPLs from *M. chelonae* (pGPL-Mc) has also been reported to increase the resistance of mice to disseminated candidiasis (292) and to enhance the immune response to influenza vaccination (293). Moreover, pGPL-Mc molecules exhibit the properties of haematopoietic growth factor (294–296).

**Biosynthesis of GPLs**

**nsGPL biosynthesis in *M. smegmatis***

The GPL biosynthetic gene cluster of *M. smegmatis* (Fig. 16B) is currently thought to encompass 24 genes. Twelve of them have been experimentally characterized using a combination of genetic approaches (210, 274, 281, 297–301). The nonribosomal peptide synthase genes, *mps1* and *msp2*, encode the enzymes responsible for the synthesis of the peptidic moiety of D-Phe-D-Ala-L-alaninol (274), a putative nonribosomal peptide synthase of the distal alanyl-containing moiety of the pseudotetrapeptide. The putative nonribosomal peptide synthases of *M. avium* share the same genetic organization consisting of two ORFs (302). The *mbtH* gene has been shown to be required for GPL production (303), but no exact function has yet been attributed to its protein product.

The main acyl residue of the *M. smegmatis* GPLs is a monounsaturated hydroxylated C₃₀ fatty acid (271, 304) whose synthesis is thought to involve the polyketide synthase product of pks1 (210) (Fig. 16B). Three glycosyltransferase genes are also found in the GPL cluster, a number in agreement with the structure of the *M. smegmatis* GPLs (Fig. 16A and 16B) (299, 301, 305). Two other genes of the cluster, namely, *rmlA*
(aka rfbA), which encodes a putative glucose-1-phosphate thymidylyltransferase, and rmlB, which encodes a putative dTDP-glucose-4,6-dehydrogenase, are likely to be involved in the synthesis of the deoxyhexoses, rhamnose and 6-deoxytalose, which would subsequently be incorporated into nsGPLs (305).

The mtf1 (aka rmt3) gene (300) encodes an S-adenosylmethionine-dependent rhamnosyl-3-O-methyltransferase (297). This enzyme is required for the O-methylation of position 3 of the rhamnosyl unit that glycosylates the alaninol. Disruption of rmt3 virtually abolishes the further methylation of the rhamnosyl unit, suggesting that this enzyme is the first methyltransferase to act on the GPL precursors (297). Three other methyltransferase genes are found in the GPL gene cluster (305). The fnt gene encodes a fatty acid O-methyltransferase that modifies the hydroxyl group of the GPL fatty acid (299), whereas rmt4 encodes a rhamnosyl-4-O-methyltransferase, and rmt2 encodes a rhamnosyl-2-O-methyltransferase (300); all of these methyltransferases have orthologs in M. avium. The gene atf1 is predicted to encode a 6-deoxytalose acetyltransferase (306). The methylation of the rhamnosyl residue occurs independently of the acetylation of the 6-deoxytalose residue, since the GPLs from an atf1 knockout mutant are normally methylated.

**ssGPL biosynthesis in M. avium**

Limited information is currently available about GPL biosynthesis in M. avium, except for the serotype 2 (ser2) ssGPLs recently reviewed by Billman-Jacobe (305) and Chatterjee and Khoo (268). The rhamnosyltransferase gene rtaA is the first gene whose function was determined experimentally (307, 308). When produced in M. smegmatis, RtfA catalyzed the addition of a rhamnosyl unit to the 6-deoxytalosyl residue of the GPL core (308), thus showing that the simpler nsGPLs can serve as biosynthetic precursors in the synthesis of ssGPLs. This result was confirmed by showing that the targeted disruption of rtaA in M. avium led to the loss of ser2-specific GPLs (307). Complementing M. smegmatis methyltransferase mutants with M. avium genes from the ser2 gene cluster, Jeevarajah et al. provided evidence of the 4-O-methyltransferase activity of the M. avium MtfC and MtfB proteins (300). In addition, they showed that MtfD displays 3-O-methyltransferase activity on the rhamnosyl residue of the M. smegmatis GPLs (300). The specificity of this methyltransferase was recently confirmed via the construction of a mtfD knockout mutant of M. avium (309, 310). Interestingly, the virulence of this mutant was attenuated in mice (309).

