Deciphering sulfoglycolipids of Mycobacterium tuberculosis

Emilie Layre, Diane Cala-de Paepe, Gerald Larrouy-Maumus, Julien Vaubourgeix, Sathish Mundayoor, Buko Lindner, Germain Puzo, Martine Gilleron

Deciphering sulfoglycolipids of *Mycobacterium tuberculosis*

Emilie Layre,[1,2,*,†] Diane Cala-De Paepe,[1,*,†] Gérald Larrouy-Maumus,[*,†] Julien Vaubourgeix,[3,*,†] Sathish Mundayoor,[§] Buko Lindner,[**] Germain Puzo,[*,†] and Martine Gilleron[4,*,†]

* CNRS; IPBS (Institut de Pharmacologie et de Biologie Structurale); 205 route de Narbonne, F-31077, Toulouse, France; UPS,[†] Université de Toulouse, UPS; IPBS; F-31077 Toulouse, France; Rajiv Gandhi Centre for Biotechnology,[§] Trivandrum, Kerala, India; and Division of Biophysics,[**] Research Center Borstel, Center for Medicine and Biosciences, Borstel, Germany

Abstract For 4 decades, in vivo and in vitro studies have suggested that sulfoglycolipids (SGLs) play a role in the virulence or pathogenesis of the tubercle bacilli. However, the SGL structure and biosynthesis pathway remain only partially elucidated. Using the modern tools of structural analysis, including MALDI-time-of-flight MS, MS/MS, and two-dimensional NMR, we reevaluated the structure of the different SGL acyl (di-, tri-, and tetra-acylated) forms of the reference strain *Mycobacterium tuberculosis* H37Rv, as well as those produced by the *mmpL8* knockout strains previously described to intracellularly accumulate di-acylated SGL. We report here the identification of new acyl forms: di-acylated SGL esterified by simple fatty acids only, as well as mono-acylated SGL bearing a hydroxyphthioceranoic acid, which were characterized in the wild-type strain. In a clinical strain, a complete family of mono-acylated SGLs was characterized in high abundance for the first time. For the *mmpL8* mutant, SGLs were found to be esterified *i*) by an oxophthioceranoic acid, never observed so far, and *ii*) at nonconventional positions in the case of the unexpected tri-acylated forms. Our results further confirm the requirement of MmpL8 for the complete assembly of the tetra-acylated forms of SGL and also provide, by the discovery of new intermediates, insights in terms of the possible SGL biosynthetic pathways.—Layre, E., D. Cala-De Paepe, G. Larrouy-Maumus, J. Vaubourgeix, S. Mundayoor, B. Lindner, G. Puzo, and M. Gilleron. **Deciphering sulfoglycolipids of *Mycobacterium tuberculosis*. *J. Lipid Res.* 2011.** 52: **1098–1110.**

Supplementary key words acylation • biosynthesis • glycolipids • lipids • mycobacteria • structure elucidation • sulfolipids

Sulfoglycolipids (SGLs) were originally discovered in 1959 by Middlebrook et al. (1) in human (H37Rv) and bovine (Vallée) virulent strains of *Mycobacterium tuberculosis*. Almost 2 decades later, Goren et al. (2) showed that this family of lipids was specific to *M. tuberculosis* and proposed that their amount correlated with the virulence of the *M. tuberculosis* strain in the guinea pig model of infection. Sulfolipids consist of multi-acylated forms of trehalose sulfate that differ by the number, structure, and location of acyl moieties (3). The major sulfolipid, SL-I, was characterized as 2,3,6,6′-tetraacyl-α-α'-trehalose-2′-sulfate acylated by two hydroxyphthioceranates (HPA), one phthioceranate (PA), and one palmitate (C_{16}) or stearate (C_{18}) (**Fig. 1A**). HPA are complex dextrorotatory fatty acids specific to the *Mycobacterium* genus and contain one hydroxyl group and several methyl groups arranged in a 2,4,6 pattern. PA are HPA analogs devoid of the hydroxyl group (3–5). Three other minor tetra-acylated forms of sulfolipids, differing from SL-I by the fatty acyl composition, namely, SL-I′, SL-II, and SL-II′, and one tri-acylated form, SL-III, have been described previously (Fig. 1B) (6). Later on, a compound called SL-IV and tentatively assigned to a 2,3-diacyltrehalose-2′-sulfate was isolated from clinical isolates of *M. tuberculosis* (7) and proposed to correspond to the more polar

Support for the Institut de Pharmacologie et de Biologie Structurale (IPBS) research group was provided by the European Union (Cluster for a tuberculosis vaccine, QLK-CT-1999-01093). The IPBS NMR equipment was financed by French research ministry, CNRS, Université Paul Sabatier, the Région Midi-Pyrénées, and the European structural funds.

Manuscript received 7 December 2010 and in revised form 28 February 2011.

Published, JLR Papers in Press, April 11, 2011
DOI 10.1194/jlr.M013482

Abbreviations: DAT, 2,3-di-acyl trehalose; F, fractions; Glc*p*, glucopyranosyl unit; HPA, hydroxyphthioceranoic acid; OPA, oxophthioceranoic acyl; PA, phthioceranoic acid; PAC_n, phthioceranoic acid with *n* carbon atoms; PIM, phosphatidyl-*myo*-inositol mannosides; QMA, quaternary methyl ammonium; RP-HPLC, reverse-phase HPLC; SGL, sulfoglycolipid; SL, sulfolipid; TAT, 2,3,6-tri-acyl trehalose; TOF, time-of-flight; WT, wild type.

[1] E. Layre and D. Cala-De Paepe contributed equally to this work.

[2] Present address of E. Layre: Harvard Medical School Brigham and Women's Hospital, 1 Jimmy Fund Way, Smith 514, Boston, MA 02115.

[3] Present address of J. Vaubourgeix: Department of Microbiology and Immunology, Weill Cornell Medical College, New York, NY 10065.

[4] To whom correspondence should be addressed.
e-mail: Martine.Gilleron@ipbs.fr

[S] The online version of this article (available at http://www.jlr.org) contains supplementary data in the form of two tables and five figures.

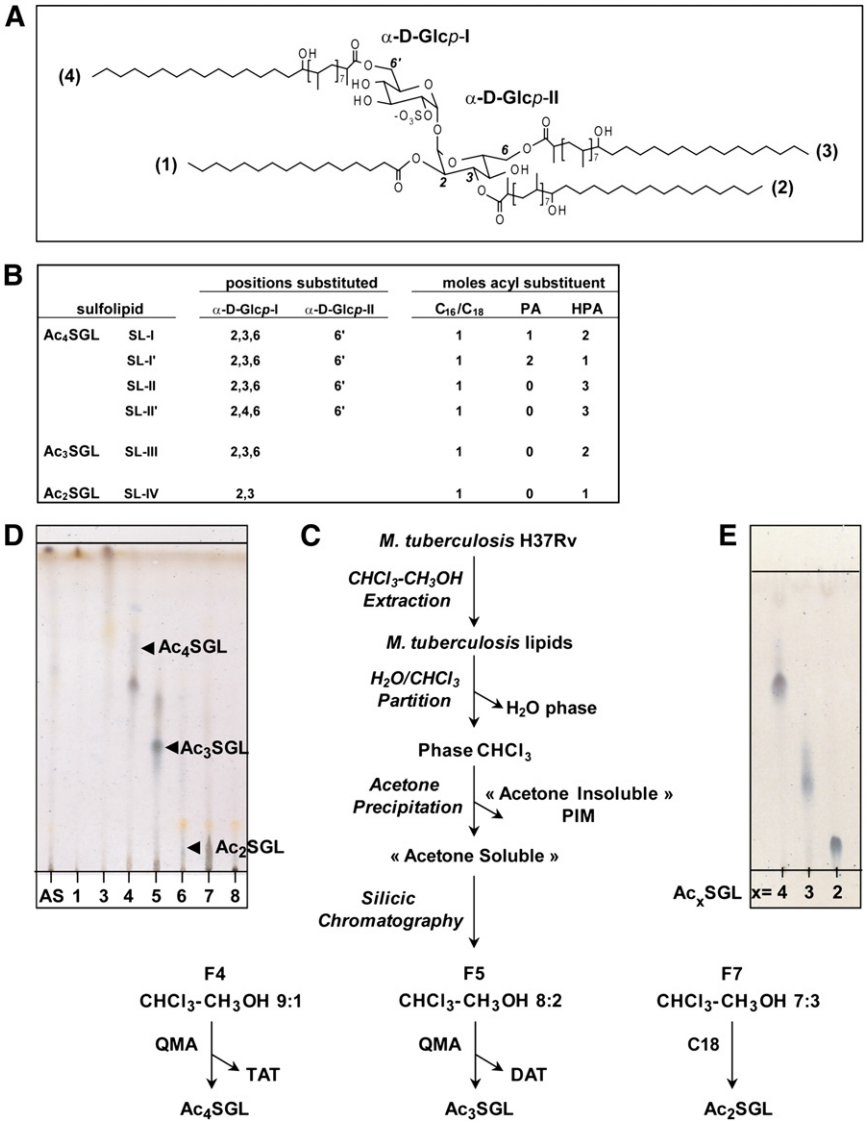


Fig. 1. Purification of Ac_2-, Ac_3-, Ac_4SGL of *M. tuberculosis* H37Rv. A: Structure of the major *M. tuberculosis* H37Rv Ac_4SGL is shown. Position (1) is always acylated by a palmitic or a stearic acid, while positions (2), (3), and (4) are acylated mainly by HPA but also by PA. Ac_2SGL are acylated in positions (1) and (2) , and Ac_3SGL are acylated in positions (1), (2), and (3). B: Structure of SL-I to SL-IV. C: Purification schemes are shown for the different acyl forms of SGL. D: TLC analysis of the different fractions issued from the silicic column eluted by $CHCl_3$ (1, 2), $CHCl_3/CH_3OH$ 9/1 (3, 4), $CHCl_3/CH_3OH$ 8/2 (5, 6), $CHCl_3/CH_3OH$ 7/3 (7, 8), and CH_3OH (9). AS, acetone-soluble fraction. F4 to F7 contained SGL. E: TLC analysis of the different acyl forms of SGL after purification. All the TLC were developed with $CHCl_3/CH_3OH$, 85:15, v/v, and sprayed with anthrone.

