Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*

P.J. Brennan*

*Department of Microbiology, B308 Microbiology Building, Colorado State University, Fort Collins, CO 80523, USA*

**Summary** Much of the early structural definition of the cell wall of *Mycobacterium* spp. was initiated in the 1960s and 1970s. There was a long period of inactivity, but more recent developments in NMR and mass spectral analysis and definition of the *M. tuberculosis* genome have resulted in a thorough understanding, not only of the structure of the mycobacterial cell wall and its lipids but also the basic genetics and biosynthesis. Our understanding nowadays of cell-wall architecture amounts to a massive “core” comprised of peptidoglycan covalently attached via a linker unit (L-Rha–D-GlcNAc-P) to a linear galactofuran, in turn attached to several strands of a highly branched arabinofuran, in turn attached to mycolic acids. The mycolic acids are oriented perpendicular to the plane of the membrane and provide a truly special lipid barrier responsible for many of the physiological and disease-inducing aspects of *M. tuberculosis*. Intercalated within this lipid environment are the lipids that have intrigued researchers for over five decades: the phthiocerol dimycocerosate, cord factor/dimycolyltrehalose, the sulfolipids, the phosphatidylinositol mannosides, etc. Knowledge of their roles in "signaling" events, in pathogenesis, and in the immune response is now emerging, sometimes piecemeal and sometimes in an organized fashion. Some of the more intriguing observations are those demonstrating that mycolic acids are recognized by CD1-restricted T-cells, that antigen 85, one of the most powerful protective antigens of *M. tuberculosis*, is a mycolyltransferase, and that lipoarabinomannan (LAM), when "capped" with short mannose oligosaccharides, is involved in phagocytosis of *M. tuberculosis*. Definition of the genome of *M. tuberculosis* has greatly aided efforts to define the biosynthetic pathways for all of these exotic molecules: the mycolic acids, the mycocerosates, phthiocerol, LAM, and the polyprenyl phosphates. For example, we know that synthesis of the entire core is initiated on a decaprenyl-P with synthesis of the linker unit, and then there is concomitant extension of the galactan and arabinan chains while this intermediate is transported through the cytoplasmic membrane. The final steps in these events, the attachment of mycolic acids and ligation to peptidoglycan, await definition and will prove to be excellent targets for a new generation of anti-tuberculosis drugs.

© 2003 Published by Elsevier Science Ltd.

---

**Introduction** Much of the early structural definition of the cell wall of *Mycobacterium* spp. was initiated at the University of Osaka, Japan, and the CNRS, France,
by Yamamura, Kato, Azuma, Lederer, and colleagues in the 1960s and 1970s. There was a long period of inactivity, but more recent developments in analytical techniques combined with definition of the M. tuberculosis genome have resulted in a thorough understanding, not only of the structure of the mycobacterial cell wall and its lipids but also the basic genetics and biosynthesis.

The cell wall is composed of two segments, upper and lower. Beyond the membrane is peptidoglycan (PG) in covalent attachment to arabinogalactan (AG), which in turn is attached to the mycolic acids with their long meromycolate and short \( \alpha \)-chains. This is termed the cell wall core—the mycolyl arabinogalactan–peptidoglycan (mAGP) complex. The upper segment is composed of free lipids, some with longer fatty acids complementing the shorter \( \alpha \)-chains, and some with shorter fatty acids complementing the longer chains. Interspersed somehow are the cell-wall proteins, the phosphatidylinositol mannosides (PIMs), the phosphotriester-containing lipids, lipomannan (LM), and lipoarabinomannan (LAM). When cell walls are disrupted, for instance extracted with various solvents, the free lipids, proteins, LM, and PIMs are solubilized, and the mycolic acid–arabinogalactan–peptidoglycan complex remains as the insoluble residue. In simplistic terms, it can be considered that these lipids, proteins, and lipoglycans are the signaling, effector molecules in the disease process, whereas the insoluble core is essential for the viability of the cell and should be addressed in the context of new drug development.