On the basis of an altered colony morphotype, Laurent et al. identified other genes likely to be involved in the biosynthesis of GPLs in M. avium (302). Two nonribosomal peptide synthase genes (psta and pspb) and a probable polyketide synthase gene located downstream of the ser2 cluster are orthologous to the mps and pks genes found in the GPL biosynthetic gene cluster of M. smegmatis mc2155, but a direct involvement of these genes in the mutant phenotypes remains to be established. The ser2 gene clusters of two ser2 strains of M. avium were also sequenced and compared with the homologous regions of M. avium ser1 strain 104, M. avium subspecies paratuberculosis, and M. avium subspecies silvaticum (311). Fifteen ORFs were identified and their putative functions in GPL biosynthesis determined: five encode glycosyltransferases (including RtfA), six encode O-methyltransferases (including MtfB, C,D), one encodes an O-acetylmethyltransferase, and three encode hexose synthetases (γ-glucose dehydrogenase, mannose dehydrogenase, and the 6-deoxy-4-keto-d-mannose reductase/epimerase). A biosynthetic model in which ser2-specific GPLs are synthesized from a serovar-1-specific GPL intermediate, itself derived from a nonspecific GPL precursor, was proposed (311).

**Regulation and Transport of GPLs**

The C-type GPL biosynthetic gene cluster begins with a triplet of transmembrane protein encoding genes possibly forming an operon: tmtPA, B, and C (now named mmps4, mmpL4a, and mmpL4b). TmtABC belongs to the MmpL and MmpS families of mycobacterial proteins (210) (Fig. 16B). Both MmpL4a and MmpL4b have 12 putative transmembrane domains, whereas the smaller Mmps4 protein displays only one. mmpL4a and mmpL4b transposon mutants have been reported to have a rough colony morphology, to lack sliding motility, and to be devoid of GPLs (210, 281); the precise role of the MmpL4 proteins in this phenotype, however, has yet to be determined. Interestingly, the biochemical characterization of an mmps4 mutant of M. smegmatis established that this protein is required for the production and export of large amounts of GPLs but is dispensable for biosynthesis per se. Cross-complementation experiments demonstrated that the Mmps4 proteins from M. smegmatis, M. avium, M. tuberculosis, and M. abscessus are exchangeable and thus not specific for a particular GPL species (312). Mmps4 requires the formation of a protein complex at the pole of the bacillus to function. It was suggested that Mmps proteins facilitate lipid biosynthesis by acting as a scaffold for a coupled biosynthetic and transport machinery.
A similar mechanism has also been proposed for the transport of PDIM and SL in *M. tuberculosis* (see “Acyltrehaloses” and “PDIMs, PGLs, and Related Compounds, above”) and is thus emerging as a common trait in the biogenesis of mycobacterial complex lipids.

While screening an *M. smegmatis* transposon mutant library for mutants with changes in cell surface properties, strains that failed to stain with Ruthenium Red were isolated (210). All of these mutants harbored a transposon insertion in the *gap* gene (Fig. 16B) and produced GPLs chemically identical to those of the wild-type strain. *gap* mutants, however, had many fewer GPLs at their surface, suggestive of a role for Gap in the export of these lipids. The precise role of Gap in the biogenesis of GPLs—particularly in relation to the MmpL4 and MmpS4 proteins—remains to be determined. Gap may be required for the transport of GPLs across the periplasmic space upon their translocation across the plasma membrane in a process involving the MmpL4-MmpS4 proteins.

Nutrient starvation was reported to induce the production of triglycosylated C-type GPLs in *M. smegmatis* mc2 155 (312). The accumulation of polar GPLs in *M. smegmatis* mc2 155 seems to be dependent on SigB, because the overexpression of *sigB* induces the production of these lipids, while the disruption of this gene leads, in contrast, to the abolition of their production (314). The expression of *gtfβ*, the glycosyltransferase responsible for the addition of the last sugar moiety of triglycosylated GPLs (301), is directly or indirectly controlled by SigB, at least during certain stages of growth (314). Another gene potentially coding for an extracytoplasmic sigma factor, *ecf*, is present in the GPL biosynthetic cluster (Fig. 16B), but evidence of the involvement of this gene in the regulation of GPL biosynthesis is lacking. *ecf* is located upstream of a gene encoding a putative sigma factor–associated protein. *M. smegmatis* displays a low frequency of spontaneous morphological variation that correlates with the production of larger amounts of GPLs (315). The transposition of insertion elements into two GPL loci accounts for these morphological changes. One locus is the *mps* operon. The other locus is the *lsr2* gene, which encodes a small basic protein that likely plays a regulatory role.

