Lipo-oligosaccharidic Antigen from *Mycobacterium gastri*. Complete Structure of a Novel C4-Branched 3,6-Dideoxy-α-*xylo*-hexopyranose†

Martine Gilleron,*,‡ Joseph Vercauteren,§ and Germain Puzo‡

Laboratoire de Pharmacologie et de Toxicologie Fondamentales du Centre National de la Recherche Scientifique, Département III, 118 route de Narbonne, 31062 Toulouse Cedex, France, and Laboratoire de Pharmacognosie, Faculté de Pharmacie, 3 Place de la Victoire, 33076 Bordeaux Cedex, France

Received February 23, 1993; Revised Manuscript Received July 13, 1993®

ABSTRACT: The following incomplete structure, α-X-(1→3)-[β-L-Xyl*p*-(1→4)]$_6$-3-*O*-Me-α-Rha*p*-(1→3)-β-D-Gal*p*-(1→3)-β-D-Glc*p*-(1→4)-2-*O*-acyl-α-D-Glc*p*-(1↔1)-4,6-di-*O*-acyl-α-D-Glc*p*, was previously established for the antigenic lipo-oligosaccharide typifying *Mycobacterium gastri*, namely, LOS-III. The partial structure of the distal monosaccharide (X) was assigned as 3,6-dideoxy-4-*C*-(1,3-di-*O*-methylpropyl)-α-hexopyranose, which corresponds to a new-found monosaccharide in nature [Gilleron M., Vercauteren J., & Puzo G. (1993) *J. Biol. Chem. 268*, 3168–3179]. This article reports the complete structure of X, which was determined from the FAB-MS and 2D NMR analysis of the peracetylated LOS-III. The comparative analyses of the native and per-*O*-acetylated LOS-III FAB-MS spectra revealed, for the monosaccharide X, a molecular mass of 370 Da and five hydroxyl groups that could be acetylated. Additionally, the 1D ^{1}H NMR spectrum of the per-*O*-acetylated LOS-III showed a dramatically increased dispersion of the protons, which resonated between 3 and 4 ppm in the spectrum of the underivatized LOS-III. Thus, thanks to 2D NMR sequences (COSY, HOHAHA, HMQC, HMQC–HOHAHA, and HMBC), the complete assignment of the ^{1}H and ^{13}C signals was achieved. Starting from the quaternary C4 resonance, the spin system of the C-alkyl chain was assigned, allowing us to propose the following structure, 3,6-dideoxy-4-*C*-(1,3-dimethoxy-4,5,6,7-tetrahydroxyheptyl)-α-*xylo*-hexopyranose. The *xylo* configuration was established from the ROESY spectrum.

The importance of cell wall glycoconjugates in the context of the immunopathology of tuberculosis has been well established since some cell wall mycobacterial glycolipids, such as phenolic glycolipids (PheGLs)[1] and lipo-arabinomannans (LAMs), have been shown to be involved in intramacrophagic survival of the mycobacteria. Moreover, phagocytized by macrophages, pathogenic mycobacteria multiply intracellularly, forming an electron-transparent zone (ETZ) (Draper & Rees, 1970) shielding the bacteria against the microbicidal functions of the host macrophages. This ETZ has been shown, in the case of *Mycobacterium lepraemurium*, to be constituted in part by glycolipids and more particularly by mycosides (Draper, 1974). Thus, the *in vitro* and *in vivo* immunological properties of a majority of cell wall mycobacterial glycolipidic components such as cord factor (CF), glycopeptidolipids (GPLs), LAMs, and PheGLs have been largely investigated.

However, little is known concerning the lipo-oligosaccharides (LOSs), a class of immunoreactive species-specific glycolipids. Nevertheless, an interesting correlation between mycobacterial virulence and cell wall LOS content has been proposed (Belisle & Brennan, 1989). Indeed, rough mutants of *Mycobacterium kansasii*, which cause a chronic systemic infection when injected into B6D2 mice, are devoid of LOSs, whereas smooth strains, which are cleared rapidly, possess LOSs in their cell walls, suggesting that LOSs are markers of avirulence. Moreover, *M. kansasii* is proving to be very closely related to another mycobacterial strain: *Mycobacterium gastri*. The *M. gastri* and *M. kansasii* cell walls were shown to contain the same phenolic glycolipid antigens, despite the unique structure of the immunodominant monosaccharide (Gilleron et al., 1990). Besides the phenolic glycolipids, LOSs have also been described in *M. kansasii* cell walls (Hunter et al., 1983, 1984, 1985). LOSs are one class of species-specific glycolipids only described in *M. kansasii*, *Mycobacterium smegmatis* (Saadat & Ballou, 1983; Kamisango et al., 1985), *Mycobacterium malmoense* (McNeil et al., 1987), *Mycobacterium szulgai* (Hunter et al., 1988), *Mycobacterium linda* (Camphausen et al., 1987), and *Mycobacterium tuberculosis* strain Canetti (Daffe et al., 1991).

In a previous work, LOSs were identified in *M. gastri* strain W471, allowing the differentiation of the two species, *M. kansasii* and *M. gastri*. Thanks to a variety of 2D NMR homonuclear (COSY, HOHAHA) and heteronuclear (HMQC, HMQC–HOHAHA, HMBC) methods applied to the native molecule, the following partial structure was established for the major representative lipo-oligosaccharide of *M. gastri*

† This work was supported by grants from the ARC (No. 6718), the Région Midi Pyrénées (RECH/9000827), and the Région Aquitaine (RECH/9203056).

* Author to whom correspondence should be addressed.

‡ Laboratoire de Pharmacologie et de Toxicologie Fondamentales du Centre National de la Recherche Scientifique.

§ Laboratoire de Pharmacognosie, Faculté de Pharmacie.

® Abstract published in *Advance ACS Abstracts,* January 1, 1994.

[1] Abbreviations: COSY, ^{1}H–^{1}H correlated spectroscopy; COSY-LR, ^{1}H–^{1}H correlated spectroscopy, long range; DQF-COSY, double quantum filtered ^{1}H–^{1}H correlated spectroscopy; EI, electronic impact; GC/MS, gas chromatography/mass spectrometry; HMBC, heteronuclear multiple bond connectivity spectroscopy; HMQC, heteronuclear multiple quantum correlation spectroscopy; HOHAHA, homonuclear Hartmann–Hahn spectroscopy; $^3J_{H,H'}$, vicinal spin–spin coupling constants; LAM, lipo-arabinomannan; LOS, lipo-oligosaccharide; LOS-III, major (quantitatively) LOS isolated from *M. gastri* W471 cell wall; FAB-MS, fast atom bombardment mass spectrometry; ^{1}H or ^{13}C NMR proton or carbon nuclear magnetic resonance; PBS, phosphate-buffered saline; PheGl K-I, major phenolic glycolipids from *M. kansasii*; RCT-1, -2, and -3, relayed coherence transfer, 1, 2, and 3 steps; ROESY, rotating frame Overhauser effect spectroscopy; Gal*p*, galactopyranose; Glc*p*, glucopyranose; Hex*p*, hexopyranose; Rha*p*, rhamnopyranose; Xyl*p*, xylopyranose.

