Structural Characterization of a Specific Glycopeptidolipid Containing a Novel *N*-Acyl-Deoxy Sugar from *Mycobacterium intracellulare* Serotype 7 and Genetic Analysis of Its Glycosylation Pathway[∇][†]

Nagatoshi Fujiwara,[1]* Noboru Nakata,[2] Shinji Maeda,[1,3] Takashi Naka,[1,4] Matsumi Doe,[5] Ikuya Yano,[4] and Kazuo Kobayashi[1]‡

Department of Host Defense, Osaka City University Graduate School of Medicine, Osaka,[1] Department of Microbiology, Leprosy Research Center, National Institute of Infectious Diseases, Tokyo,[2] Molecular Epidemiology Division, Mycobacterium Reference Center, The Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, Tokyo,[3] Japan BCG Central Laboratory, Tokyo,[4] and Department of Chemistry, Graduate School of Science, Osaka City University, Osaka,[5] Japan

The nontuberculous *Mycobacterium avium-Mycobacterium intracellulare* complex (MAC) is distributed ubiquitously in the environment and is an important cause of respiratory and lymphatic disease in humans and animals. These species produce polar glycopeptidolipids (GPLs), and of particular interest is their serotype-specific antigenicity. Structurally, GPLs contain an *N*-acylated tetrapeptide-amino alcohol core that is glycosylated at the C terminal with 3,4-di-*O*-methyl rhamnose and at the D-*allo*-threonine with a 6-deoxy-talose. This serotype nonspecific GPL is found in all MAC species. The serotype-specific GPLs are further glycosylated with a variable haptenic oligosaccharide at 6-deoxy-talose. At present, 31 distinct serotype-specific GPLs have been identified on the basis of oligosaccharide composition, and the complete structures of 14 serotype-specific GPLs have been defined. It is considered that the modification of the GPL structure plays an important role in bacterial physiology, pathogenesis, and host immune responses. In this study, we defined the complete structure of a novel serotype 7 GPL that has a unique terminal amido sugar. The main molecular mass is 1,874, and attached to the tetrapeptide-amino alcohol core is the serotype 7-specific oligosaccharide unit of 4-2′-hydroxypropanoyl-amido-4,6-dideoxy-2-*O*-methyl-β-hexose-(1→3)-α-L-rhamnose-(1→3)-α-L-rhamnose-(1→3)-α-L-rhamnose-(1→2)-α-L-6-deoxy-talose. Moreover, we isolated and characterized the serotype 7-specific gene cluster involved in glycosylation of the oligosaccharide. Nine open reading frames (ORFs) were observed in the cluster. Based on the sequence homology, the ORFs are thought to participate in the biosynthesis of the serotype 7 GPL.

About 10% of mycobacterial diseases are caused by nontuberculous mycobacteria. Among them, the closely related *Mycobacterium avium* and *Mycobacterium intracellulare* are commonly grouped as *M. avium-intracellulare* complex (MAC). Organisms of this complex are ubiquitous in nature and have been isolated from water, soil, plants, house dust, and other environmental sources. MAC infections have become increasingly common, and MAC is the most common cause of disease due to nontuberculous mycobacteria in humans (13). These organisms have distinctive laboratory characteristics, are not communicable from person to person, and are often resistant to standard antituberculosis drugs.

The mycobacterial cell wall contains numerous antigenic or immunoregulatory glycolipid molecules with great structural diversity that are considered to be involved in the bacterial

virulence through host immune responses (6, 19, 31, 33). The polar glycopeptidolipids (GPLs) produced by MAC species are of particular interest because of their serotype-specific antigenicity (9). To better understand the mechanisms of pathogenesis and drug resistance of MAC, it is necessary to elucidate the molecular structure and biochemical characteristics of the lipid components.

Structurally, GPLs contain a tetrapeptide-amino alcohol core, D-phenylalanine-D-*allo*-threonine-D-alanine-L-alaninol (D-Phe-D-*allo*-Thr-D-Ala-L-alaninol), with an amido-linked 3-hydroxy or 3-methoxy C26-C34 fatty acid at the N-terminal of D-Phe (5). The D-*allo*-Thr and terminal L-alaninol are further linked with 6-deoxy-talose (6-d-Tal) and 3,4-di-*O*-methyl-rhamnose (3,4-di-*O*-Me-Rha), respectively. This core GPL is found in all species of MAC and shows a common antigenicity (1). The serotype-specific GPLs are further glycosylated with a variable haptenic oligosaccharide at 6-deoxy-talose. At present, 31 distinct serotype-specific polar GPLs have been identified biochemically, and the complete structure of GPLs is partly defined by the serotype 1 to 4, 8, 9, 12, 14, 17, 19 to 21, 25, and 26 GPLs (9). The standard technique for classification of MAC strains has been serologic typing based on the oligosaccharide (OSE) residue of the GPL. More recently, advanced chemical synthesis of various haptenic OSEs was demonstrated, and the genes encoding the glycosylation pathways in

biosynthesis of GPL were identified and characterized (10, 17, 26).

We have established a rapid method for serodiagnosis of MAC disease using the GPL and GPL core antigens of MAC and have also shown that the levels of GPL and GPL core antibodies reflect disease activity (12, 14, 22, 23). Otherwise, it has been reported that serotype-specific GPLs participate in the pathogenesis and immunomodulation in the host (2, 18). It is reasonable to hypothesize that modification of the GPL structure plays an important role not only in antigenicity but also in host immune responses and bacterial physiology. In this study, we explored the complete OSE structure of serotype 7 GPL, which has a unique terminal amido sugar and is characterized by a serotype 7-specific gene cluster involved in the glycosylation of the OSE, using the cosmid library technique.

MATERIALS AND METHODS

Bacterial strain and preparation of GPL. *M. intracellulare* serotype 7 strain (ATCC 35847) was purchased from the American Type Culture Collection (Manassa, VA). Two clinical strains (NF 111 and NF 112) of *M. intracellulare* serotype 7 were isolated and kindly provided by Ryoji Maekura (National Hospital Organization, Toneyama National Hospital, Osaka, Japan). GPL was prepared as described previously (20, 22). Briefly, the ATCC 35847 strain of *M. intracellulare* serotype 7 was grown in Middlebrook 7H9 broth (Difco Laboratories, Detroit, MI) with 0.5% glycerol and 10% Middlebrook oleic acid-albumin-dextrose-catalase enrichment (Difco Laboratories) at 37°C for 2 to 3 weeks. Heat-killed bacteria were collected by centrifugation at $2,400 \times g$ for 15 min, sonicated, and extracted with chloroform-methanol (2:1, vol/vol). The extractable lipids were dried and hydrolyzed with 0.2 N sodium hydroxide in methanol at 37°C for 2 h. The alkaline-stable lipids were dissolved in chloroform-methanol (2:1, vol/vol). GPL was purified by preparative thin-layer chromatography (TLC) of silicagel G (Uniplate; 20×20 cm, 250 μm; Analtech, Inc., Newark, DE). The TLC plate was repeatedly developed with chloroform-methanol-water (65:25:4 and 60:16:2, vol/vol/vol) until a single spot was obtained. To reveal the GPL band, the TLC plate was exposed to iodine vapor after development. The GPL band was marked, and then the silicagels were scraped to elute the GPL with chloroform-methanol (2:1, vol/vol) through a small glass column.

Preparation of OSE moiety. The OSE moiety elongated from D-*allo*-Thr was released from GPL by alkaline borohydride reductive β-elimination (7, 20). GPL was dissolved in ethanol with 5 mg/ml sodium borohydride and 0.5 M sodium hydroxide, followed by stirring at 60°C for 16 h. The reaction mixture was decationized with Dowex 50WX8 beads (Dow Chemical Company, Midland, MI) and centrifuged at $2,400 \times g$ for 15 min. The supernatant was collected and coevaporated with 10% acetic acid in methanol under nitrogen to remove the boric acid. The dried residue was partitioned in two layers in chloroform-methanol (2:1, vol/vol) and water. The upper aqueous phase was recovered and evaporated. In these processes, the serotype 7-specific OSE was purified as an oligoglycosyl alditol.

Derivatization of perdeuteromethylated and perdeuteroacetylated OSE. Perdeuteromethylation was conducted by the modified procedure of Hakomori (16). The dried OSE was dissolved in a mixture of dimethyl sulfoxide (1 ml) and sodium hydroxide (1 mg), and 1 ml of deuteromethyl iodide was added. The reaction mixture was stirred at room temperature for 15 min, followed by the addition of 1 ml of water and 1 ml of chloroform. After centrifugation at $2,400 \times g$ for 15 min, the upper water layer was discarded. The chloroform layer was washed repeatedly with water to remove any water-soluble components and then evaporated to dryness. Perdeuteroacetylation of OSE was performed by reacting OSE with pyridine-deuteroacetic anhydride (1:1, vol/vol) for 16 h, and the product was dried completely to remove the pyridine.