uncharacterized sulfatide observed by Goren et al. (8) in the *M. tuberculosis* H37Rv strain. However, this compound was subsequently found to lack a sulfate ester group (9, 10), requiring its classification as a sulfatide to be revised (11). Nevertheless, the presence of di-acylated sulfolipids in the cell envelope of *M. tuberculosis* was definitively proven by three recent studies. Using Fourier transform ion cyclotron resonance MS to identify sulfated metabolites by virtue of their metabolic labeling with a stable sulfur isotope, Mougous et al. (12) showed evidence of several new sulfated molecules in *M. tuberculosis*, among them di-acylated sulfolipids. Studying the role of the MmpL8-mediated lipid transport in sulfolipid biogenesis, Domenech et al. (13) characterized polar sulfated molecules accumulating in the knockout mutant strain and characterized them as di-acylated sulfolipids. Finally, we identified di-acylated sulfolipids, renamed SGLs (e.g., Ac_2SGL [in Ac_XSGL, *X* refers to the total number of acyl groups, whatever the nature of the fatty acids, which could be either palmitic, stearic, hydroxyphthioceranoic, phthioceranoic, or oxophthioceranoic acids]), as new lipidic antigens presented by CD1b to T cells (14). The structure of Ac_2SGL was for the first time fully established by a combination of MS and NMR analyses (14).

In the present study, we reevaluated the structure of SGLs from the reference strain *M. tuberculosis* H37Rv, as well as those produced by the *mmpL8* knockout strains that have been described to intracellularly accumulate

Ac_2SGL (13, 15), by using MALDI-time-of-flight (TOF)-MS and MS/MS and two-dimensional NMR methods. The pattern of SGLs present in a clinical strain was compared with that in *M. tuberculosis* H37Rv, and the question of the presence of SGLs in nontuberculous mycobacteria was also addressed. Results are discussed in the context of possible SGL biosynthetic pathways.

MATERIALS AND METHODS

Bacterial strain and culture conditions

M. tuberculosis H37Rv (American Type Culture Collection no. 27294), clinical *M. tuberculosis* (Chemical Abstract Service [CAS] isolate RGTB264) (16), Δ*mmpL8* jcm108 *M. tuberculosis* Erdman strain (15), and *mmpL8*::*hyg M. tuberculosis* H37Rv Pasteur strain (13) were grown at 37°C on Sauton's medium as surface pellicles. Cells were harvested after 4 weeks, separated from the culture medium, and killed by incubation in a chloroform-methanol (2:1, v/v) solution for 2 days at room temperature.

Purification of SGL from *M. tuberculosis* H37Rv

Bacterial cells (25 g) were suspended in a chloroform-methanol (1:1, v/v) solution and filtered four times. The chloroform-methanol extract, which constituted the whole lipid extract (16.2 g), was concentrated and further partitioned between water and chloroform. Both phases were evaporated (water phase, 12g, and chloroform phase, 4.3 g). The chloroform phase was then dissolved in a minimum volume of chloroform and allowed to precipitate by the addition of acetone overnight at 4°C. The precipitate was centrifuged (3,000 *g* at 4°C for 15 min) to generate both an "acetone-soluble" phase (2 g) and an "acetone-insoluble" phase (1.7 g). Part of the acetone-soluble phase (1 g) was fractionated on a silica column (22 × 2 cm) irrigated successively with 70 ml of chloroform (fractions [F] 1 and 2) and chloroform containing 10% (F3 and F4), 20% (F5 and F6), and 30% (F7 and F8) methanol. F5 (20 mg) and F4 (90 mg) contained Ac_3- and Ac_4SGL, respectively; F7 (50 mg) contained Ac_2SGL; and F6 (75 mg) contained a mixture of Ac_2SGL and Ac_3SGL.

F4 was further purified by anion exchange quaternary methyl ammonium (QMA) chromatography, using a Sep-pak® Light cartridge (Waters Corp., Milford, CT) by eluting 3 times with 2 ml of chloroform (F4.1 to F4.3), 3 times with 2 ml of chloroform-methanol (8:2, v/v) (F4.4 to F4.6), 2 times with 2 ml of methanol (F4.7 and F4.8), and finally, 10 times with 2 ml of chloroform-methanol (1:1, v/v) containing 0.2 M ammonium acetate (F4.9 to F4.18). Then, 24 mg of Ac_4SGL was purified by pooling F4.11 with F4.17, while 2,3,6-tri-acyl trehalose (TAT) was recovered from F4.4 and F4.5.

F5 was further purified by anion exchange QMA chromatography with a Sep-pak® cartridge by eluting with 1 ml of chloroform (F5.1), two times with 2 ml of chloroform-methanol (8:2, v/v) (F5.2 and F5.3), two times with 2 ml of methanol (F5.4 and F5.5), two times with 2 ml of chloroform-methanol (8:2, v/v) containing 0.1 M ammonium acetate (F5.6 and F5.7), two times with 2 ml of chloroform-methanol (1:1, v/v) containing 0.1 M ammonium acetate (F5.8 and F5.9), four times with 2 ml of chloroform-methanol (1:1, v/v) containing 0.2 M ammonium acetate (F5.10 to F5.13), and three times with 2 ml of chloroform-methanol (1:1, v/v) containing 0.5 M ammonium acetate (F5.14 to F5.16). Less than 1 mg of Ac_3SGL was recovered from F5.12 to F5.15, whereas 2,3-di-acyl trehalose (DAT) was found in fractions F5.2 to F5.4.

F7 was further purified by reverse phase chromatography using a Sep-pak® Light C_{18} cartridge (Waters Corp.) eluted three times with 2 ml of methanol-water (9:1, v/v) (F7.1 to F7.3), methanol (F7.4 to 7.6), and chloroform-methanol (9:1, v/v) (F7.7 to F7.9). One milligram of Ac_2SGL was obtained by combining F7.4 with F7.6.

Purification was checked first by TLC on aluminum-backed silica gel plates (Alugram Sil G; Macherey-Nagel, Duren, Germany), using a migration solvent system of chloroform-methanol 9:1 (v/v). Orcinol or anthrone was used to detect carbohydrate-containing lipids. Purification was also checked by MALDI-TOF-MS analysis in both positive- and negative-ion modes.

Analysis of SGL from the clinical *M. tuberculosis* CAS isolate RGTB264

A chloroform-methanol extract was prepared as previously described, further dissolved in a minimum volume of chloroform and allowed to precipitate by the addition of acetone overnight at 4°C. The precipitate was centrifuged (3,000 *g* at 4°C for 15 min) to generate both an acetone-soluble phase and an acetone-insoluble phase. The acetone-soluble phase was concentrated and purified with a Sep-pak® Light silica cartridge and eluted six times with 1 ml of chloroform (F1 to F6) and then in chloroform containing 10% (F7 to F72), 20% (F13 to F18), and 30% (F19 to F24) methanol. F 9–10, F13–14, F16–18, and F20–22 were pooled and contained Ac_4SGL to Ac_1SGL, respectively.

Evaluation of SGL presence in strains other than *M. tuberculosis* H37Rv

A chloroform-methanol extract was prepared from each of the following strains, *Mycobacterium bovis* bacillus Calmette-Guerin Pasteur, *M. tuberculosis* H37Ra, *Mycobacterium chelonae*, *Mycobacterium fortuitum*, *Mycobacterium gastri*, *Mycobacterium kansasii*, *Mycobacterium marinum*, *Mycobacterium smegmatis* mc^2155, and *Mycobacterium xenopi*, and concentrated as described previously for *M. tuberculosis* H37Rv cells. For each extract, an acetone-soluble phase was prepared as previously explained and analyzed by MALDI-MS for SGL content.

Reverse-phase HPLC

HPLC was performed using an Atlantis reverse-phase (C_{18}) column (5 µm, 4.6 mm × 250 mm) (Waters Corp.), and lipids were eluted for a period of 1 h at a flow rate of 1 ml/min, using an elution solvent gradient in the range of 0%–100% solvent B in solvent A, with solvent A consisting of 2% water in methanol and solvent B consisting of 30% dichloromethane in methanol. A 250 µg solution of Ac_2SGL was solubilized in dichloromethane-methanol (4:1, v/v) and injected. Sixty 1 ml fractions were collected, and their contents were analyzed by MALDI-TOF-MS in negative-ion mode.

