Lipooligosaccharidic Antigen Containing a Novel C4-branched 3,6-Dideoxy-a-hexopyranose Typifies Mycobacterium gastri"

Martine Gilleron, Joseph Vercauterent, Germain Puzosv

THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 268, No. 5, Issue of February 15, pp. 3168-3179, 1993
Printed in U.S.A.

Lipooligosaccharidic Antigen Containing a Novel C4-branched 3,6-Dideoxy-α-hexopyranose Typifies *Mycobacterium gastri**

(Received for publication, July 13, 1992)

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

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

Lipooligosaccharides (LOSs) have recently been proposed as markers of mycobacterial avirulence. They have been characterized in *Mycobacterium kansasii* cell wall and were investigated in *Mycobacterium gastri* since the distinction between the two mycobacterial species remains in some question. A set of unknown LOSs was isolated from *M. gastri* W471. The highly antigenic lipooligosaccharide, LOS-III, was purified and appeared in all the *M. gastri* strains investigated regardless of their morphology. Moreover, by enzyme-linked immunosorbent assay and chromatogaphic approaches, it was found that LOS-III unambiguously distinguished *M. gastri* from the opportunistic pathogen *M. kansasii*. The LOS-III structure was established from its native form using NMR spectroscopy. This strategy revealed the presence of a supplementary monosaccharide (X) which was not characterized by routine carbohydrate analysis. Its core structure, 3,6-dideoxy-α-hexopyranose, was established from the complete assignment of the ^{1}H and ^{13}C spectra by two-dimensional homonuclear (COSY, HOHAHA) and heteronuclear ^{1}H-^{13}C heteronuclear multiple quantum correlation spectroscopy (HMQC) and HMQC-HOHAHA spectroscopy. Due to the absence of a proton at C4, the key data of the C4 side chain structure came from the heteronuclear multiple bond correlation spectroscopy (HMBC) spectrum. It was revealed to be a C-alkyl chain of partial structure 1,3-dimethoxypropyl. From the HMBC spectrum, this novel C-branched monosaccharide was located at the nonreducing end of the LOS, while the putative reducing end was found to consist of a 2′,4,6-triacylated α-α-trehalose. The following structure, based on the evidence presented in this paper, is proposed for LOS-III: Xpα(1→3)[L-Xylpβ(1→4)]$_6$3-*O*-Me-Rhapα(1→3)D-Galpβ(1→3)-D-Glcpβ(1→4)2-*O*-acyl-D-Glcp-α(1↔1)α4,6-di-*O*-acyl-D-Glcp.

Tuberculosis still remains a major human health problem, with about 3 million deaths per annum worldwide. Indeed, mycobacterial infection and disease are on the increase in the United States, due in part to the rise of HIV-1[1]-related infections (1). Among the nontuberculous mycobacteria, *Mycobacterium kansasii* ranks second behind the *Mycobacterium avium-intracellulare* complex as the causative agent of mycobacterial infection in AIDS patients (2). All pathogenic mycobacteria share the property of being intracellular pathogens with the ability to grow inside the phagosomes of phagocytic cells. It is generally accepted that the mycobacterial lipidic cell wall provides a protective capsule shielding the bacteria from the toxic metabolites of macrophages. So, considerable effort has been devoted to the isolation and characterization of the cell wall glycoconjugates which are the major mycobacterial antigens (3).

M. kansasii cell wall is endowed with two distinct classes of surface glycolipid antigens: the trehalose-containing lipooligosaccharides (LOSs) and the phenolic glycolipids (Phe Gls). LOSs have been characterized in smooth strains of *M. kansasii*, whereas the rough variants are devoid of these antigens (4). This is not the case of the Phe Gl which appear as permanent markers of the *M. kansasii* strains. Thus, an interesting correlation has been proposed between virulence and LOS content: LOSs are considered to be markers of avirulence (4). Indeed, rough variants of *M. kansasii* cause a chronic systemic infection when injected into B6D2 mice, whereas smooth strains are rapidly cleared (5). It was shown by immunochemistry that both types of glycolipids are exposed at the cell wall surface (4, 6). Thus, due to the high glycosylation of the LOSs compared to that of the Phe Gls, it has been suggested (4) that in smooth strains LOSs may mask some important cell wall glycolipids that are otherwise effective. These substances may be the phenolic glycolipids described as scavengers of oxygen radicals (7) and/or the lipoarabinomannan (LAM) that inhibits antigen-dependent

* This work was supported by Grant 6718 from the Association pour la recherche sur le Cancer and Grant RECH/9000827 from the Région Midi Pyrénées. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed. Tel.: 61-33-59-12; Fax: 61-33-58-86.

[1] The abbreviations used are: HIV, human immunodeficiency virus; AIDS, acquired immunodeficiency syndrome; COSY, ^{1}H-^{1}H correlated spectroscopy; DQF-COSY, double quantum filtered ^{1}H-^{1}H correlated spectroscopy; ELISA, enzyme-linked immunosorbent assay; GC/MS (CI and EI), gas chromatography/mass spectrometry (chemical ionization and electronic impact); HMBC, heteronuclear multiple bond correlation spectroscopy; HMQC, heteronuclear multiple quantum correlation spectroscopy; HOHAHA, homonuclear Hartmann-Hahn spectroscopy; HPLC, high-pressure liquid chromatography; $^{3}J_{H,H'}$, vicinal spin-spin coupling constants; LAM, lipoarabinomannan; LOS, lipooligosaccharide; LOS-III, quantitatively major LOS isolated from *M. gastri* cell wall; L-SIMS, liquid-secondary ion mass spectrometry; NOESY, nuclear Overhauser effect spectroscopy; Phe Gl K-I, K-II, K-III, K-IV, phenolic glycolipids from *M. kansasii*; ROESY, rotating frame nuclear Overhauser and exchange spectroscopy; TLC, thin-layer chromatography; TQF-COSY, triple quantum filtered ^{1}H-^{1}H correlated spectroscopy; Galp, galactopyranose; Glcp, glucopyranose; Rhap, rhamnopyranose; Xylp, xylopyranose; 1D, one-dimensional; 2D, two-dimensional; TPPI, time proportional phase increment.

lymphoproliferation (8) as well as γ-interferon-dependent activation of macrophages (9).

The structure of the *M. kansasii* LOS antigens has been studied in detail (10, 11). It is based on a triacylated multiglycosylated trehalose core (3). The distal monosaccharide named kansosamine, which has a particular structure (4,6-dideoxy-2-*O*-Me-3-C-Me-4-(2′-methoxypropionamido)-α-L-*manno*-hexopyranosyl) is immunodominant and constitutes the epitope with the adjacent fucosyl residue (12). *M. kansasii* phenolic glycolipid, known for many years as mycoside A, has been renamed Phe Gl K-I (13, 14). Despite its hitherto unknown structure, characterized by the particular distal monosaccharide, a 2,6-dideoxy-4-*O*-Me-α-D-*arabino*-hexopyranose, it has also been identified by serological and analytical techniques in *Mycobacterium gastri* (15). Besides Phe Gl K-I, other glycolipids in smaller amounts have been observed in the *M. kansasii* cell wall: Phe Gl K-II (15), Phe Gl K-III, and Phe Gl K-IV (6). They differ by the structure of their distal monosaccharide unit, a 2,4-di-*O*-Me-α-D-Man*p* for Phe Gl K-II and a 4-*O*-Me-α-Man*p* for Phe Gl K-IV. All these Phe Gls also occur in *M. gastri* (6). These data, as well as the mycolate and secondary alcohol structures, suggest that the strains *M. kansasii* and *M. gastri* are at least very closely related (14). However, *M. kansasii* is known to be the etiological agent of pulmonary diseases resembling tuberculosis, while *M. gastri* has never been found to be associated with human disease. So, they are an attractive model for studying mycobacterial pathogenicity at the molecular level.

In the present study, a set of unknown molecules belonging to the LOS family was found in *M. gastri* W471. These glycolipids were purified, and their structures were partially established. They were found to differ by the epitope structure compared to those described in *M. kansasii*. We report here the structural and serological analysis of their major representative referred to as LOS-III. Its structure was essentially elucidated using recent nondestructive analytical tools, homonuclear and heteronuclear NMR spectroscopy, applied to the native glycolipid. Whatever the *M. gastri* colony morphology, LOSs were detected, but with different structures suggesting a continuum between the *M. kansasii* smooth colony and the *M. gastri* species. The presence of these LOSs will be discussed in terms of avirulent factors.

EXPERIMENTAL PROCEDURES

Culture Conditions—*M. gastri* strains W471, HB 4362, and HB 4389 and *M. kansasii* strains ATCC 12478, S890175, and S890370 were grown as surface pelicules on Sauton's medium at 37 °C for 2 months.

LOS Purification—The same protocol was used for each strain but will only be presented for *M. gastri* strain W471. The bacillary mass was extracted four times in a $CHCl_3/CH_3OH$ mixture (1:1, v/v) at room temperature. The whole extract was concentrated to yield 50 g of a crude cell wall extract which was suspended in 200 ml of $CHCl_3$ and washed twice with an equal volume of H_2O. The organic phase was brought to dryness, and the obtained 20 g of crude lipid extract was solubilized in hot acetone (50 °C) and allowed to precipitate overnight at 4 °C. The suspension was then centrifuged at 4 °C for 20 min (3000 × *g*), and the supernatant was evaporated to yield 10 g of acetone-soluble and 9 g of acetone-insoluble lipid extracts. This procedure allows the elimination of 50% of the contaminants, particularly phospholipids. The acetone-insoluble lipidic fraction was dissolved in a minimum quantity of $CHCl_3$ and resuspended in CH_3OH. Then, the phospholipids precipitated, while the LOSs remained methanol-soluble (4 g).

M. gastri W471 LOSs purification was achieved by HPLC with a 5-mm spherisorb direct-phase column monitored by IR detection. The mobile phase consisted of a linear tertiary solvent $CHCl_3/CH_3OH/H_2O$ (60:30:4, v/v) flowing at 1 ml/min. The highly purified glycolipids LOS-III (25 mg) and LOS-IV (40 mg) appeared as single peaks.

LOS Methanolysis Procedure—100 μg of LOS were methanolyzed with 2 M HCl in dry methanol at 80 °C for 2 h and evaporated under nitrogen.

Analytical Methods—HPLC was conducted in a Gilson apparatus composed of a Model 303 pump, an 802 manometric module, and an 811 dynamic mixer. Elution and data acquisition were controlled with Gilson GME 714 Software and a Gilson Model 621 data module interface, connected to an OPAT Normerel computer. Elution of the products was monitored with a Waters Model 450 variable wavelength detector.

Infrared analysis was conducted on a Perkin-Elmer 1600 series FT-IR spectrophotometer.

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

Routine gas chromatography was performed on a Girdel series 30 equipped with an OV1 wall-coated open tubular, 25-m capillary column, inner diameter 0.3 mm, using nitrogen gas at a flow rate of 2.5 ml/min with a flame ionization detector. A temperature program from 100 to 250 °C at a speed of 2 °C/min and an injector temperature of 260 °C were used for trimethylsilyl methyl glycoside analysis. The temperature of the detector was 310 °C.

GC/MS was performed on a Nermag Model R 10/10 quadrupole mass spectrometer connected to a PDP-8M computer.