FAB/MS analysis of intact GPL and OSE. The molecular weight was determined by fast atom bombardment-mass spectrometry (FAB/MS) with a JMS SX102A double-focusing mass spectrometer (JEOL, Tokyo, Japan). The target gas was xenon, and the accelerating voltage was 8 kV. Intact GPL was analyzed by FAB/MS in both positive and negative ion mode with *m*-nitro-benzyl alcohol as the matrix. The perdeuteroacetylated derivative of OSE was analyzed by FAB/MS in positive ion mode with *m*-nitro-benzyl alcohol as the matrix.

GC and GC/MS analyses of OSE. To determine the glycosyl composition and linkage position, gas chromatography (GC) and GC/MS analyses of partially methylated alditol acetate derivatives were performed. Partially deuterometh-

lated alditol acetates were prepared from perdeuteromethylated OSE by hydrolysis with 2 N trifluoroacetic acid at 120°C for 2 h, reduction with 10 mg/ml sodium borodeuteride at 25°C for 2 h, and acetylation with acetic anhydride at 100°C for 1 h (8, 21). GC was performed using a 5890 series II gas chromatograph (Hewlett Packard, Avondale, PA) equipped with the fused capillary column SPB-1 (30 m, 0.25-mm inside diameter; Supelco, Inc., Bellefonte, PA). Helium was used for electron impact (EI)/MS and *iso*-butane was used for chemical ionization (CI)/MS as the carrier gas. The JMS SX102A double-focusing mass spectrometer was connected to GC as the mass detector. The molecular separator and the ion source energy were 70 eV for EI and 30 eV for CI, and the accelerating voltage was 8 kV. The DL configurations of Rha residues were determined by comparative GC/MS analysis of trimethylsilylated (*R*)-(−)-2-butyl glycosides and (*S*)-(+)-2-butyl glycosides prepared from an authentic L-Rha standard (15).

NMR analysis of OSE. The GPL was deuterium dissolved in chloroform-d (CDCl$_3$)/methanol-d$_4$ (CD$_3$OD) (2:1, vol/vol). To define the anomeric configurations of each glycosyl residue, ^{1}H and ^{13}C nuclear magnetic resonance (NMR) was employed. Both homonuclear correlation spectrometry, and ^{1}H-detected [^{1}H, ^{13}C] heteronuclear multiple-quantum correlation (HMQC) were recorded with a Bruker AVANCE-600 spectrometer (Brucker BioSpin K.K., Osaka, Japan), as described previously (9, 20, 27).

Construction of the *M. intracellulare* cosmid library. Genomic DNA of *M. intracellulare* serotype 7 strain ATCC 35847 was prepared by mechanical disruption of bacterial cells, which was accomplished by homogenizing a bacterial pellet with glass beads in phosphate-buffered saline, followed by phenol-chloroform extraction and precipitation with ethanol. Genomic DNA fragments randomly sheared to 30- to 50-kb fragments during the extraction process were fractionated and electroeluted from agarose gels using a Recochip (Takara, Bio, Inc., Kyoto, Japan). These DNA fragments were rendered blunt-ended using T4 DNA polymerase and deoxynucleoside triphosphates, followed by ligation to dephosphorylated arms of pYUB412 (XbaI-EcoRV and EcoRV-XbaI), which was kindly given by William R. Jacobs, Jr. (Department of Microbiology and Immunology, Albert Einstein College of Medicine, New York, NY). After in vitro packaging using Gigapack III Gold extracts (Stratagene, La Jolla, CA), recombinant cosmids were introduced into the *Escherichia coli* STBL2 [F$^-$ *mcrA* Δ(*mcrBC-hsdRMS-mrr*) *endA1 recA1 lon gyrA*96 *thi supE44 relA1* λ$^-$ Δ(*lac-proAB*)] and stored at −80°C in 50% glycerol.

Isolation of cosmid clones carrying the *rtfA* gene and sequence analysis. PCR was used to isolate cosmid clones carrying the rhamnosyltransferase (*rtfA*) gene with primers *rtfA*-F (5′-TTTTGGAGCGACGAGTTCATC-3′) and *rtfA*-R (5′-GTGTAGTTGACCACGCCGAC-3′). *rtfA* encodes an enzyme responsible for the transfer of Rha to 6-d-Tal in OSE (11, 26). The insert of cosmid clone 49 was sequenced using a BigDye Terminator, version 3.1, Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) and an ABI Prism 310 gene analyzer (Applied Biosystems). The putative function of each open reading frame (ORF) was identified by similarity searches between the deduced amino acid sequences and known proteins using BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) and FramePlot (http://www.nih.go.jp/~jun/cgi-bin/frameplot.pl) with the DNASIS computer program (Hitachi Software Engineering, Yokohama, Japan). To confirm the presence of ORFs in clinical strains of *M. intracellulare* serotype 7, PCRs were performed by using the primers of each ORF (Table 1).

Nucleotide sequence accession number. The nucleotide sequence reported here has been deposited in the NCBI GenBank database under accession number AB274811.

RESULTS

Purification and molecular weight of intact GPL. Serotype 7 GPL of *M. intracellulare* was purified repeatedly by silica gel TLC and showed as a single spot on the TLC plate (Fig. 1A). The R_f values of serotype 7 GPL were 0.42 and 0.24 on TLC with the developing solvents (chloroform-methanol-water at 65:25:4 and 60:16:2, vol/vol/vol, respectively). The main pseudomolecular ions of positive and negative FAB/MS were m/z 1897 for [M+Na]$^+$ and m/z 1873 for [M-H]$^-$ (Fig. 1B). These results showed that the main molecular weight of serotype 7 GPL was 1874 and differed from the molecular weights of other GPLs (9, 25), implying that it has a novel carbohydrate chain elongated from D-*allo*-Thr.

TABLE 1. Sequences of primers used for amplifying ORFs in this study

ORF	Forward sequence (5′–3′)	Reverse sequence (5′–3′)
ORF 1	GTGAAATTTGCCCTGG CGAG	TCAGCCAAAGCGCCT CGTGT
ORF 2	GTGGTATTGAATACAC GCAT	TCAAACCTCCGCCGA TTTCG
ORF 3	GTGCCCGAAGTTCCTT CCGA	TCAACGGGTGCGGTG TCGCG
ORF 4	ATGCCTGCTGAGATCC CGTT	CTATGTGCTCACTTTC TTAA
ORF 5	TTGGCAGCCTGGAGCG ACCG	TCACAGTTGCGTTCC GTCAC
ORF 6	GTGACGCGCCTTGACA CGGG	TCATGCGATTGCGCC CTGTT
ORF 7	GTGGCAATTCGCGCCG CGCC	TCACCCAAACTTGCG GCCCT
ORF 8	GTGGCGTTGGGCGCCC CTAG	TCAGCCGCTGATAAA CGCTC
ORF 9	ATGAGCGAGCCGGCTG GCCG	CTATTGGGACGGACC CCTGA

Glycosyl composition of OSE. To determine the glycosyl composition of OSE, alditol acetate derivatives from serotype 7 GPL were analyzed by GC and GC/MS. The structurally defined serotype 4 GPL was used as a reference standard (9, 28). From the comparison of the retention time and spectra on GC and GC/MS (Fig. 2), the alditol acetate derivatives of serotype 7 GPL showed the presence of 3,4-di-*O*-Me-Rha, Rha, 6-d-Tal, and an unknown sugar residue (X1) in a ratio of approximately 1:3:1:1. The alditol acetate of X1 was eluted with a retention time (22.9 min) greater than that of glucitol acetate on SPB-1 columns. The CI/MS spectrum of X1 showed [M+H]$^+$ at m/z 420 as a parent ion and m/z 360 as a loss of 60 (acetate). The fragment ions of X1 sugar showed characteristic

patterns in EI/MS. m/z 332 and 87 indicated the cleavage of C-4–C-5, and m/z 290, 272, 230, and 170 were fragmented as a loss of 42 (ketene) or 60 (acetate). Similarly, m/z 302 and 117 indicated the cleavage of C-2–C-3, and m/z 200 was fragmented as a loss of 42 and 60 (Fig. 3). These results indicated that X1 was 4,6-dideoxy-hexose (Hex). The molecular weight, 419, of X1 and fragment ions 115 and 87 were consistent with the presence of one amido group attached to 2′-hydroxypropanoic acid. Taken together, X1 was structurally determined to be 4-2′-hydroxypropanoyl-amido-4,6-dideoxy-2-*O*-methyl-Hex.