Reduction of Ac_2SGL from the *M. tuberculosis* H37Rv and MmpL8 mutant strains

A 20 µg aliquot of Ac_2SGL was reduced at room temperature for 1 h in NH_4OH-EtOH (1:1, v/v) containing 10 mg/ml $NaBD_4$. The reaction was quenched by the addition of 2 droplets of acetic acid. After evaporation of the solvent, the reduced Ac_2SGL was solubilized with chloroform and washed three times with water. The dry residue was dissolved in 20 µl of chloroform-methanol (8:2, v/v) prior to MALDI-TOF-MS analysis.

MALDI-TOF-MS analysis

Analysis by MALDI-TOF-MS was carried out with a 4700 model proteomics analyzer (Voyager DE-STR unit with TOF-TOF optics; Applied Biosystems, Framingham, MA) using the reflectron mode. Ionization was achieved by irradiation with an Nd:YAG laser (355 nm) operating at pulses of 500 picoseconds with a frequency of 200 Hz. Glycolipids were analyzed in the

negative- or positive-ion mode, as specified in the text. Spectra from 2,500 to 5,000 laser pulses were summed to obtain the final spectrum. The 2-(4-hydroxyphenylazo)benzoic acid (HABA) or 2,5-dihydroxybenzoic acid (DHB) matrix (Sigma) was used at a concentration of ~10 mg/ml in ethanol-water (1:1, v/v). In a typical experiment, 4 μl of glycolipid (5–10 μg) in chloroform-methanol (8:2, v/v) and 4 μl of the matrix solution were mixed with a micropipette, and 0.3 μl of the mixture was then deposited on the target. The measurements were internally calibrated at two points with mycobacterial phospholipids (phosphatidyl-*myo*-inositol mannosides [PIM]).

ESI-FT-MS analysis

SGLs were analyzed by high-resolution Fourier-transform ion cyclotron resonance MS using a hybrid apex-Qe FT-MS system (Bruker Daltonics, Bremen, Germany) equipped with a 7 Tesla actively shielded superconducting magnet and an Apollo dual-ESI/MALDI ion source. Samples (~10 ng • μl^{-1}) were dissolved in a 50:50:0.001 (v/v/v) mixture of 2-propanol, water, and triethylamine (pH ~8.5) and sprayed at a flow rate of 2 μl • min^{-1}. The capillary entrance voltage was set to 3.8 kV, and the drying temperature was set to 150°C. Data acquisition and analysis were performed using Apex Control version 3.0 and data analysis version 3.4 software (Bruker Daltonics, Bremen), respectively. The mass spectra were acquired in the negative-ion mode with 1 M data points in the mass range between 1,020 and 1,360 atomic mass units. Mass scale calibration was performed externally with a mixture of lipids with known structures. The mass resolution obtained was above 130,000, and the mass accuracy was in the range of 1.5 ppm.

NMR analysis

NMR spectra were recorded with an Avance DMX500 spectrometer (Bruker GmbH, Karlsruhe, Germany) equipped with an Origin 200 SGI, using Xwinnmr version 2.6 software. Native SGLs were dissolved in $CDCl_3$-CD_3OD (4:1, v/v) and analyzed in 200 × 5 mm 535 PP NMR tubes at 295 K. Proton chemical shifts are expressed in ppm downfield from the signal of chloroform (δ_H/TMS 7.27 and δ_C/TMS 77.7). All details concerning the use of ^1H-^1H COrrelation SpectroscopY (COSY), ^1H-^1H Homonuclear Hartman Hahn (HOHAHA), and ^1H-^{13}C Heteronuclear Multiple Quantum Coherence (HMQC) sequences and experimental procedures were as reported previously (17).

RESULTS

A purification method for the different SGL acyl forms of *M. tuberculosis* H37Rv

The first objective of this study was to design a protocol to prepare fractions containing pure and unique SGL acyl forms, namely the tetra-, tri-, and di-acylated SGL forms (Ac_4-, Ac_3-, and Ac_2SGL, respectively) from *M. tuberculosis* H37Rv lipid extract. The purification was monitored by negative-ion MALDI-TOF-MS. Following acetone precipitation, the three acyl forms were detected in the acetone-soluble fraction (Fig. 1C) (14). This fraction was further fractionated on a silicic acid column irrigated by chloroform containing increasing amounts of methanol. Ac_4SGL were recovered from F4, Ac_3SGL from F5 and F6, and Ac_2SGL from F6 and F7. However, F4 to F6 contained additional compounds, as revealed by TLC analysis (Fig. 1D) and positive-ion mode MALDI-TOF-MS (not shown), which highlighted the presence of neutral compounds. The neutral compounds were eliminated by QMA anion exchange chromatography and their structures were elucidated by COSY and HOHAHA NMR analysis (data not shown). The neutral product in F4, whose positive mass spectrum was dominated by two pseudo molecular ions, $[M+Na]^+$ at *m/z* 1,344 and 1,372, was assigned as a 2,3,6-triacyl trehalose (TAT). The neutral compound in F5 was characterized by the major pseudo molecular ion $[M+Na]^+$ at *m/z* 981, corresponding to 2,3-di-acyl trehalose (DAT) (18). F7 did not contain neutral lipids but did contain a dye that was separated from Ac_2SGL by reverse-phase chromatography (14).

From 8 g of *M. tuberculosis* H37Rv cells (3 liters of culture), approximately 25 mg of Ac_4SGL, which is the most abundant of the SGL acyl forms, less than 1 mg of Ac_3SGL, was obtained, and from 1 to 3 mg of Ac_2SGL was usually recovered, using this purification strategy (Fig. 1E).

MALDI-TOF-MS reveals the heterogeneity of SGL acyl forms

Negative-ion mode MALDI-TOF-MS of purified SGLs (Fig. 1E) proved to be the method of choice for characterizing the heterogeneity in acylation of the different SGL species. Indeed, for Ac_2SGL, the mass spectrum (**Fig. 2A**) exhibited a set of peaks ranging from *m/z* 1,067.7 to 1,446.1, with two major molecular ions, $[M-H]^-$ at *m/z* 1,249.9 and *m/z* 1,277.9. The relative intensities of these ions varied from one culture batch to another, depending on the culture duration (see supplementary Fig. IA, B). The mass range indicated that HPA contained 25 to 54 carbon atoms (**Table 1**). However, despite this major series of ions, peaks of much weaker intensities were also observed. For instance, the major molecular ion at *m/z* 1,277.9 dominated another one at *m/z* 1,275.9 (ratio, 15:1) (Fig. 2B), which was tentatively assigned as Ac_2SGL containing one phthioceranoic acid with 43 carbon atoms (PAC_{43}) and one palmitic acid (Table 1). The relative intensity of this series of peaks increases in the *m/z* range of 1,120–1,240 (Fig. 2C and Table 1), in agreement with those found in the study by Goren (6), in which PAC_{34} and PAC_{37} are the major PA released from SGLs. To confirm this assignment, an exact mass measurement was performed by using ESI-FT-MS analysis, where the resolution obtained was above 130,000, and the mass accuracy was in the range of 1.5 ppm. The molecular ion at *m/z* 1,277.9 showed only one molecular species at *m/z* 1,277.922, agreeing with the calculated molecular mass of 1,277.926 Da for an Ac_2SGL containing one $HPAC_{40}$ or one $HPAC_{42}$ (see below). Weak intensity of the molecular ion at *m/z* 1,275.9 prevented an accurate mass measurement; however, a fraction enriched for this species obtained by HPLC fractionation of Ac_2SGL (see below) allowed measurement of an accurate mass at *m/z* 1,275.945, in agreement with the presence of PAC_{43} within the molecule (calculated molecular mass of 1,275.947 Da) (Table 1). Nevertheless, molecular species bearing PA represent less than 15% of Ac_2SGL, with the majority of them containing HPA.

