Carbohydrate Research, 240 (1993) 1–22
Elsevier Science Publishers B.V., Amsterdam

The structure of an antigenic glycolipid (SL-IV) from *Mycobacterium tuberculosis*

Hans H. Baer

Ottawa-Carleton Chemistry Institute, Department of Chemistry, University of Ottawa, Ottawa, Ontario K1N 9B4 (Canada)

(Received February 3rd, 1992; accepted in revised form May 25th, 1992)

ABSTRACT

The structure of an antigenic glycolipid isolated recently from cell walls of various strains of *Mycobacterium tuberculosis* and believed to be a sulfolipid consisting chiefly of 2,3-di-*O*-(hexadecanoyl/octadecanoyl)-α,α-trehalose 2′-sulfate (designated as SL-IV), was reinvestigated by mass spectrometry and nuclear magnetic resonance spectroscopy. The material proved to be a complex mixture of closely related components which are indeed 2,3-di-*O*-acyltrehaloses. However, no sulfate group was found in the antigen, and revision of its designation as a sulfolipid is therefore required. The lipid substituents comprise an estimated 20% of C_{14}–C_{19} fatty acids, including tetradecanoic (myristic), 9-tetradecenoic (myristoleic), 9-hexadecenoic (palmitoleic), 9-octadecenoic (oleic), 10-methylhexadecanoic (tuberculopalmitic), and 10-methyloctadecanoic (tuberculostearic) acids in addition to hexadecanoic (palmitic) and octadecanoic (stearic) acids, and ~ 80% of higher, methyl-branched acids. Among the latter are, principally, 2,4-dimethyldocosanoic and 3-hydroxy-2,4,6-trimethyltetracosanoic acids, and a smaller proportion of 2,4,6-trimethyltetracos-2-enoic acid. Numerous further acids, related to those mentioned by methylene homology, are also present in small proportions.

INTRODUCTION

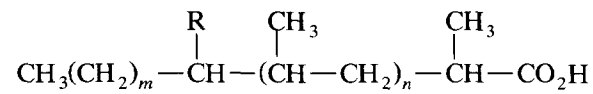
Trehalose-based glycolipids are biologically important constituents of mycobacteria[1–3]. Besides almost ubiquitous trehalose 6,6′-dimycolate (cord factor) and 6-monomycolate, a number of sulfolipids (polyacylated trehalose 2-sulfates)[2–5] were isolated from *M. tuberculosis* H37Rv by Goren[4,5] and revealed to contain, principally, hexadecanoic and octadecanoic acid and (homologous series of) phthioceranic (**1**) and hydroxyphthioceranic acids (**2**). More recently, Minnikin and co-workers[6] recognized the occurrence, in *M. tuberculosis* strains C and H37Rv, of four nonsulfated trehalose glycolipids that contain combinations chiefly of C_{16}–C_{19} fatty acids (straight-chain saturated and monounsaturated, and 10-methyl branched) as well as a 2,4-dimethyldocosanoic acid (**3**), a 2,4,6-trimethyltetracos-2-enoic acid (**4**; synonyms: C_{27} mycolipenic or C_{27} phthienoic acid), and a 3-hydroxy-2,4,6-trimethyltetracosanoic acid (**5**, C_{27} mycolipanolic acid). The positions of the acyl residues on the trehalose molecule in these glycolipids were not established[6]. The

isolation of a 2,2′,3,4′,6-penta-*O*-acyltrehalose with 80% of **4** (and homologs) and 20% of hexadecanoic/octadecanoic acid as ester groups, and of a minor component containing 18% of **5** instead of **4**, from the Canetti strain of *M. tuberculosis* was reported by Lanéelle and co-workers[7]. A similar penta-*O*-acyltrehalose and at least two further trehalose esters were obtained from the same strain and from several wild strains in the laboratory of Papa and David[8,9]. The two additional glycolipids showed[9–11] interesting immunoreactivity and potential for use in serodiagnosis of tuberculosis; the one isolated in larger proportion proved to be[9] a tetra-*O*-acyltrehalose 2-sulfate characterized by hydroxyphthioceranic residues (type **2**, mostly C_{40}) and was later stated[10,11] to be identical with the 2-*O*-hexadecanoyl/octadecanoyl-3-*O*-phthioceranoyl-6,6′-di-*O*-hydroxyphthioceranoyltrehalose 2-sulfate previously isolated by Goren[4] and designated as SL-I. For the other glycolipid, detected as a minor component and more polar in character, the probable structure of a 2,3-di-*O*-(hexadecanoyl/octadecanoyl)-α,α-trehalose 2′-sulfate was proposed[9]. The product was somewhat infelicitously termed[10–12] SL-IV in the belief that it corresponded to a minor sulfolipid previously encountered and so designated[3] by Goren *. It shows high promise as an effective and specific antigen, superior to SL-I, in serodiagnosis of tuberculosis and leprosy[10–13], and a more accurate determination of its structure was therefore desirable.

RESULTS AND DISCUSSION

Gross structure.—The glycolipid in question will here be referred to as "SL-IV", in quotation marks so as to differentiate it from the products mentioned in the preceding footnote; no identity with these is implied. Small samples extracted from *M. tuberculosis* and purified in the Pasteur Institute[8–10] were supplied by Dr. F. Papa. They were examined by spectroscopic methods, and the evidence to be presented confirmed the previous finding[9] of α,α-trehalose being esterified in positions 2 and 3. However, the most important observation made at the outset of the study was the *absence* of a sulfate ester group, requiring the classification as a sulfatide to be revised **. Thus, there was no signal in the entire range between $\delta \leqslant 90$ and $\geqslant 73.5$ of its ^{13}C NMR spectrum. Sulfatation of a secondary hydroxyl group in sugars generally causes deshielding by 6–10 ppm of the carbon atom to which it is attached, and slight shielding by 1–2 ppm of adjacent carbon atoms[15]. Thus, for D-glucose and its 2-sulfate in D_2O solution the reported[15c] shift differ-

* However, Goren referred by SL-IV to diacyltrehalose 2′-sulfates not characterized as native sulfolipids. Partial solvolysis of desulfated SL-I (termed SL-I-CF) gave 6,6′-di-*O*-hydroxyphthioceranoyltrehalose (SL-IV-CF) and 2-*O*-hexadecanoyl/octadecanoyl-3-*O*-phthioceranoyltrehalose (SL-IV′-CF). A 2,3-di-*O*-(hexadecanoyl/octadecanoyl)trehalose was not detected[3].

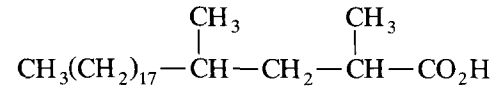


1 Phthioceranic acids

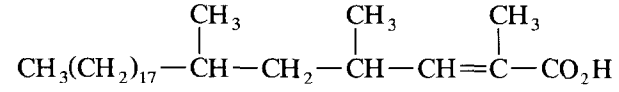
R = H, m = 14 or 16, n = 2–9 (principally 6)

2 Hydroxyphthioceranic acids

R = OH, m = 14 or 16, n = 2–10 (principally 7)

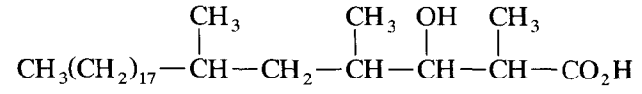


3 2,4-Dimethyldocosanoic acid



4 2,4,6-Trimethyltetracos-2-enoic acid

(C_{27}-mycolipenic acid, C_{27}-phthienoic acid)



5 3-Hydroxy-2,4,6-trimethyltetracosanoic acid

ences are −6.4 (C-2), +1.9 (C-1), and +2.1 (C-3) ppm. Similar differences should apply to α,α-trehalose and its 2-sulfate. For the unsubstituted disaccharide in Me_2SO-d_6 solution we found δ 93.4 (C-1), 72.0 (C-2), and 72.8 (C-3), values very close to those recorded[16] for a D_2O solution, and 2-sulfatation should therefore be expected to result in signals near δ 91.5 (C-1), 78.4 (C-2), and 70.7 (C-3) for the sulfated (but otherwise unsubstituted) moiety. However, the values found for "SL-IV" in Me_2SO-d_6 were δ 94.35 (C-1′), 71.4 (C-2′), and 72.9 (C-3′). The absence of a 2-sulfate group was corroborated by the fact that an authentic sample of SL-I, a polyacylated α,α-trehalose 2-sulfate kindly provided by Dr. M.B. Goren, gave a C-2 signal at δ 77.1, well downfield from its position[16] at δ 70.1 in α,α-trehalose octaacetate (for solutions in $CDCl_3$).

** This was mentioned in a preceding paper[13], following the acceptance of which we became aware of a publication by Daffé and co-workers[14] who also revised the structure. The authors found that the ^1H NMR data (200 MHz) for SL-IV do not support the presence of a sulfate group, and they confirmed its absence by ^{35}S radiolabeling studies. Concerning the fatty acids present, the authors cited mass-spectral evidence showing trehalose molecules esterified in the same glucosyl unit with octadecanoic acid and **3**, hexadecanoic acid and **4**, and octadecanoic acid and **4**, respectively, to constitute the main components of the glycolipid. On this basis, they considered it very likely that SL-IV is identical with one of the glycolipids described by Minnikin and co-workers[6], but they did not demonstrate the presence, in SL-IV, of several additional alkanoic, alkenoic, and hydroxyalkanoic acids (especially **5**) as are contained in Minnikin's compounds.