0006-2960/94/0433-1930$04.50/0

W471, namely, LOS-III: α-X*p*-(1→3)-[β-L-Xyl*p*-(1→4)]$_6$-3-*O*-Me-α-Rha*p*-(1→3)-β-D-Gal*p*-(1→3)-β-D-Glc*p*-(1→4)-2-*O*-acyl-α-D-Glc*p*-(1↔1)-4,6-di-*O*-acyl-α-D-Glc*p*. This structure presents at one end a unique monosaccharide (X), unrevealed by routine carbohydrate analysis. Its partial structure, 3,6-dideoxy-4-*C*-(1,3-di-*O*-methylpropyl)-α-hexopyranose, was previously established by 2D NMR analysis of native LOS-III.

It is well established that the species specificity of the immunoreactive glycolipids belonging to the three types, PheGLs, GPLs, and LOSs, is conferred by the particular structure of the monosaccharide localized at the nonreducing end (Vercellone et al., 1992). Thus, the complete structural elucidation of the immunodominant monosaccharide is an essential step prior to the synthesis of glycoconjugates used to identify *M. gastri*.

In the present study, we report the complete structure of the complex monosaccharide X established directly from peracetylated LOS-III by means of an analytical approach combining 2D NMR scalar correlation sequences and FAB mass spectrometry.

EXPERIMENTAL PROCEDURES

Culture Conditions. *M. gastri* strains W471, HB4362, and HB4389 and *M. kansasii* strains ATCC12478, S890175, and S890370 were grown as surface pelicules on Sauton's medium at 37 °C for 2 months.

LOS Purification. The purification and characterization of glycolipids from *M. gastri* W471 have been described in a previous work (Gilleron et al., 1993).

LOS Acetylation. LOS-III (10 mg) was dissolved in 1 mL of acetic anhydride/anhydrous pyridine (1:1) in a sealed vial overnight at 60 °C. After cooling, the mixture was evaporated under N_2. The residue was then applied to a silicic Sep-pak cartridge (Waters Associates, Inc., Milfort, MA), which was eluted successively with hexane/ethyl acetate (1:1) to remove impurities and with pure ethyl acetate to elute the peracetylated LOS-III.

Analytical Methods. TLC was run on commercial silica gel plates (DC Alurolle, Kieselgel 60 pF_{254} Merck, Darmstadt, Germany) in ethyl acetate as the migration solvent. A spray composed of 0.1% orcinol in 40% H_2SO_4 was used to locate the peracetylated LOSs.

Nuclear Magnetic Resonance Spectroscopy. Spectra were recorded on a 500-MHz Brucker AMX-500 spectrometer equipped with an Aspect X32 computer. ^1H and ^{13}C chemical shifts are given in ppm downfield from internal tetramethylsilane (0 ppm) and internal chloroform (77 ppm), respectively. The sample was dried under vacuum, dissolved in $CDCl_3$ (Spin et Techniques, Paris, France) at a concentration of 20 mg/mL, and analyzed in a 200 × 5 mm 535-PP NMR tube. All 2D NMR data sets were recorded at 30 °C without sample spinning for HOHAHA and heteronuclear experiments. The data were acquired in the phase-sensitive mode using either the TPPI method (Marion & Wüthrich, 1983) or the method of States et al. (1982). DQF-COSY (Rance et al., 1983) and HOHAHA (Bax & Davis, 1985) were recorded with standard pulse sequences. ^1H–^{13}C correlation spectra were recorded in the proton-detected mode with a Brucker 5-mm ^1H broadband tunable probe with reversed geometry. The pulse sequence used for single-bond correlation spectra (HMQC) was that of Bax and co-workers (Bax et al., 1983). The GARP sequence (Shaka et al., 1985) at the carbon frequency was used as a composite decoupling pulse during acquisition. Multiple-bond correlation spectra (HMBC) (Bax & Summers, 1986) were recorded and processed in the magnitude mode. The pulse sequence used for ^1H-detected heteronuclear relayed spectra (HMQC–HOHAHA) was that of Lerner and Bax (1986). Experimental details and processing parameters of the spectra are given in the figure captions.

The one-dimensional ^1H spectrum was measured using about a 30° tipping angle for the pulse and 1.5 s as a recycle delay between each of the 16 acquisitions of 4.3 s. The spectral width of 3788 Hz was digitized on 32K complex points that were multiplied by a Lorentz–Gauss function (LB −1.3 Hz, GB 0.3 Hz) prior to processing to 32K real points in the frequency domain, giving a digital resolution of 0.06 Hz/point.

The ^1H-decoupled ^{13}C *J*-modulated spectrum (Lecocq & Lallemand, 1981) was recorded with a recycle delay of 2.5 s, and 38 198 scans were accumulated. The spectral width used was 27 700 Hz, and the data were digitized on 32K real points, giving a digital resolution of 0.42 Hz/point. The data were multiplied by a line-broadening function (LB 2 Hz) prior to Fourier transformation.

2D HOHAHA spectroscopy of peracetylated LOS-III (10 mg) in $CDCl_3$ was realized with a spectral width of 3105 Hz, a relaxation delay of 2.0 s, and a spin lock time of 63 ms; the data matrix was 4K × 512 (States Haberkorn and Ruben) points, with 16 scans per t_1 value. For processing, a sine-bell window shifted by $\pi/3$ was applied in both dimensions, followed by expansion of the data matrix to 4K × 1K data points and transformation. The resulting digital resolution in F_2 was 0.38 Hz/point.

FAB Mass Spectrometry. FAB/MS spectra were recorded on a VG Analytical 70-250 SEQ spectrometer using a 35-kV cesium beam in the 3500–700 mass/charge range. Samples of LOS (2 μg) dissolved in $HCCl_3$ were applied to a matrix of triethylene glycol monobutyl ether/glycerol (1:1).

RESULTS

FAB-MS Analysis of the Native and Peracetylated LOS-III. Native LOS-III was analyzed by FAB-MS using a mixture of triethylene glycol monobutyl ether and glycerol as a matrix. The low mass resolution FAB-MS spectrum (R = 500) (Figure 1) is dominated, at high mass range, by an intense peak centered at m/z 2681. By adding cesium iodide, this peak was found to be shifted by 110 mass units and was assigned to the LOS molecular ion attached to sodium, revealing an average molecular mass of 2658 Da. Thus, to increase the measurement accuracy, the mass resolution was set to 2500, allowing the separation of the different isotopic peaks (Figure 1c). Thanks to the theoretical isotopic distribution (Figure 1b), the low mass monoisotopic cationized molecular ion was assigned to the peak m/z 2680.09. The nominal molecular mass of 2656 Da reveals the elementary composition $C_{121}H_{212}O_{62}$ for LOS-III. A similar analysis was applied to per-*O*-acetylated LOS-III (Figure 2). The peak at m/z 3856.7 (Figure 2b) was assigned to the cationized molecular ion, $(M + Na)^+$. Thus, the molecular mass difference of 1176 Da between the native and the peracetylated derivative revealed 28 acetyl residues. From the proposed LOS-III structure, it emerged that the C4-branched chain previously assigned to the 1,3-di-*O*-methylpropyl chain contained four supplementary hydroxyl groups that could be acetylated, supporting the fact that the proposed structure of the terminal monosaccharide was incomplete.

