Isolation and Expression of a Gene Cluster Responsible for Biosynthesis of the Glycopeptidolipid Antigens of *Mycobacterium avium*

John T. Belisle
Lisa Pascopella
*University of Montana - Missoula*, lisa.pascopella@mso.umt.edu
Julia M. Inamine
Patrick J. Brennan
William R. Jacobs Jr.

Follow this and additional works at: https://scholarworks.umt.edu/pchs_pubs

Part of the Public Health Commons

Let us know how access to this document benefits you.

**Recommended Citation**

https://scholarworks.umt.edu/pchs_pubs/14

This Article is brought to you for free and open access by the Public and Community Health Sciences at ScholarWorks at University of Montana. It has been accepted for inclusion in Public and Community Health Sciences Faculty Publications by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.
Isolation and Expression of a Gene Cluster Responsible for Biosynthesis of the Glycopeptidolipid Antigens of *Mycobacterium avium*

JOHN T. BELISLE,¹ LISA PASCOPPELLA,² JULIA M. INAMINE,¹ PATRICK J. BRENNAN,¹ AND WILLIAM R. JACOBS, JR.²*

Department of Microbiology, Colorado State University, Fort Collins, Colorado 80523,¹ and Howard Hughes Medical Institute and Department of Microbiology and Immunology, Albert Einstein College of Medicine of Yeshiva University, Bronx, New York 10461²

Received 7 June 1991/Accepted 27 August 1991

Bacteria within the *Mycobacterium avium* complex are prominent in the environment and are a source of serious disseminated infections in patients with AIDS. Serovars of the *M. avium* complex are distinguished from all other mycobacteria and from one another by the presence of highly antigenic glycolipids, the glycopeptidolipids (GPLs), on their surfaces. A genome library of DNA from serovar 2 of the *M. avium* complex was constructed in the Escherichia coli-*Mycobacterium* shuttle cosmid, pYUB18, and used to clone and express in *Mycobacterium smegmatis* the genes responsible for the biosynthesis of the oligosaccharide segment of the *M. avium* serovar 2-specific glycopeptidolipid. The responsible gene cluster was mapped to a 22- to 27-kb functional region of the *M. avium* genome. The recombinant glycolipid was also isolated by high-pressure liquid chromatography and chemically characterized, by gas chromatography-mass spectrometry and fast atom bombardment-mass spectrometry, to demonstrate that the lipopeptide core originated in *M. smegmatis*, whereas the oligosaccharide segment arose from the cloned *M. avium* genes. This first-time demonstration of the cloning and expression, in a nonpathogenic mycobacterium, of the genes encoding complex cell wall glycoconjugates from a pathogenic mycobacterium presents a new approach for studying the role of such products in disease processes.

Since the time of their recognition until recently, *Mycobacterium avium* organisms have seldom been associated with severe life-threatening, disseminated infections in humans (30). However, it is now recognized that the most frequently encountered bacterial infections in patients with AIDS are due to members of the *M. avium* complex (10, 23). The single most distinguishing feature of *M. avium* is the copious amounts of highly antigenic glycolipids, the glycopeptidolipids (GPLs), present on the bacterium’s surface (5, 20). Other than these, the cell walls of *M. avium* apparently display an architecture that is common to the genus (11). The GPLs which distinguish the 30-odd serovars of the *M. avium* complex and bear responsibility for their serotypic differences are multiglycosylated at the threonine substituent of an otherwise invariant lipopeptide core (Fig. 1). An extraordinary array of sugars within this variable oligosaccharide region confer serological specificity; the structure of the serospecific GPL of serovar 2 (ssGPL-2) (9) is shown in Fig. 1A. In addition, all members of the *M. avium* complex contain simpler, nonspecific GPLs (nsGPLs) which differ from the ssGPLs in that they are singly glycosylated at the threonine substituent with either 6-deoxytalose or 3-O-methyl-6-deoxytalose (5, 6). While considerable effort has been devoted to structural elucidation of the GPLs, their biosynthesis and roles in drug permeability and disease pathogenesis have barely been addressed (7). The fact that *Mycobacterium smegmatis*, unlike *M. avium*, is now amenable to genetic transfer systems (26, 27) provides a unique opportunity to clone the DNA responsible for the synthesis of complex secondary gene products associated with the cell walls of pathogenic *M. avium*. The added fact that *M. smegmatis* contains the simpler, singly glycosylated nsGPLs (Fig. 1B) and is devoid of the multiglycosylated ssGPLs (1, 12) also presented the opportunity to express in another mycobacterium the particular gene cluster responsible for the biosynthesis of the serovar 2-specific sugar segment of its ssGPL. These experiments have set the stage for a new generation of studies of the biosynthesis and function of not only the GPLs but also those other mycobacterial cell wall products that distinguish pathogenic from nonpathogenic mycobacteria.