The cell wall core, the mAGP complex

Structure of mAGP

A spate of intensive investigations from 1950 to 1975 allowed the definition of the insoluble cell-wall matrix as a cross-linked PG linked to AG, and esterified at the distal ends by the mycolic acids. Historically, PG is thought to consist of alternating units of N-acetylglucosamine (GlcNAc) and a modified muramic acid (Mur). The tetrapeptide side chains of PG consist of \( L \)-alanyl–\( \beta \)-isoglutaminyl–meso-diaminopimelyl–\( d \)-alanine (\( L \)-Ala–\( \beta \)-Glu–\( \alpha \)pm–\( d \)-Ala) with the Glu being further amidated. This type of PG is one of the most common found in bacteria. However, mycobacterial PG differs in two ways: some or all of the Mur residues are \( N \)-acylated with glycolic acid (MurNGly), and the cross-links include a proportion between two \( \alpha \)pm residues as well as between \( \alpha \)pm and \( d \)-Ala.

It was known, even in the 1950s, that the major cell-wall polysaccharide is a serologically active branched-chain AG with the arabinose (Ara) residues forming the reducing termini. A structural formula, which proved incorrect, was proposed consisting of repeating units of 11–16 sugar residues. There was some uncertainty about the structure of the galactan, i.e., whether \( \alpha 1 \rightarrow 4 \)-linked Galp or \( \alpha 1 \rightarrow 5 \)-linked Galp. We and others had shown that the polymer is unique not only in its elemental sugars but, unlike most bacterial polysaccharides, it lacks repeating units, comprised instead of a few distinct structural motifs. Partial depolymerization of the per-\( O \)-alkylated AG and analyses of the generated oligomers by GC-MS and FAB-MS established that: (i) the Ara and Gal residues are in the furanose form; (ii) the non-reducing termini of arabinan consists of the structural motif \([\beta \rightarrow \alpha \text{ Araf} \rightarrow (1 \rightarrow 2) \beta \rightarrow \alpha \text{ Araf} \rightarrow 3,5 \alpha \rightarrow \beta \text{ Araf} \rightarrow (1 \rightarrow 5) \alpha \rightarrow \beta \text{ Araf}]\); (iii) the majority of the arabinan chains consist of 5-linked \( \alpha \)-Araf residues with branching introduced by 3,5-\( \alpha \)-Araf; (iv) the arabinan chains are attached to C-5 of some of the 6-linked Galf residues, and approximately two to three arabinan chains are attached to the galactan core; (v) the galactan regions consist of linear alternating 5- and 6-linked \( \beta \)-Galf residues; (vi) the galactan region of AG is linked to the C-6 of some of the MurNGly residues of PG via a special diglycosyl-\( P \) bridge, \( \alpha \)-L-Rhap–(1 \rightarrow 3)–\( \beta \)-GlcNAC–(1 \rightarrow P); (vii) the mycolic acids are located in clusters of four on the terminal hexaarabinofuranoside, but only two-thirds of these are mycolated.

More recently, we obtained oligosaccharide fragments containing up to 26 glycosyl residues from which molecular weights and alkylation patterns were determined by FAB-MS. The extended non-reducing ends of the arabinan were shown to consist of a tricosarabinoside ("23mer"), with three such units attached to the galactan unit. The galactan was also isolated and was found to consist of 23 Gal residues of the repeating linear structure, \([\beta \rightarrow \alpha \text{ Galf} \rightarrow (1 \rightarrow 5) \beta \rightarrow \alpha \text{ Galf} \rightarrow (1 \rightarrow 6)]\), devoid of any branching, thereby demonstrating that the points of attachment of the arabinan chains are close to the reducing end of galactan, itself linked to PG via the linker disaccharide-\( P \).

Biosynthesis of mAGP

The importance of understanding the biosynthesis of the mycolic acid–peptidoglycan–arabinogalactan complex is more in the context of new drug development against tuberculosis (TB) and less in
defining the bacterial factors responsible for the disease process.