Other significant, though unsurprising, observations concern the constitutional homogeneity of "SL-IV". The samples examined did not consist of a single molecular species but of complex mixtures of kindred molecules, characterized by extensive homology in their fatty acid residues and also by variations with respect to the attachment of these in positions 2 and 3 of the disaccharide core. The former feature became obvious in mass spectrometry, and the latter was indicated by the ^{1}H and ^{13}C NMR spectra which, instead of showing single signals for H-2 and H-3, and for C-1, C-2, and C-3 respectively, exhibited groups of closely-spaced signals for these atoms. Evidently, their small but distinct chemical-shift differences must be due to variant substituent effects of structurally different acyl groups present at a given site, and the groups must be distinguished by features other than the simple methylene homology that applies to hexadecanoyl and octadecanoyl *. Moreover, thin-layer chromatography also revealed "SL-IV" to be inhomogeneous: It gave at least two distinct, if partially overlapping, spots having R_f 0.12 and 0.15 (solvent[6], 50:60:2.5:3 $CHCl_3$–Me_2CO–MeOH–H_2O). The upper spot was detected in UV light (254 nm), suggesting the presence of unsaturated structures, and it gave an intense reddish rim on charring with ethanolic H_2SO_4 at 120°C, as did a sample of the hydroxy acid-containing tetra-*O*-acyltrehalose[4] SL-II′-CF. By contrast, synthetic 2,3-di-*O*-hexadecanoyltrehalose was slightly more polar (R_f 0.10), UV-inactive, and colored a plain brown on charring. The conclusion was inescapable that fatty acids (probably several) other than hexadecanoic and octadecanoic acids (presumably from among the types **1**–**5**) must be quite prominently implicated in the structure.

The constituent lipid acids.—Methoxide-catalyzed methanolysis of "SL-IV" produced a complex mixture of methyl esters, the chemical-ionization mass spectrum of which showed a plethora of molecular ion and fragment peaks, up to *m/z* 497 (2.2%) corresponding to $M^+ + 1$ for the ester of a C_{31} hydroxy acid. Alkaline hydrolysis of "SL-IV" gave a similar, complex mixture of fatty acids, with a peak at *m/z* 483 (8.3%) for $M^+ + 1$ of a C_{31} hydroxy acid near the high mass end of the CI spectrum. In both instances there were only a few insignificant trace peaks above *m/z* 500, and the bulk of the most intense peaks occurred in the range below *m/z* 450. A negative-ion FAB spectrum of the carboxylic acid mixture confirmed these results; although an $M^- - 1$ peak expected at *m/z* 481 for a C_{31} hydroxy acid was absent, one at *m/z* 463 (10.5%) was interpretable as being due to a corresponding dehydration product, and again the bulk of important peaks was below *m/z* 450. Therefore, it can be concluded that such acids as the higher-molecular members of the homologous phthioceranic and hydroxyphthioceranic families (**1** and **2**, $n > 4$, C_{34} and up), found[17] in the related mycobacterial sulfolipids SL-I and SL-III,

* In none of the several 2- and 3-*O*-alkanoyl- and 2,3-di-*O*-alkanoyl-trehaloses which we have prepared in concurrent synthetic work did a replacement of hexadecanoyl by octadecanoyl have any discernible effect on the H-2 and H-3 signals (R.L. Breton and X. Wu, unpublished).

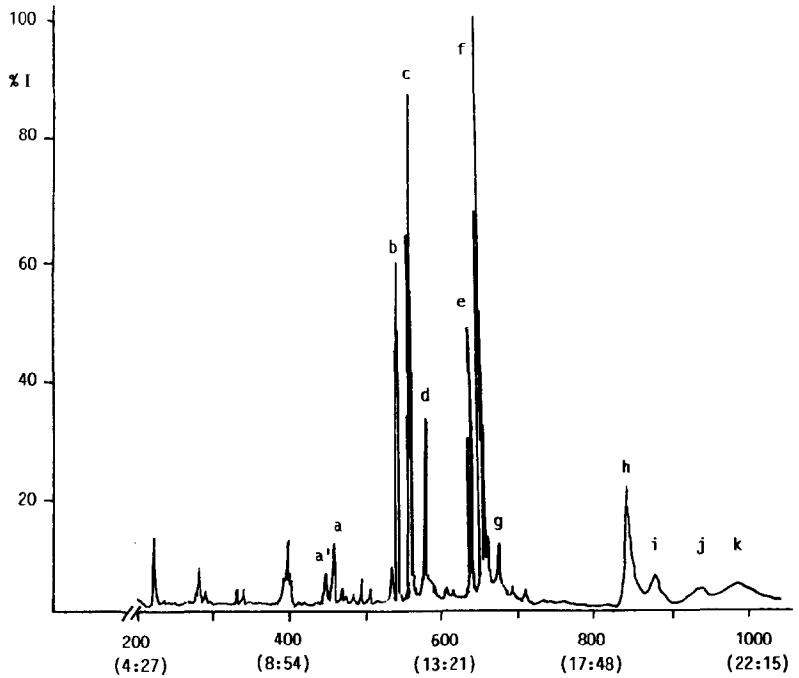


Fig. 1. Gas chromatogram of mixture of methyl esters. Abscissa: MS scan numbers, with corresponding retention times (min:s) in parentheses.

including their principal components having $n = 6$ and $n = 7$, respectively, are not present to any significant extent in "SL-IV".

For a more-detailed assessment of the nature of the acids it is convenient to begin with the lower-molecular species. The aforementioned mixture of methyl esters was examined by GLC–MS. The gas chromatogram (Fig. 1) showed a series of well-separated peaks, and electron-impact spectra were obtained from those labeled *b–i* and *k*. The spectra of *b*, *c*, *e*, and *f* displayed M^+ peaks at *m/z* 268, 270, 296, and 298, respectively, and the fragmentation patterns were identical in every respect with those obtained for comparison from authentic samples of methyl palmitoleate (*b*), palmitate (*c*), oleate (*e*), and stearate (*f*); for a discussion of these patterns, see ref. 18. Clearly, hexadecanoyl and octadecanoyl are major contributory groups to the structure of the glycolipid (as has already been stated[9]), with octadecanoyl being moderately preponderant, and their unsaturated counterparts *b* and *e* are minor but not insignificant components. The components *d* and *g* gave M^+ peaks at *m/z* 284 and 312, respectively, signifying methyl esters of C_{17} and C_{19} acids. These were identified as methyl 10-methylhexadecanoate (*d*, methyl tuberculopalmitate) and methyl 10-methyloctadecanoate (*g*, methyl tuberculostearate) by analysis of their fragmentation patterns with reference to a published spectrum[19] of the latter. Thus, because of preferred bond cleavage at a branch point their fragments of mass 171 [$^+CH_2(CH_2)_7CO_2CH_3$] and 199

$[CH_3CH^+(CH_2)_8CO_2CH_3]$, resulting from cleavage of the C-9,10 and C-10,11 bonds, were noticeably more prominent relative to other fragments in the same spectra, and in comparison to fragments of the same masses that arise from similar, but unbranched, molecules (*c* and *f*) through fission of the C-9,10 and C-11,12 bonds. Furthermore, the fragment of mass 185, which results from C-10,11 cleavage in isomers lacking a C-10 methyl branch, was characteristically absent from the spectra of *d* and *g*.

The spectrum of the rather important component *h* exhibited M^+ at *m/z* 382, indicating a methyl ester of a saturated C_{24} acid. The base peak at *m/z* 88 together with a very strong (66%) peak at *m/z* 101 required C-2 to bear a methyl branch, and C-3, to be unbranched; these strong peaks originate[18] from C-2,3 cleavage (with McLafferty rearrangement) and C-3,4 cleavage (giving the fragment $^+CH_2CH(CH_3)CO_2CH_3$), and are characteristic of methyl 2-methylalkanoates. (The components *b*–*g* yielded the corresponding fragments of *m/z* 74 and 87, as expected.) Further, there was the following good evidence for a second methyl branch in *h*, situated at C-4. Long-chain methyl alkanoates generally give all the fragments $[C_nH_{2n}CO_2CH_3]^+$ with $n = 2, 3, 4$, etc., by progressive loss of chain-end carbons, and the fragment of mass 143 ($n = 6$) tends to be markedly more intense than its homologs of masses 129 ($n = 5$), 115 ($n = 4$), and 101 ($n = 3$)[18]. Indeed, for *c*, *f*, and *g* we found the mass peaks 143, 129, 115, and 101 to have 17–24, 6–9, 3–4, and 6–10% of base peak intensity, respectively; for *d*, the values were 55, 7.3, 3.3, and 14.7%. By contrast, in *h* the *m/z* 129 was much more intense (6%) than the *m/z* 143 and 115 peaks, which were particularly weak (0.5–0.8%). This was interpreted as reflecting preferential C-4,5 cleavage in a 2,4-dimethyl structure (Fig. 2) *, as exemplified by the mass spectrum[18] of methyl 2,4-dimethylnonanoate. Consequently, *h* was designated as 2,4-dimethyldocosanoic acid (**3**) methyl ester. The possibility that it may have been a 2,4,6,...-polymethyl isomer of the phthioceranic type **1**, with three or more branches and a correspondingly shorter terminal chain, was considered unlikely on the grounds that fragments of *m/z* 143 and 171 (from cleavages adjacent to branched C-6) should then have been prominent,

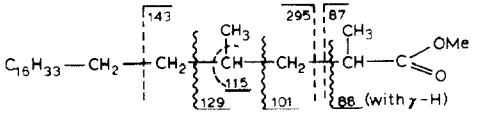


Fig. 2. Methyl 2,4-dimethyldocosanoate. Favored (~~~) and less-favored (- - - - -) mass-spectral fragmentations.