Moreover, besides the molecular ions, FAB-MS spectra of the native and the peracetylated derivative showed fragment ions, allowing oligosaccharidic sequencing (Figures 1 and 2a).

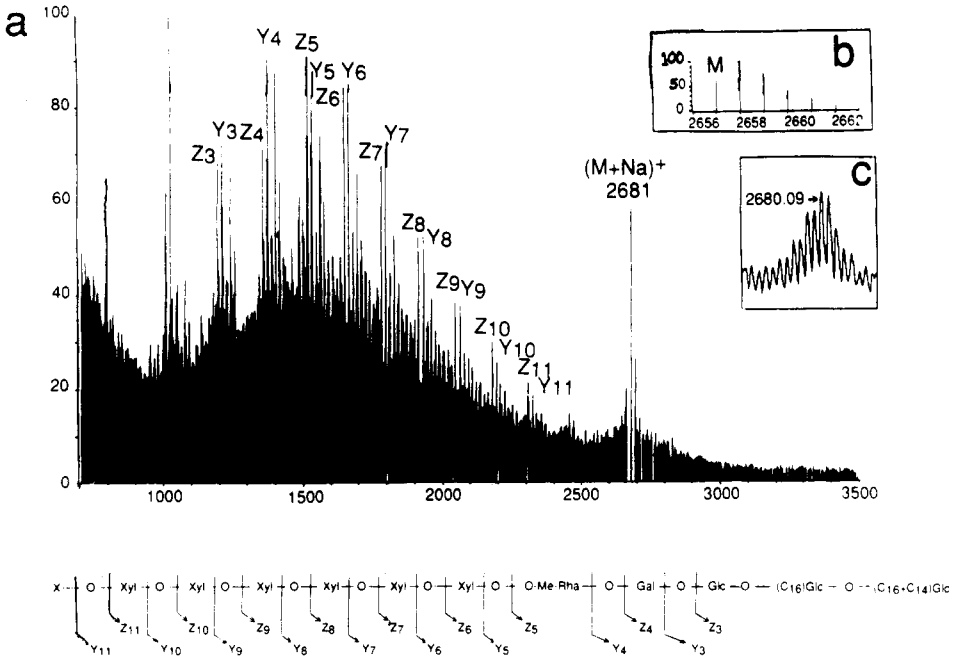

FIGURE 1: FAB-MS spectrum of the native LOS-III: (a) complete spectrum; (b) molecular ion theoretical isotopic distribution; (c) molecular ion experimental isotopic distribution.

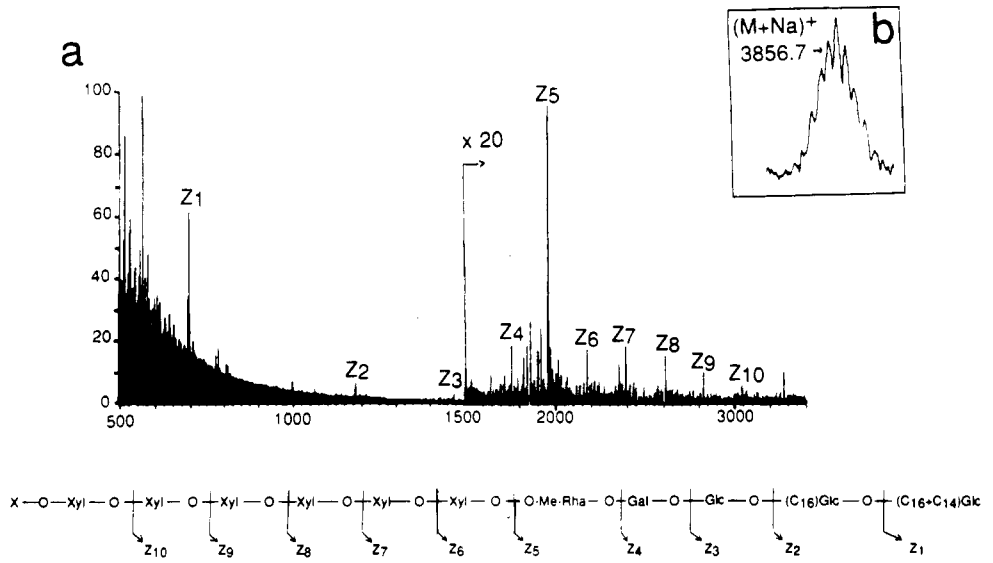

FIGURE 2: FAB-MS spectrum of the peracetylated LOS-III: (a) complete spectrum; (b) molecular ion experimental isotopic distribution.

The two fragmentation pathways (Poulter & Burlingame, 1990), namely, Y and Z, can be observed in the FAB-MS spectrum of the native LOS (Figure 1), while in the case of peracetylated LOS-III, only the Z series is present. The assignment of these fragment ions is summarized in Figures 1 and 2, in agreement with the sequence of LOS-III previously established by HMBC NMR spectroscopy on the native molecule (Gilleron et al., 1993). Moreover, the mass difference between the fragment Y_{11} and the cationized molecular ion revealed a molecular mass of 370 Da for the distal monosaccharide X. These data are in agreement with the EI mass spectrum of the peracetylated LOSs showing fragments at m/z 779 and 563 assigned to the distal disaccharidic and monosaccharidic oxonium ions, respectively. Thus, the complete structure of the distal monosaccharide X was investigated by scalar 2D NMR analysis of the LOS-III peracetylated derivative.

Identification of the Monosaccharidic Units. The protons were assigned by means of homonuclear sequences (COSY, RCT-COSY, and HOHAHA), while the ring protein relative configurations were deduced from the coupling constant values ($^3J_{H,H}$). These values were determined either from the 1D spectrum or from the phase-sensitive DQF-COSY cross peaks. Carbohydrate carbon resonance attributions were obtained from the ^1H–^{13}C shift correlation HMQC and HMQC–HOHAHA experiments and from literature data (Bock et al., 1984).

Figure 3 shows the 500-MHz ^1H 1D NMR spectrum of peracetylated LOS-III dissolved in $CDCl_3$. The three singlets at δ 3.243, 3.249, and 3.425 were easily assigned to the methoxy groups. Routinely, anomeric protons are the starting points for spin system elucidation, but in the case of peracetylated derivatives, the complexity of the anomeric region leads to ambiguous anomeric proton assignment. However, O-acetylation induces a weak deshielding effect of the substituted ^{13}C of about 2 ppm (Gagnaire et al., 1978), allowing the unambiguous anomeric carbon assignments from the 1D ^{13}C NMR spectrum and, consequently, the anomeric proton

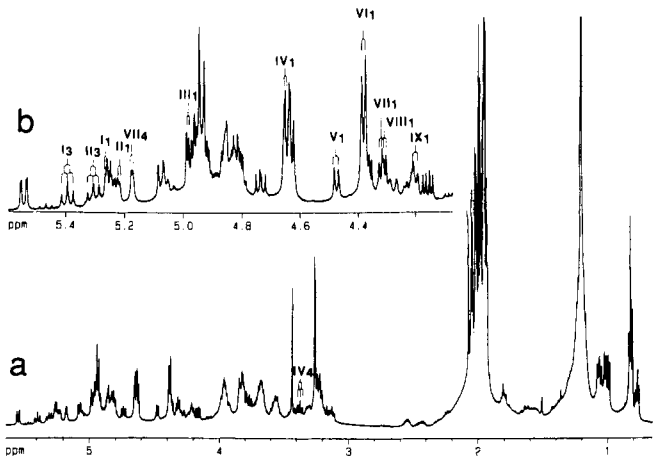

FIGURE 3: (a) 500-MHz ^{1}H NMR spectrum of peracetylated LOS-III in $CDCl_3$ and (b) expansion of the δ 4.20–5.70 region.