Linkage and sequence analyses of OSE. To determine the glycosyl linkage and sequence of OSE, GC/MS of perdeutero-methylated alditol acetates and FAB/MS of deuteroacetylated oligoglycosyl alditol from serotype 7 OSE were performed. The EI/MS spectra of perdeuteromethylated alditol acetates (Fig. 4) were assigned to the three major peaks, 1,3,4,5-tetra-*O*-deuteromethyl-2-*O*-acetyl-6-deoxytalitol (m/z 109, 132, 154, 167, and 214), 2,4-di-*O*-deuteromethyl-1,3,5-tri-*O*-acetyl-rhamnitol (m/z 121, 134, 205, 240, and 253), and 3-*O*-deuteromethyl-1,5-di-*O*-acetyl-4-2′-*O*-deuteromethyl-propanoyl-deuteromethylamido-4,6-dideoxy-2-*O*-methyl-hexitol (m/z 87, 105, 165, 222, and 300). These results demonstrated that the 6-d-Tal residue was linked at C-2, Rha was linked at C-1 and C-3, and the nonreducing terminus, 4-2′-hydroxypropanoyl-amido-4,6-dideoxy-2-*O*-methyl-Hex, was 1-linked. The FAB/MS spectrum of deuteroacetylated oligo-glycosyl alditol from serotype 7 OSE afforded the expected molecular ions [M+Na]$^+$ and [M+H]$^+$ at m/z 1398 and 1376, respectively, together with the characteristic mass increments in the series of glycosyloxonium ions formed on fragmentation at m/z 322, 558, 794, and 1030 (Fig. 5). The fragment ion, m/z 322, was in accord with a 4-2′-*O*-deuteroacetyl-propanoyl-amido-4,6-dideoxy-2-*O*-methyl-3-*O*-deuteroacetyl-hexosyl residue at the nonalditol terminus. In addition, each fragment and parent ion (m/z 558, 794, 1030, and 1376) showed three additional molecular
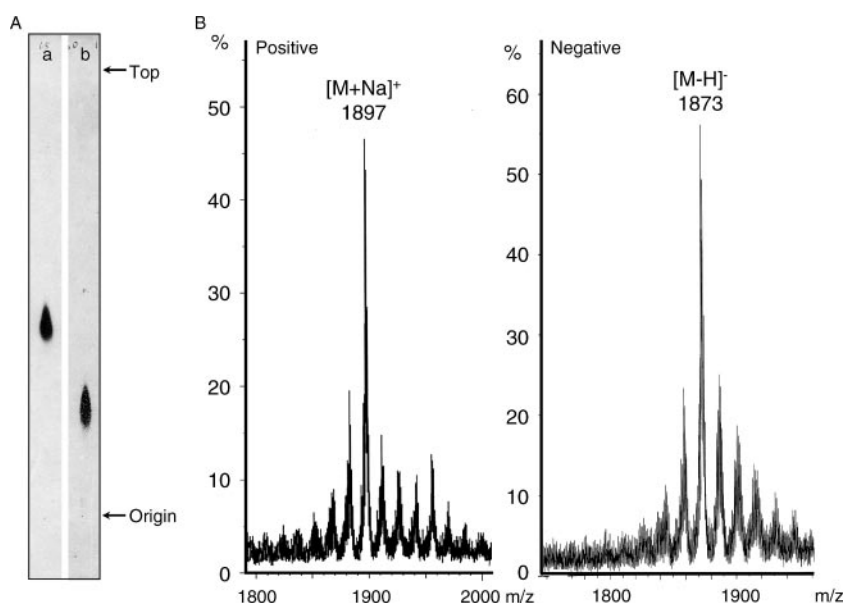

FIG. 1. TLC patterns (A) and FAB/MS spectra (B) of intact serotype 7 GPL derived from *M. intracellulare* strain ATCC 35847. The TLC solvent systems were chloroform-methanol-water (a, 65:25:4; b, 60:16:2, by volume). The TLC plate was sprayed with 10% sulfuric acid in ethanol and was charred at 180°C for 5 min. The FAB/MS spectra were acquired using an *m*-nitro-benzyl alcohol matrix, and the pseudomolecular ions were detected as [M+Na]$^+$ in positive mode and [M-H]$^-$ in negative mode.
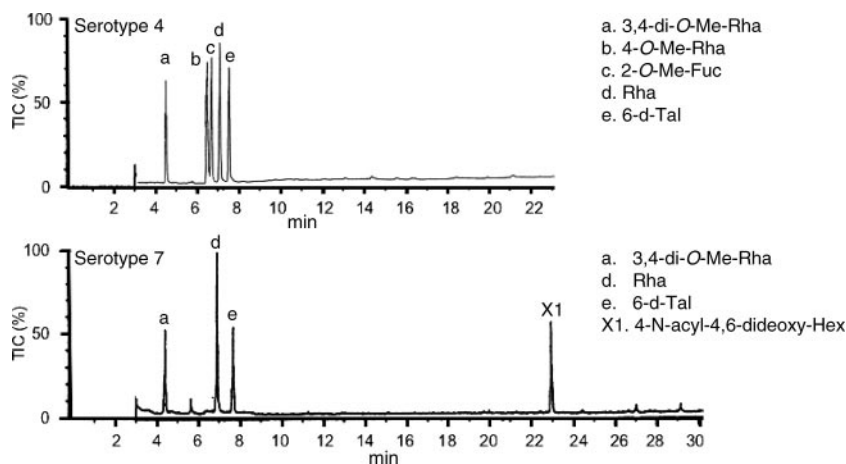
FIG. 2. GC spectra of the alditol acetate derivatives from serotype 4 and 7 GPLs. Total ion chromatograms are shown. GC was conducted on an SPB-1 fused silica column with a temperature program of 160°C for 2 min, followed by a rise of 4°C/min to 220°C and then maintained at 220°C for 13 min.

Rha and 6-d-Tal residues. Rha residues were determined to be in the L absolute configuration by comparative GC/MS analysis of trimethylsilylated (R)-$(-)$-2-butyl glycosides and (S)-$(+)$-2-butyl glycosides (see Fig. S1 in the supplemental material). Taken together, these results established the sequence and linkage arrangement: 4-2′-hydroxypropanoyl-amido-4,6-dideoxy-2-O-methyl-Hex-(1→3)-L-Rha-(1→3)-L-Rha-(1→3)-L-Rha-(1→2)-L-6-d-Tal, exclusively.

NMR analysis of serotype 7 OSE. The ^{1}H NMR and ^{1}H-^{1}H correlation spectrometry analyses of GPL revealed six distinct anomeric protons with corresponding H1-H2 cross-peaks in the low field region at δ5.01, 4.98, 4.96, 4.88, 4.69 (J_{1-2} 2–3 Hz, indicative of α-anomers), and 4.55 (a doublet, J_{1-2} 7.8 Hz, indicative of a β-hexosyl unit). When further analyzed by ^{1}H-detected [^{1}H-^{13}C] two-dimensional HMQC, the anomeric protons resonating at δ5.01, 4.98, 4.96, 4.88, 4.69, and 4.55 have

C-1s resonating at δ101.66, 101.17, 95.80, 102.40, 99.87, and 103.21, respectively. The J_{CH} values for each of these protons were calculated to be 171, 169, 171, 175, 171, and 159 Hz by the measurement of the inverse-detection nondecoupled two-dimensional HMQC shown in Fig. 6. These results are summarized (see Table S1 in the supplemental material), and establish that the terminal amido-Hex was a β configuration and the others were α-anomers.

