Glycopeptidolipid synthesis in mycobacteria

H. Billman-Jacobe

Department of Microbiology and Immunology, University of Melbourne, Victoria, 3010, Australia

Glycopeptidolipids (GPL) are abundant components of the cell walls of several species of mycobacteria. They have been implicated in the pathogenesis of the opportunistic mycobacteria belonging to the *Mycobacterium avium* complex. This article describes studies on the biochemistry and genetics of GPLs synthesis in the saprophyte *Mycobacterium smegmatis* and opportunistic pathogen *M. avium*.

MYCOBACTERIA have characteristic lipid-rich cell envelopes that afford the cells protection from desiccation, chemical disinfectants and some antibiotics. The mycobacterial cell envelope has a multilaminar structure with the outermost layer comprising an asymmetric lipid bilayer with mycolic acids on the inner leaflet. The composition of the outer leaflet varies between species of mycobacteria but generally contains a mixture of glycolipids and phospholipids. Glycopeptidolipids (GPLs) occur among the lipids of the outer layer of some non-tuberculous mycobacteria that cause opportunistic infections (*Mycobacterium avium* complex, *Mycobacterium peregrinum*, *M. chelonae*, *M. abscessus*) and saprophytic mycobacteria such as *M. smegmatis*[1,2]. Most GPLs are the alkali-stable C-type GPLs however variant forms have been observed, i.e. the alkali-labile serine-containing GPLs of *M. xenopi*[3,4]. The structure of C-type GPLs is comprised of a lipopeptide core consisting of a mixture of 3-hydroxy and 3-methoxy C_{26-34} fatty acids amidated with a tripeptide-amino alchohol (D-Phe-D *allo* Thr-D-Ala-L alaninol). GPLs vary in the glycosylation of the *allo* thr and/or alaninol. The glycosylation is responsible for the serotype specificity of GPLs in the *M. avium* complex. *M. smegmatis* has a simple range of GPLs whose *allo* Thr residue is glycosylated with 6-deoxy talose (dTal) which is in turn diacetylated. The alaninol is glycosidically attached to 3,4 di-*O*-Me rhamnose (Rha) or 2,3,4 tri-*O*-Me Rha. Four major species of GPLs occur in *M. smegmatis* mc^2155 when cultivated aerobically on nutritionally complex media. The species vary in the degree of methylation of the acyl chain (3-hydroxy or 3-methoxy C_{26-34} fatty acids) and the Rha (3,4 di-*O*-Me Rha or 2,3,4 tri-*O*-Me Rha). GPL-1 and GPL-2 have 3-methoxy C_{26-34} fatty acids but differ in the methylation of Rha. GPL-1 has 2,3,4 tri-*O*-Me Rha GPL-2 has 3,4 di-*O*-Me Rha. GPL-1a and GPL-2a have tri- and di-Me Rha as above however these GPLs have 3-hydroxy C_{26-34} fatty acids. Although these are the major species that can be detected by thin layer chromatography (TLC) in *M. smegmatis*, minor intermediates such as a GPL containing 3-*O*-Me Rha can be detected with more sensitive methods. Variations in the GPL composition can occur depending on culture conditions. Under conditions of carbon starvation, *M. smegmatis* mc^2155 can add an additional hexose on the dTal[5]. *M. avium* strains contain a more diverse range of GPL structures. The *M. avium* GPLs have the same lipopeptide core as the *M. smegmatis* GPLs but the alaninol is glycosylated with 3-*O*-Me Rha or 3,4-di-*O*-Me-Rha[6]. An oligosaccharide extending from 6-dTal varies in composition between the different serotypes of *M. avium*. For example serotype-2 *M. avium* has 2,3-di-*O*-Me-fucosyl-(1,3)-rhamnopyranosyl-(1,2)-6-deoxy talose. Belisle and colleagues identified a locus desiganted *ser*2 which contains some of the gense encoding enzymes required synthesis of the haptenic disaccharide of a serotype-2 *M. avium* (TMC724)[7]. They showed that the simpler GPLs naturally found in *M. smegmatis* could serve as intermediates in the biosynthesis of a *ser*2-like GPL in recombinant *M. smegmatis*[8].