**CAPSULAR POLYSACCHARIDES**

As mentioned above, and with a few exceptions, the “capsule”-like structures produced by *Mycobacterium* species primarily consist of polysaccharides and proteins with generally minor amounts of lipids. The ratio of protein to polysaccharide varies according to the species. While in *M. tuberculosis*, *M. kansasii*, and *M. gastri*, the major surface capsular constituents consist of polysaccharides, they mainly are proteins in *M. phlei* and *M. smegmatis* (7, 9). Capsular polysaccharides, like other capsular components, are not covalently bound to the rest of the cell envelope. The three types of capsular polysaccharides identified in the capsular material of tuberculous and nontuberculous mycobacteria are a high-molecular-weight (>1,000,000 Da) α-d-glucan composed of a → 4-α-d-Glc-1→ core branched every five or six residues by oligoglucosides; a d-arabino-D-mannan (AM) similar in structure to LAM; and a d-mannan composed of a → 6-α-d-Man-1→ core substituted at some of the 2 positions with an α-d-Man residue (6, 7, 316, 317). All are neutral compounds, devoid of acyl substituents.

The structure of AM appears to be identical to that of LAM except for the loss of the phosphatidyl-myoinositol anchor, suggesting that it may be formed from LAM by a specific hydrolytic enzyme. Likewise, the structure of d-mannan appears to be identical to that of the mannan domain of LM and LAM. It is therefore reasonable to assume that the same enzymes participate in the biosynthesis of LM/LAM and in that of the two extracellular polysaccharides. The reader is referred to reference 363 for details about this biosynthetic pathway. d-Mannan and AM are expected to share with LM and LAM common properties in their interactions with the host.

α-d-Glucan is structurally very similar to the intracellular glycogen of *M. tuberculosis* and *M. bovis* BCG, although its three-dimensional structure appears to be more compact and its molecular mass slightly higher (13 × 10^6 versus 7.5 × 10^6 Da). Capsular α-d-glucan was shown to be a ligand of the C-type lectin DC-SIGN of dendritic cells, to modulate the effector functions of monocyte-derived dendritic cells and to mediate the nonopsonic binding of *M. tuberculosis* to complement receptor 3 (318–320). It was also postulated to contribute to the antiphagocytic properties of the capsule of *M. tuberculosis*, thereby possibly controlling the interactions of *M. tuberculosis* with macrophages and promoting uptake via complement receptor 3 (321). Altogether, these biological properties may contribute to the survival of *M. tuberculosis* in the host. The structural similarity between α-d-glucan and glycogen has allowed some of the genes involved in the biosynthesis of the capsular polysaccharide to be identified, among them the α-1,4-glucosyltransferases Rv3032 and GlgA (Rv1212c), the ADP-glucose pyrophosphorylase GlgC (Rv1213), and the branching enzyme GlgB.
(Rv1326c), which is responsible for introducing α-1,6-linked branches into linear α-1,4-glucans (322) (Table 4, Fig. 17).

The phenotypic analysis of M. tuberculosis recombinant strains affected by or totally deficient in the expression of these genes confirmed their involvement in the elongation and branching of the capsular α-d-glucan and a partial redundancy between the two α-1,4-glucosyltransferases Rv3032 and GlgA. These analyses further revealed the participation of GlgC, GlgB, and Rv3032 in the biosynthesis of other intracellular M. tuberculosis α-1,4-glucans, namely, glycogen and the methylglucose lipopolysaccharides (MGLPs) (322, 323). Attempts to knock out both glgA and Rv3032 in M. tuberculosis mutants were unsuccessful, indicating that a functional copy of at least one of the two α-1,4-glucosyltransferases is required for growth. The apparent essentiality of glgB (322), in contrast, is believed to be related to the toxic accumulation of maltose-1-phosphate that follows the inactivation of this gene (324) (Fig. 17). It is important to note that mycobacterial α-1,4-glucans can also be synthesized from trehalose by a four-step pathway comprising the trehalose synthase TreS, the maltokinase Pep2, the maltose-1-phosphate maltosyltransferase GlgE, and GlgB (Fig. 17; Table 4) (324, 325). Disruption of glgE, like that of glgB, is lethal because of the toxic accumulation of maltose-1-phosphate that ensues.

As evidenced by their Rv numbers, the genes involved in the metabolism of glycogen, capsular α-d-glucan, and MGLP are clustered in four major locations on the chromosome of M. tuberculosis H37Rv (Table 4). The first cluster (Rv3030–Rv3037c) encompasses the α-1,4-glucosyltransferase gene Rv3032 and other genes likely to be involved in the modifications of MGLPs (326). The second cluster (Rv1208–Rv1213) carries glgA, glgC, and the glucosyl-3-phosphoglycerate synthase gene, Rv1208, required for the initiation of MGLPs (33).