The negative-ion mode MALDI mass spectrum of Ac_3SGL (Fig. 2D) showed a set of peaks with a major ion at *m/z* 1,868.5. The mass range varied according to the culture batch, illustrated by two examples shown in supplementary

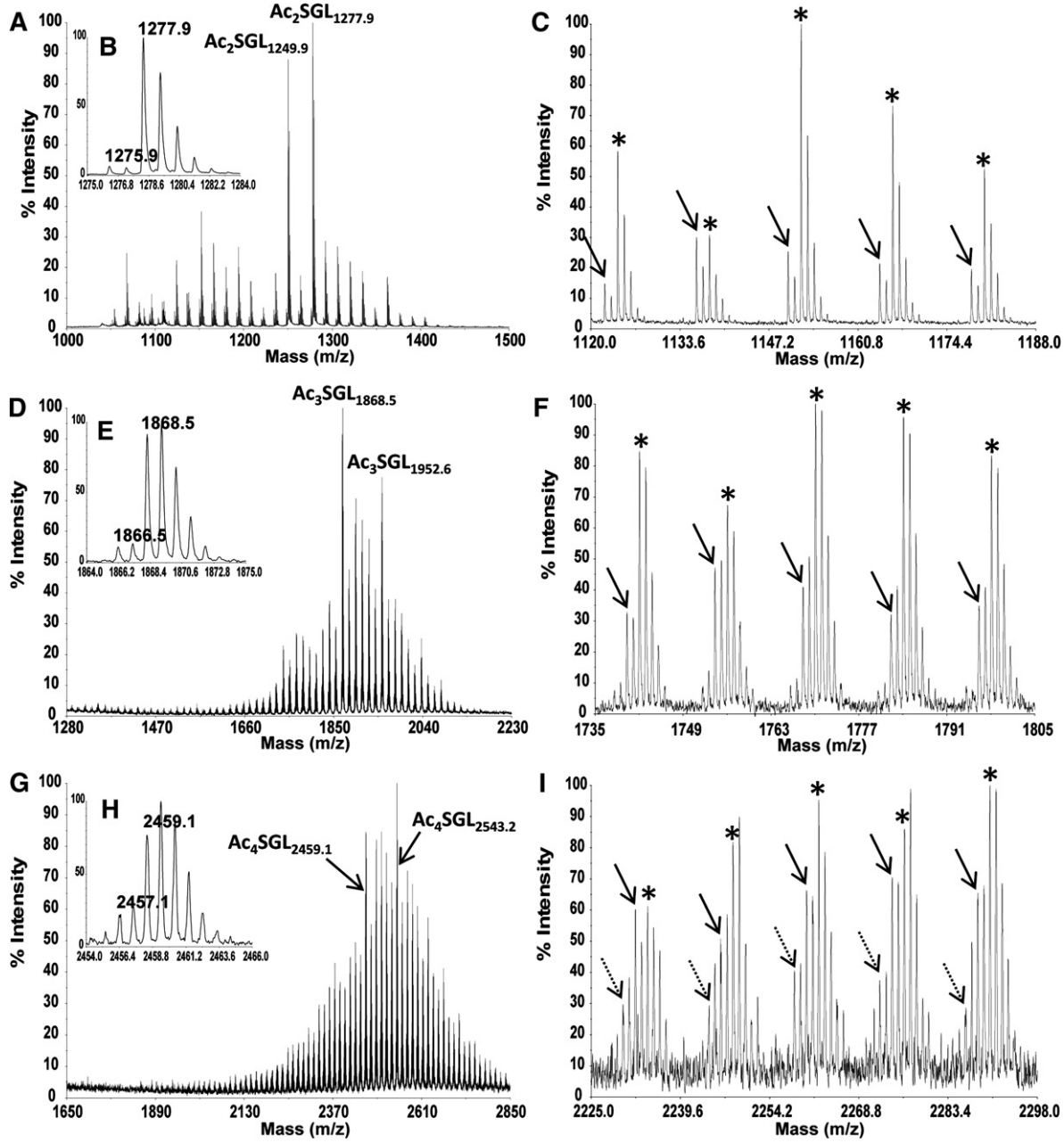


Fig. 2. MALDI MS spectra in negative-ion mode of *M. tuberculosis* H37Rv Ac_2SGL (A–C), Ac_3SGL (D–F), and Ac_4SGL (G–I). The different views correspond to different expanded areas. In panels B, E, and H, the isotopic profile of one major ion is presented. In panels A, D, and G, the arrows indicate the species targeted for MS/MS analysis shown in Fig. 3 and supplementary Fig. II. In panels C, F, and I, the species containing HPA are indicated by asterisks, the species containing one PA are indicated by arrows, and the species containing two PA are indicated by dashed arrows.

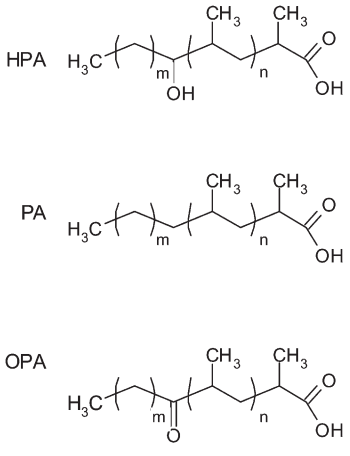
Fig. IC and D, with ions ranging from *m/z* 1,630.3 to 2,120.8 or from *m/z* 1,404.0 to 2,375.1. From the *m/z* values, we deduced that the major series of ions corresponded to Ac_3SGL bearing two HPA and one C_{16}/C_{18}. Again, however, a minor series of ions at 2 u less (Fig. 2F, arrow) was observed and tentatively assigned to molecular species containing one PA instead of an HPA, although ESI-FT-MS analysis was not performed in this case.

As for Ac_3SGL, the mass range of ions recorded for Ac_4SGL varied according to culture batches, for example, from *m/z* 1,936.6 to 2,851.6 or from *m/z* 1,756.4 to 2,725.4 (see supplementary Fig. IE, F). The main ions at *m/z* 2,248.9, 2,459.1, and 2,543.2 corresponded to Ac_4SGL bearing three HPA and one C_{16}/C_{18}. In this case, two minor series of ions at 2 and 4 u less were recovered (Fig. 2I, as indicated by solid and dashed arrows, respectively). They were tentatively assigned to the molecular species containing two HPA, one PA, and one C_{16}/C_{18} and to one HPA, two PA, and one C_{16}/C_{18}, respectively.

MS/MS uncovers an additional level of molecular complexity

A MS/MS-based approach was developed to characterize fatty acids of each acyl form of SGL. Ac_2SGL species at *m/z*

TABLE 1. *m/z* values of the different Ac_2SGL acyl forms containing one C_{16} or C_{18} at the 2-position of the acylated glucose and one HPA, one PA, or one OPA at the 3-position



2-position / 3-position	C-lenght	m	n	HPA C16	HPA C18	PA C16	PA C18	OPA C16	OPA C18
	C25	14	2	1039,7	1067,7	1023,7	1051,7	1037,7	1065,7
	C27	16	2	1067,7	1095,7	1051,7	1079,7	1065,7	1093,7
	C28	14	3	1081,7	1109,7	1065,7	1093,7	1079,7	1107,7
	C30	16	3	1109,7	1137,8	1093,7	1121,7	1107,7	1135,8
	C31	14	4	1123,8	1151,8	1107,8	1135,8	1121,8	1149,8
	C33	16	4	1151,8	1179,8	1135,8	1163,8	1149,8	1177,8
	C34	14	5	1165,8	1193,8	**1149,8**	**1177,8**	1163,8	1191,8
	C36	16	5	1193,8	1221,9	1177,8	1205,8	1191,8	1219,9
	C37	14	6	1207,9	1235,9	**1191,9**	**1219,9**	1205,9	1233,9
	C39	16	6	1235,9	1263,9	1219,9	1247,9	1233,9	1261,9
	C40	14	7	**1249,9**	**1277,9**	1233,9	1261,9	**1247,9**	**1275,9**
	C42	16	7	1277,9	1306,0	1261,9	1289,9	1275,9	1304,0
	C43	14	8	1291,9	1320,0	1275,9	1303,9	1289,9	1318,0
	C45	16	8	1320,0	1348,0	1304,0	1332,0	1318,0	1346,0
	C46	14	9	1334,0	1362,0	1318,0	1346,0	1332,0	1360,0
	C48	16	9	1362,0	1390,1	1346,0	1374,0	1360,0	1388,1
	C49	14	10	1376,0	1404,1	1360,0	1388,0	1374,0	1402,1
	C51	16	10	1404,1	1432,1	1388,1	1416,1	1402,1	1430,1
	C52	14	11	1418,1	1446,1	1402,1	1430,1	1416,1	1444,1
	C54	16	11	1446,1	1474,1	1430,1	1458,1	1444,1	1472,1

The carbon lengths of the HPA and PA are specified in the left panel.

1,249.9 ($Ac_2SGL_{1249.9}$) were therefore unambiguously assigned to contain a unique composition of one C_{16}/one $HPAC_{40}$, whereas the species at *m/z* 1,277.9 ($Ac_2SGL_{1277.9}$) corresponded to both a major form with one C_{18}/one $HPAC_{40}$ and a minor form with one C_{16}/one $HPAC_{42}$. Indeed, the sodiated precursor ion $[M+2Na]^+$ of $Ac_2SGL_{1249.9}$ at *m/z* 1,295.9 in positive mode showed fragment ions at *m/z* 653.6 and *m/z* 301.2, corresponding to sodiated cations $[RCOONa^+Na]^+$ of $HPAC_{40}$ and C_{16}, respectively, and at *m/z* 687.3 arising from the loss of $HPAC_{40}$ (**Fig. 3A**). Precursor ions of $Ac_2SGL_{1277.9}$ at *m/z* 1,323.9 gave $HPAC_{40}$ and C_{16} fragment ions and additional ions at *m/z* 681.6 and *m/z* 329.2 arising from $HPAC_{42}$ and C_{18}, respectively, with $HPAC_{40}$ and C_{18} fragment ions being the most abundant (Fig. 3B). Fragment ions at *m/z* 687.3 and 715.3 corresponding to the loss of $HPAC_{42}$ or $HPAC_{40}$, respectively, were also observed (Fig. 3B). MS/MS spectra showed additional informative fragment ions corresponding to glucose or trehalose units bearing HPA. Among them were ions attributed to (anhydro)-di-acylated-sodiated-glucose structures at *m/z* 1,031.9 (anhydro form at *m/z* 1,013.9) for $Ac_2SGL_{1249.9}$ (Fig. 3A) or at *m/z* 1,059.9 (anhydro form at *m/z* 1,041.9) for $Ac_2SGL_{1277.9}$ (Fig. 3B). These ions, arising from the loss of (anhydro)-sulfo-glucose, indicated that both fatty acyl appendages were located on the same glucopyranosyl (Glc*p*) unit, i.e., the unit that does not bear the sulfate group. Interestingly, the relative distribution of the different fatty acids on positions 2 and 3 of the Glc*p* could be deduced from negative-ion MALDI-MS/MS. Indeed, using synthetic Ac_2SGL molecules with defined fatty acids on each position, we observed that MS/MS in the negative-ion mode, in contrast to the positive-ion mode described above, generated fragment ions arising from the loss of the fatty acid in the 3-position only (not shown). For instance, the precursor ions $[M-H]^-$ at *m/z* 1,277.9 showed the loss of $HPAC_{40}$ or $HPAC_{42}$ with fragment ions recorded at *m/z* 669.3 and 641.3, respectively (Fig. 3C). However, loss of C_{16} or C_{18} was not observed for these precursor ions or for those at *m/z* 1,249.9, confirming that the short fatty acids are located on the 2-position, whereas HPA are located on the 3-position of the Glc*p* unit.