One of the great developments in recent years in the chemical definition of the cell-wall core was the recognition of the diglycosyl-P bridge lying between the linear PG and the linear galactan. We immediately speculated, based on cell-wall biosynthesis in other bacteria, that the entire linkage unit, galactan, and arabinan are synthesized as a unit on a polyprenyl-P carrier lipid. Subsequent work has provided the experimental basis of what was once speculation. The synthesis of the entire core is initiated on a decaprenyl-P with synthesis of the linker unit, and then there is concomitant extension of the galactan and arabinan chains while this intermediate is transported through the cytoplasmic membrane. Some of the enzymatic and genetic details of this complex process have been defined. For instance, the rhamnose of the diglycosyl-P bridge originates in dTDP-rhamnose, and the biosynthesis and genetics of dTDP-rhamnose synthesis and transfer of the rhamnose unit have been defined. The Galf units of the galactofuran originate in UDP-Galf, which, in turn, originates in UDP-Galp catalyzed by UDP-Galp mutase. The Rv3808c gene product is responsible for the polymerization of the bulk of the galactofuran has been worked out by Crick et al. It is not fully defined the biochemical transformations encoded by many of the genes in the mycolic acid synthetic pathway. Acetyl CoA carboxylase (Acc) gives rise to malonyl CoA. FabD exchanges CoA for ACP. FabH is responsible for the coupling of fatty acid synthases I and II (FAS I and FAS II, giving rise to the first precursor of mycolic acid synthesis, a β-ketoacyl-ACP, which then undergoes reduction by MabA, dehydration (the dehydratase has not yet been identified), enoyl ACP reduction catalyzed by inhA, the target of INH, and then another round of elongation catalyzed by KasA and KasB. Schaeffer et al. (personal communication) have already demonstrated the essentiality of several of these enzymes and configured them into high throughput screens for identification of new drugs against TB.

The polyprenyl-P carrier lipid involved in the synthesis of the cell-wall core has now been identified as decaprenyl-P. It is responsible for the synthesis of PG and also for the synthesis of the linker unit—Rha–GlcNAc–P—and subsequently galactan and arabinan formation. The carrier lipid is also responsible for the synthesis of polyprenyl-P-linked Araf, the source of arabinan. The carrier lipid may also be involved in the attachment of new mycolic acids to the cell wall. For that reason, the synthesis of the carrier lipid has been worked out by Crick et al. It is not synthesized by the mevalonic acid pathway, but by the new non-mevalonate pathway involving deoxyxylulose-5-P, a pathway that provides another good drug target.

Biosynthesis and biological functions of the free lipids of M. tuberculosis

Intercalated within the lipid environment provided by the mycolic acids of the mAGP complex are the lipids that have intrigued researchers for over five decades: the phthiocerol dimycocerosate (DIM/PDIM), cord factor/dimycolyltrehalose, the sulfolipids (SLs), the PIMs, etc. Knowledge of their roles in “signaling” events, in pathogenesis, and in the immune response is now emerging. Knowledge of the genome has greatly helped in defining the biosynthesis of these products.

Cord factor/TDM

The most notable biological attribute of cord factor is the characteristic toxicity it exerts in mice; a few repeated intraperitoneal injections of small amounts (10 μg), dissolved in paraffin oil, killed a
majority of animals; however, larger doses (50–100 µg) are rarely lethal. Death is associated with intense peritonitis and acute pulmonary hemorrhage. The biochemical mechanism of cord factor toxicity has been reviewed by Goren and Brennan.\textsuperscript{30} Apparently, cord factor stimulates host NADase activity, leading to lower levels of host NAD, especially in lung, liver, and spleen tissues and reduced activities of a range of NAD-dependent microsomal enzymes. According to Kato, cord factor intoxication is attributable to a direct physical effect on mitochondrial membranes, resulting in disruption of electron flow along the mitochondrial respiratory chain and of oxidative phosphorylation. Cord factor, especially when delivered in a variety of formats, is immunogenic, granulomagenic, and adjuvant-active. The antitumor activity of cord factor has also been studied extensively (see Goren and Brennan).\textsuperscript{30}

**Sulfolipids**

Studies by Gangadharan et al. and Goren established a clear correlation between the presence of SLs in \textit{M. tuberculosis} isolates and virulence for guinea pigs.\textsuperscript{30} The most virulent strains “were prolific in elaborating strongly acidic lipids (notably SLs), whereas the attenuated ones were notably deficient in these components.” Armstrong and Hart had observed that phagosomes containing viable virulent \textit{M. tuberculosis} of mouse peritoneal are resistant to fusion with secondary lysosomes or dense granules, leading to extensive studies by Goren and Hart and the conclusion that sulfolipids were implicated in phagosome–lysosome fusion. However, this conclusion was later questioned by Goren himself, leaving in limbo the actual mechanism of SL involvement in disease virulence.\textsuperscript{30}