The third cluster (Rv1326c–Rv1328) carries glgB, glgE, and the probable glycogen phosphorylase gene glgP (35). The fourth region (Rv0126–Rv0127) harbors pep2 and treS.

Consistent with the intracellular localization of glycogen and MGLPs, GlgA and Rv3032 are nucleotide sugar-utilizing glucosyltransferases predicted to catalyze the elongation of polysaccharides on the cytosolic face of the plasma membrane. GlgB, GlgC, GlgE, Pep2, and TreS are also predicted to be cytosolic enzymes. It is therefore reasonable to assume that capsular α-d-glucan, like glycogen and MGLPs, is synthesized in the cytoplasm. Nothing is known of the translocation machinery responsible for its export to the cell surface.

### OTHER LIPOPHILIC COMPOUNDS

Other lipophilic compounds found in the cell envelope of some mycobacterial species include siderophores known as mycobactins and the M. ulcerans toxin, mycolactone. Their structure is shown in Figures 18 and 19. The reader is referred to recent reviews for complete details of their biosynthetic pathways (327, 328).

### Mycobactins

Pathogenic and nonpathogenic mycobacteria rely on siderophores with high affinity for the ferric ion as the primary mechanism for iron acquisition (for a review, see reference 327). Two classes of siderophores are produced by mycobacteria: the exochelins and the (carboxy)mycobactins. The exochelins are water-soluble peptidic molecules and are secreted into the medium. Mycobactins and carboxymycobactins are both salicyl-capped peptide polyketide-based molecules but vary in the length of an alkyl substitution and hence in polarity and solubility (Fig. 18). The lipid-soluble mycobactins have long-chain acyl chains on the first lysine residue, whereas the water-soluble

### Table 4 M. tuberculosis H37Rv genes involved in the biogenesis of capsular α-d-glucan

<table>
<thead>
<tr>
<th>Rv number</th>
<th>Gene name</th>
<th>Function</th>
<th>Evidence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv0126</td>
<td>treS</td>
<td>Trehalose synthase; displays maltose ↔ trehalose interconverting activity and glycogen amylase activity</td>
<td>E, P</td>
<td>324, 358–360</td>
</tr>
<tr>
<td>Rv0127</td>
<td>pep2</td>
<td>Maltokinase</td>
<td>E, P</td>
<td>324, 361</td>
</tr>
<tr>
<td>Rv1212c</td>
<td>glgA</td>
<td>α-1,4-Glucosyltransferase</td>
<td>P</td>
<td>322</td>
</tr>
<tr>
<td>Rv1213</td>
<td>glgC</td>
<td>ADP-glucose pyrophosphorylase</td>
<td>P</td>
<td>322</td>
</tr>
<tr>
<td>Rv1326c</td>
<td>glgB</td>
<td>α-1,4-Glucan branching enzyme</td>
<td>E, P</td>
<td>322, 362</td>
</tr>
<tr>
<td>Rv1327c</td>
<td>glgE</td>
<td>Maltose-1-phosphate maltosyltransferase</td>
<td>E, P</td>
<td>324, 325</td>
</tr>
<tr>
<td>Rv1328</td>
<td>glgP</td>
<td>Putative glycogen phosphorylase</td>
<td>H</td>
<td>322, 323</td>
</tr>
<tr>
<td>Rv3032</td>
<td>-</td>
<td>α-1,4-Glucosyltransferase</td>
<td>E, P</td>
<td>322, 323</td>
</tr>
</tbody>
</table>

*E, P, H: see Table 1 footnote.*
carboxymycobactins have a shorter side chain that terminates with a carboxylic acid or methyl ester. Mycobactins tend to remain cell-associated, while carboxymycobactins are secreted in the culture medium. Mycobacteria fall into four groups based upon the production of these molecules. *M. tuberculosis* produces only (carboxy)mycobactins, which are essential for its virulence (327, 329); *M. vaccae* produces only the exochelin type; *M. smegmatis* produces both types, and *M. leprae* produces none. The biosynthesis of mycobacterial siderophores has been reviewed recently (327) and will therefore not be detailed here. Briefly, a cluster of 10 genes (annotated mbtA through mbtJ; Rv2377c-Rv2386c—the mbt locus) encompassing approximately 24 kb, another cluster of 6 genes referred to as mbt-2, the phosphopantetheinyl gene pptT (Rv2794c), the ess3 cluster (Rv0282-Rv0292), and the transport genes mmpS5/mmpL4 (Rv0450c-Rv0451c) and mmpS5/mmpL5 (Rv0676c-Rv0677c) encode the proteins required for the synthesis, export, utilization, and uptake of (carboxy)mycobactins (327, 329). The regulator IdeR (Rv2711) represses the expression of the mbtA-N genes.