Cloning and sequence of the serotype 7 GPL biosynthesis cluster. To isolate the serotype 7 GPL biosynthesis cluster, the genomic cosmid library of an *M. intracellulare* serotype 7 strain, ATCC 35847, was constructed. Primers were designed for amplification of a region corresponding to the *rftA* gene. DNA was extracted from each clone by boiling. By using colony PCR with *rftA* primers, more than 100 cosmid clones were tested, and the positive clone 49 was isolated from the *E. coli* transductants. The
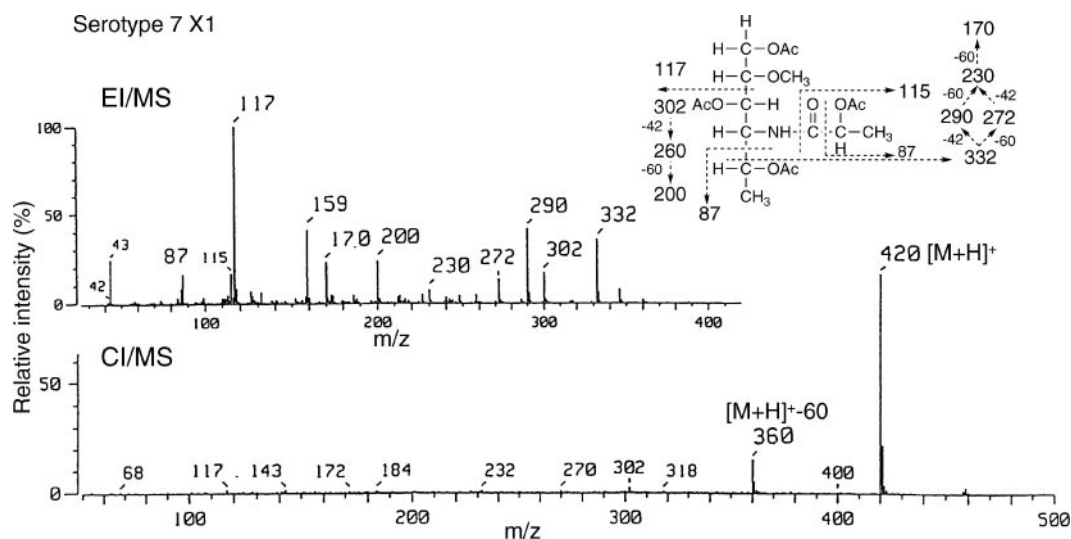

FIG. 3. EI/MS and CI/MS spectra of the alditol acetate derivative from X1. The patterns of prominent fragment ions are illustrated. The alditol acetate derivative was resolved on an SPB-1 fused silica column with a temperature program of 160°C for 2 min, followed by a rise of 4°C/min to 220°C and then maintained at 220°C for 13 min.
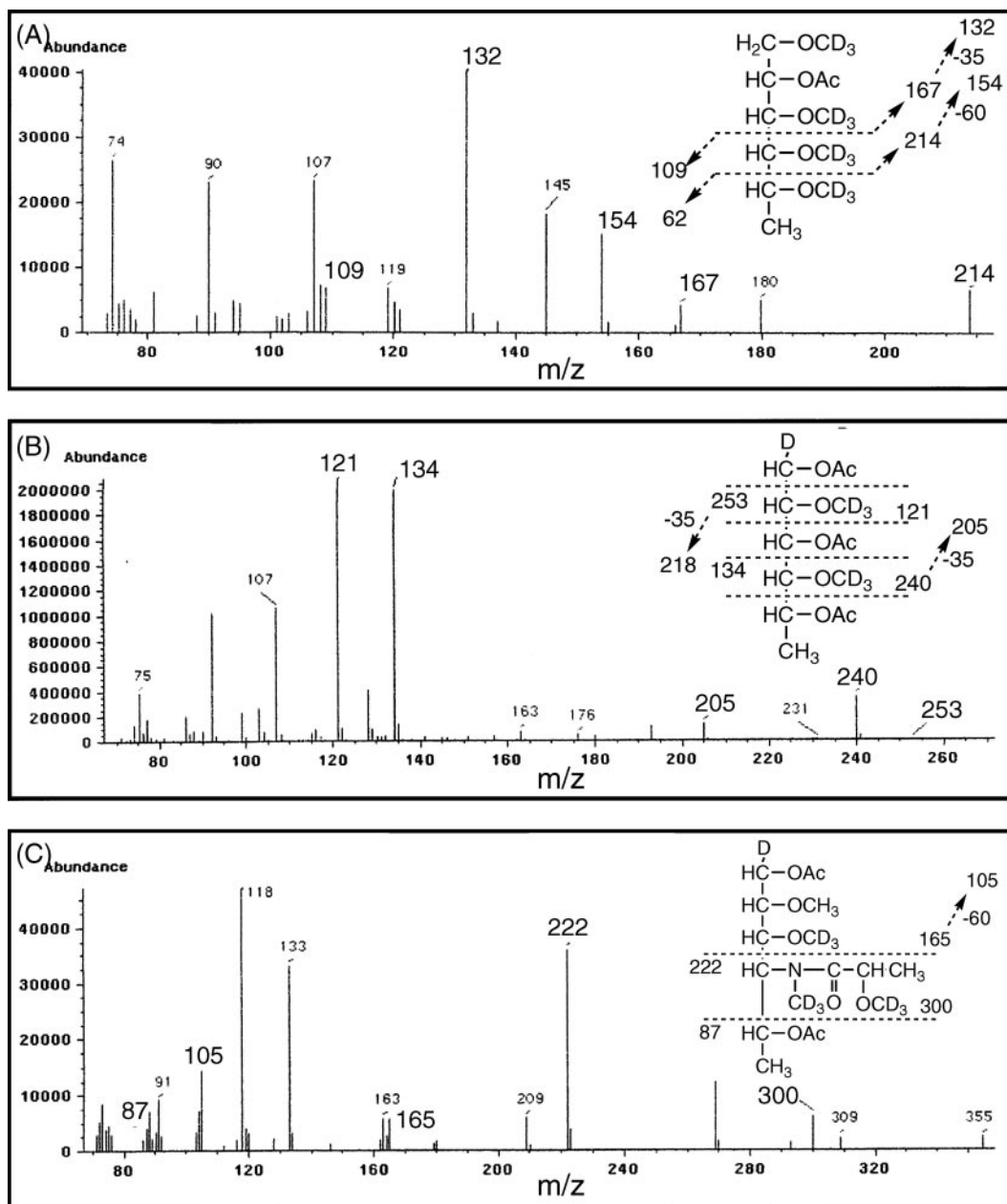
FIG. 4. EI/MS spectra of individual perdeuteromethylated alditol acetate derivative derived from serotype 7 OSE. The formation of prominent fragment ions is illustrated, and they were assigned to 1,3,4,5-tetra-*O*-deuteromethyl-2-*O*-acetyl-6-deoxytalitol (A), 2,4-di-*O*-deuteromethyl-1,3,5-tri-*O*-acetyl-rhamnitol (B), and 3-*O*-deuteromethyl-1,5-di-*O*-acetyl-4-2′-*O*-deuteromethyl-propanoyl-deuteromethyl-amido-4,6-dideoxy-2-*O*-methyl-hexitol (C).