* A noticeable peak at *m/z* 87 (10–12%) for $CH_3^+CHCO_2CH_3$ and a small one at *m/z* 295 (1–1.5%) for $^+C_{21}H_{43}$ indicated a second mode of β-cleavage to occur, by homolysis, besides the vastly predominant McLafferty-type cleavage that produces mass 88. This second mode is undoubtedly encouraged by the presence of the 2-methyl branch. The unbranched esters gave no trace of an analogous companion (*m/z* 73) to their McLafferty fragment (*m/z* 74).

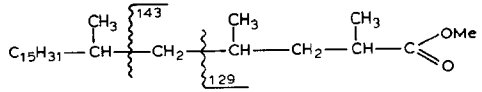


Fig. 3. Distinctive mass-spectral fragmentation of the component presumed to be methyl 2,4,6-trimethylheneicosanoate.

which was not the case. However, it may be possible that the minor peak *i* contained such an isomer.

The mass spectrum of *i* was very similar in many respects to that of **3** ester (*h*), with M^+ at *m/z* 382 and an $M^+ - C_5H_{11}$ peak at 311 (also given by *h*) suggesting isomerism, except that *m/z* 129 was more intense (14%) and *m/z* 143 (insignificant in *h*) was substantial (8.8%). These features warrant an assumption of a C-6 methyl branch (even though a fragment of *m/z* 171 was not detected) and hence suggest methyl 2,4,6-trimethylheneicosanate as a possible component (Fig. 3). However, the gas chromatogram peak *i* was evidently inhomogeneous, for the mass spectrum contained additional features by which it differed from that of **3** ester. Thus, clusters of prominent peaks were centered at *m/z* 69, 83, and 97 (representing C_5–C_7 alkenyl fragments), and at 109/110, 123/124, and 137/138 (ketenes arising by fragmentation after loss of methanol); these peaks were very weak in the saturated esters examined (including *h*), but are typical for oleate and palmitoleate although the fragments do not contain the original C=C moiety of these molecules. The presence of some alkenic species in *i* was therefore suspected. Additional peaks were seen at *m/z* 127 (13%), 128 (10%), 141 (6%), 142 (19%), 169 (7%), 217 (7%), and 294 (5%). Light was shed on their significance when the spectrum of the corresponding chromatographic fraction of trimethylsilylated acids (see later) revealed the presence of an alkenoic acid $C_{25}H_{48}O_2$ (mol wt 380). Although an M^+ peak for the corresponding methyl ester (mol wt 394) was not observed, the fragments just listed can be explained in terms of a methyl C_{25}-mycolipenate (phthienoate) structure as shown in Fig. 4, i.e., the ester of a lower homolog of **4**. Thus, ions of *m/z* 294 [$M^+ - 100$; elimination of $CH_2{=}C(CH_3)CO_2CH_3$] and 169 [$^+CH(CH_3)CH_2CH(CH_3)CH{=}C(CH_3)CO_2CH_3$ resulting from C-6,7 cleavage] are characteristic for this ester, which also yields abundant ions of *m/z* 88 and 101 that arise here by C-2,3 and C-3,4 fission after shift of the double bond[20,21]. The substantial peak at *m/z* 142 has been explained[20] as representing $(CH_3)_2C{=}CH{-}C(CH_3){=}C(OH)OCH_3$ from C-5,6 cleavage with

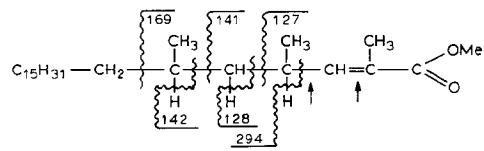


Fig. 4. Mass-spectral fragmentation of methyl 2,4,6-trimethyldocos-2-enoate (C_{25}-mycolipenate). Arrows indicate sites of cleavages occurring after shift of the double bond and hydrogen transfer, to give fragments of mass 88 and 101.

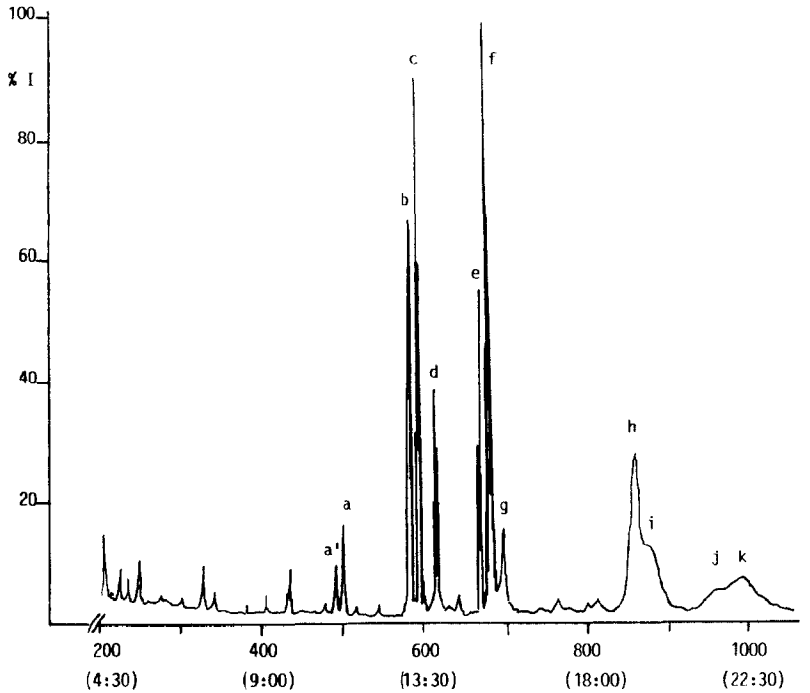


Fig. 5. Gas chromatogram of mixture of trimethylsilylated acids. Abscissa: MS scan numbers, with retention times (min:s) in parentheses.

hydrogen transfer; its weaker companion (m/z 141) probably arose by simple C-5,6 cleavage. The pair of masses 128 and 127 is explained by analogous C-4,5 cleavages. The peak at m/z 217 remains unexplained at present. It should be noted that the corresponding, saturated 2-hydroxy ester, i.e., the C_{25}-mycolipanolate (lower homolog of **5**-ester), if present, may likewise have given m/z 294 and 169 peaks of low intensity, by way of prior dehydration; its M^+ peak would have been expected to be very small or absent. However, a most characteristic and abundant fragment of m/z 117 ($^+CH(OH)CH(CH_3)CO_2CH_3$) would have been observable (as the base peak) for that hydroxy ester[21], which was not the case.

The GLC peak *k* presumably represented a mixture of molecules of higher mass. It gave a spectrum virtually identical with that of *i* in the region up to and including m/z 143 and also contained fragments of mass 169 and 217, but additionally it showed ions at m/z 147, 149, 165, 195, 197, and 346. The ion at m/z 197 must have been a tetradecyl fragment (and that at 195, an associated alkenyl), and will receive further comment below. Because of the absence of any molecular ion and paucity of high-mass fragments the spectrum afforded no further structural information.