**Phthiocerol dimyocerosate**

DIM is a major lipid of the tubercle bacillus. When immunologists refer to the tubercle bacillus as a “ball of wax,” they are referring to DIM. It is highly apolar. It was discovered by Rudolf Anderson at Yale in the 1940s. It is, in fact, a wax. It has 35 carbons as methyl or methylene groups. It does have two reactive hydroxy groups, but these are esterified with two fatty acids, the mycocerosic acids, every bit as apolar as the phthiocerol itself with 28 methyl or methylene groups. Intriguing older work has also implicated DIM in \textit{M. tuberculosis} virulence. Goren et al. (see Ref.\textsuperscript{30}) reported that DIM was found in a wide selection of patient isolates of \textit{M. tuberculosis}, of a wide spectrum of virulence, and was apparently as prominent in highly attenuated strains as in the most virulent ones, including \textit{M. tuberculosis} H37Rv. However, Goren et al. (in Ref.\textsuperscript{30}) observed that an attenuated mutant of H37Rv lost the capacity for synthesis of DIM. Goren mentioned that the “biochemical lesion might rest in an inability of the mutant to synthesize mycocerosic acids, phthiocerol,” a statement of extraordinary prescience in light of later developments (see below).

**DIM and related lipids and methyl branched fatty acids**

There has been much recent progress on the biosynthesis of DIM. One of the unexpected features of the \textit{M. tuberculosis} genome was a 50-kb fragment containing 13 genes which had the hallmarks of a type-I modular polyketide synthase, which initially left researchers puzzled, since the chemists had not shown any evidence of classical polyketides in \textit{M. tuberculosis}. However, in the midst of this operon, transcribed in the other direction, is a gene, \textit{mas}, mycocerosic acid synthase, and Rainwater and Kolattukudy,\textsuperscript{31} Azad et al.,\textsuperscript{32} Fernandes and Kolattukudy,\textsuperscript{33} and Fitzmaurice and Kolattukudy\textsuperscript{34} had shown that this encoded an enzyme with a subunit molecular weight of 238,000, an iterative polyketide synthase, that acts like a FAS I, but produces mycocerosic acid after four rounds of extension of a C\textsubscript{18} fatty acid but using methyl malonyl CoA instead of malonyl CoA, which is the source of the methyl branches of mycocerosic acid.

This early work by Kolattukudy et al. suggested that the other genes of the cluster were responsible for phthiocerol synthesis, again through a modular polyketide synthase mechanism, and they provided evidence to this effect. A FAS I-like system synthesizes a straight-chain fatty acyl group attached to its enzyme. Then module I (Pps1) contains an acyl transferase domain, a ketoacyl synthase domain and a keto reductase domain, allowing introduction of the hydroxyl group. Pps2 does the same thing. However, Pps 4 and 5 use methylmalonyl CoA, allowing the introduction of the characteristic methyl branches of phthiocerol. Finally, there are reduction and decarboxylation steps to produce the phthiocerol, and the acyl CoA synthase (FadD28) responsible for the attachment of mycocerosic acids, synthesized by Mas, to the phthiocerol synthesized by the Pps modules, has been identified.

Cox et al.\textsuperscript{35} and Camacho et al.\textsuperscript{36} have recently begun to address the role of DIM in disease
pathogenesis. Using signature-tagged transposon mutagenesis, Cox et al.\textsuperscript{35} isolated mutants with insertions upstream of \textit{fadD26} in the \textit{pps} region and within \textit{fadD28} and \textit{mmpl 7}. The French group\textsuperscript{36} isolated mutants with insertions upstream from \textit{fadD26} and within \textit{fadD28} and within \textit{drrC} and \textit{mmpl 7}. The results from both laboratories were similar and were intriguing. No DIM production or production of a related lipid was seen in the upstream \textit{fadD26}, i.e., \textit{pps} mutant or the \textit{fadD28} mutant. However, the \textit{mmpl 7} mutant could synthesize DIM, but it could not excrete DIM into the medium: it and a \textit{drrC} mutant were defective in DIM export/excretion.

Already some interesting biology has been done on the DIM-less mutants.\textsuperscript{35,36} They show attenuated growth in the mouse lung. They also show much higher cell-wall permeability, all fitting in with old work from Goren and Brennan (see Ref.\textsuperscript{30}) that a DIM-less variant of \textit{M. tuberculosis} H37Rv showed attenuated growth in the guinea pig.