**MATERIALS AND METHODS**

Bacterial strains and other materials. Escherichia coli χ2764 and χ2819 were used for transduction and amplification of the *M. avium* cosmid library (15). The *M. avium* serovar 2 strain TMC 724 was the source of genomic DNA. *M. smegmatis* mc²155 (27) was used for electroporation and subsequent expression of the recombinant serovar 2-specific GPL (R-ssGPL-2).

Restriction endonucleases, calf intestine alkaline phosphatase, and T4 ligase were purchased from Boehringer Mannheim Biochemicals (Indianapolis, Ind.), New England Biolabs, Inc. (Beverly, Mass.), and Bethesda Research Laboratories (Gaithersburg, Md.). Lysozyme, lipase, and proteinase K were purchased from Sigma Chemical Co. (St. Louis, Mo.). The Genius kit from Boehringer Mannheim Biochemicals was employed in the preparation of nonradioactive DNA probes.

**DNA and cosmid isolation.** Genomic DNA was isolated from *Mycobacterium* spp. by using slight modifications of the method of Whipple et al. (29). Cell pellets stored at
Serovar Specific GPL-2 from *M. avium* Serovar 2

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fatty acyl - NH - D - Phe - D - aThr - D - Ala - L - Alaninol - O - (3,4 - Me₂ - Rha)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-dTal</td>
<td>Rha - 2,3 - Me₂ - Fuc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Non-Specific GPL from *M. smegmatis*

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fatty acyl - NH - D - Phe - D - aThr - D - Ala - L - Alaninol - O - (2,3,4 - Me₃ - Rha)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-dTal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FIG. 1. Structures of native GPLs. (A) Structure of the multiglycosylated *M. avium* ssGPL-2 (9). (B) Structure of one of the singly glycosylated nsGPLs of *M. smegmatis* (analogous structures with different fatty acyl and sugar substitutions are present in all serovars of the *M. avium* complex) (22).

 colonies of kanamycin-resistant *M. smegmatis* were patched onto 7H10(+) agar plates containing 25 mg of kanamycin per ml and allowed to incubate for 4 to 5 days. These clones were picked and placed on Whatman no. 40 cellulose filter paper. The filters were then subjected to colony dot blot enzyme-linked immunosorbent assay (ELISA) (2) employing the serovar 2-specific CS-17 monoclonal antibody (MAB) (24).

Mapping of the ser2 gene cluster. The pJTB21 cosmid was digested with the restriction enzymes EcoRI, BamHI, and PstI (19). DNA fragments corresponding to the *M. avium* insert were purified by electroelution (13) and labeled with digoxigenin by using a Genius kit. A second probe was constructed by digesting pJTB21 with DraI plus SacI and labeling the isolated insert fragment with 32P via random oligonucleotide primers. *Escherichia coli* X2764 cells were infected with the phage particles containing the *M. avium* library and grown at 30°C on Luria-Bertani plates containing 25 mg of kanamycin per ml (15). Approximately 7,000 *E. coli* recombinant clones were lifted onto nitrocellulose filters, lysed, and hybridized with the pJTB21 probes (19). Positive clones were isolated and grown in superbroth with 25 mg of kanamycin per ml at 30°C. The recombinant cosmid was isolated as circular DNA and mapped with restriction enzymes DraI, CiaI, HindIII, and SspI. Additionally, each recombinant cosmid was electrophorotized into *M. smegmatis* mc2155 and the *M. smegmatis* transformants were screened for serovar 2 GPL expression by colony dot blot ELISA and gas chromatography (GC) of the GPL-associated sugars.