Gas chromatography of the (trimethylsilylated) mixture of fatty acids obtained from "SL-IV" by alkaline hydrolysis gave the picture portrayed in Fig. 5, which quite faithfully corresponded to that of the methyl esters (Fig. 1). The mass spectra

for the GLC peaks *b–h* confirmed the previous assignments, by showing M^+ and (stronger) $M^+ - 15$ peaks 58 and 43 mass units higher than M^+ of the respective methyl esters, as well as by showing readily assignable fragmentation peaks. For *b–g*, the latter were at *m/z* 117 (93–100%, CH_2=C(OH)OSi$^+$$Me_2$), 129 (45–67%, $CH_2CHCO_2Si^+Me_2$), 132 (20–27% for *b*, *e*; 50–63% for *c*, *d*, *f*, *g*; [CH_2=C(OH)OSiMe_3]$^+$), and 145 (22–55%, $^+CH_2CH_2CO_2SiMe_3$). Significantly the GLC peaks *d* and *g* containing 10-methyl branched acids gave fragments at *m/z* 229 and 257 that were more intense than in the unbranched acids, attesting to preferred fission of the C-9,10 and C-10,11 bonds. The GLC peak *a* (whose counterpart in Fig. 1 had not been analyzed) showed M^+ and $M^+ - 15$ at *m/z* 300 and 285, and all of the fragments just mentioned for *c* (hexadecanoic acid) and *f* (octadecanoic acid), identifying it as their C_{14} homolog, myristic acid. The minor satellite peak *a'* was not analyzed, but in view of its position and intensity, and by virtue of the presence of palmitoleic (*b*) and oleic acid (*e*) it doubtless represented myristoleic acid. Similarly, the minor unlabeled components of even higher chromatographic mobility seen in Figs. 1 and 5 were almost certainly lower homologs whose presence in small proportions were already indicated by the mixture spectra. The trimethylsilylated 2,4-dimethyldocosanoic acid **3** (*h*) exhibited M^+ and $M^+ - 15$ at *m/z* 440 and 425, and fragment ions at 146 (98%, with a companion at 145 [15%]), 159 (53%), and 187 (3.3%) from β-, γ-, and δ-cleavage, mirroring the masses 88 (with companion 87), 101, and 129 seen for the methyl ester of **3** (compare *h* in Fig. 1). The peaks for masses 173 and 201 were very weak (0.5–1%), as were the corresponding masses 115 and 143 in the methyl ester spectrum. A prominent peak at *m/z* 143 (23%) had no counterpart in the methyl ester; it was interpreted as CH_2=C(CH_3)$CO_2Si^+Me_2$ resulting from the dimethylsilyl species $M^+ - 15$ by γ-cleavage with hydrogen transfer.

The GLC peak *i* (Fig. 5) was ill separated from peak *h* and evidently contained a mixture of compounds, presumably trimethylsilylated **3** as well as its putative 2,4,6-trimethylheneicosanoic isomer; it showed M^+ and $M^+ - 15$ at *m/z* 440 and 425 attributable to both, and a fragmentation pattern closely resembling that of *h*. An additional pair of M^+ and $M^+ - 15$ peaks occurred at *m/z* 452 and 437, indicating an enoic acid $C_{25}H_{48}O_2$ (mol wt 380) for whose methyl ester the fragmentation data have been discussed above. Counterparts of the methyl ester fragments of masses 142, 141, 128, and 127 were seen at *m/z* 200, 185; 199, 184; 186, 171; and 185, 170. The spectra of the GLC peaks *j* and *k* of the trimethylsilylated acids were in the region up to *m/z* 187 almost superposable on those of *h* and *i*, and were therefore indicative of similar, 2,4-branched structures. Both of the fractions contained more than one compound, owing to poor chromatographic separation. The highest mass in *j* was 453, assignable to $M^+ - 15$ for an alkanoic acid $C_{26}H_{52}O_2$ (mol wt 396), probably a phthioceranic acid (**1**, with $n = 3$ and $m = 12$, uncommon). Although fragments with masses of 201, 229, 243, and 271 expected from cleavage of the C-5,6, C-6,7, C-7,8, and C-8,9 bonds, respectively, were not discernible, a peak at *m/z* 197 (3%) for a tetradecyl ion $^+C_{14}H_{29}$ did

support the proposed structure. Unbranched, long-chain alkanoates such as hexadecanoates or octadecanoates do not give such a fragment, and its formation therefore seems to signify favored cleavage at a branch point, i.e., of the C-8,9 bond in a 2,4,6,8-tetramethyldocosanoic species. (Phthioceranates are known to incur loss of the terminal n-alkyl group[4]). Peak *k* gave a pair of M^+ and $M^+ - 15$ ions at *m/z* 468 (2%) and 453 (3.3%) as well as the *m/z* 197 ion (3%), attributable to the foregoing C_{26} acid. It gave furthermore a peak at *m/z* 467 (2.3%) regarded as the $M^+ - 15$ ion of a homologous alkanoic acid $C_{27}H_{54}O_2$ (mol wt 410), presumably also a phthioceranic acid (**1**, $n = 2$, $m = 16$). Finally, *k* showed an $M^+/M^+ - 15$ pair at *m/z* 480 (3.1%)/465 (4.2%) corresponding to an alkenoic acid $C_{27}H_{52}O_2$ (mol wt 408). There were clusters of fragment ions (3–6%) at *m/z* 201–199 and 187–185 which, as counterparts of the previously mentioned methyl ester fragments (see Figs. 3 and 4), were taken to indicate preferential C-5,6 and C-4,5 fissions in 2,4,6-trimethyl substituted acids. A small peak at *m/z* 253 (1%) suggested a terminal n-$C_{18}H_{37}$ fragment arising from C_{27} acids. There can be little doubt that the unsaturated component of *k* was 2,4,6-trimethyltetracos-2-enoic acid (**4**), which has repeatedly been encountered in related, mycobacterial glycolipids[6–8,21].

As regards the presence of hydroxy acids of the hydroxyphthioceranic (**2**) or mycolipanolic (**5**) types, the CI mass spectra of the free-acid and methyl ester mixtures referred to before contained numerous peaks attributable to $M^+ + 1$ ions of acids $C_nH_{2n}O_3$ ($n = 19–31$) and their esters, and the ester $M^+ + 1$ peaks were accompanied by $M^+ - 31$ peaks for the corresponding acylium ion fragments. Most prominent was a C_{27} species, and C_{20}, C_{24}, and C_{28} homologs were also fairly abundant. It is assumed that the failure of these hydroxy derivatives to become manifest in the GLC–MS experiments just described was due to their suffering dehydration in the chromatographic column at the temperatures of operation (150–320°C). This was convincingly borne out when spectra were procured by use of the technique of ion-spray flow injection at room temperature. A spectrum of the carboxylic acid mixture taken in the negative-ion mode is shown in Fig. 6. A strong anion peak at *m/z* 425 indicates an acid $C_{27}H_{54}O_3$ to be a major constituent, second only to stearate (*m/z* 283). Other significant hydroxy acids, though of less importance, are revealed by peaks at *m/z* 439 (C_{28}), 411 (C_{26}), 397 (C_{25}), and 383 (C_{24}); small peaks for further homologs are also visible. The small peak at *m/z* 215 appears to represent a hydroxydodecanoic acid. Of particular interest was the unexpectedly low intensity of the *m/z* 407 ion attributed to C_{27}-mycolipenic acid (**4**), which relegates this component to a position of minor importance. This acid had *appeared* to be a *major* component when acid or ester mixtures were analyzed by the chemical ionization method, but under these conditions it evidently arose in large part as an artefact from the abundant C_{27} hydroxy acid. Incidentally, this facile dehydration accords with the view that the enoic acid in question is an α,β-unsaturated one, i.e., that it is indeed **4** (and its precursor, therefore, **5**), rather than the dehydration product of an isomeric

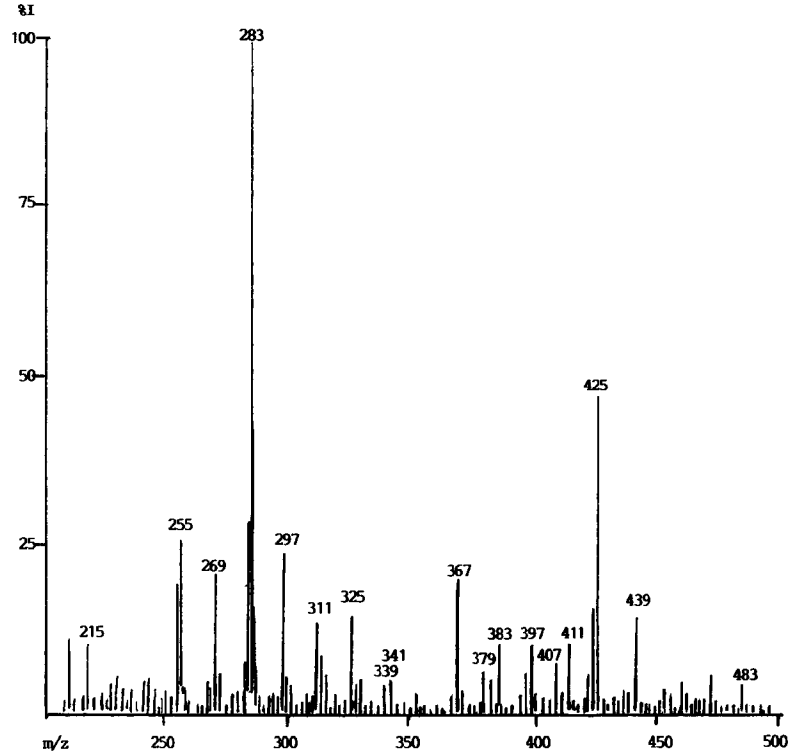


Fig. 6. Negative-ion mass spectrum of carboxylic acid mixture, obtained by ion-spray flow injection at room temperature.

hydroxyphthioceranic acid (**2**, $m = 16$, $n = 2$). This was in agreement with an apparent absence of acids of type **2** from "SL-IV". According to Goren's detailed mass-spectroscopic studies[17], methyl esters of **2** ($m = 14$) are readily identified by very strong twin peaks at m/z M − 240, M − 239, resulting from loss of palmitaldehyde; we did not encounter such fragments in our studies, nor fragments M − 268, M − 267 for C_2 homologs with $m = 16$.