Reports have also recently come from Kolattukudy’s lab on the generation of mutants devoid of SLs.\textsuperscript{37} These are trehalose derivatives, largely defined by Goren, but with a sulfate group at the 2 position of trehalose (see Ref.\textsuperscript{30}). They also have methyl-branched fatty acids, called the phospho-ceramic or hydroxy-phthioceranic acids. A mas-like \textit{pks} gene responsible for their synthesis was identified which was successfully disrupting by homologous recombination.\textsuperscript{37} In light of the controversial role of SLs in TB pathogenesis, it will be intriguing in the future to see the phenotype of these mutants.

**PIMs, LM, and LAM**

Another group of free lipids that are benefiting from the genome in terms of biosynthesis and function are the PIMs, LM, and LAM.

The PIMs themselves were described by C.E. Ballou and colleagues in the 1960s (see Ref.\textsuperscript{30}). They are based on phosphatidylinositol (PI) and attached to the inositol there may be from one to six mannoses. The PIMs with two mannoses are the most common in \textit{M. tuberculosis}. Some years ago, we discovered that molecules described as LM, lipomannan, and LAM, lipoarabinomannan, are extensions of PIM, in that, in the case of LM, the mannan chain was extended and branching was introduced; in the case of LAM, an additional arabinan is attached.\textsuperscript{38}

Chatterjee and Khoo\textsuperscript{38} and Puzo and colleagues\textsuperscript{39} are mostly responsible for the modern-day structure of LAM.

There is intense interest in the role of LAM in TB. LAM from \textit{M. tuberculosis} has short mannose-containing oligosaccharide “caps” that allow it to bind to the mannose receptor on macrophages, unlike the product from \textit{M. smegmatis}, which has no such mannose caps. Also, LAM can bind to Toll receptors and can physically insert itself into membranes, inducing all sorts of signaling events important in the host response in TB. Thus, understanding the biosynthesis of LAM and the creation of LAM-less mutants are major priorities.

Based on structures, one would predict that the first event would be synthesis of PI, then the addition of mannose to give PIM\textsubscript{1}, PIM\textsubscript{2}, and so on. However, in the original genome sequence, the genes for phosphatidylserine and phosphatidylethanolamine synthesis were clearly annotated. However, all other genes for phospholipid synthesis, including the one that may be responsible for PI synthesis, could not be distinguished, because all of them use CDP-diacylglycerol as the source of diglyceride, and its binding motif dominated the structures of the genes. Hence, these were named \textit{pgs} A, \textit{pgs} A\textsubscript{2}, \textit{pgs} A\textsubscript{3}. M. Jackson from Dr. B. Gicquel’s laboratory, working initially with us and then back home, overexpressed each one of these in \textit{M. smegmatis} and established that only \textit{pgs} A (Rv2611c) was responsible for the synthesis of PI.\textsuperscript{40} \textit{Pgs} A is part of a small operon containing Rv2609c, Rv2610c, and Rv2611c. Jackson et al.\textsuperscript{41} have now shown that \textit{pim} A (Rv2610c) adds on the first mannose to PI, the Rv2611c then probably adds on a fatty acid. Previously, Schaeffer et al. had shown that \textit{Pim} B, half a chromosome away, adds on the second mannose.\textsuperscript{42} So, \textit{Pgs} A, the PI synthase, allows the condensation of inositol and the diglyceride of the CDP-diacylglycerol derivative. \textit{Pim} A allows the addition of mannose 1 followed by an acyl function. \textit{Pim} B allows the addition of the second mannose. Gurcha et al.\textsuperscript{43} recently described a \textit{Pim} C that attaches the third mannose; they speculated that this is the direct precursor of LM and LAM.

Even though the early genes in the PIM/LAM biosynthetic pathway are essential and required for growth, it should now be possible to generate LAM-less mutants in order to explore the biological role of LAM.

**Acknowledgements**

Research in the author’s laboratory was supported by grants AI-18357 and AI-46393 from the National Institute of Allergy and Infectious Diseases, National Institutes of Health.
References


34. Fitzmaurice AM, Kolattukudy PE. An acyl-CoA synthase (acoA) gene adjacent to the mycocerosic acid synthase