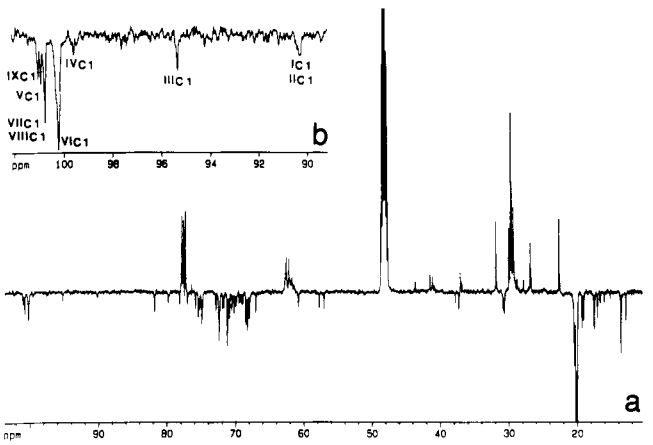

FIGURE 4: (a) 125-MHz ^{13}C NMR spectrum of peracetylated LOS-III in $CDCl_3$ and (b) expansion of the δ 10–104 region.

assignments by means of the HMQC spectrum. The ^{13}C NMR spectrum (Figure 4) recorded in the *J*-modulated mode shows seven anomeric carbon resonances at δ 90.3, 95.3, 99.56, 100.23, 100.77, 100.95, and 101.06. By means of the HMQC spectrum (Figure 5), these anomeric carbon resonances were found to correlate with nine anomeric proton resonances. The 90.3 and 100.77 ppm carbon resonances each correlate with two protons at δ 5.26 (I_1), 5.23 (II_1) and at δ 4.31 (VII_1), 4.30 ($VIII_1$), respectively. The carbon resonances at δ 95.3, 99.56, 100.23, 100.95, and 101.06 correlate with the anomeric protons at δ 4.99 (III_1), 4.65 (IV_1), 4.47 (V_1), 4.37 (VI_1), and 4.19 (IX_1), respectively. Only some of these protons resonated independently, allowing the coupling value determination of the following systems, V ($J_{1,2}$ = 7.3 Hz), VI ($J_{1,2}$ = 7.3 Hz), VII ($J_{1,2}$ = 7.6 Hz), VIII ($J_{1,2}$ = 6.9 Hz), and IX ($J_{1,2}$ = 8.2 Hz), indicating β-linked glycosyl residues. The remaining $J_{1,2}$ coupling constants were determined from the DQF-COSY spectrum. The most intense signal at δ 100.23 (Figure 4), which correlated with the anomeric proton (δ 4.37) and integrating for four protons (Figure 3), is consistent with the anomeric carbon resonance of β-Xyl*p* residues, in agreement with the GC/MS data (Gilleron et al., 1993).

The downfield resonance I_1 at δ 5.26 (Figure 3) was assigned to one of the two expected α-D-Glc*p* anomeric protons. From the COSY spectrum (Figure 6), the cross-peak connectivities allow proton assignments up to the two H6's (δ H1 = 5.26, δ H2 = 4.88, δ H3 = 5.40, δ H4 = 5.06, δ H5 = 3.83, δ H6 = 3.92, δ H6′ = 4.01) (Table 1). The coupling constant values $J_{2,3}$ = 9.7 Hz and $J_{3,4}$ = 9.7 Hz were unambiguously determined

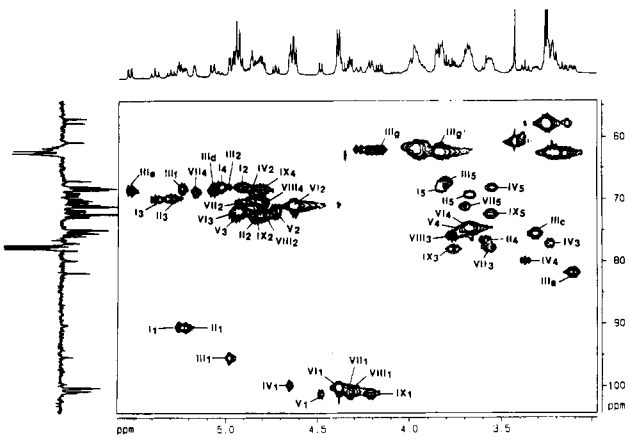

FIGURE 5: Expanded area of the phase-sensitive, ^{13}C-decoupled, ^{1}H-detected multiple quantum correlation spectrum (^{1}H{^{13}C} HMQC) of the peracetylated LOS-III at 500 MHz. The data matrix was 4K × 512 (TPPI) complex points with 96 scans per t_1 value. The spectral window was 27 700 Hz in the F_1 dimension (^{13}C) and 3030 Hz in the F_2 dimension (^{1}H). The relaxation delay was 2.5 s, and the delay after the BIRD pulse was 380 ms. For processing, a sine-bell window shifted by π/2 was applied in both dimensions, followed by expansion of the data matrix to 4K × 1K real matrix. The resulting digital resolutions were 0.35 Hz/point in F_2 (^{1}H) and 13.56 Hz/point in F_1 (^{13}C). Peaks are labeled with a Roman numeral identifying the residue followed by the number or letter assigning the carbon atom.

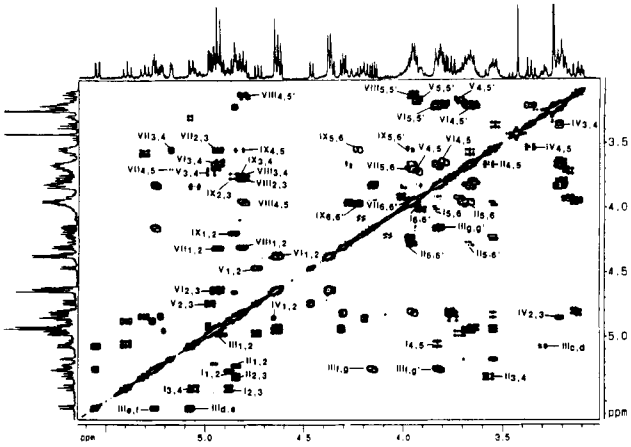

FIGURE 6: Expanded area of the 2D (^{1}H,^{1}H) COSY spectrum of peracetylated LOS-III (10 mg) in $CDCl_3$. The spectral width was 3105 Hz, and the relaxation delay was 1.5 s; the data matrix was 4K × 512 (TPPI) points, with 16 scans per t_1 value. Sine-bell apodizations with a π/2 phase shift were used in the t_2 and t_1 dimensions. Data were zero-filled in the F_1 dimension to obtain a final data matrix with 2K × 1K real points.