We initially embarked on a project to find cell wall mutants by screening an *M. smegmatis* Tn611 mutant[9] library using three criteria. First, susceptibility to a lytic mycobacteriophage; second, changes in colony morphology and third, composition of the chloroform : methanol extractable lipids of the cell wall. The rationale for phage susceptibility was that if the receptor of the phage was absent or obscured then a mutant may become phage-resistant. GPLs have been shown to be the receptor of mycobacteriophage[10] D4. However, in this study we were not specifically looking for GPL mutants so we used a lytic mutant of mycobacteriophage L5 called L5cd5 whose receptor is not known[11]. Populations of Tn611 mutants were mixed with L5cd5 and the survivors were recovered on agar plates. A large proportion of survivors had a rough colony morphology, the smooth colony forms were susceptible to lysis by phage. Further investigation of the rough forms showed that they could absorb phage and were susceptible to lysis therefore they must have retained the receptor for the phage[12]. Our hypothesis is that the rough cells, which are very hydrophobic and aggregate in large clumps, escaped phage challenge by aggregation. In the second screening we examined the colony morphology of many Tn611 mutants without mycobacteriophage challenge. The rationale in this screening was based on

e-mail: hbj@unimelb.edu.au

the well-established principle that changes in the composition of the cell surface may correlate to altered colony morphology. The most obvious variant colony form was rough colony morphology as previously observed during phage susceptibility screening. Rough colony mutant of *M. smegmatis* mc^2155 arise spontaneously at a rate of 1 rough mutant/10,976 colonies screened whereas in Tn611 mutagenesis they were recovered at a rate of ~ 3 per 1000 transposon mutants. Analysis of the crude cell wall extract from these mutants revealed that they all had defects in GPL synthesis[12].

Another notable mutant isolated on the basis of having altered colony morphology was the one that had a defect in synthesis of PIM6 although it retained PIM2 and accumulated PIM4 (unpublished results). In an effort to find more PIM mutants we devised a miniature biosynthetic labelling assay whereby mutants were grown in 96-well plates and were pulsed with $+^3$H-Man then the lipids were extracted and resolved by HPTLC and autoradiography. Despite screening over 2000 strains, no further PIM mutants were isolated. We have since concentrated on creating PIM mutants by targetted mutagenesis in specific genes predicted to encode enzymes involved in PIM synthesis.

Peptide synthetase

Genetic characterization of the GPL mutants was initially performed by using inverse PCR and/or linker mediated PCR[13] and to obtain sequences flanking the sites of Tn611 insertion in the *M. smegmatis* chromosome. The majority of the mutants had Tn611 inserted into a 17.9 kb open reading frame that resembled a peptide synthetase[12]. This family of proteins are multimodular enzymes that catalyse the non-ribosomal synthesis of peptides. Each module in the enzymes determines the incorporation of a specific amino acid in the nascent peptide. In the case of the *M. smegmatis* peptide synthetase, the enzyme had four modules of which the first three contained racemase domains. The nature and arrangement of modules was consistent with the structure of the GPL lipopeptide which contains D-phe-D *allo* thr-D ala-L alaninol. The peptide synthetase gene was named *mps* and all *mps* :: Tn611 mutants were devoid of GPLs but contained the normal complement of PIMs and other cell wall components. Other mutants with transposon insertions in *mps* have been reported by Recht *et al.*[14]. The mutants in that study had rough colony morphology, lacked sliding motility and were devoid of GPLs.

Rhamnosyl 3-*O*-methyltransferase

Another rough Tn611 mutant contained some GPLs although the profile on HPTLC showed that mature GPLs were absent. The transposon had inserted into the *mtf*1 gene that encoded a methyltransferase[15]. Compositional analysis of the mutant revealed that all of the GPLs were glycosylated with 6dTal and Rha however the majority of Rha was not methylated. Other cell wall components were the same as the wildtype controls. We were able to show that Mtfl was a S-adenosyl methionine-dependent methyltransferase which appeared to methylate the OH at C3 of Rha and proposed that it catalysed the initial methylation of Rha. A novel GPL with 4-*O*-Me Rha accumulated in minor amounts in the mutant. This GPL species was never detected in the parent strain of *M. smegmatis* therefore we think that it does not represent a normal biosynthetic intermediate. Complementation restored mature GPL synthesis to the *mtf1* :: Tn611 mutant however the complemented colonies retained the rough morphology. The total amount of GPLs in the complemented mutant only represented about 65% of the cellular GPL content of wildtype cells possibly due to suboptimal expression of *mtf*1 encoded by the complementation plasmid. It appears that an absence or a reduction in GPL content can result in the rough colony morphology.