Isolation and purification of GPLs. *M. avium* serovar 2 was grown as previously described (20). *M. smegmatis*:pYUB18 and *M. smegmatis*:m. avium recombinant clones were grown on 7H10(+) agar plates (15 by 150 mm) containing 25 mg of kanamycin per ml for 5 to 6 days at 37°C. Cells were scraped from plates, autoclaved, and lyophilized. Lipids were extracted with CHCl₃·CH₃OH (2:1) and subjected to a mild alkali treatment which, by destroying nonspecific acylglycerols, served as a purification step for the GPLs (20). The resulting lipids were applied to columns (25 by 1.5 cm) of silicic acid-Celite (2:1) and eluted with increasing concentrations of CH₃OH in CHCl₃ (8); the total GPL population from each strain was collected in the 25% CH₃OH eluate. Total GPL (10 mg), in CHCl₃, was loaded onto an Alltech Econosphere high-pressure liquid chromatography (HPLC) column (4.6 by 250 mm) packed with 5 μm of silica. The individual ns- and ssGPLs were eluted with increasing
concentrations of CH$_3$OH in CHCl$_3$ (20). Fractions were examined for purity by thin-layer chromatography in CHCl$_3$-CH$_3$OH (9:1) (3).

Other analytical methods. The amino acid compositions of the various total and purified GPL fractions were determined by GC-mass spectrometry (MS) of the N$_2$-(O)-heptfluorobutyryl isobutyl derivatives (14, 17) on an HP-1 capillary column by using a Hewlett-Packard model 5890 gas chromatogram and model 5970 mass detector (20). The total and purified GPL fractions were also analyzed for their sugar compositions by GC-MS of the alditol acetate derivatives (20) on a DB-23 capillary column. Fast atom bombardment (FAB)-MS was performed on 10-μg samples of selected GPL fractions on a 3-nitrobenzyl alcohol matrix spiked with 1 μl of 0.1 M KCl by using a VG 7070 extra-high-frequency mass spectrometer and an ion saddle field gun operated at 7 to 8 kV and 1 mA with xenon gas (16, 21).

FIG. 2. Screening of M. smegmatis transformants with the M. avium serovar 2-specific MAb CS-17. (A) Initial screening of M. smegmatis transformed with the pYUB18:M. avium library, resulting in identification of the MA21-21 clone. (B) Secondary screening of M. smegmatis transformed with pJTB21. The negative (−) and positive (+) controls are M. smegmatis:pYUB18 and M. avium serovar 2, respectively.

RESULTS

Construction and screening of the M. avium cosmid library. A major obstacle in expressing recombinant secondary gene products in mycobacteria has been the inability to clone the required complex of genes. In the present instance, the capability to clone genes for glycosyl- and methyltransferases and possibly for enzymes responsible for de novo synthesis of specific sugar residues had to be developed. In order to clone the expected large multigene segment, we chose the E. coli-Mycobacterium shuttle cosmid vector pYUB18 (26, 27). The genomic library of M. avium serovar 2 DNA was constructed first in E. coli by utilizing bacteriophage λ-mediated packaging. The transduction of E. coli produced approximately 6,700 kanamycin-resistant E. coli clones containing recombinant cosmids with M. avium inserts of 32 to 40 kb; the efficiency of this transduction was

FIG. 3. Functional mapping of the ser2 gene cluster. Restriction endonuclease maps of the cloned M. avium inserts are presented with a summary of results from the application of colony dot blot ELISA and GC to the GPLs from the recombinant clones that helped define the functional ser2 region. Abbreviations: C, Clal; H, HindIII; S, SspI; N/D, not done. Symbols: −, no R-ssGPL-2 detected; +, weak R-ssGPL-2 expression; ++, strong R-ssGPL-2 expression.
calculated to be $1.8 \times 10^4$ transformants per µg of insert DNA. Covalently closed circular recombinant cosmids were isolated from E. coli and electroporated into M. smegmatis mc²155. Upon screening of 1,200 kanamycin-resistant M. smegmatis clones, only one (MA21-21) reacted with the CS-17 serovar 2-specific MAb (Fig. 2A). This recombinant clone contained a 44-kb cosmid which was termed pJTB21. After isolation from MA21-21 and amplification in E. coli x2764, pJTB21 was electroporated back into M. smegmatis mc²155, resulting in a 100% cotransformation frequency of kanamycin resistance with serovar 2 activity (Fig. 2B). This exercise demonstrated that the pJTB21 cosmid contained those M. avium serovar 2-specific genes necessary for synthesis of the antigenic segment of the ssGPL-2, i.e., the distal 2,3-di-O-methyl-α-L-fucopyranosyl-(1→3)-α-L-rhamnopyranosyl[2,3-Me₂-α-L-Fucp-(1→3)-α-L-Rhap] disaccharide (8, 9, 24).