Immunological Procedures and Enzyme-linked Immunosorbent Assay—Rabbit antiserum was obtained as previously described (14), and ELISA tests were performed as detailed in Ref. 15.

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 expressed in ppm downfield from internal tetramethylsilane (0 ppm) and internal chloroform (77 ppm), respectively. The sample was dried under vacuum, dissolved in $CDCl_3/CD_3OD$ (Spin et Techniques, Paris, France; 1:1, v/v) 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. Data were acquired in the phase-sensitive mode using the TPPI method (17) or the method of States *et al.* (18). DQF-COSY (19) and HOHAHA (20) were recorded with standard pulse sequences. ^1H-^{13}C correlation spectra were recorded in the proton-detected mode with a Bruker 5-mm ^1H-broadband turnable probe with reversed geometry. The pulse sequence used for single-bond correlation spectra (HMQC) was that of Bax (21). Globally optimized alternating phase rectangular pulse sequence (22) at the carbon frequency was used as a composite pulse decoupling during the acquisition. Multiple-bond correlation spectra (HMBC) (23) 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 (24). Experimental details and processing parameters of the spectra are given in the figure legends.

The one-dimensional ^1H spectrum was measured using a 30–90° tipping angle for the pulse and 0.2 s as a recycle delay between each of the 64 acquisitions of 13.7 s. The spectral width of 2400 Hz was digitized on 64,000 complex points that were multiplied by a Lorentz-Gauss function (LB −1.3 Hz; GB 0.28) prior to processing to 32,000 real points in the frequency domain, giving a digital resolution of 0.075 Hz/point.

The proton-decoupled ^{13}C-modulated spectrum was recorded with a recycle delay of 2.3 s, and 39761 scans were accumulated. The spectral width used was 27700 Hz, the data were digitized on 32,000 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.

RESULTS

Purification of an Antigenic Lipooligosaccharide from M. gastri W471—From a $CHCl_3/CH_3OH$ extract of *M. gastri*, the crude acetone-soluble fraction contained the phenolic glycolipids as previously described (6). However, the acetone-insoluble fraction contained, besides phospholipids, a set of molecules revealed by anthrone, but unrevealed by ninhydrin or dittmer. Most of the phospholipids were CH_3OH-precipitated, while the glycolipids of interest (called LOS) were solubilized. This methanol-soluble fraction, which was still contaminated by phospholipids was then chromatographed

on either a DEAE or QMA ion-exchange column. The QMA column (acetate) was first irrigated by $CHCl_3$ to eliminate the more apolar neutral compounds. Then the LOS were eluted by $CHCl_3/CH_3OH$ (1:2, v/v) giving, on silicic acid TLC, a homogeneous family of glycolipids revealed by anthrone. The fractionation of these glycolipids was successfully accomplished using an HPLC device equipped with a 5-mm spherisorb semipreparative column monitored by a refractory detector. The mobile phase consisted of a tertiary solvent: $HCCl_3/CH_3OH/H_2O$ (60:30:4, v/v). In this way, the two quantitatively major LOSs, namely LOS-III and LOS-IV were purified until silicic acid TLC homogeneity (Fig. 1). This approach led to 25 mg of LOS-III and 40 mg of LOS-IV from 50 g of total $CHCl_3/CH_3OH$ extractable lipids. The general feature of the Fourier transform infrared spectrum of LOS-III and the absorption at 1740 cm^{-1} (ester function) and at 3000–3500 cm^{-1} (alcohol function) confirmed that these glycolipids were indeed structurally related to the LOS class.

Conventional plate ELISA using rabbit polyclonal antiserum raised against *M. gastri* W471 showed that only LOS-III and LOS-IV were antigenic among the LOSs (data not shown).

Glycosyl Composition of LOS-III—The monosaccharide composition was established by a routine approach. The native glycolipid was methanolyzed, then the products of the reaction were pertrimethylsilylated and analyzed by GC/MS using CI and EI ionization modes (25). Both CI and EI mass spectrum analysis and GC cochromatography with authentic standards indicated the presence of 3-*O*-Me-Rha (*peaks 1* and *2*), Xyl (*peaks 3* and *4*), Gal (*peaks 6* and *7*), and Glc (*peaks 9* and *10*) (Fig. 2). The two remaining peaks (5 and 8) were assigned to 2,4-dimethyltetradecanoic methyl ester and palmitic methyl ester, respectively. The presence of these fatty methyl esters released by methanolysis confirmed that this glycolipid belongs to the LOS group (10, 11). Based on quantitation GC analysis, it can be advanced that the LOS-III is composed of Glc, Gal, Xyl, and 3-*O*-Me-Rha in the ratio 3:1:6:1.

The absolute configurations of the glycosyl residues were determined according to the method of Gerwig *et al.* (26). The LOS-III was solvolyzed in the presence of (+)-2-butanol to yield butyl glycoside diastereoisomers. GC analysis of the trimethylsilylated derivatives showed that the xylosyl residues belong to the "L" series, while the galactosyl and glucosyl residues are "D."

LOS-III NMR Studies—The ^1H NMR spectrum of native LOS-III (Fig. 3) in a mixture of $CDCl_3/CD_3OD$ (2:1, v/v) at 500 MHz shows a complex anomeric proton region with multiple peaks having both different coupling constants and intensities. The three singlets at δ 3.355, 3.385, and 3.461 were assigned to three methoxy groups.

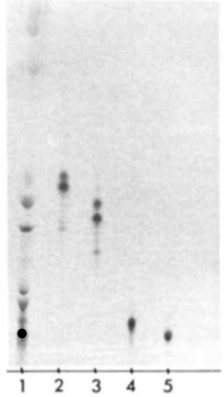

FIG. 1. **Thin-layer chromatography in $CHCl_3/CH_3OH/H_2O$ (60:35:8, v/v), using the orcinol-H_2SO_4 spray.** *1*, total unfractionated lipid fraction; *2*, LOS-I; *3*, LOS-II; *4*, LOS-III; *5*, LOS-IV.

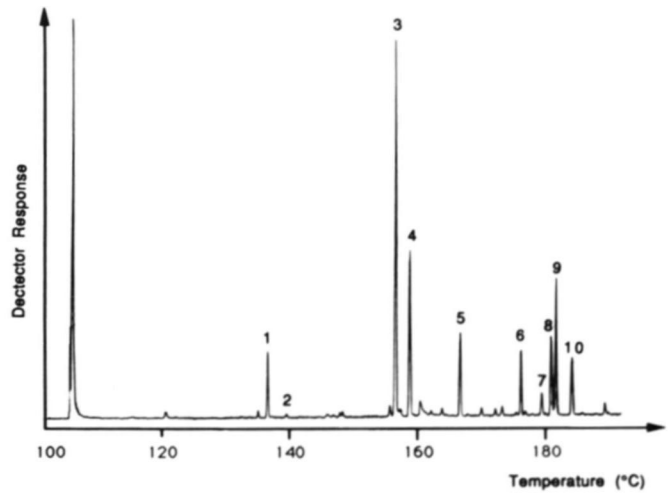

FIG. 2. **Capillary GC chromatogram of the trimethylsilyl derivatives of the glycoside ester arising from the LOS-III.** *Peaks 1* and *2*: methyl-2,4-di-*O*-trimethylsilyl-3-*O*-Me-α- and β-Rha*p* (CI: $(M + NH_4)^+$ *m/z* 354; $(M + NH_4)^+ - 32$ *m/z* 322; $(M + H)^+ - 32$ *m/z* 335; EI: *m/z* 146; *m/z* 217; *m/z* 75). *Peaks 3* and *4*: methyl-2,3,4-tri-*O*-trimethylsilyl-α- and β-Xyl*p* (CI: $(M + NH_4)^+$ *m/z* 398; $(M + H)^+$ *m/z* 381; $(M + NH_4)^+ - 32$ *m/z* 366; $(M + H)^+ - 32$ *m/z* 349; EI: *m/z* 204; *m/z* 217; *m/z* 133). *Peak 5*: 2,4-dimethyltetradecanoic methyl ester (CI: $(M + NH_4)^+$ *m/z* 288; EI: Mac Lafferty *m/z* 88; $(M - 31)$ *m/z* 239; $(M - 43)$ *m/z* 227; $(M - 71)$ *m/z* 199). *Peaks 6* and *7*: methyl-2,3,4,6-tetra-*O*-trimethylsilyl-α- and β-Gal*p* (CI: $(M + NH_4)^+$ *m/z* 500; $(M + H)^+$ *m/z* 483; $(M + NH_4)^+ - 32$ *m/z* 468; $(M + H)^+ - 32$ *m/z* 451; EI: *m/z* 204; *m/z* 217; *m/z* 133). *Peak 8*: palmitic methyl ester (CI: $(M + NH_4)^+$ *m/z* 288; EI: Mac Lafferty *m/z* 74). *Peaks 9* and *10*: methyl-2,3,4,6-tetra-*O*-trimethylsilyl-α- and β-Glc*p* (CI: $(M + NH_4)^+$ *m/z* 500; $(M + H)^+$ *m/z* 483; $(M + NH_4)^+ - 32$ *m/z* 468; $(M + H)^+ - 32$ *m/z* 451; EI: *m/z* 204; *m/z* 217; *m/z* 133).

The anomeric region presents 11 signals, nine which appear as doublets and pseudo-doublets were tentatively assigned to anomeric protons. From the chemical shifts and coupling constant values, four of these resonances, δ 5.177 ($J_{1,2}$ 3.6 Hz), δ 5.053 ($J_{1,2}$ 3.5 Hz), δ 4.984 ($J_{1,2}$ 2.4 Hz), and δ 4.963 ($J_{1,2}$ 3.8 Hz) were found to correspond to H1 of α-linked glycosyl residues. The five other resonances δ 4.447 ($J_{1,2}$ 7.8 Hz), δ 4.416 ($J_{1,2}$ 7.7 Hz), δ 4.338 ($J_{1,2}$ 7.1 Hz), δ 4.314 ($J_{1,2}$ 7.2 Hz), and δ 4.255 ($J_{1,2}$ 7.4 Hz) were attributed to H1 of the β-linked glycosyl residues. In agreement with the nine anomeric proton resonances, the ^{13}C NMR spectrum, recorded in the *J*-modulated mode, shows nine anomeric carbon resonances at δ 92.99, δ 95.32, δ 101.36, δ 103.87, δ 104.15, δ 104.42, δ 104.69, δ 105.44, and δ 106.44. The nonanomeric proton resonances at δ 4.784 and δ 4.732, which appear as a triplet and a doublet of doublets, respectively, were assigned to gem acyl protons in agreement with fatty acyl residues binding the carbohydrate part. Indeed, in the HMQC spectrum, these protons show connectivities with nonanomeric carbon resonances at 72.52 and 74.61 ppm, while the anomeric proton resonances are found to correlate with anomeric carbon resonances (92–107 ppm) (Fig. 4). Based on the ^{13}C chemical shift literature data and proton integration, some anomeric resonances can be assigned. The upfield resonances at δ 92.99 and 95.32 typify a trehalose unit (δ C1 94.0 (27)). Thus, from the HMQC experiment (Fig. 4), the proton resonances at δ 5.177 and 5.053 can be assigned to the anomeric protons of the trehalose unit. Moreover, the proton anomeric resonance at 4.255 ppm which integrates for four protons can be assigned to β-Xyl according to the GC/MS data. This resonance cor-