from the well-identified H3 signal, which resonates independently as a triplet on the 1D spectrum (Figure 3), while the $J_{1,2}$ coupling constant value (3.9 Hz) was determined from the (H1,H2) cross peak on the DQF-COSY spectrum. These values typify the spin system of the α-glucopyranosyl unit. Likewise, the spin system (Figure 6) starting with the anomeric proton II_1 at δ 5.23 ($J_{1,2}$ = 3.7 Hz) was assigned to the second α-D-Glc*p*. Correlation between these two anomeric protons at δ 5.26 and 5.23 and the upfield resonance at 90.3 on the HMQC spectrum demonstrated a trehalose unit [δ C1 92.15 (Gagnaire et al., 1976)]. The H4 upfield resonance (δ 3.57) indicates a hydroxyl that could not be acetylated, proving that unit II is involved in the glycosidic linkage while unit I is the distal monosaccharide of LOS-III (Table 1). For both units, a cross-section through the H3 resonance in the 2D HOHAHA spectrum shows cross peaks for all of the resonances up to the two H6's, supporting the two complete

Table 1: Proton Chemical Shifts (in ppm) and Coupling Constant Values (in Hz) Characterizing the Peracetylated LOS-III Monosaccharides[a]

sugar unit	H-1 ($^3J_{H1-H2}$)	H-2 ($^3J_{H2-H3}$)	H-3 ($^3J_{H3-H4}$)	H-4 ($^3J_{H4-H5}$)	H-5	H-6	others
3,6-dideoxy-α-Hex*p*(III) 1 → 3	4.99	4.92 (6/10)	1.79		3.79	0.98	
β-Xyl*p* (VIII) 1 → 4	4.30 (6.9)	4.80 (8.9)	3.77 (10)	4.82 (4.3/7.5)	3.95/3.13		
[β-Xyl*p*(1→4)]$_4$ (VI)	4.37 (7.3)	4.64	4.95	3.68	3.82/3.21		
β-Xyl*p*(V) 1 → 4	4.47 (7.3)	4.74 (7.5)	4.98 (9.2)	3.71 (7.1/5.3)	3.92/3.18		
3-*O*-Me-α-Rha*p*(IV) 1 → 3	4.65 (3.6)	4.85 (3.6)	3.21 (9.0)	3.36 (9.0)	3.54	1.17	3.25
β-Gal*p*(VII) 1 → 3	4.31 (7.6)	4.95 (10.7)	3.55 (3.5)	5.18 (weak)	3.69	3.95/3.97	
β-Glc*p*(IX) 1 → 4	4.19 (8.2)	4.86 (9.0)	3.75 (9.0)	4.82 (10.7)	3.55	3.97/4.22	
α-Glc*p*(II) 1 ↔ 1	5.23 (3.7)	4.84 (9.7)	5.31 (9.7)	3.57 (10.7)	3.68	3.96/4.28	
α-Glc*p*(I)	5.26 (3.9)	4.88 (9.7)	5.40 (9.7)	5.06 (10.7)	3.83	3.92/4.01	

[a] Measured at 303 K in $CDCl_3$ using tetramethylsilane as internal reference.

Table 2: Carbon Chemical Shift Values (in ppm) Characterizing the Peracetylated LOS-III Monosaccharides[a]

sugar unit	C-1	C-2	C-3	C-4	C-5	C-6	others
3,6-dideoxy-α-Hex*p*(III) 1 → 3	95.3	68.01	29.88	76.52	67.05	12.86	
β-Xyl*p*(VIII) 1 → 4	100.77	72.58	75.96	70.57	62.61		
[β-Xyl*p*]$_4$(VI) 1 → 4	100.23	71.21	72.45	74.98	62.61		
	100.4[b]	71.0[b]	71.9[b]	75.0[b]	62.6[b]		
β-Xyl*p*(V) 1 → 4	100.95	71.77	72.96	75.51	62.8		
3-*O*-Me-α-Rha*p*(IV) 1 → 3	99.56	69.23	77.11	79.89	68.01	17.00	57.095
β-Gal*p*(VII) 1 → 3	100.77	70.57	77.70	69.00	70.25	?	
β-Glc*p*(IX) 1 → 4	101.06	70.26	78.10	68.27	72.45	?	
α-Glc*p*(II) 1 ↔ 1	90.3	72.96	69.93	75.52	69.52	?	
α-Glc*p*(I)	90.3	68.01	70.01	68.01	68.26	?	
	92.15[c]	70.05[c]	70.10[c]	68.70[c]	68.35[c]	61.90[c]	

[a] Measured at 303 K in $CDCl_3$ using chloroform (δ 77) as internal reference. [b] Peracetylated methyl glycoside chemical shift taken from (Bock et al., 1984). [c] Peracetylated methyl glycoside chemical shift taken from (Gagnaire et al., 1976).

spin system assignments. Carbon resonances were identified up to C5 by the HMQC experiment (Table 2). The C6 resonance was not assigned due to the close resonances of C5 and C6 from the other glycosyl residues. Units I and II were previously found to be naturally acylated at C4 and C6 for unit I and at C2 for unit II (Gilleron et al., 1993).

The spin system starting with the anomeric proton IV$_1$ at δ 4.65 ($J_{1,2}$ = 3.6 Hz) is easily followed on the COSY spectrum (Figure 6) (δ H2 = 4.85, δ H3 = 3.21, δ H4 = 3.36, δ H5 = 3.54, δ C6H$_3$ = 1.17), leading to a C6 methyl group typifying a 6-deoxyhexose (Table 1). A cross-section through the methyl group at 1.17 ppm in the HOHAHA spectrum supports these assignments. The *manno* configuration was established from the coupling constant values ($J_{3,4}$ = 9.0 Hz, $J_{4,5}$ = 9.0 Hz) determined from the H4 proton signal, which resonates as a triplet on the 1D spectrum (Figure 3). The H3 resonance at δ 3.21 suggests the 3-*O*-Me-α-Rha*p* structure, in agreement with the HMBC spectrum showing a correlation between the C3 (77.11 ppm) and the *O*-methyl protons at 3.25 ppm. Moreover, the H4 chemical shift, which is unaffected upon peracetylation ($\delta_{\text{native LOS}}$ 3.573, $\delta_{\text{peracetylated LOS}}$ 3.36), supports the C4 glycosidic linkage. The 3-*O*-Me-Rha*p* ^{13}C resonances were unambiguously identified up to C5 by the HMQC experiment; the C3 chemical shift was confirmed by the observation of (H3,C3), (H4,C3), and (H5,C3) cross peaks from the HMQC–HOHAHA spectrum. The C6 resonance (δ 17.00) was characterized from its correlation with that of H4 on the HMBC spectrum.

System VI was attributed to the four β-Xyl*p*'s ($J_{1,2}$ = 7.3 Hz). The anomeric signal (δ 4.37) integrates for four protons, and the entire system (Figure 6) is easily followed up to H5 and H5′ (Table 1) due to the cross-peak intensities. This assignment was supported by a cross-section through the H2 proton at 4.64 ppm in the HOHAHA spectrum. Chemical shifts of the native and the peracetylated LOS-III (H4: $\delta_{\text{native LOS}}$ 3.594, $\delta_{\text{peracetylated LOS}}$ 3.68) support the 1→4 interglycosidic linkage. The carbon resonances were easily revealed from the *J*-modulated spectrum (Figure 4) due to their intensities. Attribution (δ C1 100.23, δ C2 71.21, δ C3 72.45, δ C4 74.98, δ C5 62.61) (Table 2) was achieved from the HMQC spectrum (Figure 5).