In summary, our studies thus far discussed have demonstrated that "SL-IV" is composed of trehalose esterified principally with the following acids: (a), C_{14}–C_{19} fatty acids comprising mainly octadecanoic and hexadecanoic acid and smaller proportions of tetradecanoic, 9-tetradecenoic, 9-hexadecenoic, and 9-octadecenoic acids as well as 10-methylhexadecanoic and 10-methyloctadecanoic acid; and (b), higher, methyl-branched acids and hydroxy acids, with 2,4-dimethyldocosanoic acid (**3**) and 3-hydroxy-2,4,6-trimethyltetracosanoic acid (**5**) as major components together with a small proportion of 2,4,6-trimethyltetracos-2-enoic acid (**4**). Extensive methylene homology exists, with traces of numerous lower and higher homologs of most of the main constituents being present. It is obvious that the composition of "SL-IV" is strikingly similar to, if not identical with, that of the polar glycolipid

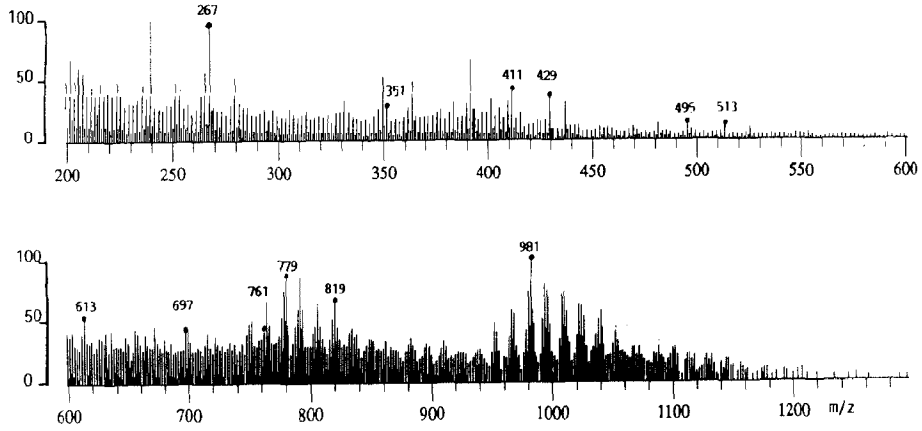


Fig. 7. Positive-ion FAB mass spectrum of "SL-IV" from glycerol–thioglycerol matrix with added NaCl. The molecular-ion peak of dimethyldocosanoyloctadecanoyltrehalose ($[M+23]^+ = 981$) and the peaks attributed to its fragmentation (see Scheme 1) are marked (•).

fractions extracted by Minnikin and co-workers[6] from *M. tuberculosis* strains C and H37Rv.

The position of the lipid acids on the trehalose core.—The $^+$FAB mass spectrum of intact "SL-IV" (with Na^+ added) shows a series of clusters of $M^+ + 23$ ions separated by 14 mass units (methylene homology), chiefly in the range of *m/z* 950–1150, with maximum intensity at *m/z* 981 (Fig. 7). This range corresponds to di-*O*-acyltrehaloses containing almost all of the possible combinations of the acids previously identified. In Table I are given the most prominent peaks and suggested assignments which are illustrative, not exhaustive; other combinations can of course be written as possible contributors to individual peaks. The most important aspect of the pattern is that its center of gravity lies in the range C_{40}–C_{45} for the two acyl groups combined. Peaks corresponding to such combinations as $C_{16} + C_{16}$, $C_{16} + C_{18}$, or $C_{18} + C_{18}$ are very weak and it is therefore clear that a majority of molecules contain at least one of the higher acids of types **3**–**5**. Minor peaks above *m/z* 1100 suggest the presence of two of these acids in at least a small proportion of molecules. That proportion may actually be not insignificant, considering that generally the intensity of M^+ peaks tends to decrease with increasing molecular mass, and the relatively great abundance of higher acids was also supported by ^{1}H NMR spectroscopy (see later). The notion expressed earlier[9] that "SL-IV" is mainly a trehalose di(hexadecanoate/octadecanoate) can therefore no longer be upheld. (See also ref. 14). However, the conclusion[9] that two acyl residues are attached to the same glucose unit of trehalose while the other glucose unit is unsubstituted was confirmed in the present study. Thus, in the region of *m/z* 750–900 the mass spectrum exhibited fragmentation peaks $M^+ + 23 - 202$, representing diacylglucosyl ions resulting from loss of (neutral) glucosyloxy sodium, and companion peaks lower by 18 mass units which arise from them by dehydration occurring either in the sugar ring or in a hydroxyacyl group, if present. These peaks

TABLE I

The most prominent [a] molecular ion peaks in the FAB mass spectrum [b] of "SL-IV"

$M^+ + 23$ (m/z)	Intensity (%) [c]	Fitting acyl group combinations [d] (number of carbon atoms)	$M^+ + 23$ (m/z)	Intensity (%) [c]	Fitting acyl group combinations [d] (number of carbon atoms)
1053	33	$19 + 27_{OH}$	997	46	$18 + 24_{OH}$
1051	41	16 + 31	995	74	16 + 27, 19 + 24, $18_e + 24_{OH}$
1049	30	$16_e + 31$	993	80	$16 + 27_e$, $16_e + 27$
1039	57	$18 + 27_{OH}$	991	51	$16_e + 27_e$
1037	51	$18_e + 27_{OH}$, 19 + 27, 18 + 28, 16 + 30	983	46	$17 + 24_{OH}$, $14 + 27_{OH}$
1035	44	$19 + 27_e$	981	100	18 + 24
1025	53	$17 + 27_{OH}$, $20 + 24_{OH}$	979	73	$18_e + 24$
1023	61	18 + 27	977	42	$18_e + 24_e$
1021	63	$18 + 27_e$, $18_e + 27$	967	55	$16_e + 24_{OH}$
1011	57	$16 + 27_{OH}$, $19 + 24_{OH}$	965	58	$17 + 24_e$, $14 + 27_e$
1009	73	$16_e + 27_{OH}$, 17 + 27, 16 + 28, 18 + 26	963	33	$14_e + 27_e$
1007	71	$17 + 27_e$	953	41	16 + 24
1005	33		951	48	$16_e + 24$

[a] From 30–100% intensity. Peaks at m/z 1101, 1099, 1087, 1071, 1067, and 1065 had ~25% intensity. [b] From glycerol–thioglycerol matrix, with added NaCl. [c] Relative to m/z 981. [d] Acyl groups are indicated by the number of their carbon atoms. Subscripts e and OH denote alkenoyl and hydroxyalkanoyl groups, respectively. Some of the peaks corresponding to combinations that include higher alkenoyls may not represent (or represent solely) true molecular ions but may indicate dehydration products originating from hydroxyalkanoates.

were observed for all of the molecular ions listed in Table I. Additionally there were peaks $M^+ + 23 - 162$ for sodiated diacylglucoses originating from elimination of neutral 2-hydroxyglucal, and usually companion peaks lower by 18 mass units due to dehydration. These fragmentation modes are well established in FAB mass spectrometry of oligosaccharides[22], and are illustrated in Scheme 1 for the example of dimethyldocosanoyloctadecanoyltrahalose ($M^+ + 23 = 981$). Further fragmentations by elimination of one substituent as a carboxylic acid or ketene molecule gave monoacylated di- and mono-saccharidic fragments as shown in Scheme 1 and highlighted in Fig. 7 for the m/z 981 parent ion. In the lower-mass region peaks appeared for the various substituents split off as acylium ions, for example, m/z 351 (dimethyldocosanoylium) and 267 (stearoylium), and there were strong peaks for unsubstituted glycosylium ions (m/z 163, 145, and 127, see Scheme 1). Peracetylated "SL-IV" gave a strong peak at m/z 331 (65% of base peak at 169) for tetra-*O*-acetylglucopyranosylium ion, furnishing additional proof that one glucose moiety was unsubstituted prior to acetylation. Proof for the other moiety to be substituted in positions 2 and 3 was obtained by NMR spectroscopy.