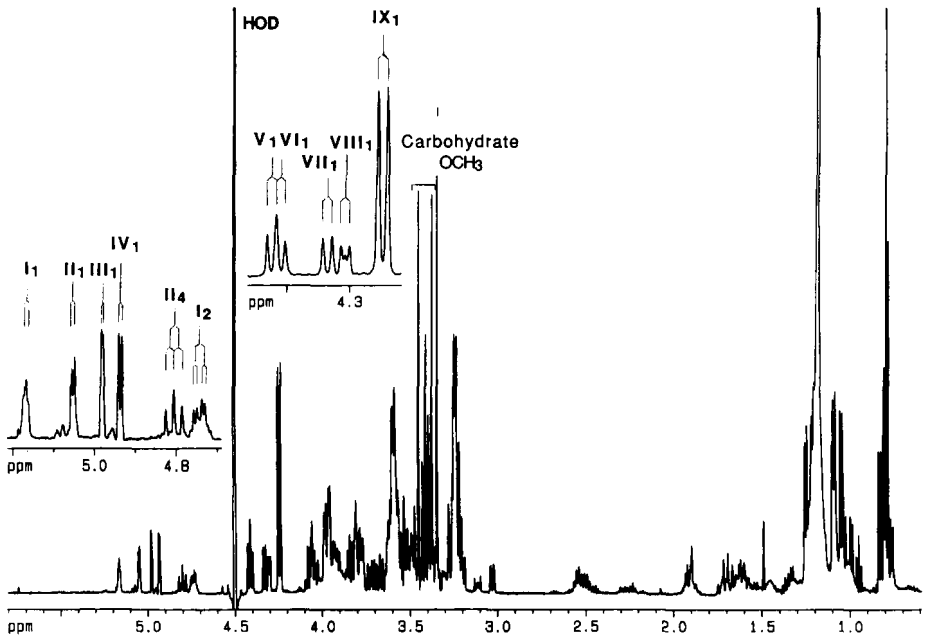

FIG. 3. **500-MHz ^{1}H NMR spectrum of LOS-III in $CDCl_3/CD_3OD$ (1:1, v/v).**

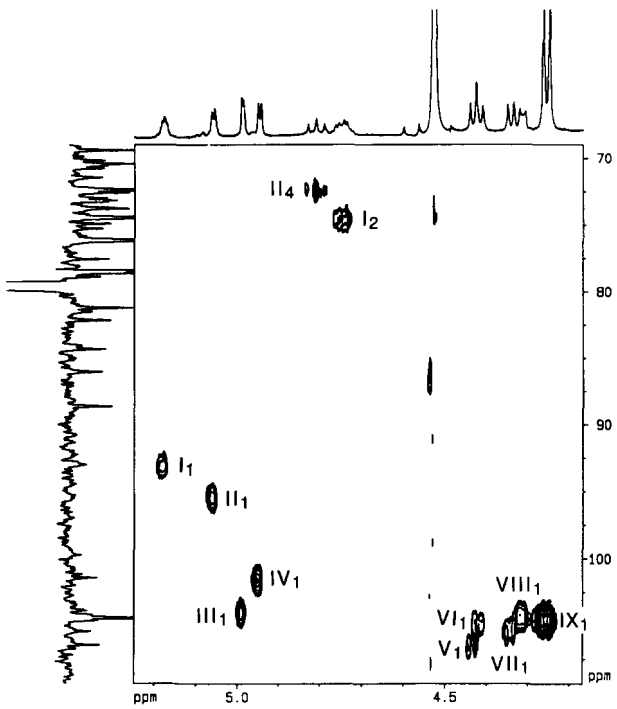

FIG. 4. **Anomeric zone of the phase-sensitive, ^{13}C-decoupled, ^{1}H-detected multiple quantum correlation spectrum (^{1}H -^{13}C HMQC) of native LOS-III (10 mg) in $CDCl_3/CD_3OD$ (1:1, v/v) at 500 MHz.** The data matrix was 4000 × 256 (TPPI) complex points with 64 scans per t_1 value. The spectral window was 27 700 Hz in the F_1 dimension (^{13}C) and 2900 Hz in the F_2 dimension (^{1}H). Relaxation delay was 1.5 s, and delay after BIRD pulse was 380 ms. For processing, a sine bell window shifted by $\pi/2$ was applied in both dimensions, followed by expansion of the data matrix to a 4000 × 512 real matrix. The resulting digital resolutions were 0.35 Hz/point in F_2 (^{1}H) and 13.53 Hz/point in F_1 (^{13}C). Peaks are labeled with a roman numeral identifying the residue followed by the number assigning the carbon atom.

relates, in the HMQC spectrum, with the most intense signal at δ 104.42. Except for the resonances at δ 104.42, each of the other anomeric resonances integrate for one proton suggesting that the carbohydrate part of LOS-III is composed of 12 monosaccharide residues, which is not in agreement with the GC assay indicating the presence of only 11 monosaccharides. Thus, the structure of the extra monosaccharide unit was investigated through the following 2D NMR experiments.

The monosaccharidic units were identified after complete assignment of all their respective proton and carbon resonances in the NMR spectra. Starting from the well separated proton anomeric resonances, the spin systems of each monosaccharidic unit were identified by tracing cross-peak connectivities on the COSY contour map. Coupling constant values (compiled in Table I), determined either from the 1D spectrum or from the cross-peaks in the phase-sensitive DQF-COSY spectrum (Fig. 5), allowed the determination of the anomeric configuration as well as the relative stereochemical configuration of the ring protons. Complete carbohydrate carbon resonance attribution was accomplished using the two following heteronuclear sequences: HMQC and ^{1}H detected heteronuclear relayed spectroscopy HMQC-HOHAHA. This last sequence permitted the detection of connectivity between not directly coupled heteronuclei belonging to the same spin system (24). So, from a well resolved proton like the anomeric proton, the ^{13}C resonance chemical shifts were determined for each unit. In this way, the ^{13}C resonances were assigned from the HMQC spectrum and from literature data of the ^{13}C chemical shifts (28, 29). The analysis of these heteronuclear spectra not only allowed the confirmation of the proton chemical shifts but was also required for the determination of the sequence by HMBC spectrum analysis.

The pyranose ring size, for all the monosaccharide units, was deduced from different but convergent observations: (i) the anomeric carbon chemical shift agreed with a pyranose form (27), (ii) the $J_{1,2}$ coupling constants higher than 7 Hz unambiguously typify pyranose rings in which H1 and H2 are both diaxial (units V, VI, VII, VIII, IX) (30), (iii) an intraresidue correlation was seen to occur between H1 and C5 on the HMBC spectrum for units I, II, III, and IV (31).

To simplify the presentation, the labeling of the sugar residues in roman numerals and their proton in arabic numerals is used here for the entire assignment procedure.

The downfield anomeric resonance I_1 at δ 5.177 was assigned to the α-Glc*p* of the trehalose unit (C1 δ 92.99). It is

TABLE I

^1H NMR chemical shifts data for LOS-III

Shift measured at 303 K in $CDCl_3/CD_3OD$ (1:1, v/v) using tetramethylsilane as internal reference.

Sugar unit	H-1 $^3J_{H1-H2}$	H-2 $^3J_{H2-H3}$	H03 $^3J_{H3-H4}$	H-4 $^3J_{H4-H5}$	H-5	H-6	Others
3,6-Dideoxy-α-hex*p* (IV) 1 → 3	4.963	3.934	1.916/1.706		4.111	1.043	
	3.8	*4.9/12.2*					
β-Xyl*p* (VIII) 1 → 4	4.314	3.391	3.391	3.564	3.203/3.900		
	7.2	*~8*	*?*	*?*			
[β-Xyl*p*]$_4$ (IX) 1 → 4	4.255	3.210	3.397	3.594	3.256/3.973		
	7.4	*~9*	*~9*	*11.5*			
β-Xyl*p* (VII) 1 → 4	4.338	3.217	3.404	3.573	3.181/3.950		
	7.1	*~8*	*?*	*?*			
3-*O*-Me-α-Rha*p* (III) 1 → 3	4.984	4.071	3.542	3.573	3.813	1.253	3.385 (OMe)
	2.4	*2.8*	*9.4*	*7.5*			
β-Gal*p* (V) 1 → 3	4.447	3.669	3.495	3.819	?	?	
	7.8	*10*	*3.0*	*~1*			
β-Glc*p* (VI) 1 → 4	4.416	3.347	3.491	3.593	3.310	3.819	
	7.7	*9.4*	*8.0*	*?*			
α-Glc*p* (I) 1 ↔ 1	5.177	4.732	4.052	3.612	3.913	3.786	
	3.6	*10.3*	*9.4*	*8.5*			
α-Glc*p* (II)	5.053	3.507	3.794	4.784	3.863	3.998	
	3.5	*10.4*	*9.8*	*9.8*			

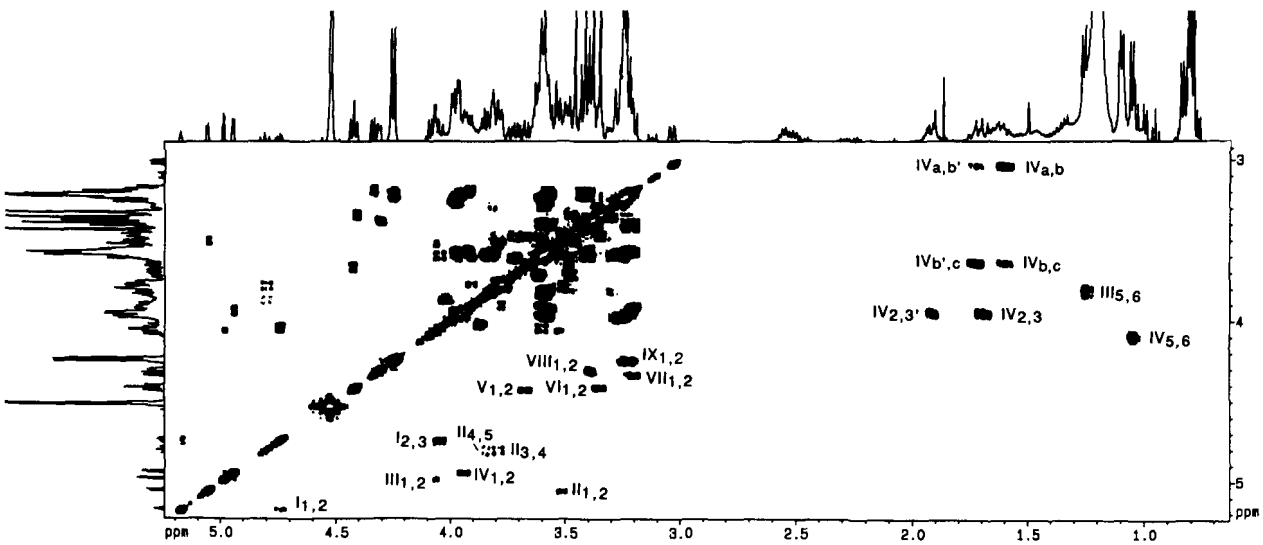

FIG. 5. **Partial 2D (^1H-^1H) phase-sensitive COSY spectrum (δ 0.60–5.25/δ 2.85–5.25) of native LOS-III (10 mg) in $CDCl_3/CD_3OD$ (1:1, v/v) at 500 MHz.** The spectral width was 2800 Hz, and the relaxation delay 1.5 s; the data matrix was 4000 × 512 (TPPI) points, with 64 scans per t_1 value. Sine bell apodizations with $\pi/2$ 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 2000 × 2000 real points.