38.4-kb insert of cosmid clone 49 was sequenced. The 19.7-kb region in this clone was deposited in the NCBI GenBank database (accession no. AB274811). The *rtfA* gene in cosmid clone 49 from *M. intracellulare* serotype 7 had 98.4% DNA identity with *M. intracellulare* strain 5509-Borstel (serotype 13) and around 84% DNA identity with *M. avium* strains. The understanding of gene function was based on the comparison of this sequence information to homologous regions in the genome sequence data from *M. avium* strain A5 (serotype 4, NCBI GenBank AY130970) (24). The gene order from *mtfB* (encoding methyltransferase) to *gtfB* (encoding glycosyltransferase) including *rtfA* was identical to that of the *M. avium* strain A5 cluster (Fig. 7A). The DNA region between ORF 1 and ORF 9 (11.0 kb) was unique to *M. intracellulare* serotype 7 strain ATCC 35847, and nine ORFs were observed. In this region, the insertion of IS elements and transposons was not detected. To confirm that the region of ORF 1 to ORF 9 is specific to *M. intracellulare* serotype 7, we showed the presence of the nine novel ORFs in two clinical isolates of *M. intracellulare* serotype 7 by PCR using the nine ORFs as primers. All ORFs in the ATCC 35847 strain were conserved in the clinical serotype 7 strains, although ORF 6 in the NF 112 strain was detected with a band about 1.3 kb longer (Fig. 7B). Sequence
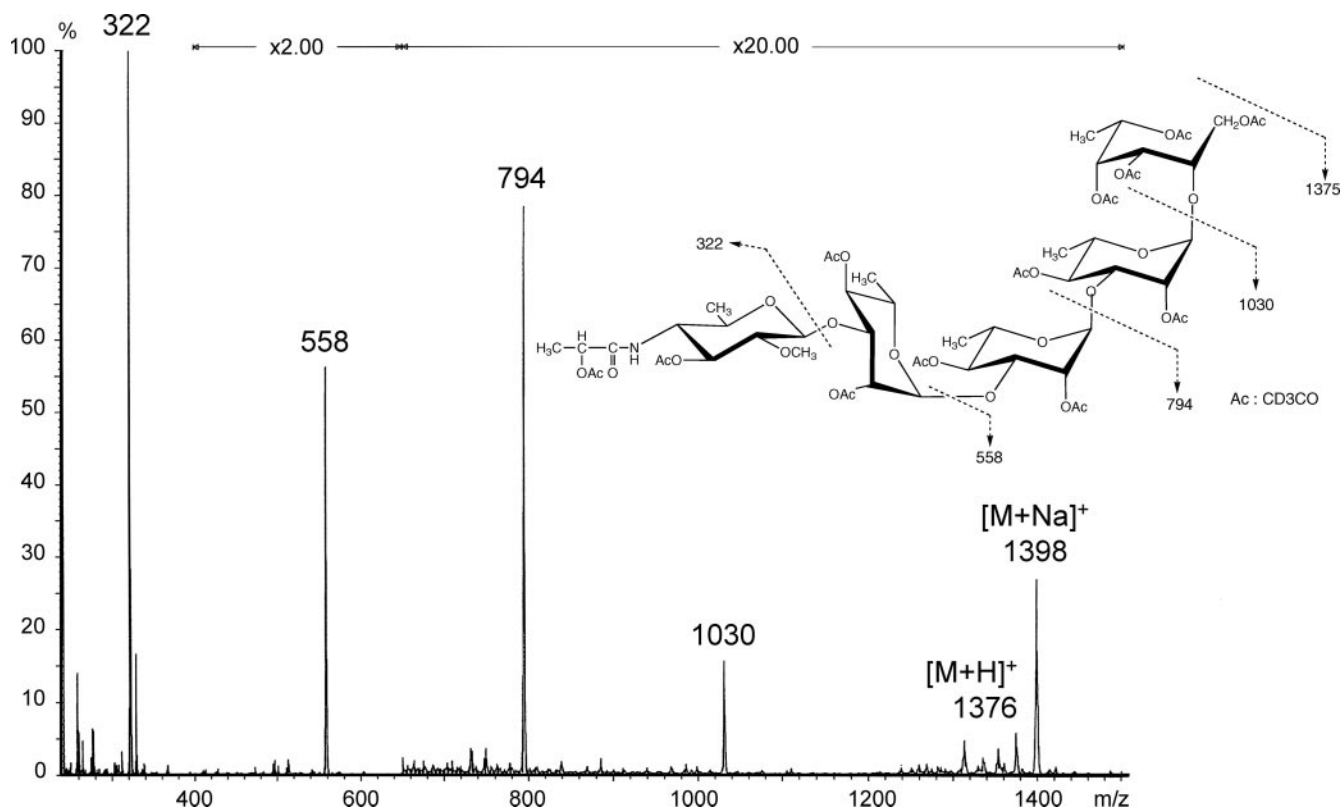
FIG. 5. FAB/MS spectrum of the deuteroacetylated derivative of serotype 7 OSE. The formation of the characteristic increment of fragmentation ions is illustrated. The matrix was *m*-nitro-benzyl alcohol.

analysis of the ORF 6 region in the NF 112 strain showed the insertion of a transposon. Almost none of the ORFs of serotype 7 strains were detected in the reference strains of serotypes 2, 4, 8, 12, 16, and 17, although the PCR bands of some ORFs were positive in other serotype strains. Based on the alignments of genetic maps and comparison to the database provided by NCBI GenBank, these ORFs were homologous to the regions involved in the glycosylation of OSE (Table 2). The deduced amino acid sequence of ORF 1 had much similarity to three putative glycosyltransferases, GtfA, GtfB, and RtfA, of *M. avium* strains. The similarity of the deduced amino acid sequences suggested the possibility that the function of ORF 2 and ORF 3 was to encode methyltransferase and glycosyltransferase Rv1516c (GtfTB). Moreover, from the similarity of the deduced amino acid sequences, ORFs 4, 8, and 9 seem to be aminotransferase, acyltransferase, and glycosyltransferase Rv1518, respectively. Although the homology score was not very high, the deduced amino acid sequences of ORFs 5, 6, and 7 were similar to those for carbamoyl phosphate synthase, nucleotide sugar epimerase, and metallophosphoesterase, respectively. These results suggest that this region of DNA is responsible for the biosynthesis of serotype 7-specific GPL.

DISCUSSION

Nontuberculous mycobacteria, including the pathogenic species belonging to MAC, have serotype-specific GPLs that are important components of the outer layer of the lipid-rich cell walls (6). Structural analyses of some serotype-specific GPLs derived from the predominant clinical isolates have been reported (28), but further structural analyses remain to be performed. The present study describes the chemical structure of the serotype 7 GPL derived from *M. intracellulare*.

We determined the glycosyl composition, linkage positions, and anomeric and ring configurations of glycosyl residues in serotype 7 GPL, which suggested that its OSE is 4-2′-hydroxypropanoyl-amido-4,6-dideoxy-2-*O*-methyl-β-Hex-(1→3)-α-L-Rha-(1→3)-α-L-Rha-(1→3)-α-L-Rha-(1→2)-L-6-d-Tal (Fig. 8). Chatterjee and Khoo have classified the structures of surface GPLs of MAC based on divergent biosynthesis (9). Serotype 7 GPL was assigned to polar GPL group 2 by its chemical structure. Serotype 12, 17, and 19 GPLs have been classified into group 2 GPL, which is commonly composed of *R*→α-L-Rha-(1→3)-α-L-Rha-(1→2)-L-6-d-Tal (*R*, variable region). The external sugar of serotype 7 GPL was a characteristic amido sugar that is composed of 4-2′-hydroxypropanoyl-amido-4,6-dideoxy-2-*O*-methyl-Hex. The presence of an amido sugar has been reported in only four GPLs, serotypes 12, 14, 17, and 25 (8, 9). Bozic et al. demonstrated that the OSE structure of serotype 12 GPL is 4-2′-hydroxypropanoyl-amido-4,6-dideoxy-3-*O*-methyl-β-D-Glu-(1→3)-4-*O*-methyl-α-L-Rha-(1→3)-α-L-Rha-(1→3)-α-L-Rha-(1→2)-α-6-d-Tal (4), which closely resembles the serotype 7 GPL. Comparing the detailed carbohydrate structure of serotype 7 GPL to that of serotype 12 GPL revealed that the acylated-amido group and linkage position-bound terminal sugar were the same, but the position of the *O*-methyl group
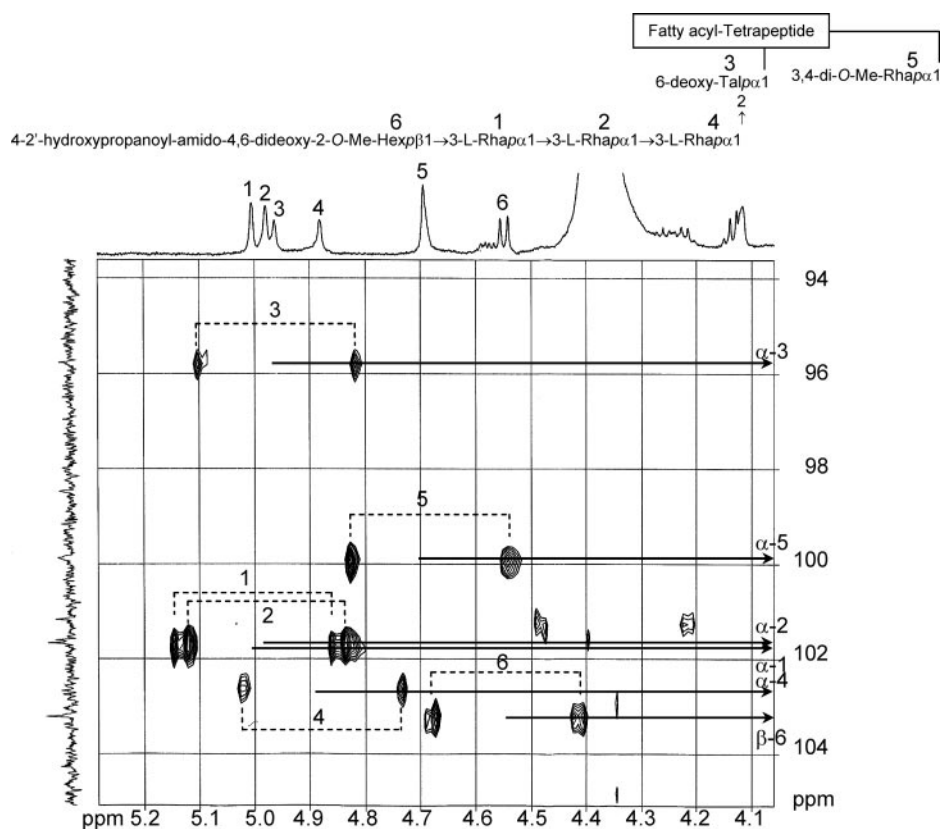
FIG. 6. Nondecoupled ^{1}H-detected [^{1}H-^{13}C] HMQC spectrum of serotype 7 GPL. Cross-peak labels correspond with those shown on the structure.

was different. Moreover, α-L-Rha was next to the terminal Hex in serotype 7 GPL, whereas this position was occupied by 4-*O*-Me-α-L-Rha in serotype 12 GPL. It was difficult to determine the species of acylated-amido sugar because no reference standard was available. In a previous study by Bozic et al., the terminal amido sugar of serotype 12 was assigned as glucose by the proton configuration in ^{1}H-NMR (4). It is possible that the terminal Hex of serotype 7 GPL is a gluco- or galacto-configuration.