The 600-MHz, ^1H NMR spectrum of "SL-IV" is shown in Fig. 8. Assignments were verified by COSY plots, and the carbohydrate proton resonances are listed in

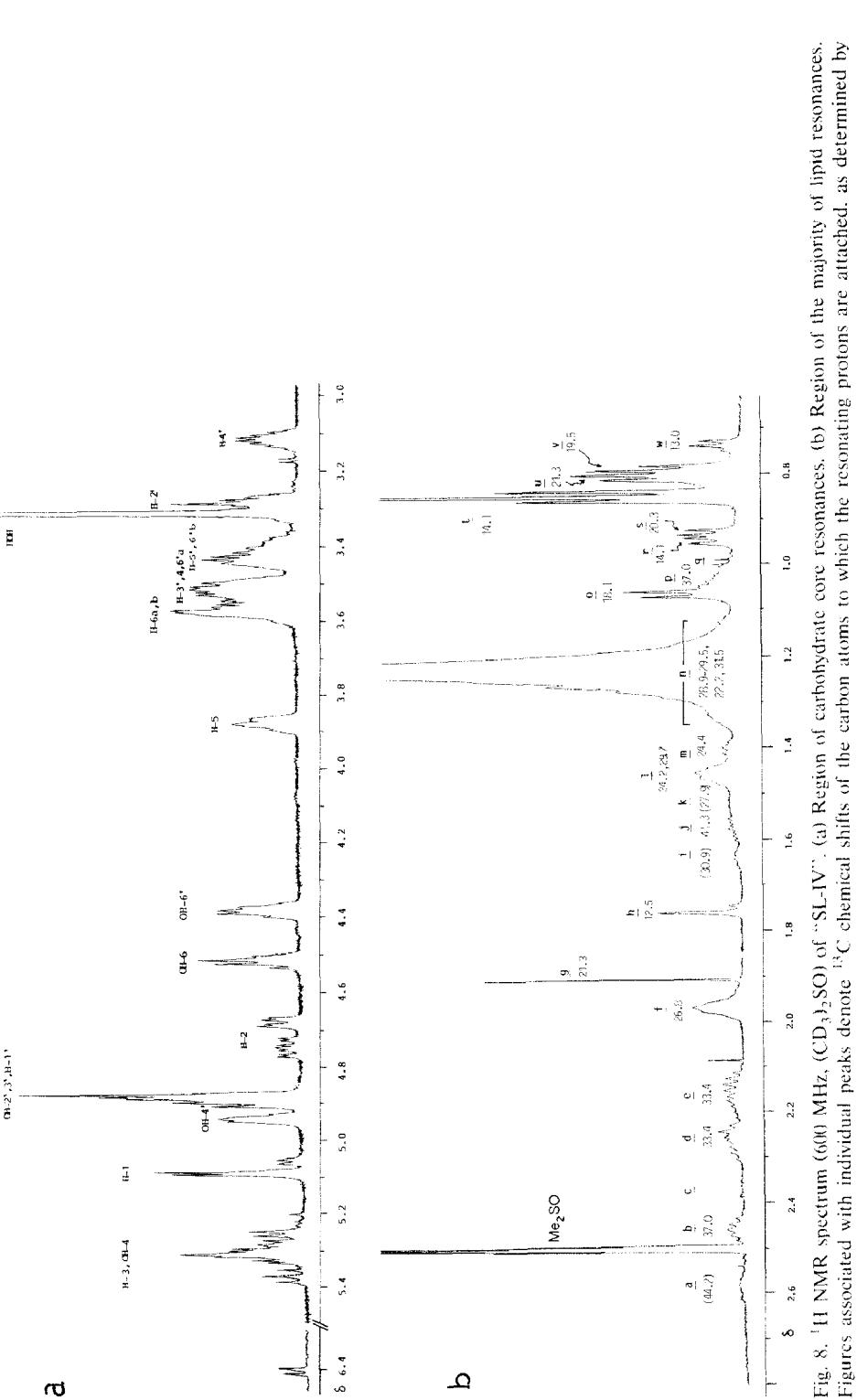
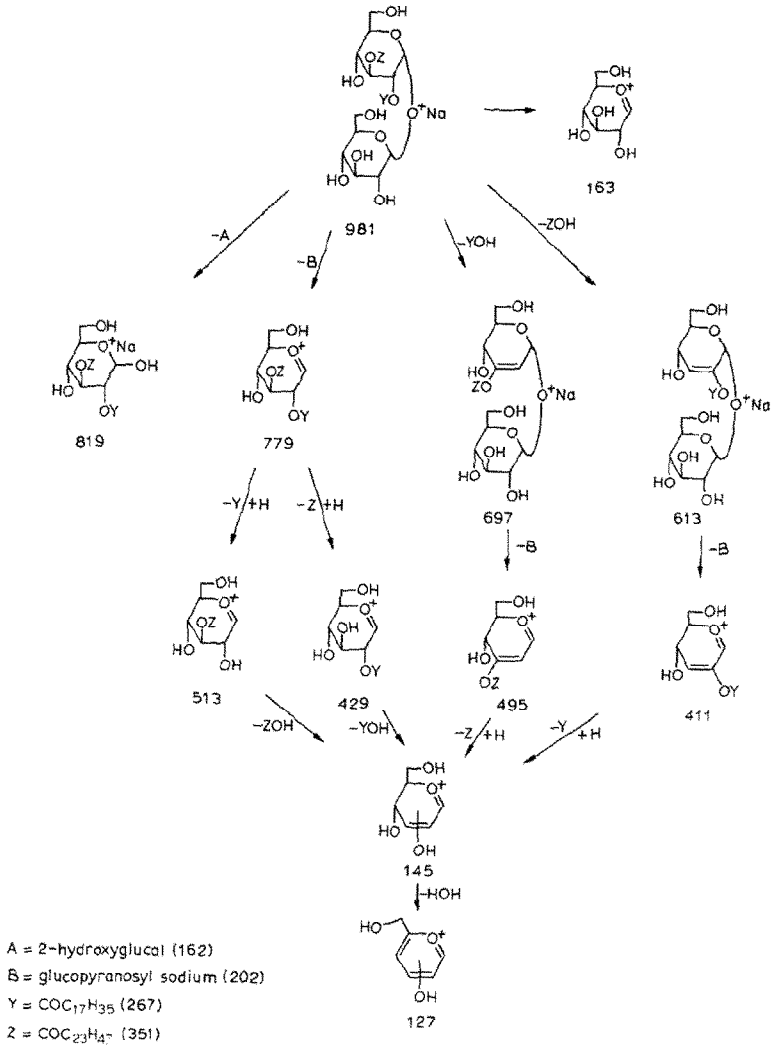


Fig. 8. ^{1}H NMR spectrum (600 MHz, $(CD_3)_2SO$) of "SL-IV". (a) Region of carbohydrate core resonances. (b) Region of the majority of lipid resonances. Figures associated with individual peaks denote ^{13}C chemical shifts of the carbon atoms to which the resonating protons are attached, as determined by HETCOR; figures in parentheses were obtained by a separate experiment, performed with a $CDCl_3$ solution.



Scheme 1.

Table II, together with those of synthetic 2,3-di-*O*-octadecanoyl-α,α-trehalose and unsubstituted α,α-trehalose for comparison. One notes that H-2 and H-3 appear well downfield from the other ring proton resonances, which signifies acylation of OH-2 and OH-3; small deshielding effects (0.2–0.4 ppm) are experienced by the proximate H-1, H-4, and H-5 protons, but not by the more remote protons H-6,a,b. It is noteworthy, furthermore, that both H-2 and H-3 gave clusters of signals with nearly identical coupling constants but slightly different chemical shifts, which was taken to indicate that "SL-IV" is a mixture of closely similar components bearing different acyl groups in each of these positions, as already mentioned. The COSY experiment established unambiguously that the three H-2 multiplets (which jointly

TABLE II

^{1}H NMR spectral data for the carbohydrate core of "SL-IV", 2,3-di-*O*-octadecanoyl-α,α-trehalose, and α,α-trehalose

Compound	Chemical shifts (δ) and *J* values (Hz, in parentheses)										
	H-1	H-2	H-3	H-4	H-5	H-6a	H-6b	OH-2	OH-3	OH-4	OH-6
"SL-IV" [a], acylated moiety	5.09d (3.4)	4.77–4.67 [b] (3.5, 10.3)	5.37–5.31 [c] (9.5–9.7)	3.52 [d]	3.88m	⊢—3.58 [d]—⊣				5.30bs	4.52t (5.9)
Unsubstituted moiety	4.88d (3.6)	3.28m	~ 3.55 [d]	3.11m	~ 3.4 [e]	~ 3.55 [d]	~ 3.4 [e]	⊢—4.89m[2H]—⊣		4.95d (5.0)	4.38t (5.0)
2,3-Di-*O*-octadecanoyl-α,α-trehalose [f], acylated moiety	5.09d (3.4)	4.68dd (3.4, 10.3)	5.31t (9.0, 10.3)	3.6–3.4 [g]	3.87m	⊢—3.6–3.4 [g]—⊣				5.32d (6.3)	4.55t (5.8)
Unsubstituted moiety	4.88d (3.65)	3.28m	3.6–3.4 [g]	3.12m	⊢—3.6–3.4 [g]—⊣			⊢—4.95d, 4.94d—⊣ (5.9)	(4.9)	5.00d (5.1)	4.43t (5.6)
α,α-Trehalose [f]	4.88d (3.7)	3.23sp (3.7, 6.3, 9.8)	3.55m (4.8, 8.8, 9.8)	3.14ddd (5.25, 8.6, 10.0)	3.65m	⊢—3.5m—⊣		4.61 (6.3)	⊢—4.78—⊣ (4.8, 5.3)		4.35t (5.9)
Per-*O*-acetyl-"SL-IV" [h], both moieties	5.25	5.0	5.50	5.0	3.9	4.2	3.9				
α,α-Trehalose octaacetate [h,i]	5.29	5.03	5.49	5.04	4.04	4.23	4.04				

[a] In $(CD_3)_2SO$ at 600 MHz. [b] Three dd; combined intensity, 1 H. [c] Overlapping triplets. [d] Part of unresolved 5H-multiplet at δ 3.6–3.5. [e] Part of unresolved 2H-multiplet at δ 3.43–3.35. [f] In $(CD_3)_2SO$ at 300 MHz. [g] Part of unresolved 7H-multiplet. [h] In $CDCl_3$ at 300 MHz. [i] Values taken from ref. 23.