coherent with the cross-peak connectivities allowing proton assignments up to H6 (δ H1 = 5.177, δ H2 = 4.732, δ H3 = 4.052, δ H4 = 3.612, δ H5 = 3.913, δ H6 = 3.786) (Fig. 5). Moreover, the coupling constant value $J_{1,2}$ = 3.6 Hz and the axial-axial coupling $J_{3,4}$ = 9.4 Hz indicate the α-glucopyranosyl configuration. As expected the nonanomeric proton which resonates at δ 4.732 was identified as a H2. Its downfield resonance is coherent with the presence of the fatty acyl appendage limited to C2. The spin system of unit I characterizes a 2-*O*-acyl-α-Glc*p*. The carbon resonances of this spin system were assigned. A cross-section taken through the anomeric proton resonance (δ 5.177) in the HMQC-HOHAHA spectrum shows cross-peaks with the three carbons at δ 71.51, δ 74.61, δ 81.63 but also with the anomeric carbon previously assigned (see Fig. 7). Likewise, HMQC-HOHAHA through the H2 which resonates independently at δ 4.732 reveals a correlation with the carbons at 71.51, 74.61, and 81.63 ppm but also with an as yet unidentified carbon at 73.05 ppm (see Fig. 7). The analysis of the HMQC spectrum allows the attribution of these carbons (Table II) δ 74.61 to C2, δ 71.51 to C3, δ 81.63 to C4, δ 73.05 to C5, and the characterization of the C6 at 62.54 ppm. These attributions are supported by the H1,C2 and H1,C3 correlations observed on the HMBC spectrum.

Likewise, the spin system starting with the anomeric resonance II_1 at δ 5.053 ($J_{1,2}$ = 3.5 Hz) assigned to the second α-Glc*p* of the trehalose unit (C1 δ 95.32), shows a downfield resonance of H4 (δ 4.784). This proton resonates independently as a triplet ($J_{3,4} = J_{4,5}$ = 9.8 Hz) supporting the glucosyl structure of the unit. As previously mentioned, the downfield shift of the H4 resonance (Δδ ~1 relative to the expected value for an H4 resonance) agrees with an *O*-acyl group on C4. The remaining protons resonated at normal chemical shifts. All these resonances are coherent with a 4-*O*-acyl-α-Glc*p* (unit II). Since H4 resonates independently, the correlation with the C4 at 72.52 is easily observed by means of the HMQC spectrum. The cross-section taken through the anomeric proton in the HMQC-HOHAHA spectrum shows con-

TABLE II

^{13}C NMR chemical shifts data for LOS-III

Shifts were measured at 303 K in $CDCl_3/CD_3OD$ (1:1, v/v) using chloroform (δ 77) as internal reference.

Sugar unit	C-1	C-2	C-3	C-4	C-5	C-6	Others
3,6-Dideoxy-α-hex*p* (IV) 1 → 3	101.36	67.44	34.01	78.37	69.47	14.95	
	99.9[a]	*65.6*[a]	*31.0*[a]	*77.4*[a]	*68.1*[a]	*13.8*[a]	
β-Xyl*p* (VIII) 1 → 4	104.15	73.95	85.71	70.67	67.39		
[β-Xyl*p*]$_4$ (IX) 1 → 4	104.42	74.61	76.25	78.63	65.47		
β-Xyl*p* (VII) 1 → 4	105.44	75.41	76.37	78.91	65.14		
	105.1[b]	*74.0*[b]	*76.9*[b]	*70.4*[b]	*66.3*[b]		
3-O-Me-α-Rha*p* (III) 1 → 3	103.87	69.37	81.33	81.33	70.19	19.38	58.75 (OMe)
	101.9[b]	*71.0*[b]	*71.3*[b]	*73.1*[b]	*69.4*[b]	*17.7*[b]	
β-Gal*p* (V) 1 → 3	106.44	73.21	82.21	70.67	?	63.27	
	104.5[b]	*71.7*[b]	*73.8*[b]	*69.7*[b]	*76.0*[b]	*62.0*[b]	
β-Glc*p* (VI) 1 → 4	104.69	74.61	88.68	70.80	78.49	63.27	
	104.0[b]	*74.1*[b]	*76.8*[b]	*70.6*[b]	*76.8*[b]	*61.8*[b]	
α-Glc*p* (I) 1 ↔ 1	92.99	74.61	71.51	81.63	73.05	62.54	
α-Glc*p* (II)	95.32	73.31	70.48	72.52	72.80	64.37	
	100.0[b]	*72.2*[b]	*74.1*[b]	*70.6*[b]	*72.5*[b]	*61.6*[b]	

[a] Methyl glycoside chemical shifts in italics are taken from Ref. 33.
[b] Methyl glycoside chemical shifts in italics are taken from Ref. 29.

nectivity with the carbon at 73.31 ppm (see Fig. 7). Likewise, H4 exhibits a correlation with another carbon at 64.37 ppm assigned to the C6 from its chemical shift (29). The remaining carbon resonances were assigned from the HMBC spectrum analysis. Indeed, a cross-section through H4 leads to the characterization of two supplementary carbons at 70.48 and 72.80 ppm. Again, the HMQC spectrum allows the unambiguous identification of C2 (δ 73.31) and C5 (δ 72.80), while C3 was attributed by deduction to the resonance at 70.48 ppm.

The location of the fatty acyl groups on the trehalose was supported by the HMBC spectrum analysis which revealed long-range J_{CH} coupling of the degenerate carbonyl carbons of the fatty acids (δ 178.4) with the proton resonances I_2 and II_4 of the trehalose unit. Moreover, the carbonyl group also correlates with II_6, revealing that the unit is also acylated on C6. So, we can conclude that the trehalose unit is composed of 2-*O*-acyl-α-Glc*p* and 4,6-di-*O*-acyl-α-Glc*p*.

Starting with the next anomeric proton III_1 at δ 4.984, assigned to α-Rha*p* (C1 δ 103.87) from the weak $J_{1,2}$ value of 2.4 Hz, resonances up to the upfield methyl proton resonance at δ 1.253 were assigned in the COSY spectrum. Assignment of the proton resonances of residue III was confirmed from 2D HOHAHA experimental data (Fig. 6). Although a cross-section taken through the anomeric proton resonance in the 2D HOHAHA spectrum shows cross-peaks only up to H4 (δ 3.573), all the connectivities up to the H1 (δ 4.984) resonance were seen in the cross-section through the upfield methyl doublet (δ 1.253) (not shown). The weaker $J_{1,2}$ coupling constant (2.4 Hz) indicates that H2 is in equatorial arrangement, while the stronger $J_{3,4}$ (9.4 Hz) and $J_{4,5}$ (7.5 Hz) coupling values obtained as active coupling constants in the H3/H4 and H4/H5 cross-peaks, indicate that H5 and H3 are trans-diaxial toward H4. This spin system resonance is coherent with the 3-*O*-Me-α-Rha*p* structure. From the HMQC-HOHAHA spectrum, five carbon atoms of spin system III were found to resonate at δ 103.87 and 70.19 (Fig. 7) and at δ 81.33, 69.37, and 19.38 (not shown). According to their chemical shifts (Table II), two resonances were easily assigned: δ 103.87 to C1 and δ 19.38 to C6. The HMQC spectrum supports these attributions and allowed the assignment of C2 (69.37 ppm) and C5 (70.19 ppm). Then the remaining resonance at 81.33 ppm is attributed to both C3 and C4. This assignment is supported by the HMBC spectrum analysis since the δ 81.33 resonance correlates with both H6 and H1. It also correlates with the methoxy proton at 3.385 ppm supporting the methoxy location on C3 previously established by GC/MS. Likewise, the connectivity between the H3 at 3.542 ppm and the methoxy carbon at 58.75 ppm supports the previous assignment.

Anomeric resonances V_1 and VI_1 at δ 4.447 and 4.416, respectively, show a similar H1/H2 cross-peak with partial cancellation of the central components indicating strong, equal $J_{1,2}$ and $J_{2,3}$ coupling constants. In both cases, by means of the COSY spectrum, unambiguous assignment was possible only up to the H3 proton, but, from the HOHAHA spectrum (Fig. 6), assignment was determined up to H4 and to H6 for systems V and VI, respectively. Thus, for system V, the $J_{3,4}$ and $J_{4,5}$ coupling constants measured from the (H3,H4) cross-peak on the DQF-COSY spectrum indicate a small $J_{4,5}$ coupling constant of about 1 Hz in agreement with the absence of connectivity beyond H4 from the COSY spectrum. These observations indicate a β-Gal*p* unit, which is supported by the analysis of the HMQC-HOHAHA spectrum (Fig. 7). A cross-section taken through the anomeric proton resonance of system V shows connectivities with two carbon atoms at δ 106.44 and δ 82.21 which were, respectively, attributed by HMQC to C1 and C3. From the cross-section taken through the C3 resonance, we characterized two other correlations with H2 and H4. The cross-peak fine structure which involves the H4 reveals only one coupling constant around 3 Hz, in agreement with the H4 multiplicity of a Gal*p* residue (unit V). A second HMQC-HOHAHA experiment was run with a different mixing time (τ_m 90 ms). The cross-section taken through the anomeric resonance shows, besides the connectivities previously described from the HMQC-HOHAHA spectrum recorded with a τ_m of 65 ms, correlations with two other carbons at 73.21 ppm and 70.67 ppm respectively assigned by the HMQC spectrum to C2 and to C4. Since the Gal*p* spin system was unambiguously identified, it was deduced that spin system VI corresponds to a β-Glc*p*. The coupling constants were measured ($J_{1,2}$ 7.7 Hz and $J_{3,4}$ 8.0 Hz) from the DQF-COSY spectrum (Fig. 5) and indicate a β-glucopyranosyl configuration. From the HMQC-HOHAHA spectrum, the H1 exhibits connectivities with four carbons at δ 104.69, attributed to C1, δ 88.68, δ 74.61, and δ 70.80. The remaining carbon resonances were characterized from the H2 and H4 at δ 78.49 and δ 63.27. Four carbons were assigned from the HMQC experiment: C2 at δ 74.61, C3 at δ 88.68, C5 at δ 78.49, C6 at δ 63.27, while the C4 was assigned to the remaining resonance at δ 70.80.