When applied to Ac_3SGL and Ac_4SGL, this method revealed the level of molecular complexity of each *m/z* species. Concerning Ac_3SGL, we specifically focused on the species detected in negative-ion mode at *m/z* 1,868.5 ($Ac_3SGL_{1868.5}$) and *m/z* 1,952.6 ($Ac_3SGL_{1952.6}$) (Fig. 2D). As expected, positive MS/MS analysis of the sodiated precursor ions $[M+2Na]^+$ of $Ac_3SGL_{1868.5}$ at *m/z* 1,914.5 showed $[RCOONa^+Na]^+$ fragment ions corresponding to both HPA and short fatty acids (C_{16} and C_{18}). The major HPA ions (see supplementary Fig. IIA) were not only at *m/z* 653.6 ($HPAC_{40}$), as ions of lower intensity at *m/z* 569.5 ($HPAC_{34}$), 611.5 ($HPAC_{37}$), 667.6 ($HPAC_{41}$), 681.6 ($HPAC_{42}$), 695.6 ($HPAC_{43}$), 737.7 ($HPAC_{46}$) were also present. Therefore, $Ac_3SGL_{1868.5}$ corresponds not only to SGL acylated by one C_{18} and two $HPAC_{40}$ or one C_{16}/one $HPAC_{40}$/one $HPAC_{42}$ but also to SGL acylated by one C_{16}/two $HPAC_{41}$, one C_{18}/one $HPAC_{34}$/one $HPAC_{46}$, and one C_{18}/one $HPAC_{37}$/one $HPAC_{43}$. For $Ac_3SGL_{1952.6}$, $HPAC_{40}$ was again the major HPA species detected alongside ions with lower intensities, indicating $HPAC_{41}$, $HPAC_{42}$, $HPAC_{43}$, $HPAC_{45}$, and $HPAC_{46}$ (see supplementary Fig. IIB) in addition to the short fatty acids C_{16} and C_{18}. $Ac_3SGL_{1952.6}$ was, thus, also a combination of at least five different forms

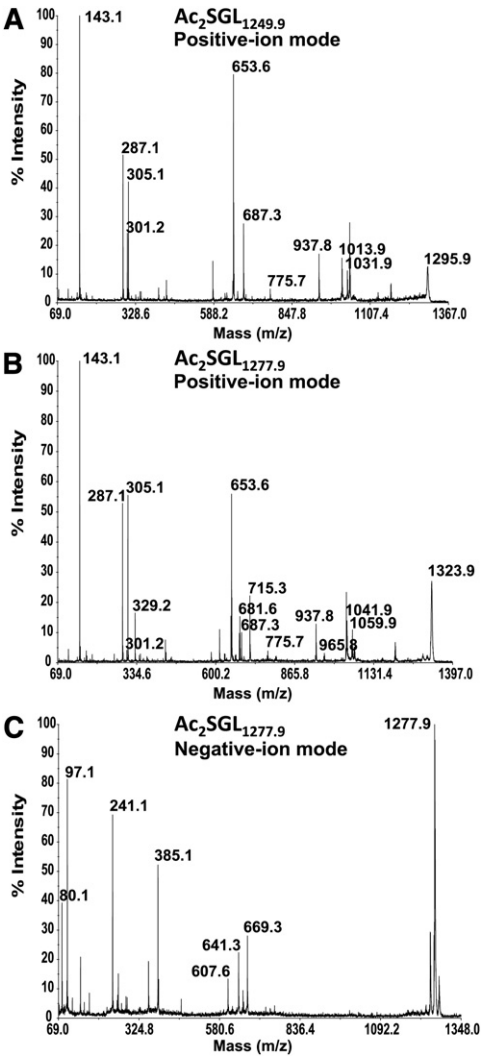


Fig. 3. MALDI-TOF MS/MS spectra of *M. tuberculosis* H37Rv Ac_2SGL are shown. A: Positive-ion mode MS/MS spectrum is shown of precursor ions $[M+2Na]^+$ at *m/z* 1,295.9 of $Ac_2SGL_{1249.9}$. B: Positive-ion mode MS/MS spectrum is shown of precursor ions $[M+2Na]^+$ at *m/z* 1,323.9 of $Ac_2SGL_{1277.9}$. C: Negative-ion mode MS/MS spectrum is shown of precursor ions $[M-H]^-$ at *m/z* 1,277.9.

containing either one C_{16}/one $HPAC_{40}$/one $HPAC_{46}$; one C_{16}/one $HPAC_{41}$/one $HPAC_{45}$; one C_{16}/2 $HPAC_{43}$; one C_{18}/1 $HPAC_{41}$/one $HPAC_{43}$; or one C_{18}/two $HPAC_{42}$.

As for Ac_4SGL, MS/MS analysis targeted the major species detected in negative-ion mode at *m/z* 2,459.1 ($Ac_4SGL_{2459.1}$) and *m/z* 2,543.2 ($Ac_4SGL_{2543.2}$) (Fig. 2G). As we expected, we observed the same fragment ion signature in the MS/MS spectra (see supplementary Fig. IIC, D) as those observed for Ac_3SGL. In fact, $Ac_4SGL_{2459.1}$ and $Ac_4SGL_{2543.2}$ corresponded to $Ac_3SGL_{1868.5}$ and $Ac_3SGL_{1952.6}$ acylated by an additional $HPAC_{40}$, respectively. Indeed, the major HPA ions were detected at *m/z* 653.6 ($HPAC_{40}$), with ions of lower intensity corresponding to $HPAC_{34}$, $HPAC_{37}$, and $HPAC_{41}$ to $HPAC_{47}$, resulting in many possible combinations for each acyl form. For example, considering $Ac_4SGL_{2459.1}$ (see supplementary Fig. IIC) and only the forms containing C_{18}, we can propose seven different acyl forms containing either one $HPAC_{34}$/one $HPAC_{40}$/one $HPAC_{46}$; one $HPAC_{34}$/one $HPAC_{42}$/one $HPAC_{44}$; one $HPAC_{34}$/two $HPAC_{43}$; two $HPAC_{37}$/one $HPAC_{46}$; one $HPAC_{37}$/one $HPAC_{40}$/one $HPAC_{43}$; one $HPAC_{37}$/one $HPAC_{41}$/one $HPAC_{42}$; or three $HPAC_{40}$.

New SGL molecular species revealed by HPLC-based fractionation of Ac_2SGL acyl forms

The data described above clearly show that even the simplest SGLs, i.e., Ac_2SGL, is actually a complex mixture of molecular species. We therefore explored the possibility of resolving this complexity by (C_{18}) reverse-phase HPLC (RP-HPLC) monitored by off-line negative MALDI-TOF-MS analysis (**Fig. 4**). An increasing gradient of dichloromethane in methanol was used as mobile phase to successively elute the different acyl forms, from the most polar to the most apolar molecular species. The first SGL eluted were observed in F7 and F8 and corresponded to unexpected species at *m/z* 897.5 and 925.6 and were assigned by MS/MS experiments to SGL esterified by short fatty acids (two C_{16} and one C_{16}/one C_{18}, respectively), as well as a compound at *m/z* 1,011.6 assigned by positive-ion mode MS/MS experiments (not shown) to a previously unknown Ac_1SGL esterified by a single $HPAC_{40}$. Indeed, in the MS/MS spectrum, only fragment ions corresponding to $HPAC_{40}$ at *m/z* 653.6 and no fragment ions corresponding to C_{16} or C_{18} at *m/z* 301.2 or 329.2 were detected. Moreover, in negative-ion mode, we observed fragment ions arising from the loss of $HPAC_{40}$, proving that $HPAC_{40}$ is located on the 3-position of the glucose. Next, in fractions F9 to F40, a series of compounds characterized by MALDI-TOF-MS with peaks at *m/z* from 1,039.7 to 1474.1 were assigned by MS/MS experiments to Ac_2SGL containing one C_{16}/C_{18} and one $HPAC_{25}$ to $HPAC_{55}$ (Table 1). We could not separate Ac_2SGL populations containing HPA from those containing PA. Indeed, intermediate fractions, such as F24, F31, or F38 contained Ac_2SGL with long HPAs together with Ac_2SGL with shorter PAs. Ac_2SGLs with PAC_{31} to PAC_{48} were observed from F22 to F49.