**Mycolactones**

Mycolactones are a family of lipophilic macrocyclic polyketide molecules that is the primary virulence factor produced by *M. ulcerans*, the etiologic agent of Buruli ulcer in humans, and some closely related aquatic mycobacteria (Fig. 19). Mycolactones display cytotoxic, analgesic, and immunosuppressive activities (328). They are found in abundant quantities in the extracellular matrix surrounding *M. ulcerans* under certain in vitro growth conditions and during host infection (330). A 174-kb megaplasmid named pMUM001 in the *M. ulcerans* strain Agy99 carries all of the genes required for mycolactone synthesis (331). These include two very large type I polyketide synthase genes, *mlsA1* (51 kb) and *mlsA2* (7 kb), responsible for the synthesis of the upper side chain and macrolactone core; another giant type I polyketide synthase gene, *mlsB* (42 kb), involved in the elongation of the acyl side chain; *mup_045*, thought to encode the transferase linking the acyl side chain and the macrolactone core; *mup_053c*, a P450 monooxygenase gene most likely responsible for hydroxylation of the mycolactone acyl side chain at C12; and *mup_038*, a type II thioesterase gene predicted to
play a role in maintaining the fidelity of the polyketide synthases by removing acyl chains from modules where synthesis has stalled (328).

CONCLUSIONS AND FUTURE PROSPECTS
As illustrated in this article, knowledge of cell envelope biosynthesis in M. tuberculosis has greatly benefited from the publication of the complete genome sequence of this bacterium and developments in the genetic manipulation of mycobacteria in the late 1990s. As more genomes from slow- and fast-growing mycobacteria are sequenced, this impetus is progressively extending to other tuberculous and nontuberculous mycobacterial species, with the result that the processes leading to the biosynthesis of more and more species-specific cell envelope constituents are now being elucidated. Beyond the opportunities offered by some of these pathways for drug development, interest in the biosynthesis of species-specific cell envelope constituents stems from their antigenicity and potential for serodiagnosis and as biomarkers. GPL, DAT, LOS, PGL, pHBAD, and TDM in particular are potent B-cell antigens and have been the object of extensive studies aimed at assessing their potential for the diagnosis of tuberculosis, leprosy, and other mycobacterial diseases (14, 223, 256, 257, 332–334). Key to their widespread application for therapeutic or diagnostic purposes, however, will be a precise understanding of their role in the physiology and virulence of the bacterium and regulatory processes governing their synthesis during the various stages of the lifecycle of the producing mycobacterium. While it is now well established that mycobacteria adjust the composition of their cell envelope in response to the nutrients available in the environment (e.g., carbon and nitrogen source, iron concentration, etc.), physical conditions to which they are exposed (pH, oxygen tension), and age of the culture, knowledge of the regulatory processes involved is more limited.

FIGURE 18 Representative structures of mycobactins and carboxymycobactins from M. tuberculosis. See text for details. Mycobactins: $R_1 = H; R_2 = (\text{CH}_2)_n\text{CH}_3, n = 16-19$; $(\text{CH}_2)_n\text{CH} = \text{CH} (\text{CH}_2)_n\text{CH}_3, x + y = 14-17$. Carboxymycobactins: $R_1 = H, \text{CH}_3; R_2 = (\text{CH}_2)_n\text{COOCH}_3/\text{COOH}, n = 1-7$; $(\text{CH}_2)_n\text{CH} = \text{CH}(\text{CH}_2)_n\text{COOCH}_3/\text{COOH}, x + y = 1-5$. doi:10.1128/microbiolspec.MGM2-0021-2013.f18

FIGURE 19 Representative structure of a mycolactone from M. ulcerans. The genes involved in the biosynthesis of the various constituents of mycolactone are indicated on the structure. doi:10.1128/microbiolspec.MGM2-0021-2013.f19
Yet these changes affect all major cell envelope constituents including phospholipids, PIMs, triglycerides, capsular polysaccharides, lipoglycans, PG, arabinogalactan, and mycolic acids (14, 15, 141, 180, 335–341). Regulation appears to occur both at the transcriptional and the posttranslational levels. The two-component transcriptional regulator PhoP-PhoR, for instance, stands out as a major regulator of polymethyl-branched acyl-trehalose production in M. tuberculosis (SL, DAT, and PAT) (143, 177). Other important regulators controlling cell division and cell envelope biogenesis are found among the serine/threonine kinase family (for a review, see reference 342). These enzymes regulate through phosphorylation the activity of multiple enzymes and transporters involved in the biosynthesis of mycolic acids, PG, arabinogalactan, LAM, and PDIM.