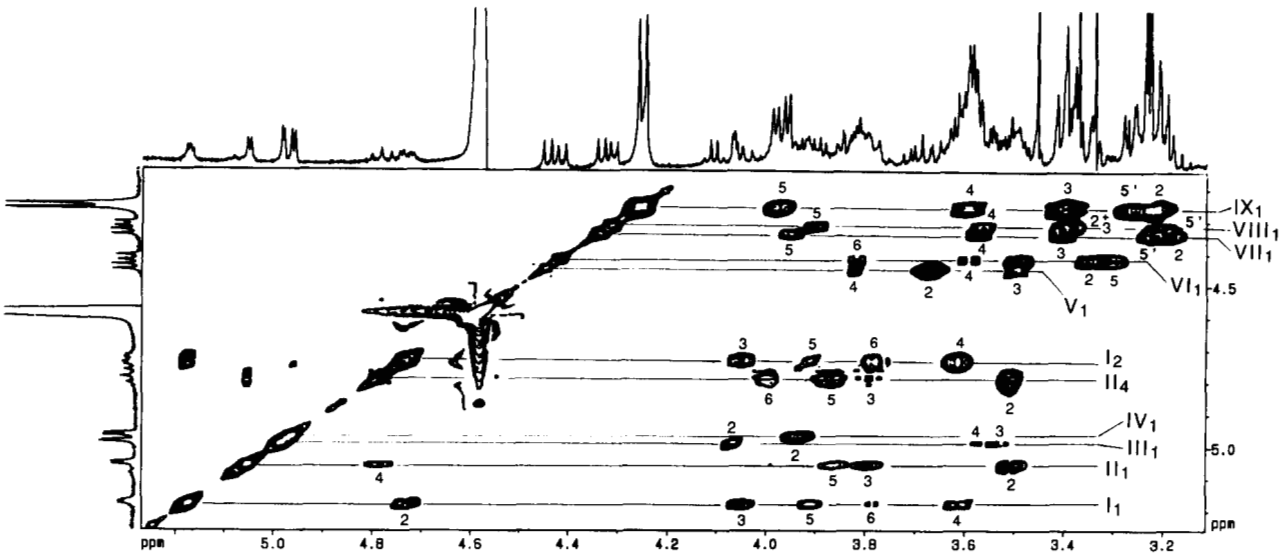

FIG. 6. **Partial 2D HOHAHA spectrum (δ 3.10–5.28/δ 4.15–5.25) of native LOS-III (10 mg) in $CDCl_3/CD_3OD$ (1:1, v/v) at 500 MHz.** The spectral width was 2800 Hz, the relaxation delay 1.7 s, and the spin lock time was 125 ms; the data matrix was 4000 × 512 (States Haberkorn and Ruben) points, with 16 scans per t_1 value. For processing, a sine bell window shifted by $\pi/2$ was applied in both dimensions, followed by expansion of the data matrix to 2000 × 2000 data points and transformation. The resulting digital resolution in F_2 was 1.36 Hz/pt.

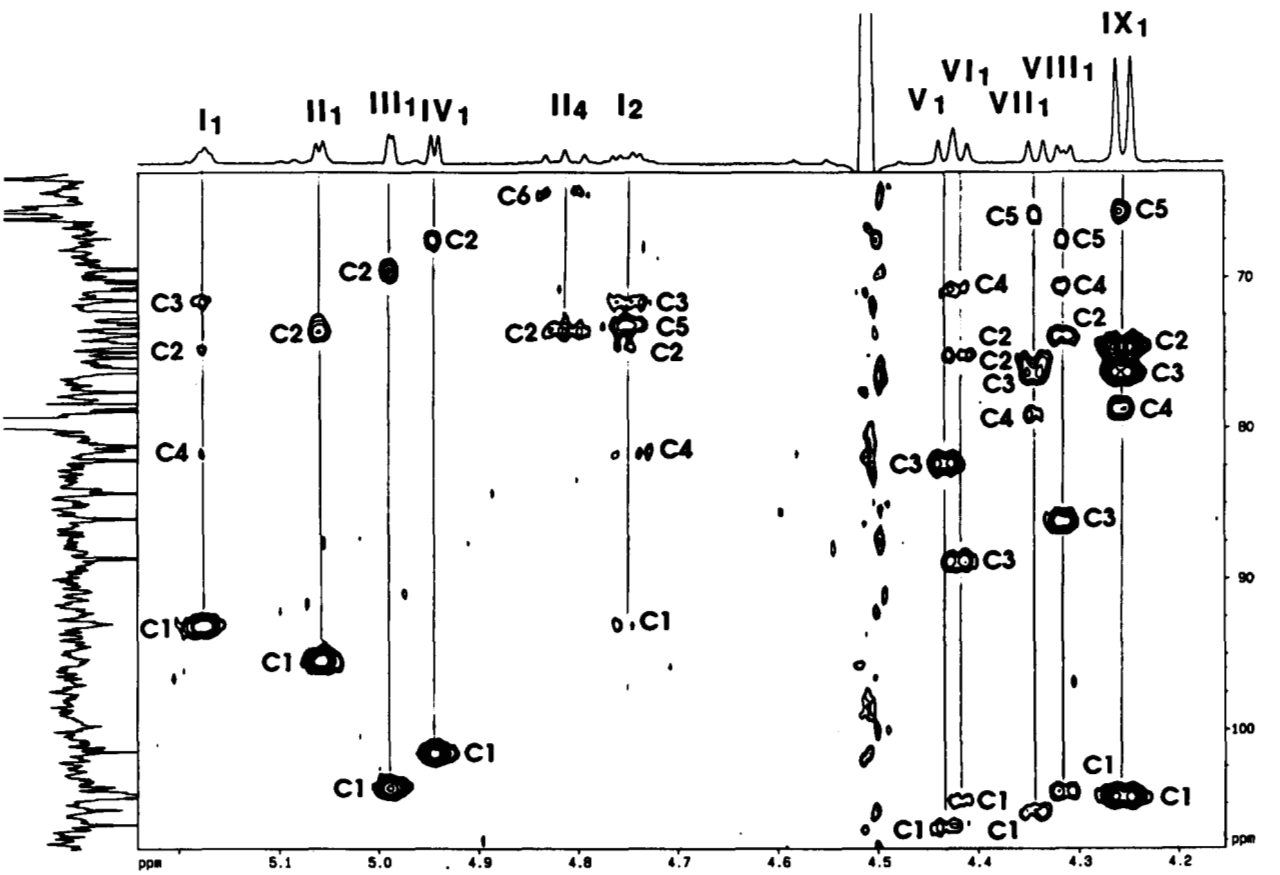

FIG. 7. **Partial 2D HMQC-HOHAHA spectrum (δ 4.12–5.25/δ 63–108) of native LOS-III (10 mg) in $CDCl_3/CD_3OD$ (1:1, v/v) at 500 MHz.** The data matrix was 2000 × 256 (TPPI) points with 128 scans per t_1 value. The spectral window was 16,700 Hz in the F_1 dimension (^{13}C) and 2,500 Hz in the F_2 dimension (^1H). Relaxation delay was 1.5 s, and the mixing time was 67 ms. For processing, a sine bell window shifted by $\pi/2$ was applied in both dimensions, followed by expansion of the data matrix to a 2,000 × 512 real matrix. The resulting digital resolutions were 0.61 Hz/point in F_2 (^1H) and 8.15 Hz/point in F_1 (^{13}C). The data are presented in the magnitude mode.

At this point, all the expected units identified by GC/MS have been defined except for the xyloses which therefore correspond to units VII, VIII, and IX. The system IX was attributed to four β-Xyl*p*s. Its anomeric proton (δ 4.255) integrates for four protons, and the system which was easily followed due to the intensities of the cross-peaks showed H5

and H5′ protons. The anomeric protons VII_1 (δ 4.338) and $VIII_1$ (δ 4.314) correlate with H2 at 3.217 and 3.391 ppm, respectively (Fig. 5). The remaining protons, unassignable from the 2D COSY spectrum due to an overcrowded area, were characterized from the 2D HOHAHA spectrum (Fig. 6) and attributed from relayed COSY experiments (not shown). A cross-section through the anomeric proton resonance of units VII, VIII, and IX in the 2D HOHAHA spectrum (Fig. 6) shows cross-peaks for all the resonances up to H5, and reveals that the protons of system VII resonate closely to those of system IX. The attribution of H5 was confirmed for both units by the TQF-COSY spectrum analysis which only showed cross-peaks for three or more mutually coupled spins (32). Thus, in the case of xylopyranosyl residues, we observed the three expected cross-peaks (H4,H5, H4,H5′, and H5,H5′) (data not shown). For the three spin systems VII, VIII, and IX, correlations in the HMQC spectrum between the H5 and the methylene carbons resonating around 66 ppm unambiguously characterize pentose residues. This assignment is also supported by the H1/C5 cross-peaks (C5: VII, 65.14; VIII, 67.39; IX, 65.47) observed in the HMQC-HOHAHA spectrum (Fig. 7) since similar cross-peaks were missing for hexose units I, II, V, and VI. The carbon resonances for the three spin systems VII, VIII, and IX were characterized as follows. From the HMQC-HOHAHA spectrum, each anomeric proton (VII_1, 4.338 ppm; $VIII_1$, 4.314 ppm; IX_1, 4.255 ppm) shows correlations with five carbons. The chemical shifts of the carbons belonging to system IX were easily determined since the four xylopyranoses appear as a unique spin system. The ^{13}C resonances are characterized by an intense signal at δ 104.42, δ 78.63, δ 76.25, δ 74.61, and δ 65.47 and were assigned to C1, C4, C3, C2, and C5 by means of the HMQC spectrum. The carbon resonances of system VII were assigned by analogy to the chemical shifts of the carbons belonging to unit IX. Indeed, as previously indicated, the protons of spin system IX resonate closely to those of system VII (Table I), thus hindering the cross-peaks of unit VII in the HMQC spectrum. However, concerning system VIII, which shows a different network to that of the system IX, the carbon resonances were routinely attributed from the HMQC spectrum.

The spin systems of the 11 monosaccharide units identified by GC/MS are now established. The remaining spin system IV (H1 δ 4.963) assigned to the extra monosaccharide unit unrevealed by the GC/MS experiment is investigated below.

Partial Structure of the 3,6-Dideoxyhexose (Unit IV)—The anomeric proton IV_1 resonates independently as a doublet at δ 4.963 with a coupling constant $J_{1,2}$ of 3.8 Hz. This proton shows, from the COSY spectrum, resonances only up to H3 (Fig. 5). Similar behavior was observed in two- and three-step-relayed COSY as well as in the HOHAHA spectrum starting either at the H1 or H2 proton resonances. None of these experiments managed to reveal connectivity beyond the H3 resonances due to either a weaker coupling constant between H3 and H4 or to the absence of a proton attached to C4. The interesting feature of this spin system is the presence of a deoxy group characterized by two H3 vicinal protons located at δ 1.706 and δ 1.916. This methylene unit is supported by the coupling constants and their sign deduced from the cross-peak fine structure in DQF-COSY ($J_{2,3e}$, 4.9 Hz; $J_{2,3a}$, 12.2 Hz; and $J_{3,3'}$, −12 Hz) but also from the correlation between the two H3 protons and the methylene ^{13}C resonance at 34.01 ppm in the HMQC spectrum. In the COSY spectrum, the remaining resonance at δ 1.043 is connected to the resonance at δ 4.111. Based on the chemical shifts, this partial spin system could characterize the methyl group and the vicinal H5 of a 6-deoxyhexose. Taken together, these two spin systems characterize a 3,6-dideoxy-α-hexose with probably no proton linked to C4. The HMQC-HOHAHA spectrum shows connectivities between the H1, H2, H3, and H3′ protons and the following carbon resonances at δ 101.36, assigned to the anomeric carbon, δ 34.01, previously attributed to the C3 methylene and δ 67.44, assigned to C2 (Fig. 7). Likewise, H5 and H6 show correlations with C6 (14.95 ppm) and C5 (69.47 ppm), respectively. This analysis corroborates the absence of H4 and so does not allow the resonance assignment of C4 which was investigated from the HMBC spectrum. The cross-peaks between H3, H5, H6, and an as yet unassigned carbon atom resonance at 78.37 ppm lead to the characterization of C4, the quaternary structure of which is supported by the *J*-modulated 1D ^{13}C spectrum. Moreover, the connectivities between H1/C5, H3/C5, and H5/C1 unambiguously demonstrate that the two spin systems are connected supporting the structure 3,6-dideoxy-α-hex*p* devoid of hydrogen on C4. From the C4 resonance, the spin system of the C4-linked groups was investigated.