**Mapping of the ser2 gene cluster.** In order to map the minimal contiguous M. avium DNA fragment necessary for expression of the serovar 2-specific oligosaccharide, probes were derived from the M. avium insert of pJTB21. Application of these probes yielded 22 positive E. coli transformants containing individual pYUB18:M. avium cosmids. After isolation from E. coli, each cosmid was transformed into M. smegmatis, and the resulting M. smegmatis:M. avium clones were screened with the CS-17 MAb. Of the 22 clones, 3 (pJTB22, pJTB41, and pJTB211) produced a positive reaction. However, as summarized in Fig. 3, the degree of reactivity with the CS-17 MAb varied among these clones, suggesting altered levels of oligosaccharide expression. The alkali-stable GPLs were isolated from each clone and examined by GC for the presence of the serovar 2-specific sugars. As expected, the clones containing pJTB21, pJTB22, and pJTB41 possessed the serovar 2-specific sugars, Rha and 2,3-Me₂-Fuc (Fig. 3). Small quantities of these sugars were also observed in the GPLs from a clone that did not react with the CS-17 MAb; this clone contained cosmid pJTB11 (Fig. 3). Functional mapping of the DNA of pJTB21 and the additional 22 cosmids resulted in the recognition of an M. avium genomic region of approximately 60 kb, of which a 48-kb segment is depicted in Fig. 3. The insert in the original serovar 2 GPL-producing cosmid, pJTB21, was estimated to

**FIG. 4. GC of alditol acetates derived from the sugars of the various GPL-containing fractions.** (A) M. avium serovar 2; (B) M. smegmatis:pYUB18; (C) MA21-21. Details of derivatizations and chromatography conditions are provided in the text. Istd, internal standard of 6-O-methylgalactose.
be 32 kb, and, of this, 22 to 27 kb appeared to be necessary for the production of the serovar 2-specific segment of its GPL molecule (Fig. 3). We have termed this region the ser2 gene cluster. A comparison of the ELISA and GC results with maps of the inserts (Fig. 3) indicates that pJTB11, pJTB22, and pJTB211 are devoid of those sequences on the right side of the gene cluster deemed necessary for increased expression of the ssGPL-2. Expression seems not to be influenced by vector sequences, since the two inserts of pJTB22 and pJTB211 were cloned in opposite orientations in pYUB18. Thus, the right segment of this gene cluster may be a regulatory region.

Chemical evidence for the generation of a recombinant GPL. In order to lend chemical weight to the immunological evidence for the presence of the serovar 2-determinant in recombinant M. smegmatis, the total GPLs from M. avium serovar 2, the transformed M. smegmatis containing pJTB21 (MA21-21), and control M. smegmatis containing pYUB18 were isolated and analyzed by a number of chemical means. GC-MS revealed that the GPLs of the recombinant MA21-21 possessed those sugars that typify the M. smegmatis nsGPLs, namely, 2,3,4-Me2-Rha, 3,4-Me2-Rha, and 6-dTal, as well as 2,3-Me2-Fuc and Rha (Fig. 4), i.e., the sugars characteristic of the serovar 2-specific oligosaccharide. Also, large amounts of 3-Me-Rha were evident, the origin of which is not known at this time. Individual GPLs from each organism were purified by HPLC. The presence of the tetrapeptide core was confirmed by the observation that all contained D-Phe, D-allo-Thr, D-Ala, and L-alaninol in equimolar quantities (5, 6). Thin-layer chromatography demonstrated that the purified ssGPL-2 migrated directly below its nonspecific relatives (Fig. 5, lanes 1 and 2), consistent with previous observations (8, 9). The purified R-ssGPL-2 from MA21-21 was shown to be less polar than the corresponding M. avium ssGPL-2 (Fig. 5, lanes 1 and 3), most likely reflecting differences in the fatty acyl content. GC-MS analysis of the sugars from the purified R-ssGPL-2 confirmed the presence of 2,3,4-Me2-Rha, 6-dTal, Rha, and 2,3-Me2-Fuc (data not shown) and thus the presence of the serovar 2-specific oligoglycosyl unit in the recombinant glycolipid.