The proton chemical shifts of spin system V were established mainly by means of the HOHAHA spectrum from the well-characterized H2 at 4.74 ppm. On the basis of the identification of the H5 and H5′ resonances, it was assigned to a β-Xyl*p* ($J_{1,2}$ = 7.3 Hz). Moreover, from the H4 proton chemical shift (H4: $\delta_{\text{native LOS}}$ 3.573, $\delta_{\text{peracetylated LOS}}$ 3.71), unit V was found to be, as expected, glycosylated through C4. The anomeric carbon resonance was assigned by HMQC (Figure 5), while the other ring carbon resonances up to C5 were easily characterized by the HMQC–HOHAHA spectrum by a cross-section through the well-defined H2 proton (Table 2).

System VIII was attributed to the expected sixth β-Xyl*p* unit from the coupling constant values ($J_{1,2}$ = 6.9 Hz, $J_{2,3}$ = 8.9 Hz, $J_{3,4}$ = 10.0 Hz) and also from H5 and H5′ proton resonance assignments. All of the protons were identified (Table 1) by means of the HOHAHA spectrum from a cross-section through the H5′ proton at 3.13 ppm. The ^{13}C resonances were determined from the HMQC spectrum

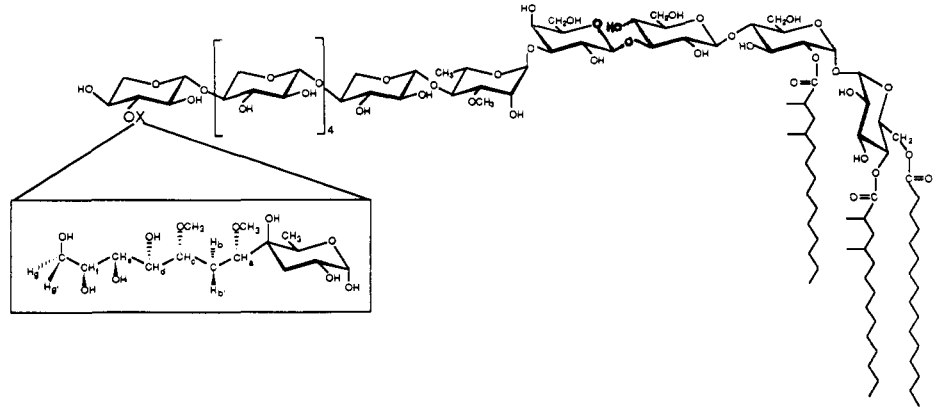

FIGURE 7: *M. gastri* LOS-III structure. The C-alkyl chain relative configurations and the monosaccharidic absolute configuration remain undetermined.

(Figure 5). The xylopyranosyl spin system VIII differs significantly from xylopyranosyl systems V and VI, notably by the chemical shifts of H3,H4 and C3,C4 (Tables 1 and 2). This unit is glycosylated through C3, in agreement with the fact that the H3 resonance is not significantly shifted in the peracetylated derivative (H3: $\delta_{\text{native LOS}}$ 3.391, $\delta_{\text{peracetylated LOS}}$ 3.77).

Starting with the anomeric proton VII_1 at 4.31 ppm ($J_{1,2}$ = 7.6 Hz) resonances up to H4 were easily followed (δ H2 = 4.95, δ H3 = 3.55, δ H4 = 5.18) (Figure 6). H4 resonates as a pseudodoublet ($J_{3,4}$ = 3.5 Hz and $J_{4,5}$ weak), and consequently the (H4,H5) cross peak appears very weak on the COSY spectrum (Figure 6). However, this correlation was unambiguously observed by means of the COSY-LR sequence, allowing the assignment of the remaining protons (δ H5 = 3.69, δ H6 = 3.95, δ H6′ = 3.97). This spin system characterizes β-Gal*p*. A cross-section through the well-resolved H4 proton at 5.18 ppm in the HOHAHA spectrum leads to the identification of correlations only for the H1, H2, and H3 protons consistently with a weak coupling constant between H4 and H5, supporting the β-D-Gal*p* structure. As expected, glycosylation of this unit at position 3 was established from the H3 chemical shift values (H3: $\delta_{\text{native LOS}}$ 3.495, $\delta_{\text{peracetylated LOS}}$ 3.55).

The anomeric resonance IX_1 (δ 4.19) was the starting point of the β-D-Glc*p* spin system. The characterization of H6 and H6′ protons and a C5 that resonates as a methine (CH) at 72.45 ppm supports this assignment. By means of the COSY spectrum (Figure 6), the protons were only assigned up to H5 due to the H5 superimposition with the H3 of unit VII. This ambiguity was overcome with HOHAHA spectrum analysis since closely spaced resonances can be assigned. Indeed, a cross-section through the H4 of system VII and through the anomeric proton of system IX, respectively, indicates chemical shifts of three different protons of each spin system. The higher intensity of the (H1,H5) cross peak suggests, prior to coupling constant value measurements, a β-Glc*p* residue instead of a β-Gal*p* residue. Anyway, the coupling constant values measured on the DQF-COSY spectrum ($J_{2,3}$ = 9.0 Hz, $J_{3,4}$ = 9.0 Hz, $J_{4,5}$ = 10.7 Hz) establish the *gluco* configuration. Analysis of the chemical shifts confirmed the glycosylation of this unit at position 3 (H3: $\delta_{\text{native LOS}}$ 3.491, $\delta_{\text{peracetylated LOS}}$ 3.75). The remaining spin system III, corresponding to the last expected residue, 3,6-dideoxy-α-hexopyranose, previously highlighted by the study of the native LOS (Gilleron et al., 1993), is now investigated.

Complete Structure of the Distal Monosaccharide. From the COSY spectrum, the anomeric resonance III_1 at δ 4.99 shows connectivities with resonances only up to H3 at 1.79 ppm (not shown). Similarly, cross-sections through the H1 as well as through the H2 in the HOHAHA spectrum failed to reveal any connectivity beyond this H3. By means of the HMQC sequence, it was found that H3 correlates with a methylene ^{13}C resonance at 29.88 ppm. Moreover, from the COSY spectrum (Figure 6), the resonance at 0.98 ppm correlates with the resonance at δ 3.79. Thus, this spin system characterizes the 3,6-dideoxyhexose core of the distal monosaccharide. the HMQC–HOHAHA spectrum reveals connectivities between the H1, H2, H3, and the ^{13}C resonance at 29.88 ppm assigned to C3. Likewise, H5 and H6 show correlations with both C5 (67.05 ppm) and C6 (12.86 ppm). The HMQC spectrum allows the observation of the connectivities (H5,C5) and (H2,C2), supporting the chemical shift of C5 at 67.05 ppm and thus assigning the chemical shift of C2 at 68.01 ppm, in agreement with the (C2,H3) correlation observed with the HMBC spectrum. The absence of correlation between the atoms at positions 1, 2, and 3 and at positions 5 and 6 in the HMQC–HOHAHA spectrum supports the absence of a proton on C4. So, the C4 resonance assignment was determined at 76.52 from the HMBC spectrum by the observation of connectivities between this carbon and the previously characterized protons H3, H5, and H6. As expected, this C4 resonated on the *J*-modulated spectrum (Figure 4) as a quaternary carbon. Moreover, the observation of (H3,C5) and (H5,C1) connectivities on the HMBC spectrum unambiguously established that these two systems are connected. Among these protons, only the chemical shift of H2 was affected upon acetylation (H2: $\delta_{\text{native LOS}}$ 3.934, $\delta_{\text{peracetylated LOS}}$ 4.92), suggesting, in agreement with the proposed structure, that the remaining carbons of this unit are devoid of any hydroxyl groups that could be acetylated.