A cross-section through the C4 carbon (78.37 ppm) in the HMBC spectrum shows a cross-peak with an unidentified proton resonance at δ 3.034. This proton, named Ha, resonates independently as a doublet of doublets ($J \sim 2$ and 9.5 Hz) in the 1D ^1H spectrum. From the COSY spectrum (Fig. 5), this Ha proton correlates with two vicinal protons Hb and Hb′ at δ 1.735 and δ 1.596. Both protons Hb and Hb′ correlate with Hc at 3.637 ppm. In the HOHAHA spectrum, a cross-peak is observed between Ha and Hc. Unfortunately, due to overcrowding, analysis of the COSY as well as of the HOHAHA spectra did not allow progress to be made in the structural elucidation of the aliphatic chain. The heteronuclear 2D spectra led to the assignment of the different carbons of the C4 substituent (Ca, 84.50 ppm; Cb, 33.14 ppm; Cc, 79.42 ppm). The Cb resonance indicates a methylene unit. Moreover, the HMBC spectrum shows correlation between carbons Ca and Cc and the methoxy proton resonances at 3.461 and 3.355 ppm, respectively, indicating the presence on Ca and Cc of methoxy substituents which is consistent with the chemical shifts of the two carbons (34). So from these data, the partial structure 3,6-dideoxy-4-(2′,3′-dimethoxypropyl)-α-hex*p* can be tentatively proposed for unit IV in which the structure of *R* is still unknown (Fig. 8).

Glycosyl Linkage and Sequence Analysis of LOS-III—The sequence of the glycosyl residues and the positions of the carbons involved in the interglycosidic linkage were established by HMBC experiments from scalar couplings $^3J_{CH}$ between the ^{13}C and ^1H involved in the glycosidic linkage (35). As the resonance of a carbon atom which is either involved in an interglycosidic linkage or linked to a methoxy group is shifted downfield by 4 to 10 ppm, the glycosylated carbons are easily characterized from their chemical shifts. So, from the literature data (29, 33) summarized in Table II, it can be advanced that C3, for units V, VI, and VIII, and C4, for units I, III, VII, and IX, are involved in the glycosidic linkage.

Fig. 9 shows the partial HMBC spectrum used for the sequencing. It can be observed from this spectrum, that the H1 of β-units only correlates with the ^{13}C involved in the

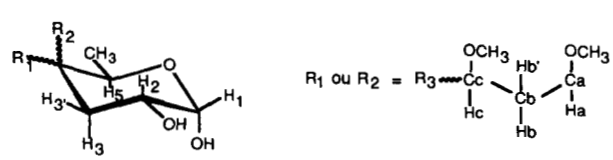

FIG. 8. **Partial structure of monosaccharide (IV) (R_3 = undetermined data).**

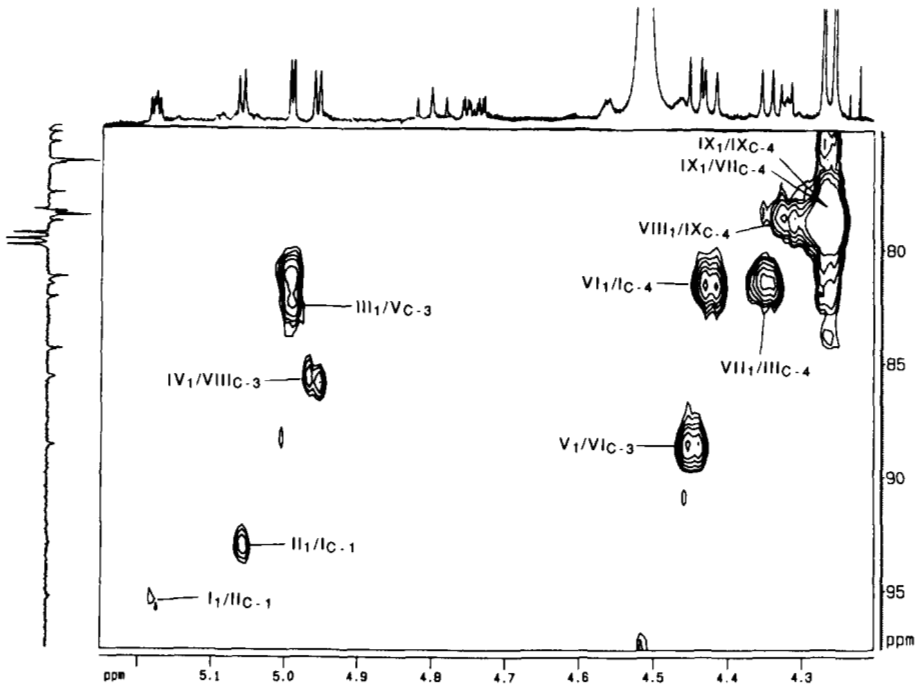

FIG. 9. **Expanded zone (δ 4.20–5.25/δ 75–98) of the ^{13}C decoupled, ^{1}H-detected multiple bond correlation spectrum (^{1}H [^{13}C] HMBC) of native LOS-III (10 mg) in $CDCl_3$/CD_3OD (1:1, v/v) at 500 MHz.** The data matrix was 2,000 × 256 complex points with 144 scans per t_1 value. The spectral window was 27,700 Hz in the F_1 dimension (^{13}C) and 2,500 Hz in the F_2 dimension (^{1}H). Relaxation delay was 0.8 s. For processing, a sine bell window shifted by $\pi/2$ was applied in both dimensions, followed by expansion of the data matrix to a 2,000 × 512 real matrix. The resulting digital resolutions were 0.61 Hz/point in F_2 (^{1}H) and 13.53 Hz/point in F_1 (^{13}C). The data are presented in the magnitude mode.

glycosidic linkage, whereas the H1 of the α-residues also exhibits intraresidue long range correlations to C3 ($^{3}J_{CH}$) and in some cases to C2 ($^{2}J_{CH}$) and C5 ($^{3}J_{CH}$).

The trehalose is first identified without fail. Indeed, 2-*O*-acyl-α-Glc*p* (unit I) H1 (δ 5.177) shows, beside long range intraresidue correlation to C3 and C2 resonances, an interresidue connectivity to C-1 (95.32 ppm) of 4,6-di-*O*-acyl-α-Glc*p* (unit II). Likewise, H1 of unit II (δ 5.053) shows long range correlation to C2, C3, and C5 resonances but also to C1 of unit I at 92.99 ppm. The two cross-peaks between α-Glc*p* (unit II) H1 and α-Glc*p* (unit I) C1 and the second between α-Glc*p* (unit I) H1 and α-Glc*p* (unit II) C1 show connectivity across the glycosidic linkage characterizing the α-α trehalose unit which is located at one extremity. Precise analysis of the correlation peaks shows that the C4 of unit I at 81.63 correlates with the VI_1 (4.416 ppm), establishing the sequence β-Glc*p*(1→4)2-*O*-Acyl-α-Glc*p*. The C4 chemical shift is in agreement with glycosylation at position 4, since the literature (28) gives a value of 79.9 ppm for the C4 involved in the interglycosidic linkage of the disaccharide β-Glc*p*(1→4)α-Glc*p*. The C3 (88.68 ppm) of the β-Glc*p* (unit VI) shows connectivity across the glycosidic linkage to the proton resonance of the Gal*p* (unit V) (4.447 ppm) allowing the determination of the sequence β-Gal*p*(1→3)β-Glc*p*(1→4)2-*O*-Acyl-α-Glc*p*. Again, the C3 chemical shift at 88.68 ppm is in agreement with literature data indicating that the C3 of the Glc*p* from the disaccharide β-Glc(1→3)β-Glc resonates at 86.0 ppm (28). The C3 of β-Gal*p* (unit V) (82.21 ppm) shows then interresidue connectivity to the anomeric proton resonance of the 3-*O*-Me-α-Rha*p* (unit III) at δ 4.984 indicating the sequence 3-*O*-Me-α-Rha*p*(1→3)β-Gal*p*, in agreement with the C3 chemical shift of a β-Gal*p* from the disaccharide α-Rha(1→3)β-Gal which is found at 81.8 ppm (28). As shown in Table II, the most deshielded carbon of the 3-*O*-Me-Rha*p* (unit III) compared to literature data resonates at 81.33 ppm and corresponds to both C3 and C4. Since C3 is methoxylated, the carbon involved in the interglycosidic linkage is C4. A long range inter-residue connectivity between this carbon and the VII_1 resonance (4.338 ppm) allows the determination of the linkage β-Xyl*p*(1→4)3-*O*-Me-α-Rha*p*. The chemical shift is coherent with the literature data (28) which indicate that the C4 of a Rha*p* unit resonates at 81.5 ppm in the trisaccharide α-Rha(1→3)β-Xyl(1→4)α-Rha(1→2). The partial sequence established at this point is VII(1→4)III(1→3)V(1→3)VI(1→4)I(1↔1)II.

The C4 of unit VII (δ 78.91) correlates with the anomeric protons of both units IX (δ 4.255) and VIII (δ 4.314) preventing the sequencing process. The sequence was therefore investigated from the anomeric proton of unit IV. Indeed, three correlation peaks from the H1 resonance of the C4 branched 3,6-dideoxy-α-hex*p* (unit IV) (δ 4.963) were observed in the HMBC spectrum, one of which is to the C3 (85.71 ppm) of the β-Xyl*p* (unit VIII) revealing the disaccharide sequence 3,6-dideoxy-α-hex*p*(1→3)β-Xyl*p*. The other two ^{13}C resonances were identified as intraresidue connectivities to C3 (34.01 ppm) and C5 (69.47 ppm). The remaining system IX assigned to 4 β-Xyl*p*s shows an intense inter-residue (H1,C4) cross-peak indicating a unique 1→4 linkage between the four Xyl*p*s composing system IX. Moreover, the anomeric proton of system VIII (δ 4.314) show inter-residue connectivity with the C4 of system IX (δ 78.63) or VII (δ 78.91) thus indicating the sequence 3,6-dideoxy-α-hex*p*(1→3)β-Xyl*p*(1→4)[β-Xyl*p*(1→4)]$_4\beta$-Xyl*p*. The chemical shifts of the Xyl*p* carbon atoms involved in the interglycosidic linkages are consistent with the literature data (28): a xylopyranose unit glycosylated in position 3 resonates at 84.9 ppm in the trisaccharide β-Xyl(1→3)β-Xyl(1→4)β-XylOMe (the C3 of unit VIII resonates at δ 85.71). Moreover, the C4 of a 1→4-linked xylane resonates at 77.7 ppm against δ 78.63 for C4 of unit IX and δ 78.91 for C4 of unit VII. The established sequence as well as the carbons involved in the interglycosidic linkages are in full agreement with both L-SIMS MS and methylation data.[2] Thus, the following LOS-III structure can be proposed in which *X* corresponds to the C4-branched 3,6-dideoxy-α-hex*p*, including the absolute configuration and ring size of each monosaccharidic unit (Fig. 10) as follows: *X*pα(1→3)[L-Xyl*p*β(1→4)]$_6$3-*O*-Me-Rha*p*α(1→3)D-Gal*p*β(1→3)D-Glc*p*β-

[2] M. Gilleron, J. Vercauteren, and G. Puzo, unpublished results.