The presence of the 2,3,4-Me2-Rha unit in R-ssGPL-2 indicated that the core originated in M. smegmatis (Fig. 1B), whereas chemical proof of the presence of Rha and 2,3-Me2-Fuc supported the antigenic evidence that the distal segment of the oligosaccharide originated in the M. avium serovar 2 genome (Fig. 1A). In order to confirm the proposed structure of the R-ssGPL-2, FAB-MS, with its potential to provide structural information about the holistic molecule, was employed (16, 21). A (M + K)⁺ pseudomolecular ion at m/z 1509 was observed (Fig. 6A), supporting the evidence for a GPL endowed with the serovar 2-specific oligosaccharide: a 2,3,4-Me2-Rha unit linked to L-alaninol and an amino-terminal-linked saturated C28 mono-hydroxy fatty acid (1). One of the major nsGPLs isolated from the host M. smegmatis also contained 2,3,4-Me2-Rha and 6-dTal and showed an (M + K)⁺ ion of m/z 1189 (Fig. 6B), demonstrating the presence of an nsGPL with an amide-linked saturated C28 mono-hydroxy fatty acid in M. smegmatis, presumably the template for serovar 2-directed glycosylation.

**DISCUSSION**

This landmark study represents the first reported cloning of genes responsible for the synthesis of complex cell wall structures of mycobacteria. Thus, genes from a member of the M. avium complex were successfully expressed in M. smegmatis, and the resulting products were shown to reflect the serotypic identity of the parental M. avium. The responsible genes were mapped to a 22- to 27-kb contiguous M. avium genomic fragment, the so-called ser2 gene cluster. Biochemical and immunological analyses of M. smegmatis clones possessing the pJTB21 and pJTB208 cosmids indicate that the left terminus of ser2 is contained within a 4-kb region, whereas, according to analysis of pJTB11 and pJTB22, the right terminus of ser2 mapped to a 1.5-kb region. Also, the decreased expression of the serovar 2-specific oligosaccharide with increased deletions of the right terminus of the ser2 gene cluster points to a possible region of regulation for gene expression.

Although the ser2 cluster must possess the genes required for serovar 2-specific oligosaccharide biosynthesis, the 22- to 27-kb fragment is larger than expected for the two to three glycosyltransferases and methyltransferases expected to facilitate synthesis of the oligoglycosyl unit. Thus, this gene cluster may contain multiple operons or it may encompass the genes required for the de novo synthesis of fucos and rhamnose, perhaps analogous to the Salmonella rfb operon (4, 28, 31). Conceivably, the ser2 gene cluster could contain the genes required for synthesis of the GPL core. However, if this was so, the R-ssGPL-2 should possess an M. avium-type core rather than an M. smegmatis-type core. Chemical analysis of the R-ssGPL-2, by GC-MS and FAB-MS, demonstrated the presence of a 2,3,4-Me2-Rha unit and a saturated C28 mono-hydroxy fatty acyl unit, the same substituents attributed to the simple GPLs of M. smegmatis (1, 12, 22) but not to those of M. avium (3, 6, 22). The chemical characteristics of the R-ssGPL-2 also indicate that the simpler monoglycosylated GPLs act as biosynthetic precursors of the multiglycosylated GPLs of M. avium, a relationship that had been assumed but never demonstrated. The chemical evidence also suggests that the genes encoding the synthesis of the components of the GPL core, namely,
Sequence analysis, once specific genes within the ser2 gene cluster are identified, should be of considerable benefit in this sense. *M. smegmatis*, with a growth rate 10 times higher than that of mycobacterial pathogens (25), was chosen in the present study as the preferred host for cloning *M. avium* genes because its cell wall is similar in several key respects to that of *M. avium*, and, unlike *M. avium*, it is genetically manipulable. Application of this strategy to other major mycobacterial pathogens, *Mycobacterium leprae* and *Mycobacterium tuberculosis*, should greatly facilitate an analysis of the pathways leading to the biosynthesis of their own distinctive cell wall entities (7, 20). In addition, this initial study of genetic mechanisms underlying the biosynthesis of the exceptionally complex glycoconjugates of mycobacterial cell walls should provide new approaches to the age-old issue of the roles of such entities in pathogenesis and to the problem of intractable chemotherapy of *M. avium* infections.

ACKNOWLEDGMENTS

This work was supported by grants AI-26170, AI-18357, and AI-30189 from the National Institute of Allergy and Infectious Diseases, National Institutes of Health.

We thank Barry R. Bloom and Michael R. McNeil for encouragement and intellectual and technical input.

REFERENCES