A 1,3-dimethoxypropyl chain has been proposed previously as the partial structure of the C4 appendage. The two geminal protons, Hb and Hb′, belonging to the methylene unit of this chain, easily identified at 1.25 and 1.63 ppm, were the starting points for the spin system identification (Figure 7). These two protons correlate in the COSY spectrum (Figure 6) with the proton (Ha) at 3.12 ppm ($J_{a,b}$ = 10.7 Hz, $J_{a,b'}$ = 4.3 Hz) and with the proton (Hc) at 3.30 ppm ($J_{b,c}$ = 9.0 Hz, $J_{b',c}$ = 5.4 Hz). As described for the native molecule analysis (Gilleron et al., 1993), the cross peak between Ha and Hc was missing on the COSY spectrum, while it was observed in the HOHAHA spectrum. As expected, these two protons were not deshielded by the peracetylation process, supporting the presence of two methoxyl groups.

Contrary to the native LOS-III COSY spectrum, other correlations from Hc were observed, allowing the C-alkyl chain spin system assignment to be continued. COSY spectrum analysis (Figure 6) revealed the following sequence: Hc correlates with a well-characterized doublet of doublets (Hd) at 5.08 ppm ($J_{c,d}$ = 2.5 Hz), which in turn correlates with the proton He at 5.56 ppm, also resonating as a doublet of doublets ($J_{d,e}$ = 9.8 Hz). The proton He correlates with Hf at 5.25 ppm ($J_{e,f}$ = 1.8 Hz), which is found to be connected to two vicinal protons, Hg (δ 3.82, $J_{f,g}$ = 7.2 Hz) and Hg′ (δ 4.16, $J_{f,g'}$ = 5.4 Hz). Cross peaks (Hd,Hf) and (He,Hg) were observed in the RCT-COSY spectrum. The HMQC spectrum shows that Hg and Hg′ are limited to the same carbon resonance at δ 62.2, which is also supported by the *J*-modulated spectrum (Figure 4), revealing that this carbon resonates as a methylene carbon. The carbons bearing Ha and Hc resonate at 81.9 and 75.51 ppm, respectively, which is in agreement with one methoxyl group borne by each. The downfield chemical shifts of Hd, He, Hf, Hg, and Hg′, which resonate between 3.8 and 5.0 ppm, suggest that their carbons bear four hydroxyl groups, in agreement with the FAB-MS data.

Thus, a 2,3-dimethoxy-4,5,6,7-tetrahydroxyheptyl structure can be tentatively proposed for the C4 alkyl chain (Figure 7). From the mass values of the alkyl side chain and the core 3,6-dideoxy-α-Hex*p*, it can be calculated that 17 mass units are missing to obtain 370 Da, which is the molecular mass previously established for this monosaccharide. In agreement with the fact that tertiary hydroxyl groups are incapable of being acetylated (Hunter et al., 1984), the remaining C4 substituent was assigned to a hydroxyl group. Thus, 3,6-dideoxy-4-*C*-(1,3-dimethoxy-4,5,6,7-tetrahydroxyheptyl)-α-Hex*p* was proposed as the structure for the distal monosaccharide (Figure 7). The C4 configuration was investigated by ROESY spectrum analysis. An ROE contact between Ha and H5 agrees with the equatorial orientation of the lateral chain, indicating a *xylo* configuration.

DISCUSSION

The partial structure of the *M. gastri* W471 LOS-III has been previously established (Gilleron et al., 1993) from its native form using a new analytical approach in the field of mycobacterial glycolipids based on a panel of 2D NMR scalar coupling techniques for ^{1}H and ^{13}C homonuclear (COSY, HOHAHA, and RCT-1, -2, and -3 COSY) and heteronuclear (HMQC, HMBC, and HMQC–HOHAHA) sequences. The *M. gastri* LOS-III has been shown to be composed of 12 monosaccharides, including at one end a trehalose triacylated on carbons 2′, 4, and 6 and at the other end a new type of monosaccharide. The LOS-III core has been found to be composed of one β-D-Glc*p*, one β-D-Gal*p*, one 3-*O*-Me-α-Rha*p*, and six β-L-Xyl*p*'s. The originality of the *M. gastri* LOS-III structure arises from the unique feature of the terminal monosaccharide. Its structure was previously partially established as a 3,6-dideoxy-α-Hex*p* with a C4-linked 1,3-dimethoxypropyl chain. The drawback of the analytical approach we used for the elucidation of the distal glycosyl unit results from the resonance of 62 ring protons over a 1.1 ppm range, leading to overcrowding of the 2D spectra. In order to spread out the ring proton resonances, the LOS-III was peracetylated since it is well known that the geminal protons of acetoxy groups are then shifted at lower field (Gagnaire et al., 1978; Gasa et al., 1986). Thus, the ring protons of the peracetylated derivative resonated over a 2.7 ppm range, leading to homonuclear and heteronuclear spectra with less cross peak overlapping.

The first analysis step was the determination of each glycosyl spin system by homonuclear (COSY, RCT-COSY, and HOHAHA) and heteronuclear (HMQC, HMQC–HOHAHA, and HMBC) sequences. The anomeric resonances are currently the starting points for the characterization of each spin system. Nevertheless, in the case of peracetylated derivatives, the anomeric proton region becomes more complex due to the resonance of the geminal acetoxy protons in this area. However, by means of the HMQC experiment, the anomeric protons were easily identified from the anomeric carbons since the carbon chemical shifts are almost unaffected by the acetylation (Gagnaire et al., 1978).

The NMR analysis of the peracetylated LOS-III derivative is in agreement with the previously proposed structure, notably with the incomplete structure of the distal monosaccharide assigned to 3,6-dideoxy-4-*C*-(1,3-di-*O*-methylpropyl)-α-Hex*p*. Moreover, as mentioned above, LOS-III peracetylation, which spreads out the sugar ring proton resonances, allows the complete establishment of the C4 alkyl chain structure: 1,3-dimethoxy-4,5,6,7-tetrahydroxyheptyl. The present NMR study agrees with the absence of any hydrogen atom linked to C4. Thus, the remaining C4 substituent, assigned to a hydroxy group, was revealed from the FAB-MS studies of the native and peracetylated LOS-III and was supported by means of the EI mass spectra of the peracetylated and perdeuterioacetylated LOS-III derivatives. Besides the molecular mass of 2656 Da for LOS-III determined from this study and allowing the establishment of the following LOS-III elementary composition, $C_{121}H_{212}O_{62}$, a molecular mass of 370 Da was determined for the C4-branched 3,6-dideoxyHex*p*.