Next, we attempted to elucidate the biosynthetic mechanism of GPL by molecular genetics, because very little is known of how the carbohydrate chain elongates in sereotype-specific GPLs. Belisle et al. first identified the *ser2* gene cluster responsible for biosynthesis of OSE in the serotype 2 GPL derived from *M. avium* serotype 2 strain TMC 727, which is mapped to a 22- to 27-kb functional region (3). Krzywinska and Schorey isolated and sequenced a 27.5-kb DNA fragment responsible for the carbohydrate portion of serotype 4 GPL from *M. avium* strain A5 (24). Recently, enzymatic characterization of the glycosyltransferase and methyltransferase in *M. smegmatis*, which can synthesize only nonpolar GPLs, has been reported (29, 30).

As in the serotype-specific polar GPL biosynthesis of MAC, only the *rtfA* gene present in the *ser2* gene cluster was functionally clarified to encode the rhamnosyltransferase responsible for the transfer of L-Rha to 6-d-Tal (11). The precise gene loci correlated to *O*-methylation and glycosylation are poorly understood. In this study, we cloned the serotype 7 GPL bio-

synthetic cluster and analyzed its sequence. The genetic map of the serotype 7 GPL biosynthetic cluster was compared to that of serotype 4 GPL from *M. avium* strain A5 (24). Although the *mtfB-gtfB* region was fully conserved, significant differences appeared in the neighborhood of this conserved region. Nine novel ORFs were detected only in the serotype 7 strains containing clinical isolates, which strongly suggested that this region is related to the biosynthesis of serotype 7-specific GPL. On the other hand, ORF 6 may not be necessary to serotype 7 GPL biosynthesis because of the insertion of a transposon in a clinical isolate, the NF 112 strain. Based on the analysis of sequence homology (Table 2 and Fig. 7A), the ORFs may be responsible for the glycosylation of serotype 7-specific GPL. From the structural analysis of the serotype 7 GPL and sequence of cosmid clone 49, it is possible to predict the relationship between the biosynthesis of serotype 7 GPL and the function of each ORF (Fig. 8).

rtfA functions to catalyze only the addition of Rha to 6-d-Tal (26), and which gene cluster transfers additional sugars to L-Rha elongated from 6-d-Tal is unclear. ORFs 1, 3, and 9 have high homology to the glycosyltransferases GtfA, Rv1516c, and Rv1518, respectively. We have analyzed similar gene clusters in *M. intracellulare* serotype 12, 16, and 17 strains in addition to the *M. intracellulare* serotype 7 strain. The sequence homology of the region of ORF 1 to ORF 9 was highly conserved between only *M. intracellulare* serotype 7 and 12 strains (unpublished data). ORFs 1, 3, and 9 may lead to transfer of the two
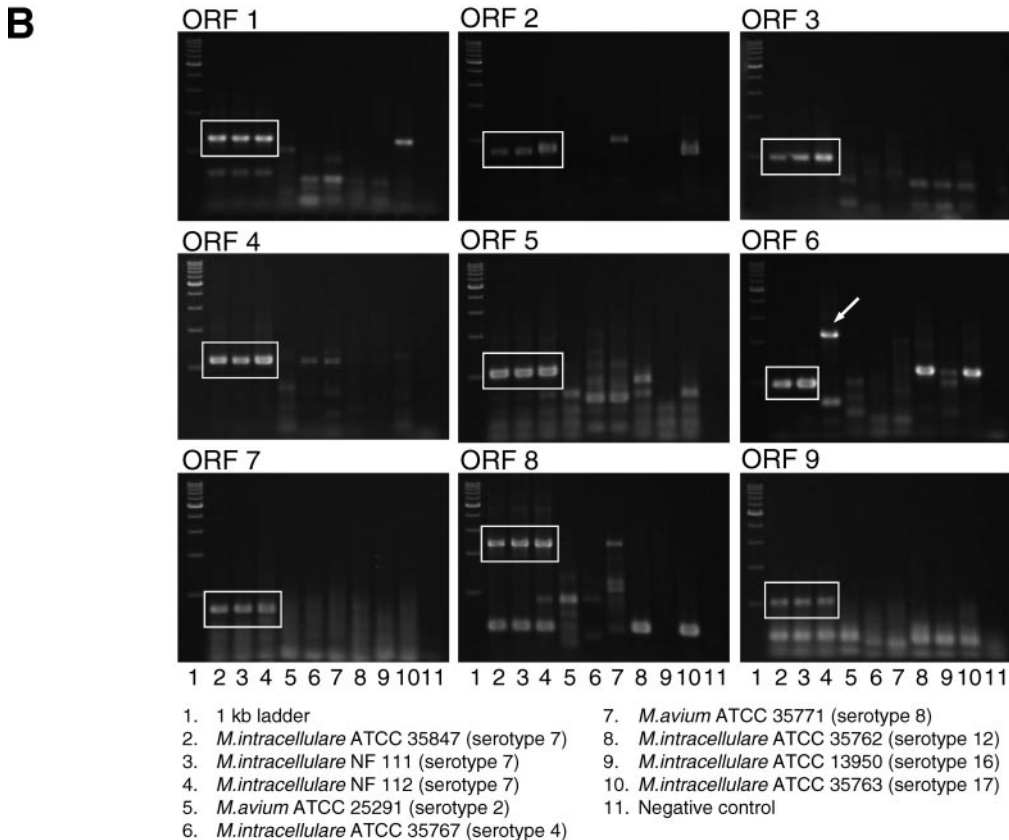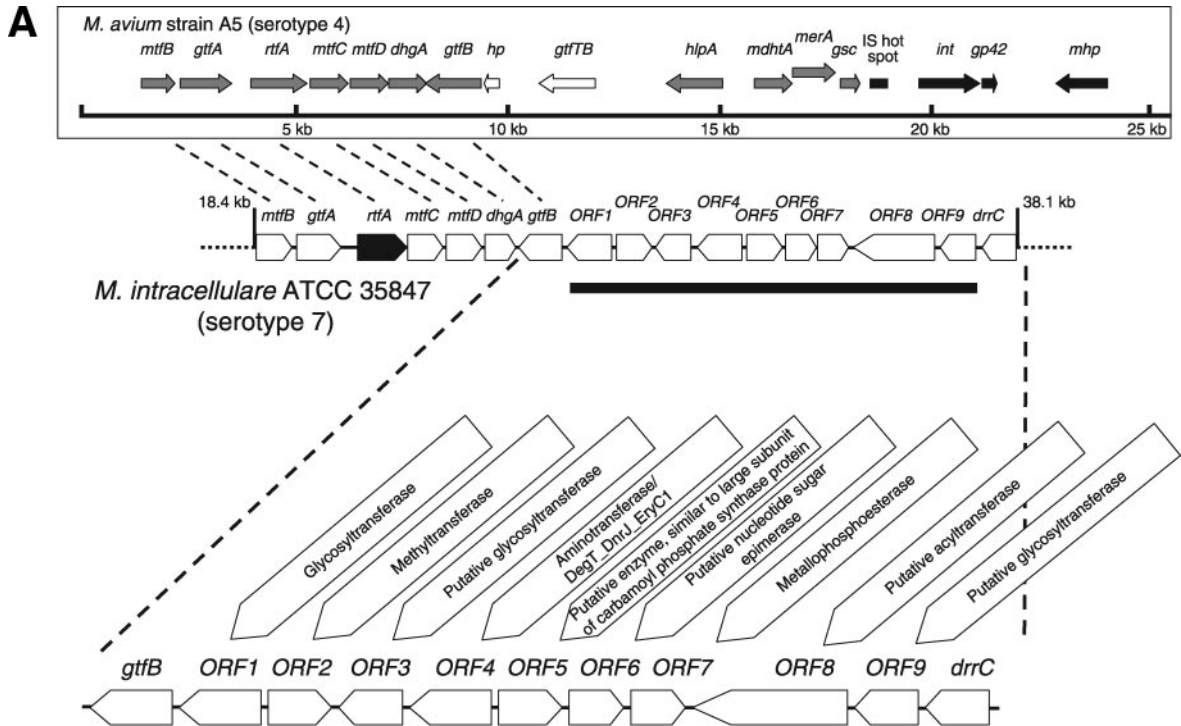
FIG. 7. Genetic map of the GPL biosynthetic cluster and detection of novel serotype 7-specific ORFs. Nine ORFs were observed, and they were homologous to the regions involved in the glycosylation of OSE from the alignment of genetic maps and comparison to the database provided by NCBI GenBank (A). The ORFs in *M. intracellulare* serotype 7 strain ATCC 35847 were serotype 7 specific and preserved in clinical strains (B). The arrow indicates the band of ORF 6 in strain NF 112 that was inserted a transposon. For the genetic map of *M. avium* strain A5 (serotype 4), see the study of Krzywinska and Schorey (24).