SGL of strains other than the *M. tuberculosis* H37Rv laboratory strain

SGLs were investigated in strains other than *M. tuberculosis* H37Rv, either from the tuberculosis complex, such as *Mycobacterium bovis* bacillus Calmette-Guérin or *M. tuberculosis* H37Ra, or not, such as *Mycobacterium chelonae*, *Mycobacterium fortuitum*, *Mycobacterium gastri*, *Mycobacterium kansasii*, *Mycobacterium marinum*, *Mycobacterium smegmatis*, and *Mycobacterium xenopi*. For each extract, the acetone-soluble phase was prepared and analyzed by negative-ion mode MALDI-TOF-MS for SGL content. None of these strains was found to produce SGLs. The absence of SGLs in *M. tuberculosis* H37Ra has already been correlated with a point mutation in the PhoP regulator (19, 20). These data are in agreement with those of previous studies showing that SGLs are specific glycolipids from *M. tuberculosis* species (21, 22).

Because *M. tuberculosis* H37Rv is a laboratory strain, one could argue that the acyl profiles could be merely a result of the passage of this strain for over 70 years. Then, the SGL acyl profile of a clinical isolate belonging to the Central

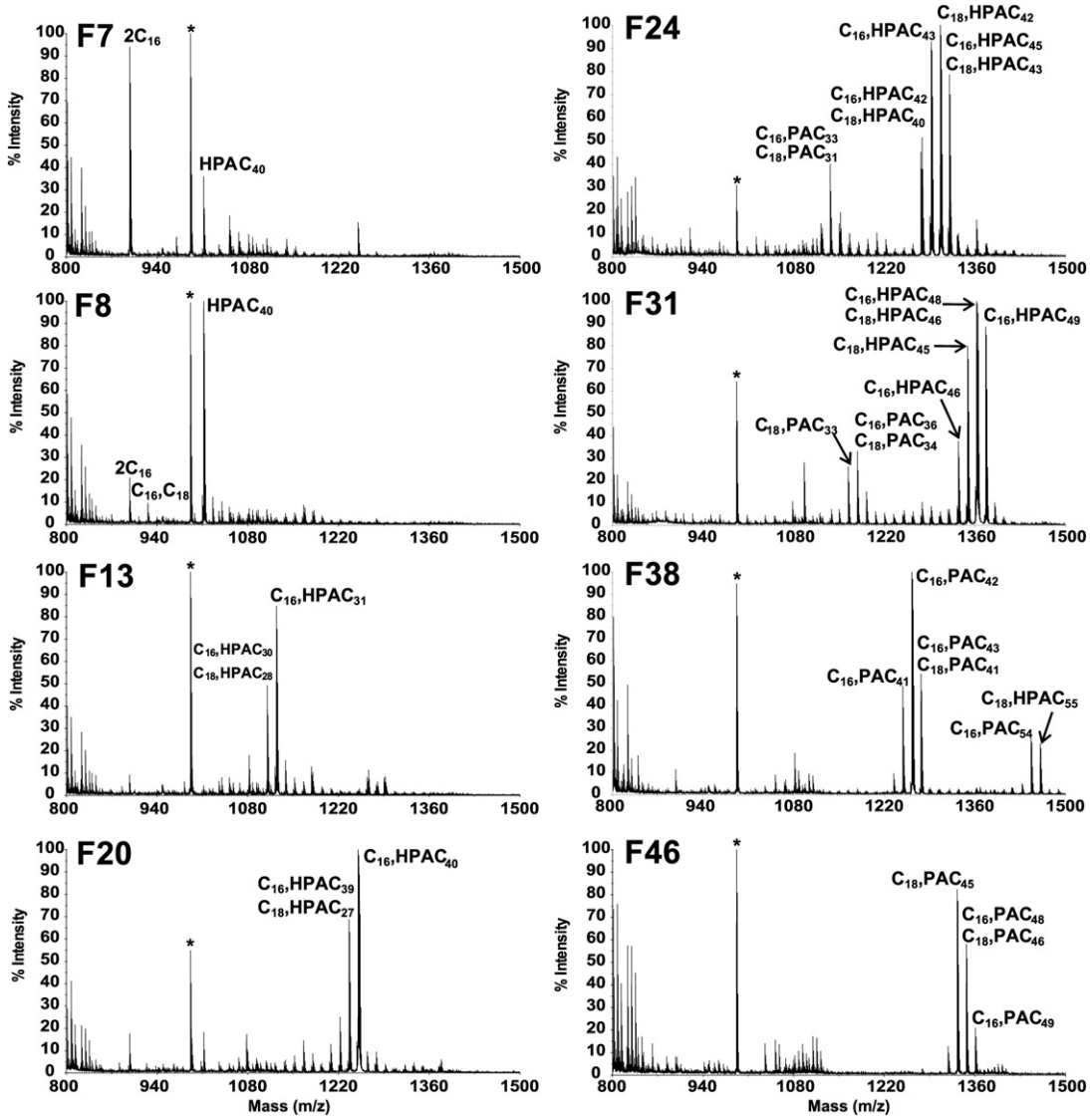


Fig. 4. RP-HPLC purification of Ac_2SGL, monitored by off-line negative-ion mode MALDI-TOF-MS is shown. MALDI-MS mass spectra are shown of specific 1 ml fractions collected after injection of 50 μg of Ac_2SGL on C_{18} Atlantis column. The fatty acid composition of Ac_2SGL species is annotated when deduced from MS/MS experiments. Stars indicate the matrix ion.

Asian India clade (16) was also studied. The complete SGL profile was first analyzed by negative-ion mode MALDI-TOF-MS of the acetone-soluble phase (**Fig. 5A**). Surprisingly, the previously characterized *M. tuberculosis* H37Rv Ac_1SGL species at *m/z* 1,011.7 was very abundant, and the ions corresponding to Ac_4SGL were very weak compared with those of the MALDI-TOF-MS assay results of the *M. tuberculosis* H37Rv acetone-soluble phase (14). Then, the different acyl forms were purified and analyzed by MALDI-TOF-MS (Fig. 5B–E). In all cases, the major SGL species remained essentially the same as the ones described in *M. tuberculosis* H37Rv. The MS profile of Ac_4SGL (Fig. 5B) was very similar to that of *M. tuberculosis* H37Rv (see supplementary Fig. IE). As for Ac_3SGL, ions ranging from *m/z* 1,249.9 to 2,190.9 were observed (Fig. 5C), clearly indicating two families of Ac_3SGL: one family with two C_{16}/C_{18} and one HPA and one family with one C_{16}/C_{18} and two HPA, as confirmed by MS/MS analysis. The Ac_2SGL mass spectrum (Fig. 5D) exhibited a set of peaks ranging from *m/z* 1,039.7 to 1404.1, with peaks ranging between *m/z* 1,249.9 and 1,404.1 predominating, as previously observed for *M. tuberculosis* H37Rv Ac_2SGL (see supplementary Fig. IB). The new result concerns the purification of an entire family of mono-acylated SGL observed in the mass spectrum from *m/z* 801.4 to *m/z* 1,179.8 (Fig. 5E). This was confirmed by MS/MS analysis, which revealed that these SGLs are acylated by one $HPAC_{25}$ to $HPAC_{52}$. Moreover, in this fraction, species at *m/z* 897.5, 925.6, and 953.6 assigned by MS/MS experiments to SGLs esterified by short fatty acids (two C_{16}, one $C_{16}/1C_{18}$, and two C_{18}, respectively) were also observed (Fig. 5E).

Unexpected SGLs in *M. tuberculosis mmpL8* knockout strains

M. tuberculosis mmpL8 knockout mutants (from H37Rv and Erdman strains) have been reported to show both a dramatic defect in Ac_4SGL synthesis and an intracellular accumulation of Ac_2SGL (13, 15). We purified Ac_2SGL