TABLE III

^{13}C NMR chemical shifts (δ) for the carbohydrate core of "SL-IV", 2,3-di-*O*-octadecanoyl-α,α-trehalose, and α,α-trehalose

Compound	C-1	C-2	C-3	C-4	C-5	C-6
"SL-IV" [a], acylated moiety	90.8–90.6 [b]	~ 70.6 [b]	72.7, 72.4, 71.0	67.8, 67.6	72.4	60.0
Unsubstituted moiety	94.35	71.4	72.9	70.0	73.5	60.8
2,3-Di-*O*-octadecanoyl-α,α-trehalose [c], acylated moiety	90.7	70.5	72.1	67.5	72.5	59.9
Unsubstituted moiety	94.3	71.3	72.8	69.9	73.4	60.7
α,α-Trehalose [c]	93.4	72.0	72.8	70.5	73.5	61.2
Per-*O*-acetyl-"SL-IV" [d], both moieties	92.1–91.7 [e]	70.0–69.25 [f]		68.5–68.1 [b]		61.9–61.5 [g]
α,α-Trehalose octaacetate [h]	92.2	70.1	70.1	68.7	68.4	61.9

[a] In $(CD_3)_2SO$ at 150 MHz. [b] Three lines. [c] In $(CD_3)_2SO$ at 75.4 MHz. [d] In $CDCl_3$ at 75.4 MHz. [e] Four lines. [f] Three lines in ~3:2:1 intensity ratio. [g] Three lines in ~1.7:1:0.7 intensity ratio. [h] In $CDCl_3$ at 62.9 MHz; values taken from ref. 23.

integrated to 1 proton) were all coupled with H-1 (and not H-1′) and with the cluster of H-3 signals which, in turn, showed coupling with H-4 (and not H-4′); hence, all the ester groups were located in the same sugar unit. After peracetylation (acetic anhydride–pyridine, 18 h at 25°C) the pattern of sugar proton signals corresponded fully to that of α,α-trehalose octaacetate (see Table II).

The ^{13}C NMR data (150 MHz; assignments confirmed by a HETCOR experiment) for the sugar carbon atoms of "SL-IV", are given in Table III, also together with those for α,α-trehalose and its synthetic 2,3-dioctadecanoate. Note the slight shielding, by 2.3–3.5 ppm, of C-1 and C-4 in the 2,3-substituted moiety. The C-5 shift in "SL-IV" is little affected relative to free trehalose, whereas shielding of C-5 is increased by several ppm after peracetylation, as is the case for trehalose, as a result of substituents introduced in the 4- and 6-positions. Further, it is significant that C-1, C-2, C-3, and C-4 all give closely-spaced, multiple signals, which again suggests a nonuniform occupation of positions 2 and 3. Similar signal multiplication was observed in the peracetylated compound (see Table III). Somewhat surprising, however, was the presence of more than two signals for C-6 and C-6′ in the latter *.

In the field region mainly populated by carbohydrate proton signals (Fig. 8a), the ^1H NMR spectrum of "SL-IV" displayed some peculiar, additional signals, which had to be ascribed to certain structural elements of the ester groups. Thus, two small doublets at δ 6.40 (0.23 H, J 9.3 Hz) and 5.06 (0.17 H, J 5.8 Hz) doubtless represented the alkenic β-protons of geometrical isomers of phthienoyl residues (compare formula **4**). Their proportionate (1.5 : 1) and combined (0.4 H) intensities were mirrored in two upfield singlets (h and g in Fig. 8b) at δ 1.76 (0.75 H) and 1.91 (0.5 H), attributable to the α-methyl branches (together 0.4 CH_3) in these residues. Further, the group of overlapping signals for H-3 and OH-4 centered at δ 5.3 actually integrated to 2.6 H, and exhibited on its upfield border a multiplet that was revealed by COSY to be coupled with a multiplet at δ 1.97 (f in Fig. 8b) and was, therefore, extraneous to the carbohydrate core. These signals represented, respectively, the alkenic (H-9,10) and allylic (H-8,8′,11,11′) protons of the minor oleyl and palmitoleyl constituents; indeed, the allylic protons (δ 1.97) were coupled with their methylenic neighbors which resonated with the bulk of lipid CH_2 groups in the massive peak centered at δ 1.23 (n in Fig. 8b). From the intensities of the signals mentioned it was estimated that alkenoic acids contribute about 0.2–0.3 equiv to the 2 equiv of acids present in "SL-IV". The ^{13}C NMR resonances of the $-CH_2-CH{=}CH-CH_2-$ segment occurred at δ 130.0 and 26.7 in agreement with expectations (oleic acid: δ 129.9 and 27.2 in $CDCl_3$)[25]. Finally, the region of ill-resolved carbohydrate proton multiplets at 3.6–3.4 (H-3′,4,5′,6a,6b,6′a, and 6′b) integrated to 7.4 H, i.e., slightly more than the seven protons enumerated;

* Inspection of a molecular model constructed in accord with the established[24] C_2 symmetry of α,α-trehalose permits a possibility that C-6′ may experience different degrees of steric crowding depending on the structure of the C-2 acyloxy substituent.

it must have contained a minor, noncarbohydrate component because coupling existed with a lipid proton resonating at δ 2.45 (*b* in Fig. 8b). Similarly, the massive DOH peak at δ 3.31 apparently obscured an additional, small signal from a noncarbohydrate proton coupled to a minor lipid constituent resonating at δ 2.60 (*a* in Fig. 8b). These extra signals in the δ 3.5–3.3 range pointed to carbinol groupings incorporated in some of the ester chains. The presence of hydroxyacyl groups, already recognized by mass spectrometry, was also supported by NMR spectra of *acetylated* "SL-IV", not only through interpreting ^{1}H signals in the δ 5.50–5.25 region that were coupled with lipid resonances in the δ 2.0–1.0 region, and through extra ^{13}C signals (δ 71.9 and 70.8) attributable to secondary carbinol carbons, but more simply through the pattern of acetyl ^{1}H signals. This accorded with the presence of a certain proportion of (acetylated) hydroxy acids, with the integral for the δ 2.07–2.00 region indicating 20 protons, i.e., an excess over the number expected for a hexaacetate, and resolution enhancement of the cluster of acetyl signals revealed several minor peaks among the major peaks for the sugar acetyl groups *.

Approximate proportions of the various kinds of lipid ester groups.—The majority of lipid protons, except for the relatively small proportions of alkenic and carbinolic protons already mentioned, resonated between δ 2.6 and 0.7 and are labeled *a–w* in Fig. 8b. Also recorded in the Figure are the ^{13}C chemical shifts of the carbon atoms to which the protons are appended, as indicated by a HETCOR experiment. Most conspicuous is the large peak *n* centered at δ 1.23, enveloping the resonances of the many "inner" methylene groups (^{13}C: δ 29.2 ± 0.3) and including those of the penultimate and ultimate CH_2 groups (^{13}C: δ 31.5 and 22.2, respectively) of the lipid chains. The α-and β-methylene protons of the unbranched acyls, deshielded by proximity of ester carbonyl, gave multiplets near δ 2.2 (*d* and *e*; ^{13}C for C-2: δ 33.4) and 1.45 (*l*; ^{13}C for C-3: δ 24.2), respectively, showing geminal and vicinal coupling. (All the chemical shifts were practically identical with those found for synthetic trehalose hexadecanotes and octadecanotes). Signal *f* has already been discussed, and *a–c*, *i–k*, *m*, and *p* were assigned to methine protons in branched acyl groups (–C*H*Me–). The various methyl groups produced signals as follows. A large triplet (*t*) coupled with *n* represented the alkyl chain terminals (–CH_2CH_3). Seven doublets (*o*, *q*, *r*, *s*, *u*, *v*, and *w*) were due to methyl branches on saturated carbon [–CH(CH_3)–]. Of these signals, *r* and *s* were coupled with *b* and *a*, respectively, which implies that they belonged to hydroxy acids, whereas *u*, *v*, and *w* were coupled with higher-field methine protons in the range of δ 1.65–1.33 (*i* to downfield shoulder of *n*); doublet *o*

* A distinctive ^{1}H signal (< 2H) at δ 3.67 and a corresponding ^{13}C signal at δ 64.1 were suggestive of a CH_2OAc grouping additional to the trehalose terminals. This particular feature appeared variable, however; it was weak in one of the batches examined. It may have indicated an ω-hydroxyalkanoic acid (perhaps the putative hydroxydodecanoic acid revealed in the ion-spray flow injection mass spectrum, see above) or a hitherto unrecognized, hydroxymethyl-branched acid.

apparently represented two coincident methyl resonances as it showed coupling with methine signals *p* and *b*, with the implication that *b*, too, must have been a double signal. Finally, there were the two methyl singlets *g* and *h* attributed to the segment $-CH=C(CH_3)-CO-$ of phthienoyl residues as previously mentioned.