TABLE 2. Similarity to protein sequences of ORFs in cosmid clone 49 derived from *M. intracellulare* serotype 7 strain ATCC 35847

ORF	Predicted molecular mass (kDa)	Predicted pI	Similar protein (GenBank accession no.)	E value	Amino acid identity (% matched/total no. of residues)[a]
MtfB	30.5	5.09	Methyltransferase MtfB (Q9WVW3)	E-143	90.4/272
GtfA	46.5	5.84	Glycosyltransferase GtfA (O68999)	0.0	90.7/421
RtfA	46.9	7.76	Rhamnosyltransferase A (Q6U848)	0.0	98.6/419
MtfC	30.0	5.61	Methyltransferase MtfC (Q8GEA1)	E-142	92.1/265
MtfD	32.9	4.89	Methyltransferase MtfD (Q9WW56)	E-147	91.5/272
DhgA	27.3	5.96	Dehydrogenase DhgA (AY130970)	E-120	85.8/260
GtfB	45.5	6.35	Glycosyltransferase GtfB (Q9WW66)	0.0	85.4/419
ORF 1	45.1	6.10	Glycosyltransferase GtfA (O68999)	E-137	58.6/420
ORF 2	30.9	8.14	Methyltransferase (Q885B6)	5E-47	44.3/237
ORF 3	36.6	6.01	Putative glycosyltransferase Rv1516c (P71795)	E-118	66.4/307
ORF 4	40.4	4.92	Aminotransferase/DegT_DnrJ_EryC1 (Q50723)	E-154	75.5/351
ORF 5	36.1	6.43	Putative enzyme, similar to large subunit of carbamoyl phosphate synthase protein (Q92VH3)	1E-39	36.1/321
ORF 6	33.8	5.48	Putative nucleotide sugar epimerase (Q92VH2)	2E-32	30.0/297
ORF 7	27.6	5.91	Metallophosphoesterase (Q9HMV3)	5E-16	27.5/218
ORF 8	78.5	8.91	Putative acyltransferase (Q73SR3)	0.0	49.1/721
ORF 9	37.6	8.77	Putative glycosyltransferase Rv1518 (Q50590)	E-101	61.8/304
DrrC	28.5	11.69	Daunorubicin resistance protein C (Q9XCF9)	E-123	83.7/263

[a] Identity values were calculated by using BLASTX searches.

additional molecules of L-Rha and terminal amido-Hex. ORF 2 was assigned to methyltransferase and may be correlated with the synthesis of the *O*-methyl group at the C-2 position in the terminal amido-Hex. ORFs 4, 5, 7, and 8 were homologous to aminotransferase, carbamoyl phosphate synthase protein, metallophosphoesterase, and acyltransferase, respectively, and possibly relate to the biosynthesis of 2′-hydroxypropanoyl-amido in the terminal Hex. Taken together, this gene cluster may participate in the biosynthetic pathway of serotype 7 GPL, but further study will be required to define the function of each ORF that we have shown for the first time in this study.

GPL is one of the immunologically active molecules characteristic of MAC. Tassel et al. have reported that the core GPL seems to play a role in suppression of a mitogen-induced blastogenic response in spleen cells (35), and our previous study has shown that sera of patients with MAC disease contain immunoglobulin G (IgG), IgA, and IgM antibodies against

the core of the GPL molecule (23). In addition, the immuno-modulating activity of GPL on macrophage functions is serotype dependent (18, 34). The serotype 4 GPL promotes phagocytosis and inhibits phagosome-lysosome (P-L) fusion, whereas the GPLs of serotypes 9 and 16 exhibit no effect on phagocytosis and P-L fusion. The serotype 8 GPL shows concomitant stimulation of both phagocytosis and P-L fusion. The OSE of GPL may be involved in the mechanism of inhibition of P-L fusion, which is mediated through mannose receptors of macrophages (32). The serotype 4 GPL inhibited lymphoproliferative response to mitogens (18). Thus, host responses to GPLs vary with the MAC serotype.

The pathogenicity of GPL may comprise both a common peptide core and an OSE elongated from 6-d-Tal. GPL is a pleiotropic molecule and participates in the pathogenesis of MAC disease. Elucidation of the structure-activity relationship of GPL is required to better understand the pathogenesis.
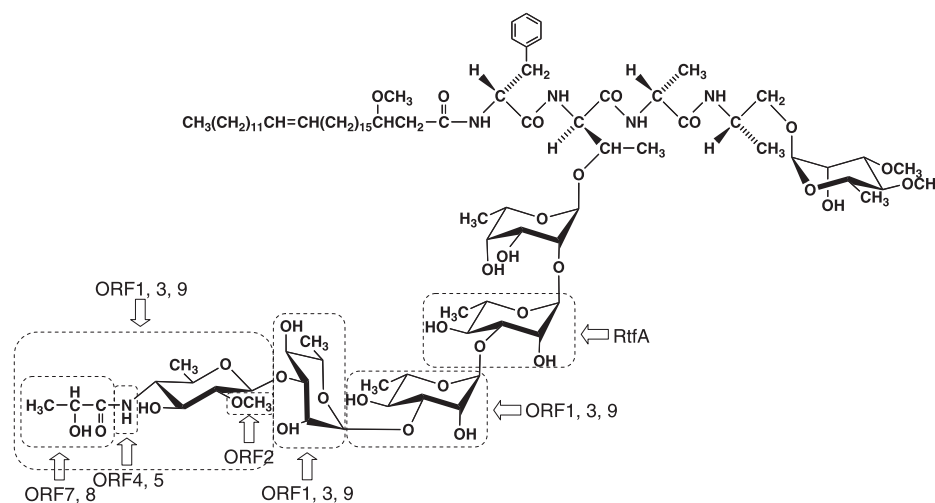

FIG. 8. Proposed structure of serotype 7 GPL and its predicted relationship to the genetic cluster.