Despite the considerable advances made in deciphering the metabolic pathways of mycobacterial cell envelope constituents, most of them are not yet complete. The clustered genetic organization of many of these pathways (e.g., lipoglycans, acyltrehaloses, GPLs, LOSs, PDIMs, etc.) raises the hope that some of the missing genes will be found on the basis of their colocalization with known clusters. In the case of a few other minor or species-specific cell envelope constituents, however, biosynthetic pathways have hardly started to be explored. This is, for instance, the case for the glycosyl DAGs described by Hunter et al. (343) and the mycobacterial carotenoids, whose structural definition extends back to the work of E. Chargaff in 1930 (259). Likewise, although there is at present no chemical evidence of LOS in M. tuberculosis, preliminary data indicate that some of the unannotated glycosyltransferases of the GT-A, -B, or -C classes may participate in the synthesis of chemically undetectable amounts of these products (204, 207).

Beyond the identification of the missing biosynthetic and regulatory proteins will be the identification of the transporters required for the translocation of biosynthetic precursors or end products from their site of production, for the most part cytoplasmic, to their final periplasmic, outer membrane or capsular locations. More than 148 transport-associated proteins belonging to 33 major transporter families were identified in the genome of M. tuberculosis H37Rv (http://www.membranetransport.org/). More transporters are typically found in environmental Mycobacterium species. The latest M. tuberculosis genome annotation was updated with 134 bioinformatically predicted outer membrane proteins (344). The transporters required for the building of the cell envelope are thus most likely to be found in this long and diverse list of candidate genes with or without homologs in other prokaryotes. Indeed, searches for mycobacterial transporters sharing sequence similarity with known (lipo)polysaccharide or glycolipid transporters from Gram-positive or Gram-negative bacteria typically yield limited if any meaningful candidates.

Yet it is becoming increasingly evident that, similar to the biogenesis of other prokaryotic (lipo)polysaccharides, the biosynthesis and translocation of many mycobacterial cell envelope constituents (e.g., mycolic acids, PDIMs, acyltrehaloses, mycobactins, lipoglycans, and arabinogalactan) are temporally and spatially coupled by multiprotein complexes that possibly span the cell envelope. Reasons for their elusive nature may be found in the unusual structure and composition of the mycobacterial cell envelope, which may have driven mycobacteria to evolve specialized transport systems mechanistically related to but structurally divergent from those of other prokaryotes. Examples of these include the type VII secretion system ESX-3 for the uptake of mycobactins (345), the Mce4 proteins for the import of cholesterol (346), the RND-like inner membrane transporters of the MmpL family involved in the export of complex lipids and mycobactins (145, 170, 171, 210, 229, 235, 252, 281, 329, 347), the periplasmic Lol-like lipoprotein carriers LppX and LprG for the export of PDIM and PIM/lipoglycans (38, 253), and the SMR-like (small multidrug resistance) lipid-linked arabinose translocase Rv3789 (348).

Finally, another area where much remains to be done is understanding the genetic bases underlying the cell envelope’s constant remodeling (including degradation and recycling), which accompanies cell division or any changes in the metabolism of the bacterium following host infection or exposure to various environmental stresses.

ACKNOWLEDGMENT
The authors’ research on cell envelope biogenesis is supported through NIH/NIAID grants AI064798, AI063054, AI049151, and AI097550.

We thank Hugo Gramajo (Universidad Nacional de Rosario, Rosario, Argentina) for sharing unpublished results, and Victoria Jones (CSU), Michael McNeil (CSU), Gilles Etienne, Anne Lemassu, and Lucie Spina (IPBS-CNRS, Toulouse) for critical reading of the manuscript and their help with the preparation of the figures.

REFERENCES
In tuberculous mycobacteria, as it goes into a dormancy-like state in culture.


Genetics of Capsular Polysaccharides and Cell Envelope (Glycolipids)


biosynthesis in Mycobacterium abscessus and M. chelonae. BMC Genomics 8:114.


Daffé et al.