GPL biosynthetic locus

The *mtf*1 and *mps* genes are located within a 41.4 kb locus on the *M. smegmatis* chromosome. The locus contains 15 genes that are predicted to encode enzymes involved in GPL biosynthesis (Figure 1). The cluster begins with a triplet of transmembrane protein genes (*tmpt* A, B and C) whose function has yet to be determined. Both TmptB and TmptC have 12 putative transmembrane domains whereas the smaller TmptA protein has a single transmembrane domain. A transposon mutant of *tmpt*C has been reported[14]. The mutant had rough colony morphology, lacked sliding motility and was devoid of GPLs. The role of TmptC in this phenotype has not been determined. One other transposon mutant generated by these investigators displayed an intermediate sliding phenotype. The mutant had a disrupted *atf*1 gene (ORF 10) which was

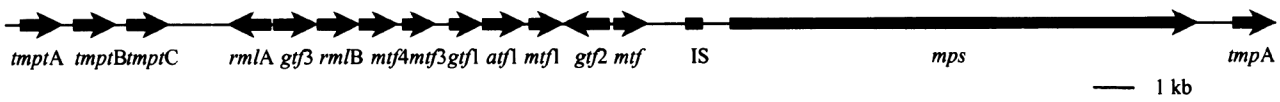

Figure 1. Glycopeptidolipid biosynthetic gene cluster of *M. smegmatis*.

predicted to encode an acetyltransferase[16]. The composition of GPLs from the mutant indicated that the GPLs was not acetyated, consistent with the expected phenotype of an *atf*1 mutant.

ORFs 4 and 6 are *rml*A, encoding a putative glucose-1-phosphate thymidylyl transferase and *rml*B, encoding a putative dTDP glucose 4,6 dehydrogenase. These enzymes pontentially involved in the synthesis of the deoxyhexoses, Rha and 6dTal which are subsequently incorporated into GPLs. There are three putative glycosyltransferases (ORF5, *gtf*3; ORF9, *gtf*1 and ORF12c, *gtf*3) although only a talosysl- and a rhamnosyl-transferase are required for synthesis of the GPLs described above. The additional glycosyltransferase may be have a role in modifying GPLs under specific conditions. Ojha *et al.*[5] showed that if *M. smegmatis* is grown in carbon-limited culture then a novel polar GPL species accumulated in the cell wall in addition to the usual species. They were able to show that the novel polar GPL was a hyperglycosylated derivative of the apolar species[5]. It is interesting speculate that expression of one or more of the *gtf* genes may be regulated by in response to environmental signals.

Other genes in the locus include four methyltransferases (*mtf*1-4; ORFs 11, 13, 8 and 7 respectively). The function of Mtf1 was discussed above. We have engineered mutants by disrupting each of the three other *mtf* genes. The *mtf*2 mutant is unable to methylate the hydroxyl of fatty acid of the GPL[17] whereas *mtf*3 encodes a rhamnosyl 4-*O*-methyltransferase and *mtf*4 encodes a rhanmosyl 2-*O*-methyltransferase (manuscript in preparation).

The largest ORF of the locus is *mps*, encoding peptide synthetase (described above). The final gene I have included in the locus at this time is *tmp*A which encodes a 272aa ploypeptide with 6 putative transmembrane domains. We have a single transposon mutant of *tmp*A that has rough colony morphology and accumulates only mature GPL (i.e. versions with 2,3,4-tri-*O*-methylated Rha) however no further investigations have been conducted on this mutant.

Elucidation of the function of some of the enzymes involved in GPL biosynthesis in *M. smegmatis* has allowed us to propose a biosynthetic pathway. Synthesis of the lipopeptide from a lipid precursor may proceed by sequential addition of amino acids by the *mps* peptide synthetase or the tripeptide-amino alcohol may be synthesized and added as a single unit to the lipid. The only evidence to support the latter process is provided by David *et al.*[18] who showed that phenylalanine analogue inhibited GPL synthesis and radioactively labelled phenylalanine was incorporated into GPLs. Once the lipopeptide has been completed then the *allo* thr and alaninol are glycosylated with dTal and Rha respectively. Which glycosyltransferases catalyse these events has yet to be determined. Acetylation of dTal by the *atf*1 gene product then takes place. After acetylation, two populations of GPLs appear which differ in the methylation of the hydroxyl of the lipid. It is not possible, at this stage, to confidently predict the exact order of the catalytic events described above because the simplest GPL intermediates that have been detected possess both sugars, acetyl groups and some have a methoxy lipid[14]. We are confident of the sequence of methylation of the Rha proceeds with modification of the hydroxy of C3, C4 then C2.