This last structural information was the key data suggesting a hydroxyl group for the remaining C4 substituent and a C-linkage of the alkyl chain. Since the tertiary hydroxyl group is unable to be acetylated, the six hydroxyl groups determined from the FAB-MS study of the native and peracetylated LOS-III are in agreement with the NMR data, indicating that four of them are localized in the branched chain, the other two being in the sugar ring. From the downfield chemical shift of their geminal protons in the peracetylated derivative, the carbons linked to these groups were easily identified. These data taken together allows us to propose a 3,6-dideoxy-4-*C*-(1,3-dimethoxy-4,5,6,7-tetrahydroxyheptyl)-α-*xylo*-hexopyranose structure, revealing that LOS-III contains a monosaccharide that is new-found in nature. The *xylo* configuration was proposed from ROESY spectrum analysis. The structure of this complex monosaccharide, which was not revealed by routine sugar analysis such as GC/MS, was elucidated from the analysis of the whole molecule by means of 2D NMR and FAB-MS. This analytical approach, which does not require monosaccharide purification, could be applied to the structural elucidation of the immunodominant monosaccharide of other antigenic glycolipids. The relative configuration of the different carbons of the C4 substituent chain remains unknown.

The 3,6-dideoxyhexoses have considerable biological significance, as they have been shown to contribute to the serological specificity of many immunologically active polysaccharides. Four of the eight possible 3,6-dideoxyhexoses have been isolated from various lipo-polysaccharides elaborated by Gram-negative bacteria (Hanessian, 1966): abequose (D-xylo), tyvelose (D-arabino), paratose (D-ribo), and colitose (L-xylo). To our knowledge (Lindberg, 1990), the only two known examples of C-branched 3,6-dideoxyHex*p*'s occur in *O*-antigens of *Yersinia* species [*pseudotuberculosis* (Gorshkova et al., 1984) and *enterocolitica* (Gorshkova et al., 1987)]. They share the same sugar ring structure, 3,6-dideoxy-4-*C*-

(1-hydroxyethyl)-D-Xyl*p*, but differ in the carbon configuration of the side chain.

By analogy to the LPS dideoxyhexoses, the distal LOS-III monosaccharide, 3,6-dideoxy-4-*C*-(1,3-dimethoxy-4,5,6,7-tetrahydroxyheptyl)-α-*xylo*-hexopryanose, must typify the *M. gastri* species, since it has been previously shown that LOS-III is present in all of the *M. gastri* strains investigated. Thus, a molecular marker has now been identified, allowing the unambiguous differentiation of *M. gastri* from *M. kansasii*, two strains which have long been considered, by some bacteriologists, as synonymous. Moreover, *M. kansasii* is known to be the etiologic agent of pulmonary disease, while *M. gastri* is considered to be nonpathogenic. Thus, in order to characterize *M. gastri* by a serological probe, the molecular specificity of the anti-LOS-III antibodies must now be investigated.

ACKNOWLEDGMENT

We thank Dr. Marie-Antoinette Lanéelle for the kind gift of the mycobacterial cultures and Dr. Bernard Escoffier and Danielle Ithier for recording the FAB mass spectra.

REFERENCES

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–60.
Bax, A., & Summers, M. F. (1986) *J. Am. Chem. Soc. 108*, 2093–2094.
Bax, A., Griffey, R. H., & Howkins, B. L. (1983) *J. Magn. Reson. 55*, 301–315.
Belisle, J. T., & Brennan, P. J. (1989) *J. Bacteriol. 171*, 3465–3470.
Bock, K., Pedersen, C., & Pedersen, H. (1984) *Adv. Carbohydr. Chem. 42*, 193–225.
Camphausen, R. T., McNeil, M., Jardine, I., & Brennan, P. J. (1987) *J. Bacteriol. 169*, 5473–5480.
Daffé, M., McNeil, M., & Brennan, P. J. (1991) *Biochemistry 30*, 378–388.
Draper, P. (1974) *J. Gen. Microbiol. 83*, 431–433.
Draper, P., & Rees, R. J. W. (1970) *Nature 228*, 860–861.
Gagnaire, D. Y., Taravel, F. R., & Vignon, M. R. (1976) *Carbohydr. Res. 51*, 157–168.
Gagnaire, D. Y., Mancier, D., & Vicendon, M. (1978) *Org. Magn. Reson. 11*, 344–349.
Gasa, S., Nakamura, M., Makita, A., Ikura, M., & Hikichi, K. (1986) *Eur. J. Biochem. 155*, 603–11.
Gilleron, M., Venisse, A., Fournié, J. J., Rivière, M., Dupont, M. A., Gas, N., & Puzo, G. (1990) *Eur. J. Biochem. 189*, 167–173.
Gilleron, M., Vercauteren, J., & Puzo, G. (1993) *J. Biol. Chem. 268*, 3168–3179.
Gorshkova, R. P., Zubkov, V. A., Isakov, V. V., & Ovodov, Y. S. (1984) *Carbohydr. Res. 126*, 308–312.
Gorshkova, R. P., Zubkov, V. A., Isakov, V. V., & Ovodov, Y. S. (1987) *Bioorg. Khim. 13*, 1146–1147.
Hanessian, S. (1966) *Adv. Carbohydr. Chem. 21*, 143–201.
Hunter, S. W., Murphy, R. C., Clay, K., Goren, M. B., & Brennan, P. J. (1983) *J. Biol. Chem. 258*, 10481–10487.
Hunter, S. W., Fujiwara, T., Murphy, R. C., & Brennan, P. J. (1984) *J. Biol. Chem. 259*, 9729–9734.
Hunter, S. W., Jardine, I., Yanagihara, D. L., & Brennan, P. J. (1985) *Biochemistry 24*, 2798–2805.
Hunter, S. W., Barr, V. L., McNeil, M., Jardine, I., & Brennan, P. J. (1988) *Biochemistry 27*, 1549–1556.
Kamisango, K., Saadat, S., Dell, A., & Ballou, C. E. (1985) *J. Biol. Chem. 260*, 4117–4121.
Lecocq, C., & Lallemand, J. Y. (1981) *J. Chem. Soc., Chem. Commun.* 150–152.
Lerner, L., & Bax, A. (1986) *J. Magn. Reson. 69*, 375–380.
Lindberg, B. (1990) *Adv. Carbohydr. Chem. Biochem. 48*, 279–317.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
McNeil, M., Tsang, A. Y., McClatchy, J. K., Stewart, C., Jardine, I., & Brennan, P. J. (1987) *J. Bacteriol. 169*, 3312–3320.
Poulter, L., & Burlingame, A. L. (1990) *Methods Enzymol. 193*, 661–689.
Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.
Saadat, S., & Ballou, C. E. (1983) *J. Biol. Chem. 258*, 1813–1818.
Shaka, A. J., Barker, P. B., & Freeman, R. (1985) *J. Magn. Reson. 64*, 547–552.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.
Vercellone, A., Rivière, M., Fournié, J. J., & Puzo, G. (1992) *FEBS Lett. 303*, 22–26.