By careful evaluation of the integrals for the lipid protons in several independent 300- and 600-MHz ^1H NMR spectra we determined an 18: ~ 75 ratio of methyl protons to secondary (CH_2) plus tertiary ($-CH-$ and $-CH=$) protons, equivalent to 6 methyl groups (2 chain terminals and 4 branches) and ~ 37.5 nominal methylene groups (with $-CH-$ and $-CH=$ each counted as one-half CH_2). This ratio tells us that the combined branched-chain acids must be present in greater than equimolar proportion relative to the combined C_{14}–C_{19} acids, as the following argument will show. If one takes $CO(CH_2)_{15}CH_3$ as an approximation for the *average* composition (*A*) of the lower fatty acid residues (which appears well justified in light of the mass spectral observations; the small proportions of C-10 methyl branches are neglected), then an equimolar proportion of combined, higher acyl groups would have to have an average composition (*B*) of $CO(CHCH_3)_4(CH_2)_{21}CH_3$, with 5 CH_3 groups and 23 nominal CH_2 groups, corresponding to a tetramethylheptacosanoic acid, to account for the observed integrals. Although small amounts of such a C_{31} species were seen by MS to be present, the strongly preponderant components are the C_{24} dimethyl acid **3** and the C_{27} trimethyl acid **5**, present in comparable proportions. An average composition (*C*) of these would be $CO(CHCH_3)_{2.5}(CHOH)_{0.5}(CH_2)_{18}CH_3$, with 3.5 CH_3 and 19.5 nominal CH_2 groups. A combination of 0.4 equiv of *A* (0.4 CH_3 + 6 CH_2) and 1.6 equiv of *C* (5.6 CH_3 + 31 CH_2) closely reflects the observed proton ratio.

EXPERIMENTAL

Materials.—Three samples (5–10 mg each) of different batches of "SL-IV", isolated and purified[8–10] in the Pasteur Institute, were received as colorless, viscid films and used as supplied. Methanolysis was performed in 1:1 MeOH–$CHCl_3$ containing a catalytic amount of NaOMe (16 h at 25°C); the mixture was then diluted with MeOH, deionized with Amberlite IR-120(H^+) cation-exchange resin, and evaporated. Alkaline hydrolysis of the glycolipid was effected with 2 M KOH in aq 80% EtOH (16 h at 90°C); the cooled hydrolyzate was carefully neutralized with HCl, concentrated, adjusted to pH 3 with formic acid, and evaporated to dryness.

Spectroscopy.—Chemical-ionization mass spectra of the methanolyzed and hydrolyzed samples were obtained with a VG Instruments (Manchester) mass spectrometer, model 707OE, using ether as the ionizing gas. Gas chromatography–mass spectrometry was performed with the same instrument, in conjunction with a 30-m, DB5 Megabore column (J & W Scientific, Inc.), operating at a temperature gradient from 150 to 320°C; peak analysis was conducted by electron impact

ionization at 70 eV, with 6 kV acceleration. The mixture of carboxylic acids was treated with *N*,*O*-bis(trimethylsilyl)acetamide prior to injection. For FAB mass spectrometry, a Kratos Concept 2H instrument operating at 8 kV acceleration was used, employing Cs ion bombardment at 12 kV acceleration. Ion-spray flow injection spectra were obtained at room temperature using an API-III spectrometer of Perkin–Elmer Sciex Instruments, Thornhill, ON.

The 600-MHz ^{1}H (and 150-MHz ^{13}C) NMR spectra of "SL-IV" were run on a Bruker AMX-600 instrument, and other spectra were taken with a Varian XL-300 spectrometer.

ACKNOWLEDGMENTS

This study was undertaken at the suggestion of Dr. F. Papa, Chef de Laboratoire, Unité de la Tuberculose et des Mycobactéries, Institut Pasteur, Paris, who kindly supplied the glycolipid. The author is indebted to Dr. C. Kazakoff for producing most of the mass spectra in the mass-spectroscopic facility of this Department; ion-spray spectra were obtained by Perkin–Elmer Sciex Instruments, courtesy of Dr. B. Shushan. Dr. A. Williams and Mr. R. Capoor of this Department are thanked for taking the NMR spectra, and the National Research Council of Canada, Ottawa, is thanked for allowing access to its 600-MHz instrument operated by Dr. J.R. Brisson. The author also greatly appreciated the interest expressed in this work by Dr. M.B. Goren, Denver, Colorado, who provided valuable samples of various mycobacterial glycolipids that proved useful for purposes of comparison.

REFERENCES

1 C. Asselineau and J. Asselineau, *Prog. Chem. Fats Other Lipids*, 16 (1978) 59–99.
2 M.B. Goren, in G.P. Kubica and L.G. Wayne (Eds.), *The Mycobacteria – A Source Book*, Marcel Dekker, New York, 1984, pp. 379–415.
3 M.B. Goren, in M. Kates (Ed.), *Handbook of Lipid Chemistry*, Vol. 6, Plenum, New York, 1990, pp. 363–461.
4 M.B. Goren, O. Brokl, P. Roller, H.M. Fales, and B.C. Das, *Biochemistry*, 15 (1976) 2728–2735.
5 M.B. Goren, B.C. Das, and O. Brokl, *Nouv. J. Chim.*, 2 (1978) 379–384.
6 D.E. Minnikin, G. Dobson, D. Sesardic, and M. Ridell, *J. Gen. Microbiol.*, 131 (1985) 1369–1374.
7 M. Daffé, C. Lacave, M.-A. Lanéelle, M. Gillois, and G. Lanéelle, *Eur. J. Biochem.*, 172 (1988) 579–584.
8 F. Papa, A. Laszlo, and H.L. David, *Ann. Inst. Pasteur, Microbiol.*, 139 (1988) 535–545; F. Papa, A. Laszlo, H.L. David, and M. Daffé, *Proc. 4th Internat. Colloq. Mycobacteria*, in *Acta Leprol.* 7, Suppl. 1, (1989) 98–101.
9 M. Daffé, F. Papa, A. Laszlo, and H.L. David, *J. Gen. Microbiol.*, 135 (1989) 2759–2766.
10 F. Papa, P. Cruaud, and H.L. David, *Res. Microbiol. (Inst. Pasteur)*, 140 (1989) 569–578.
11 P. Cruaud, C. Berlie, J. Torgal Garcia, F. Papa, and H.L. David, *Zbl. Bakteriol.*, 271 (1989) 481–485.
12 P. Cruaud, J.T. Yamashita, N. Martin Casabona, F. Papa, and H.L. David, *Res. Microbiol. (Inst. Pasteur)*, 141 (1990) 679–694.
13 A. Laszlo, H.H. Baer, M.B. Goren, V. Handzel, L. Barrera, and I.N. deKantor, *Res. Microbiol. (Inst. Pasteur)*, 143 (1992) 217–233.

14 A. Lemassu, M.-A. Lanéelle, and M. Daffé, *FEMS Microbiol. Lett.*, 78 (1991) 171–176.
15 (a) S. Honda, H. Yuki, and K. Takiura, *Carbohydr. Res.*, 28 (1973) 150–153; (b) G.K. Hamer and A.S. Perlin, *ibid.*, 49 (1976) 37–48; (c) P.J. Archbald, M.D. Fenn, and A.B. Roy, *ibid.*, 93 (1981) 177–190; (d) N. Lupescu, S. (M.) Arad, S. Geresh, M.A. Bernstein, and R. Glaser, *ibid.*, 210 (1991) 349–352.
16 K. Bock, C. Pedersen, and H. Pedersen, *Adv. Carbohydr. Chem. Biochem.*, 42 (1984) 193–225.
17 M.B. Goren, O. Brokl, B.C. Das, and E. Lederer, *Biochemistry*, 10 (1971) 72–81.
18 G. Odham and E. Stenhagen, in G.R. Waller (Ed.), Biochemical Applications of Mass Spectrometry. Wiley, New York (1972), pp. 211–228; G. Odham, *ibid.*, Suppl. 1 (1980), pp. 153–210.
19 L. Larsson, P.-A. Mårdh, and G. Odham, *J. Chromatogr., Biomed. Appl.*, 163 (1979) 221–224.
20 R. Ryhage, S. Ställberg-Stenhagen, and E. Stenhagen, *Arkiv Kemi*, 18 (1961) 179–194.
21 L. Coles and N. Polgar, *J. Chem. Soc., C*, (1968) 1541–1549.
22 A. Dell, *Adv. Carbohydr. Chem. Biochem.*, 45 (1987) 19–72.
23 D.Y. Gagnaire, F.R. Taravel, and M.R. Vignon, *Carbohydr. Res.*, 51 (1976) 157–168.
24 G.M. Brown, D.C. Rohrer, B. Berkin, C.A. Beevers, R.O. Gould, and R. Simpson, *Acta Crystallogr.*, B 28 (1972) 3145–3158; T. Taga, M. Senma, and K. Osaki, *ibid.*, B 28 (1972) 3258–3263.
25 L.F. Johnson and W.C. Jankowski, *Carbon-13 NMR Spectra*, John Wiley & Sons, New York (1972), spectrum No. 480.