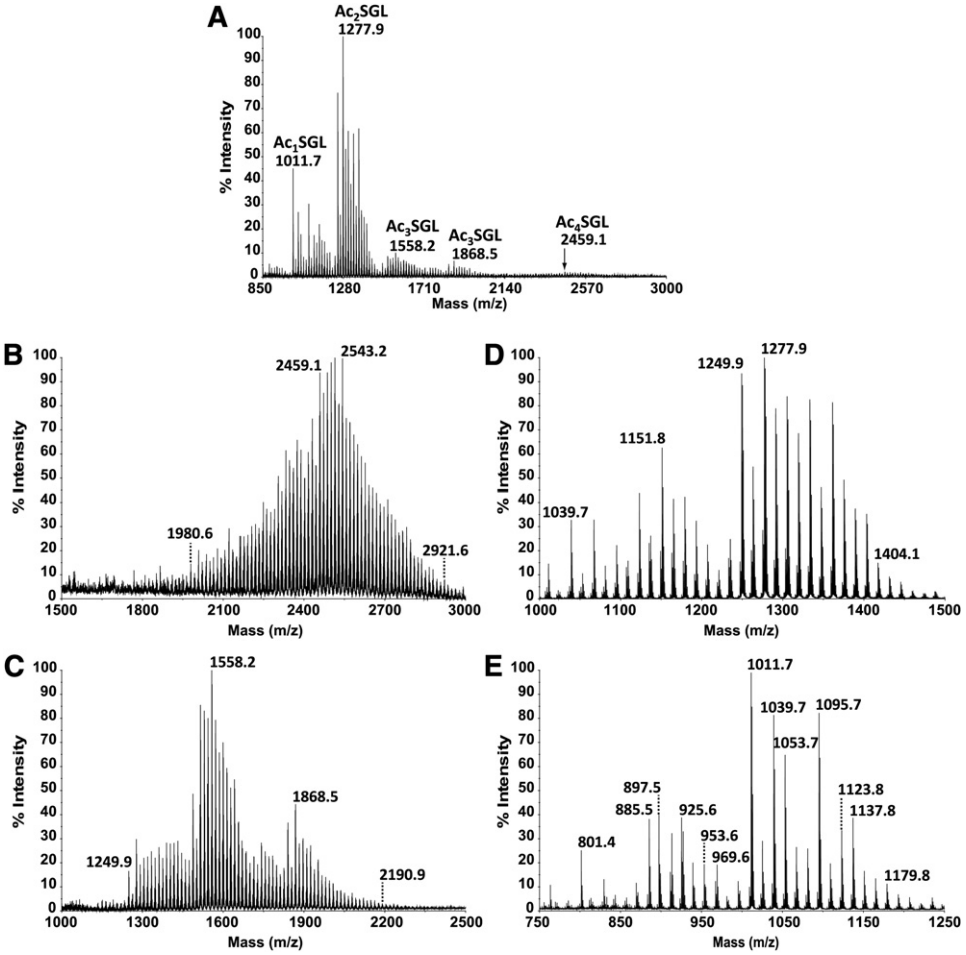


Fig. 5. SGLs of a clinical *M. tuberculosis* CAS isolate. MALDI-TOF-MS spectra in negative-ion mode of the "Acetone-Soluble" fraction (A), and of purified Ac_4SGL (B), Ac_3SGL (C), Ac_2SGL (D) and Ac_1SGL (E).

from both mutant strains (called $m1Ac_2SGL$ for the mutant generated by Domenech et al. [13] and $m2Ac_2SGL$ for the mutant generated by Converse et al. [15]). Both of these strains gave a similar negative-ion mode MALDI-TOF-MS profile (**Fig. 6A,B**). However, their profiles differed from that of the *M. tuberculosis* H37Rv wild-type (WT) strain Ac_2SGL (Fig. 6C) by the presence of a second series of peaks with the same intensities but with a difference of 2 u less (Fig. 6D, E, and F). This difference could be due to the presence of either *i*) a double bond on HPA; *ii*) a PA instead of an HPA, as a fatty acid devoid of hydroxyl group (meaning PA instead of HPA, −16) and of one CH_2 unit longer (+14) results in a final difference of −2; or *iii*) a ketonic instead of an alcoholic function on the fatty acid. The first hypothesis was ruled out by ^1H NMR analysis (not shown), which did not reveal any signal at approximately 6 ppm, indicative of a double bond. To investigate the two last hypotheses, we first performed an exact mass measurement by ESI-FT-MS of $m1Ac_2SGL$, which gave single molecular species at *m/z* 1,275.909 and 1,277.921. As expected, the latter measurement was in good agreement with the calculated molecular mass of 1,277.926 Da for an Ac_2SGL containing an HPA, whereas the former ion corresponded better to the presence of a ketonic function on the acyl group (i.e., oxophthioceranoic acyl [OPA], with a calculated molecular mass of 1,275.910 Da) rather than the presence of a PA on the molecule (calculated molecular mass of 1,275.947 Da). To confirm this assumption, $m1Ac_2SGL$ and $m2Ac_2SGL$ were submitted to reduction in the presence of sodium borodeuteride ($NaBD_4$). Based on the premise that only a ketonic group and not an alcoholic group would be sensitive to the reduction, we expected this treatment to displace the molecular species containing one OPA of 3 u but not those containing a PA or an HPA. As expected, no shift was observed with WT *M. tuberculosis* Ac_2SGL (see supplementary Fig. IIIA,B), whereas the species at *m/z* 1,275.9 in $m1Ac_2SGL$ was displaced to *m/z* 1,278.9 (see supplementary Fig. IIIC,D). The intensity of the remaining signal at *m/z* 1,275.9 after reduction (see supplementary Fig. IIID) reveals the proportion of Ac_2SGL containing one PA, as observed for the WT strain (Fig. 2B). The fatty acid composition of the Ac_2SGL molecular species at *m/z* 1,275.9 was confirmed by positive MS/MS analysis (not shown). $OPAC_{40}$ and $OPAC_{42}$ were detected at *m/z* 651.6 and 679.6, respectively, in addition to the presence of C_{16} and C_{18}, indicating that the Ac_2SGL species is esterified mainly by one $OPAC_{40}$/one C_{18} and less frequently by one $OPAC_{42}$/one C_{16} (Table 1).

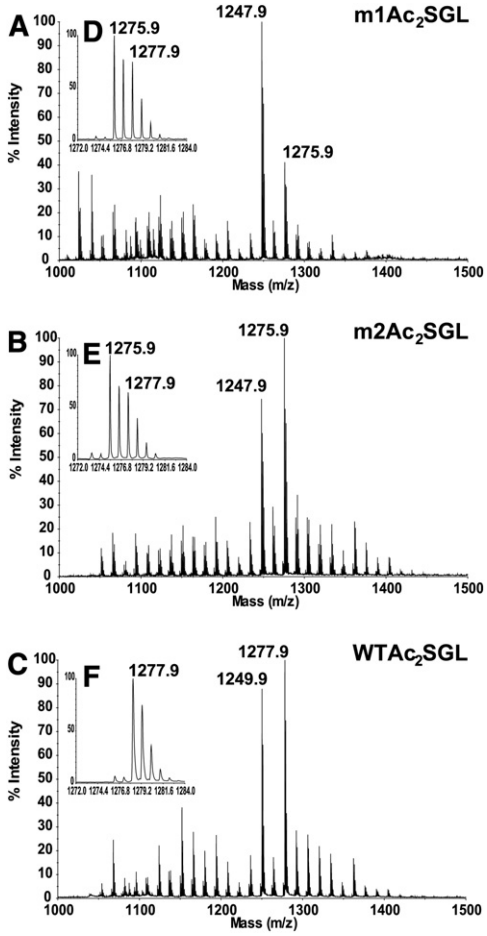


Fig. 6. Comparative MALDI-MS analyses of Ac_2SGL purified from both Δ*mmpL8* mutants of *M. tuberculosis* and *M. tuberculosis* H37Rv are shown. Negative-ion mode MALDI-TOF-MS spectra of Ac_2SGL of *mmpL8*::*hyg M. tuberculosis* H37Rv Pasteur strain (A) and Δ*mmpL8* jcm108 *M. tuberculosis* Erdman strain (B) and *M. tuberculosis* H37Rv (C) are shown. Expanded area (*m/z* 1,272–1.,284) is centered on the *m/z* 1,277.9 ion (D–F).

Only trace levels of Ac_3SGL and Ac_4SGL could be purified from the *M. tuberculosis* H37Rv *mmpL8* mutant ($m1Ac_3SGL$ and $m1Ac_4SGL$). Whereas 1 to 1.5 mg of Ac_2SGL was purified from 1 liter of *M. tuberculosis* H37Rv mutant culture, only 200 to 300 μg of Ac_3SGL and 20 to 50 μg of Ac_4SGL were obtained, respectively. Surprisingly, $m1Ac_3SGL$ and $m1Ac_4SGL$ analyses also gave unexpected negative-ion mode MALDI mass spectra. Indeed, their molecular masses were lower than that of the SGL present in the WT strain (see supplementary Fig. IV). However, because $m1Ac_4SGL$ were present at trace levels, their analysis by MS/MS and their purification for NMR studies could not be completed. Concerning $m1Ac_3SGL$, the major species were observed at *m/z* 1,514.2 and 1,528.2 and were shown esterified by one OPA and two short fatty acids (C_{16}/C_{18}). Indeed, as demonstrated by positive MS/MS analysis, they are mainly esterified by one C_{16}/one C_{18}/one $OPAC_{40}$ or by two C_{16}/one $OPAC_{42}$ and one C_{16}/one C_{18}/one $OPAC_{40}$, respectively (not shown). It is noteworthy that an Ac_3SGL population bearing two simple fatty acids can be detected in some batches of WT *M. tuberculosis* H37Rv (see supplementary Fig. ID). Indeed, positive MS/MS analysis of WT $Ac_3SGL_{1516.2}$, observed in the MALDI mass spectrum presented in supplementary Fig. ID, revealed an Ac_3SGL esterified by two C_{16}/one $HPAC_{42}$ or by one C_{16}/one C_{18}/one $HPAC_{40}$. Moreover, interestingly, the $m1Ac_3SGL$ structure differed from that of the WT Ac_3SGL, as deduced from ^1H-^1H COSY, ^1H-^1H HOHAHA, and ^1H-^{13}C HMQC NMR analyses (see supplementary Fig. V). Indeed, the presence of an acyl group could be easily deduced from ^1H chemical shifts, as the germinal acyloxy proton is deshielded compared with the proton of a nonacylated position. The carbon chemical shifts are almost unaffected by the acylation (23). Considering WT Ac_2SGL, as previously published (14), the downfield chemical shifts of Glc*p*-II H2 and H3 at 4.74 and 5.24 ppm, respectively, proved that fatty acyl appendages are located on C2 and C3 of Glc*p*-II (see supplementary Fig. VA and supplementary Table I). For WT Ac_3SGL, H6 of Glc*p*-II is deshielded at 4.26 ppm, compared with δ 3.49/3.82 for Ac_2SGL (see supplementary Fig. VB), in accordance with the presence of a fatty acid on C6 of Glc*p*-II, as previously described (8). For WT Ac_4SGL, H6 of Glc*p*-I is deshielded at 4.15 ppm (see supplementary Fig. VC), compared with δ 3.59/3.55 for Ac_3SGL, in accordance with the presence of another fatty acid on C6 of Glc*p*-I. In the case of $m1Ac_3SGL$, ^1H chemical shifts proved that they are acylated on C2 and C3 of the nonsulfated α-D-Glc*p*-II and on C6′ of the sulfated α-D-Glc*p*-I (see supplementary Fig. VD and supplementary Table I). This acylation pattern was further confirmed by positive-ion mode MS/MS analysis of $m1Ac_3SGL_{1528.2}$, which showed fragment ions at *m/z* 543.2, assigned as a palmitoyl-sulfo-Glc*p* unit, instead of the ions at *m/z* 305.0, corresponding to the sulfo-Glc*p* usually detected for the WT Ac_3SGL (not shown).