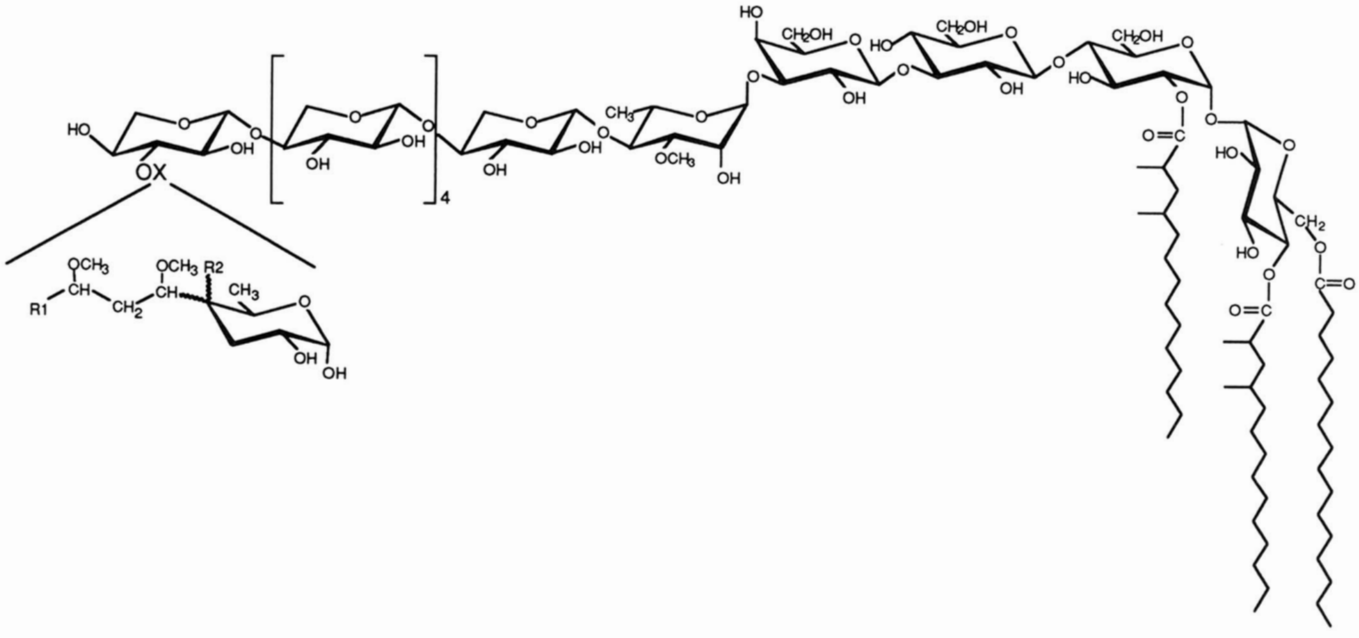

FIG. 10. ***M. gastri* LOS-III structure (R_1 and R_2 = undetermined data).**

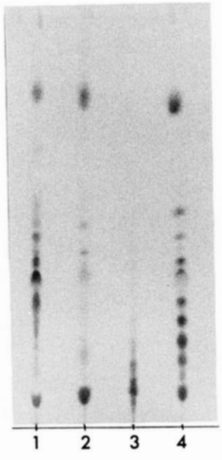

FIG. 11. **Thin-layer chromatography in $CHCl_3/CH_3OH/H_2O$ (60:35:8, v/v), using the orcinol-H_2SO_4 spray of QMA fractions eluted by $CHCl_3/CH_3OH$ (2:1, v/v).** *1, M. kansasii,* strain ATCC 12478 (smooth morphology); *2, M. kansasii,* strain S890175 (rough morphology); *3, M. gastri,* strain W471 (rough morphology); *4, M. gastri,* strains HB4362 (rough morphology) and HB4389 (smooth morphology).

(1→4)2-*O*-acyl-D-Glc*p*α(1↔1)α4,6-di-*O*-acyl-D-Glc*p*.

Occurrence of LOS in Some M. gastri Strains—Beside *M. gastri* W471, two other *M. gastri* strains, HB4362 and HB4389, of rough and smooth morphology, respectively, were analyzed in order to determine the presence of LOSs. Their cell wall glycolipids were extracted, as previously mentioned, and the QMA fractions eluted by $HCCl_3/CH_3OH$ (2:1, v/v) were monitored by silicic acid TLC (Fig. 11). TLC analysis revealed the presence of a set of LOSs with an identical chromatographic behavior in both strains. This set of LOSs is spread over a wide range of R_F values compared to those of *M. kansasii* and *M. gastri* W471 (Fig. 11). Based on the TLC analysis, the presence of the *M. kansasii* and *M. gastri* W471 LOSs can be tentatively proposed in the *M. gastri* HB4362 and HB4389 strains. In order to support this assumption, a rabbit polyclonal serum anti-LOS-k (LOS described in *M. kansasii*) binding the distal disaccharidic epitope kansosamine(1→3)D-Fuc*p* was prepared. By ELISA, it was shown that the antibodies cross-react with the glycolipidic fraction of *M. gastri* strains HB4362 and HB4389 revealing the LOS-k epitope in these strains. As expected, no binding reactions were detected with the LOS-III. Likewise, by the inhibition ELISA technique using the polyclonal anti-*M. gastri* W471 antibodies, it was demonstrated that the *M. gastri* strains HB4362 and HB4389 also contain the LOS-III epitope. These serological data were supported by routine chemical approaches corroborating the presence together of LOS-k and LOS-III in *M. gastri* strains HB4362 and HB4389.

DISCUSSION

Trehalose containing lipooligosaccharides (LOSs) have been described as glycosylated extensions of a sometimes methylated trehalose invariably tri-*O*-acylated at C3, C4, and C6 of its terminal Glc*p* by particular fatty acids, *i.e.* 2,4-dimethyltetradecanoic acids (3). Their amphipathic structure suggests an organization in which the lipid moiety is involved in the mycocerosic layer, while the glycosidic moiety is exposed at the cell wall surface. The LOSs are immunoreactive, and the restricted specificity of their antibodies arises from the unique structure of the terminal immunodominant monosaccharide (11). It is either a recently discovered monosaccharide in nature, like the kansosamine described in the LOSs of *M. kansasii*, or a ubiquitous monosaccharide but with acetyl (10) or methyl (36, 37) residues located on limited carbons.

M. gastri LOS-III is described in this study as being composed of 12 monosaccharides with trehalose tri-*O*-acylated on carbons 2′, 4, and 6 at one end and, at the other, a new type of monosaccharide. The core is 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 its terminal monosaccharide. It is a 3,6-dideoxy-α-hex*p* with a C4-linked 1′,3′-dimethoxy-propyl chain with a structure that has not yet been fully defined.The structure of *M. gastri* LOS-III was established, from its native form, using a new analytical approach in the field of mycobacterial glycolipids, based on a panel of 2D NMR techniques. The utility of NMR spectroscopy in bacterial polysaccharide

structure determination has already been well established (38). However, to our knowledge, it is the first example of the structural elucidation by NMR of an unknown complex lipo-oligosaccharide containing 12 monosaccharide units, one of which is unique in nature, bearing three methoxy and three fatty acyl appendages. The NMR approach required both proton and carbon carbohydrate resonance assignments and the determination of the proton vicinal coupling constants in order to establish the monosaccharidic structures. The data were obtained from the combination of homonuclear (COSY, RCT, HOHAHA) (38) and heteronuclear (HMQC, HMQC-HOHAHA, HMBC) two-dimensional NMR experiments (39). The proton resonance assignment was established from COSY and long-range COSY experiments. Indeed, in most cases, COSY allows attribution only up to H2 since most of the remaining protons of the 12 monosaccharides resonate in a very restricted zone, which is the cause of overcrowding in 2D COSY spectra. Thus, these protons were assigned from long-range correlation experiments (relayed COSY and HOHAHA). From the HOHAHA spectrum (Fig. 6) only the cross-sections taken through the anomeric proton resonances were easily analyzed, and proton assignments were supported by one-, two-, and three-step-relayed COSY experiments. Unfortunately this approach did not allow assignment of protons having identical chemical shifts as, for example, H2 and H5 of xylose units VII and IX. This analytical problem was overcome with TQF-COSY experiments which only show cross-peaks of at least three mutually coupled spins. Thus, the set of three cross-peaks (H4,H5, H4,H5′, H5,H5′) leads to the assignment of H5 and H5′ of each xylose unit once H4 has been found. The combination of these 2D homonuclear NMR techniques allowed carbohydrate proton assignment and consequently carbon assignment from the HMQC and HMQC-HOHAHA spectra.

In order to define all the monosaccharidic units, the proton vicinal coupling constant $^3J_{HH}$ values were determined either from the 1D NMR spectra or from the COSY spectra processed in the phase-sensitive mode. From the cross-peak fine structure, this technique yielded information on the size of the scalar coupling constants leading to the establishment of the ring proton stereochemistry as well as the anomeric configuration.

The NMR approach notably allowed the characterization of the C4 branched 3,6-dideoxy-α-hex*p* which was not revealed using routine carbohydrate analysis procedures based on GC/MS analysis of LOS-III methanolysis products. Due to the absence of a C4 proton, HMBC was the key experiment in the proposal of a partial structure of the C4 alkyl chain by revealing the C4 resonance and thus the Ha alkyl proton resonance. It was then easy to assign the two other carbons Cb and Cc as well as their Hb, Hb′, and Hc proton resonances. Moreover, HMBC cross-peaks between proton methoxy resonances and the Ca and Cc carbon resonances revealed a 1,3-di-*O*-methoxypropyl chain. Likewise, the acyl group positions were defined from the HMBC spectrum starting from the carbonyl resonances. From the downfield resonance of the geminal acyl protons two fatty acyls were revealed located at 2′ (I_2) and 4 (II_4) of the trehalose unit. The HMBC approach supports these data and reveals a third fatty acyl appendage leading to the 4,6-di-*O*-acyl-α-D-Glc*p* structure for the distal monosaccharide unit. These methoxyl and acyl groups, when they are involved in the epitope, must be identified and localized precisely since they play a key role in the recognition process of the antigen by the immunoglobulins (16). In the present work, we show that proton chemical shifts used to reveal the presence of acyl appendages can lead to incomplete structure. The proposed approach using HMBC appears to be more powerful. Moreover, the HMBC spectrum through three-bond coupling correlations across the glycosidic linkage also allowed the sequence and the carbons involved in the interglycosidic linkage to be established. Usually, these structural data are obtained either from 2D NOESY or ROESY spectra but the drawback of this approach is that the magnetic transfer depends on the glycolipid conformation.