M. avium

Studies of the genetics of GPL biosynthesis in *M. avium* are more difficult than in *M. smegmatis* because most strains of *M. avium* are difficult to genetically manipulate. Earlier studies by Belisle and co-workers showed that a serotype 2 strain of *M. avium*, had a large locus containing a number of genes that were predicted to encode enzymes for the biosynthesis of the GPL core and possibly for synthesis of the haptenic oligosaccharide 2,3-di-*O*-Me-fucosyl-(1 → 3)-rhamnosyl-(1 → 2)-6-d talose which is characteristic of serotype 2 strains of *M. avium*[7]. The *ser*2 locus can be lost from the *M. avium* chromosome through homologous recombination between two copies of IS1601 which flank the gene cluster[19]. Loss of *ser*2 results in loss of GPL synthesis and rough colony morphology. A rhamnosyltransferase gene (*rtf*A) is the only gene whose function as been determined experimentally. The *rtf*A gene was into *M. smegmatis* and transformants were able to synthesize a novel GPL that had an additional Rha attached to dTal thereby showing that the simpler GPLs, like those of *M. smegmatis*, can serve as biosynthetic precursors in the synthesis of serotype-specific GPLs[8]. Krzywinska and Schorey[20] have made a careful comparison of the GPL biosynthetic cluster of 3 serotypes of *M. avium*. These studies show that the 5 region of the gene cluster is highly conserved between all strains tested however the 3 region has a diverged in difference strains though deletions and insertions of sequences including mobile elements such as insertion sequences. The organization of the *M. avium* and *M. smegmatis* GPL gene clusters would suggest that the 5 regions are reasonably conserved. However the functional characterization of the enzymes encoded by these genes will be required before we can rely upon the predictions based on sequence alignments.

1. Brennan, P. J. and Nikaido, H., *Annu. Rev. Biochem.*, 1995, **64**, 29–63.
2. Daffe, M. and Draper, P., *Adv. Microbial. Physiol.*, 1998, **39**, 131–203.
3. Riviere, M. and Puzo, G., *Biochemistry*, 1992, **31**, 3575–3580.
4. Besra, G. S., Khoo, K. H., Belisle, J. T., McNeil, M. R., Morris, H. R., Dell, A. and Brennan, P. J., *Carbohydr. Res.*, 1994, **251**, 99–114.
5. Ojha, A. K., Varma, S. and Chatterji, D., *Microbiology*, 2002, **148**, 3039–3048.
6. Daffe, M., Laneelle, M. A. and Puzo, G., *Biochim. Biophys. Acta*, 1983, **751**, 439–443.

7. Belisle, J. T., Pascopella, L, Inamine, J. M., Brennan, P. J. and Jacobs, Jr. W. R., *J. Bacteriol.*, 1991, **173**, 6991–6997.
8. Eckstein, T. M., Silbaq, F. S., Chatterjee, D., Kelly, N. J., Brennan, P. J. and Belisle, J. T., *J. Bacteriol.*, 1998, **180**, 5567–5573.
9. Guilhot, C., Otal, I., Vanrompaey, I., Martin, C. and Gicquel, B., *J. Bacteriol.*, 1994, **176**, 535–539.
10. Furuchi, A. and Tokunaga, T., *J. Bacteriol.*, 1972, **111**, 404–411.
11. Donnelly-Wu, M. K., Jacobs, Jr. W. R. and Hatfull, G. F., *Mol. Microbiol.*, 1993, **7**, 407–417.
12. Billman-Jacobe, H., Mc Conville, M. J., Haites, R. E., Kovacevic, S. and Coppel, R. L., *Mol. Microbiol.*, 1999, **33**, 1244–1253.
13. Prod'hom, G., Lagier, B., Pelicic, V., Hance, A. J., Gicquel, B. and Guilhot, C., *FEMS Microbiol. Lett.*, 1998, **158**, 75–81.
14. Recht, J., Martinez, A., Torello, S. and Kolter, R., *J. Bacteriol.*, 2000, **182**, 4348–4351.
15. Patterson, J., McConville, M., Haites, R., Coppel, R. and Billman-Jacobe, H., *J. Biol. Chem.*, 2000, **275**, 24900–24906.
16. Recht, J. and Kolter, R., *J. Bacteriol.*, 2001, **183**, 5718–5724.
17. Jeevarajah, D., Patterson, J. H., McConville, M. J. and Billman-Jacobe, H., *Microbiology*, 2002, **148**, 3079–3087.
18. David, H. L., Rastogi, N., Clavel-Seres, S. and Clement, F., *Curr. Microbiol.*, 1988, **17**, 61–68.
19. Eckstein, T. M., Inamine, J. M., Lambert, M. L. and Belisle, J. T., *J. Bacteriol.*, 2000, **182**, 6177–6182.
20. Krzywinska, E. and Schorey, J. S., *Vet. Microbiol.*, 2003, **91**, 249–264.