ACKNOWLEDGMENTS

REFERENCES

1. **Aspinall, G. O., D. Chatterjee, and P. J. Brennan.** 1995. The variable surface glycolipids of mycobacteria: structures, synthesis of epitopes, and biological properties. Adv. Carbohydr. Chem. Biochem. **51:**169–242.
2. **Barrow, W. W., T. L. Davis, E. L. Wright, V. Labrousse, M. Bachelet, and N. Rastogi.** 1995. Immunomodulatory spectrum of lipids associated with *Mycobacterium avium* serovar 8. Infect. Immun. **63:**126–133.
3. **Belisle, J. T., L. Pascopella, J. M. Inamine, P. J. Brennan, and W. R. Jacobs, Jr.** 1991. Isolation and expression of a gene cluster responsible for biosynthesis of the glycopeptidolipid antigens of *Mycobacterium avium.* J. Bacteriol. **173:**6991–6997.
4. **Bozic, C. M., M. McNeil, D. Chatterjee, I. Jardine, and P. J. Brennan.** 1988. Further novel amido sugars within the glycopeptidolipid antigens of *Mycobacterium avium.* J. Biol. Chem. **263:**14984–14991.
5. **Brennan, P. J., and M. B. Goren.** 1979. Structural studies on the type-specific antigens and lipids of the *Mycobacterium avium·Mycobacterium intracellulare·Mycobacterium scrofulaceum* serocomplex. *Mycobacterium intracellulare* serotype 9. J. Biol. Chem. **254:**4205–4211.
6. **Brennan, P. J., and H. Nikaido.** 1995. The envelope of mycobacteria. Annu. Rev. Biochem. **64:**29–63.
7. **Camphausen, R. T., R. L. Jones, and P. J. Brennan.** 1986. Structure and relevance of the oligosaccharide hapten of *Mycobacterium avium* serotype 2. J. Bacteriol. **168:**660–667.
8. **Chatterjee, D., G. O. Aspinall, and P. J. Brennan.** 1987. The presence of novel glucuronic acid-containing, type-specific glycolipid antigens within *Mycobacterium* spp. Revision of earlier structures. J. Biol. Chem. **262:**3528–3533.
9. **Chatterjee, D., and K. H. Khoo.** 2001. The surface glycopeptidolipids of mycobacteria: structures and biological properties. Cell Mol. Life Sci. **58:**2018–2042.
10. **Eckstein, T. M., J. T. Belisle, and J. M. Inamine.** 2003. Proposed pathway for the biosynthesis of serovar-specific glycopeptidolipids in *Mycobacterium avium* serovar 2. Microbiology **149:**2797–2807.
11. **Eckstein, T. M., F. S. Silbaq, D. Chatterjee, N. J. Kelly, P. J. Brennan, and J. T. Belisle.** 1998. Identification and recombinant expression of a *Mycobacterium avium* rhamnosyltransferase gene (*rtfA*) involved in glycopeptidolipid biosynthesis. J. Bacteriol. **180:**5567–5573.
12. **Enomoto, K., S. Oka, N. Fujiwara, T. Okamoto, Y. Okuda, R. Maekura, T. Kuroki, and I. Yano.** 1998. Rapid serodiagnosis of *Mycobacterium avium-intracellulare* complex infection by ELISA with cord factor (trehalose 6, 6′-dimycolate), and serotyping using the glycopeptidolipid antigen. Microbiol. Immunol. **42:**689–696.
13. **Falkinham, J. O., III.** 1996. Epidemiology of infection by nontuberculous mycobacteria. Clin. Microbiol. Rev. **9:**177–215.
14. **Fujita, Y., T. Doi, R. Maekura, M. Ito, and I. Yano.** 2006. Differences in serological responses to specific glycopeptidolipid-core and common lipid antigens in patients with pulmonary disease due to *Mycobacterium tuberculosis* and *Mycobacterium avium* complex. J. Med. Microbiol. **55:**189–199.
15. **Gerwig, G. J., J. P. Kamerling, and J. F. G. Vliegenthart.** 1978. Determination of the D and L configuration of neutral monosaccharides by high-resolution capillary G.L.C. Carbohydr. Res. **62:**349–357.
16. **Hakomori, S.** 1964. A rapid permethylation of glycolipid, and polysaccharide catalyzed by methylsulfinyl carbanion in dimethyl sulfoxide. J. Biochem. (Tokyo). **55:**205–208.
17. **Heidelberg, T., and O. R. Martin.** 2004. Synthesis of the glycopeptidolipid of *Mycobacterium avium* serovar 4: first example of a fully synthetic C-mycoside GPL. J. Org. Chem. **69:**2290–2301.
18. **Kano, H., T. Doi, Y. Fujita, H. Takimoto, I. Yano, and Y. Kumazawa.** 2005. Serotype-specific modulation of human monocyte functions by glycopeptidolipid (GPL) isolated from *Mycobacterium avium* complex. Biol. Pharm. Bull. **28:**335–339.
19. **Kaufmann, S. H.** 2001. How can immunology contribute to the control of tuberculosis? Nat. Rev. Immunol. **1:**20–30.
20. **Khoo, K. H., D. Chatterjee, A. Dell, H. R. Morris, P. J. Brennan, and P. Draper.** 1996. Novel *O*-methylated terminal glucuronic acid characterizes the polar glycopeptidolipids of *Mycobacterium habana* strain TMC 5135. J. Biol. Chem. **271:**12333–12342.
21. **Khoo, K. H., E. Jarboe, A. Barker, J. Torrelles, C. W. Kuo, and D. Chatterjee.** 1999. Altered expression profile of the surface glycopeptidolipids in drug-resistant clinical isolates of *Mycobacterium avium* complex. J. Biol. Chem. **274:**9778–9785.
22. **Kitada, S., R. Maekura, N. Toyoshima, N. Fujiwara, I. Yano, T. Ogura, M. Ito, and K. Kobayashi.** 2002. Serodiagnosis of pulmonary disease due to *Mycobacterium avium* complex with an enzyme immunoassay that uses a mixture of glycopeptidolipid antigens. Clin. Infect. Dis. **35:**1328–1335.
23. **Kitada, S., R. Maekura, N. Toyoshima, T. Naka, N. Fujiwara, M. Kobayashi, I. Yano, M. Ito, and K. Kobayashi.** 2005. Use of glycopeptidolipid core antigen for serodiagnosis of *Mycobacterium avium* complex pulmonary disease in immunocompetent patients. Clin. Diagn. Lab. Immunol. **12:**44–51.
24. **Krzywinska, E., and J. S. Schorey.** 2003. Characterization of genetic differences between *Mycobacterium avium* subsp. *avium* strains of diverse virulence with a focus on the glycopeptidolipid biosynthesis cluster. Vet. Microbiol. **91:**249–264.
25. **Maekura, R., Y. Okuda, A. Hirotani, S. Kitada, T. Hiraga, K. Yoshimura, I. Yano, K. Kobayashi, and M. Ito.** 2005. Clinical and prognostic importance of serotyping *Mycobacterium avium-Mycobacterium intracellulare* complex isolates in human immunodeficiency virus-negative patients. J. Clin. Microbiol. **43:**3150–3158.
26. **Maslow, J. N., V. R. Irani, S. H. Lee, T. M. Eckstein, J. M. Inamine, and J. T. Belisle.** 2003. Biosynthetic specificity of the rhamnosyltransferase gene of *Mycobacterium avium* serovar 2 as determined by allelic exchange mutagenesis. Microbiology **149:**3193–3202.
27. **McNeil, M., H. Gaylord, and P. J. Brennan.** 1988. *N*-Formylkansosaminyl-(1–3)-2-*O*-methyl-D-rhamnopyranose: the type-specific determinant of serovariant 14 of the *Mycobacterium avium* complex. Carbohydr. Res. **177:**185–198.
28. **McNeil, M., A. Y. Tsang, and P. J. Brennan.** 1987. Structure and antigenicity of the specific oligosaccharide hapten from the glycopeptidolipid antigen of *Mycobacterium avium* serotype 4, the dominant *Mycobacterium* isolated from patients with acquired immune deficiency syndrome. J. Biol. Chem. **262:**2630–2635.
29. **Miyamoto, Y., T. Mukai, N. Nakata, Y. Maeda, M. Kai, T. Naka, I. Yano, and M. Makino.** 2006. Identification and characterization of the genes involved in glycosylation pathways of mycobacterial glycopeptidolipid biosynthesis. J. Bacteriol. **188:**86–95.
30. **Patterson, J. H., M. J. McConville, R. E. Haites, R. L. Coppel, and H. Billman-Jacobe.** 2000. Identification of a methyltransferase from *Mycobacterium smegmatis* involved in glycopeptidolipid synthesis. J. Biol. Chem. **275:**24900–24906.
31. **Porcelli, S. A., and R. L. Modlin.** 1999. The CD1 system: antigen-presenting molecules for T cell recognition of lipids and glycolipids. Annu. Rev. Immunol. **17:**297–329.
32. **Shimada, K., H. Takimoto, I. Yano, and Y. Kumazawa.** 2006. Involvement of mannose receptor in glycopeptidolipid-mediated inhibition of phagosome-lysosome fusion. Microbiol. Immunol. **50:**243–251.
33. **Smith, I.** 2003. *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. Clin. Microbiol. Rev. **16:**463–496.
34. **Takegaki, Y.** 2000. Effect of serotype specific glycopeptidolipid (GPL) isolated from *Mycobacterium avium* complex (MAC) on phagocytosis and phagosome-lysosome fusion of human peripheral blood monocytes. Kekkaku **75:**9–18. [In Japanese.]
35. **Tassell, S. K., M. Pourshafie, E. L. Wright, M. G. Richmond, and W. W. Barrow.** 1992. Modified lymphocyte response to mitogens induced by the lipopeptide fragment derived from *Mycobacterium avium* serovar-specific glycopeptidolipids. Infect. Immun. **60:**706–711.