The structures of the LOS from other mycobacterial species have been established by routine carbohydrate analysis. More recently, molecular weights and oligosaccharide sequences were determined by either fast atom bombardment-mass spectrometric analysis of the deacetylated LOSs or by plasma desorption mass spectrometry of the peracetylated derivatives (11, 40). However, the presence of either stereoisomers or repeated monosaccharides require the determination of their sequence by partial hydrolysis and HPLC purification followed by MS analysis (41). The LOS-III structure was established from its native form by NMR spectrocopy. It is likely that the NMR approach developed for the structural elucidation of LOS-III can be successfully applied to other large lipopolysaccharides.

The 3,6-dideoxyhexoses are nonubiquitous monosaccharides in nature and are restricted to paratose, abequose, ascarylose, colitose, and tyvelose. They have been found in lipopolysaccharides of Gram-negative cell envelopes where they have been identified as essential antigenic determinants contributing to the serological specificity of many immunologically active polysaccharides (42). Only two C4-branched 3,6-dideoxyhexoses have been found as components of lipopolysaccharides from *Yersinia* species (43). Their structures were assigned to 3,6-dideoxy-4-C-(1-hydroxyethyl)-D-Xylo-hex*p* and differ by the configuration of the side chain. Thus, the immunodominant component of LOS-III, the C4 branched 3,6-dideoxy-hex*p* is a recently discovered monosaccharide in nature.

From bacteriological (44), serological (15, 45), and molecular (46, 47) studies, the distinction between the two species *M. kansasii* and *M. gastri* remains in question. However, if *M. kansasii* is known to be the etiologic agent of pulmonary disease resembling tuberculosis, *M. gastri* is considered as nonpathogenic. When phagocytized by macrophages, *M. kansasii* multiplies intracellularly and forms an electron transparent zone shielding the bacteria against the microbicidal functions of the host macrophages, whereas *M. gastri* does not multiply and lacks the capacity to form an electron transparent zone (48). It is well established that cell wall glycolipids such as phenolic glycolipids and LAM are involved in intramacrophagic mycobacterial survival, while it has been suggested that LOSs in the context of *M. kansasii* could be considered as avirulent factors. In this way, the glycolipidic composition of the *M. kansasii* and *M. gastri* cell wall may account for the functional diversity of the envelopes, in relation to their capacity for intracellular growth. It has been shown that both species, whether rough or smooth, contain identical phenolic glycolipids (6). In the present work, we report that the *M. gastri* strains as well as the *M. kansasii* smooth strains investigated contain LOSs. *M. gastri* W471 LOS-III differs significantly from the LOSs of other species of mycobacteria (3) but, more to the point, differs from those of *M. kansasii* (10) by the immunodominant monosaccharide a C4-alkylated 3,6-dideoxy-α-hex*p*, α-D-Gal*p* and the absence of Fuc*p*. This glycolipid (LOS-III) has also been identified in two other *M. gastri* strains HB4362 and HB4389 of rough and smooth morphology, respectively. Thus, LOS-III is a specific marker of *M. gastri* allowing it to be differentiated from *M.*

kansasii. The other interesting feature is that the two other *M. gastri* strains (HB4362 and HB4389) contain both the set of LOSs described in *M. kansasii* smooth variants and the LOSs of *M. gastri* W471. These two *M. gastri* strains appear to make up a continuum between *M. gastri* W471 and the smooth *M. kansasii* strains. Moreover, the fact that *M. gastri* strains are rapidly cleared by the macrophages (like the smooth *M. kansasii* strains) compared to the rough *M. kansasii* is in agreement with the presence of LOSs in all the *M. gastri* strains investigated. It can be assumed that rather than having a masking effect, LOSs decrease the rigidity of the outer layer and thus the permeability barrier,[3] so it can be speculated that the mycobacteria become more susceptible to macrophage metabolites. Nevertheless, the LOSs, as a factor of mycobacterial avirulence, remain an open question since they have been found in *Mycobacterium tuberculosis* strain Canetti (37), *Mycobacterium malmoense* (49), and *Mycobacterium szulgai* (36). Thus, the role of the LOSs in the immunopathogenesis of mycobacterial infection remains to be more precisely defined.

Acknowledgment—We thank Dr. Marie-Antoinette Lanéelle for the kind gift of the mycobacterial cultures.

REFERENCES

1. Nunn, P. P., and Mc Adam, K. P. W. J. (1988) *Br. Med. Bull.* **44,** 801–813
2. Horsburgh, C. R., Jr., and Selik, R. M. (1989) *Am. Rev. Resp. Dis.* **139,** 4–7
3. Brennan, P. J., Hunter, S. W., McNeil, M., Chatterjee, D., and Daffé, M. (1990) *Microb. Determinants Virulence Host Response* (Ayoub, E. M., Cassell, G. H., Branche, W. C., and Henry, T. J., eds) Chap. 4, pp. 55–75, American Society for Microbiology, Wash., D. C. 20005
4. Belisle, J. T., and Brennan, P. J. (1989) *J. Bacteriol.* **171,** 3465–3470
5. Collins, F. M., and Cunningham, D. S. (1981) *Infect. Immun.* **32,** 614–624
6. Gilleron, M., Venisse, A., Fournié, J. J., Rivière, M., Dupont, M. A., Gas, N., and Puzo, G. (1990) *Eur. J. Biochem.* **189,** 167–173
7. Chan, J., Fujiwara, T., Brennan, P. J., McNeil, M., Turco, S. J., Sibille, J. C., Snapper, M., Aisen, P., and Bloom, B. R. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 2453–2457
8. Kaplan, G., Gandhi, R. R., Weinstein, D. E., Levis, W. R., Patarroyo, M. E., Brennan, P. J., and Cohn, Z. A. (1987) *J. Immunol.* **138,** 3028–3034
9. Sibley, L. D., Hunter, S. W., Brennan, P. J., and Krahenbuhl, J. L. (1988) *Infect. Immun.* **56,** 1232–1236
10. Hunter, S. W., Murphy, R. C., Clay, K., Goren, M. B., and Brennan, P. J. (1983) *J. Biol. Chem.* **258,** 10481–10487
11. Hunter, S. W., Jardine, I., Yanagihara, D. L., and Brennan, P. J. (1985) *Biochemistry* **24,** 2798–2805
12. Hunter, S. W., Fujiwara, T., Murphy, R. C., and Brennan, P. J. (1984) *J. Biol. Chem.* **259,** 9729–9734

[3] M. Gilleron, J. Vercauteren, and G. Puzo, personal communication.

13. Fournié, J. J., Rivière, M., and Puzo, G. (1987) *J. Biol. Chem.* **262,** 3174–3179
14. Fournié, J. J., Rivière, M., Papa, F., and Puzo, G. (1987) *J. Biol. Chem.* **262,** 3180–3184
15. Vercellone, A., Rivière, M., Fournié, J. J., and Puzo, G. (1988) *Chem. Phys. Lipids* **48,** 129–134
16. Rivière, M., Fournié, J. J., and Puzo, G. (1987) *J. Biol. Chem.* **262,** 14879–14884
17. Marion, D., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **113,** 967–974
18. States, D. J., Haberkorn, R. A., and Ruben, D. J. (1982) *J. Magn. Reson.* **48,** 286–292
19. Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **117,** 479–485
20. Bax, A., and Davis, D. G. (1985) *J. Magn. Reson.* **65,** 355–60
21. Bax, A., Griffey, R. H., and Howkins, B. L. (1983) *J. Magn. Reson.* **55,** 301–315
22. Shaka, A. J., Barker, P. B., and Freeman, R. (1985) *J. Magn. Reson.* **64,** 547–552
23. Bax, A., and Summers, M. F. (1986) *J. Am. Chem. Soc.* **108,** 2093–2094
24. Lerner, L., and Bax, A. (1986) *J. Magn. Reson.* **69,** 375–380
25. Rivière, M., Fournié, J. J., Monsarrat, B., and Puzo, G. (1988) *J. Chromatogr.* **445,** 87–95
26. Gerwig, G. J., Kamerling, J. P., and Vliegenthart, J. F. G. (1978) *Carbohydr. Res.* **62,** 349–357
27. Bradbury, J. H., and Jenkins, G. A. (1984) *Carbohydr. Res.* **126,** 125–156
28. Bock, K., and Pedersen, C. (1984) *Adv. Carbohydr. Chem. Biochem.* **42,** 193–225
29. Bock, K., and Pedersen, C. (1983) *Adv. Carbohydr. Chem. Biochem.* **41,** 27–66
30. Angyal, S. J., and Pickles, V. A. (1972) *Aust. J. Chem.* **25,** 1965–710
31. Bax, A. (1984) *J. Carbohydr. Chem.* **3,** 593–611
32. Piantini, U., Sorensen, O. W., and Ernst, R. R. (1982) *J. Am. Chem. Soc.* **104,** 6800–6801
33. Zubkov, V. A., Sviridov, A. F., Gorshkova, R. P., Shashkov, A. S., and Ovodov, Y. S. (1989) *Bioorg. Khim.* **15,** 192–198
34. Gagnaire, D., Mancier, D., and Vincendon, M. (1978) *Org. Magn. Res.* **11,** 344–349
35. Abeygunawardana, C., Bush, C. A., and Cisar, J. O. (1990) *Biochemistry* **29,** 234–248
36. Hunter, S. W., Barr, V. L., McNeil, M., Jardine, I., and Brennan, P. J. (1988) *Biochemistry* **27,** 1549–1556
37. Daffe, M., McNeil, M., and Brennan, P. J. (1991) *Biochemistry* **30,** 378–388
38. Dabrowski, J. (1989) *Methods Enzymol.* **179,** 122–156
39. Lerner, L., and Bax, A. (1987) *Carbohydr. Res.* **166,** 35–46
40. Jardine, I., Scanlan, G., McNeil, M., Brennan, P. J. (1989) *Anal. Chem.* **61,** 416–422
41. McNeil, M., Chatterjee, D., Hunter, S. W., and Brennan, P. J. (1989) *Methods Enzymol.* **179,** 215–242
42. Hanessian, S. (1966) *Adv. Carbohydr. Chem.* **21,** 143–201
43. Gorshkova, R. P., Zubkov, V. A., Isakov, V. V., and Ovodov, Y. S. (1984) *Carbohydr. Res.* **126,** 308–312
44. Wayne, L. G., Andrade, L., Froman, S., Kappler, W., Kubala, E., Meissner, G., and Tsukamura, M. (1978) *J. Gen. Microbiol.* **109,** 319–327
45. Lind, A., and Ridell, M. (1984) in *The Mycobacteria: A Sourcebook* (Kubica, G. P., and Wayne, L. G., eds) Part A, pp. 67–104, Marcel Dekker, New York
46. Imaeda, T., Broslawski, G., and Imaeda, S. (1988) *Int. J. Syst. Bact.* **38,** 151–156
47. Rogall, T., Flohr, T., and Böttger, E. C. (1990) *J. Gen. Microbiol.* **136,** 1915–1920
48. Rastogi, N., and David, H. (1988) *Biochimie (Paris)* **70,** 1101–1120
49. McNeil, M., Tsang, A. Y., McClatchy, J. K., Stewart, C., Jardine, I., and Brennan, P. J. (1987) *J. Bacteriol.* **169,** 3312–3320