DISCUSSION

In order to gain better insight into the structure/function relationships underlying the CD1 presentation of SGL to T cells (14), we made great attempts to improve both their purification and structural characterization.

The first objective of this study was to prepare fractions containing pure and unique SGL acyl forms, namely Ac_2SGL, Ac_3SGL, and Ac_4SGL. To achieve this goal, two steps were crucial: *i*) to get rid of the phospholipids, particularly mycobacterial PIM, from the lipidic extract, thus making the acetone-precipitation an inevitable step; and *ii*) to eliminate neutral lipid contaminants such as TAT and DAT from SGL fractions, using anion exchange chromatography.

Goren et al. previously characterized five SGL families, four of which were tetra-acylated (SL-I, SL-I′, SL-II, and SL-II′), and one was tri-acylated (SL-III) (3–6). Those authors used an elegant strategy based on chemical degradation/modification and subsequent MS and infrared spectroscopy analyses that allowed *i*) identification of trehalose-2-sulfate, *ii*) determination of the acylated positions on the trehalose core, *iii*) identification of the acyl substituents (C_{16}/C_{18}, HPA, and PA), and *iv*) the determination that the 2-position of SL-I is acylated by one C_{16} or C_{18} and that the 3-position is occupied almost exclusively by

one PA (6). In this way, they demonstrated that the major species (SL-I) was acylated on the 3-, 6-, and 6′-positions by one PA and two HPA; SL-I′ was acylated by two PA and one HPA; and SL-II was acylated by three HPA. SL-II′ was also described as being acylated by three HPA on the 4-, 6-, and 6′-positions. The tri-acylated SL-III was shown to be acylated on the 3- and 6-positions by two HPA.

In this study, we developed a strategy in which we first purified the different acyl forms before completely characterizing them by MALDI-TOF-MS and NMR. We confirmed most of the structural points described by Goren et al., such as the presence of C_{16}/C_{18} exclusively on the 2-position; the presence in Ac_3SGL of the fatty acids on the 2-, 3-, and 6-positions of the nonsulfated glucose; and the fact that $HPAC_{40}$ is the principal HPA component. Here, for the first time, we realized a complete NMR analysis of each acyl form, and in contrast to the findings of Goren, we never observed acylation on the 4-position of the glucose unit. We cannot exclude the presence of this acyl form, if present at trace level. We showed that Ac_2SGL contains mainly one C_{16}/C_{18} and one HPA on the 2- and 3-positions, respectively, and, in accordance with the study by Goren et al., that Ac_3SGL contains one C_{16}/C_{18} and two HPA on the 2-, 3-, and 6-positions. On the basis of a biosynthetic linkage between all these acyl forms and considering our result showing that Ac_4SGL contains mainly three HPA, it is obvious that the 3-position of Ac_4SGL is occupied almost exclusively by one HPA. However, Goren et al. found that the 3-position of SL-I (the major tetra-acylated species, acylated by one C_{16} or C_{18}, one PA, and two HPA) was occupied almost exclusively by one PA. This species, in minor abundance in our case, most probably derives from the 15% of Ac_2SGL species containing one PA (see below). Moreover, Goren et al. described SL-I to be acylated by HPA from C_{31} to C_{46}. Using MALDI, we derived a precise profile of the different molecular species composing each SGL acyl form. We concluded that Ac_2SGL contained HPA in lengths from C_{25} to C_{54}. Furthermore, analysis of the isotopic masses gave the proportions of PA versus HPA. We deduced that Ac_2SGL was composed of 15% of species containing one PA and 85% containing one HPA. Ac_4SGL was distributed in 73% of species with three HPA, 24% with one PA, and 3% with two PA.

Negative-ion mode MALDI-TOF-MS proved to be the method of choice for characterizing the acylation pattern of the different SGL species. Surprisingly, we observed that this pattern, particularly the molecular weight range, varied from one culture batch to another, although the culture conditions were essentially the same, except for the duration of the culture. The lipid chain length of SGL has been recently shown to depend on the availability of precursors such as methyl malonyl CoA (24), which might be confounded by the growth phase, a parameter that is difficult to control because bacteria were grown as cell surface pellicles. Nevertheless, the major SGL species remained essentially the same for the different SGL acyl forms.

Fractionation of the Ac_2SGL mixture by RP-HPLC allowed us to not only separate SGL species containing HPA or PA but also to reveal previously uncharacterized *M. tuberculosis* H37Rv SGL species, i.e., Ac_2SGL esterified by simple fatty acids only, namely di-palmitoyl-SGL, di-stearoyl-SGL, mono-palmitoyl-mono-stearoyl-SGL, and Ac_1SGL bearing a single HPA. This last species was even more abundant in the clinical strain, and a complete family of Ac_1SGL acylated by one HPA of variable length was unambiguously characterized. The identification of this mono-acylated species was unexpected, as the first acylation step is considered to be a palmitoylation/stearoylation on the 2-position of the sulfated trehalose (Fig. 1A) and because PapA1, the acyl-transferase that adds the HPA on C2-palmitoyl/stearoyl-sulfated trehalose, has been shown to harbor no activity toward sulfated trehalose (25). More clues about the biosynthetic pathway of SGLs arose from an in-depth analysis of SGLs produced by Δ*mmpL8* mutants. Our present knowledge indicates that sulfation of trehalose, catalyzed by the sulfotransferase Stf0, is the first step in SGL synthesis (26). Palmitate or stearate acyl chains derive from fatty-acid synthase-I (FAS-I), whereas HA and HPA are synthesized by polyketide synthase 2 (Pks2) (27) working in conjunction with the FadD23 fatty-acyl-CoA ligase (28, 29). Then, two acyl-transferases, PapA1 on one hand and PapA2 associated with Pks2 on the other hand, would act sequentially (25, 30), where PapA2 converts sulfated trehalose into 2-palmitoyl/stearoyl-sulfated trehalose, and PapA1 further elaborates Ac_2SGL by adding a methyl-branched fatty acid on the 3-position of trehalose. The MmpL8 protein, predicted to be a lipid transporter, is described as playing a role in mediating the transport of the Ac_2SGL precursor forms from the cytosol into the cell envelope (13, 15). Indeed, characterization of the Δ*mmpL8* mutants showed that synthesis of mature Ac_4SGL was halted and that Ac_2SGL concomitantly accumulated within the cell. This strongly suggests that transport and biogenesis of SGLs are coupled and that the final step in SGL biosynthesis might be the extracellular acylation of the C6 position of both α-D-Glc*p* units (13). Here, we show that the Δ*mmpL8* mutant was still able to generate Ac_3SGL but with a structure different from that of the WT strain. Indeed, the third added fatty acid was not only C_{16}/C_{18}, rather than HPA/HA, but was in a different position: C6′ of the sulfated α-D-Glc*p* instead of the C6 of the 2,3-di-acylated α-D-Glc*p* unit. This result proves that Ac_2SGL must be transported (or translocated) across the cell membrane by MmpL8 in order to be correctly transformed into the final Ac_4SGL (**Fig. 7**). Another intriguing finding was that approximately half of the Δ*mmpL8* mutant SGLs carried a methyl-branched fatty acid containing a ketonic function (OPA) instead of the hydroxyl group, a form that does not exist in the WT. This result strongly suggests that MmpL8 functions in conjunction with enzymes involved in the elaboration of the final methyl-branched fatty acids.

Two hypotheses are currently proposed for acylation steps that generate Ac_4SGL from Ac_2SGL: fatty acid addition to both the C6 and C6′ positions of the trehalose unit could occur either intracellularly or extracellularly (31, 32). In the first case, Ac_4SGL would be elaborated from Ac_2SGL in the cytosol, and Ac_4SGL would be transported

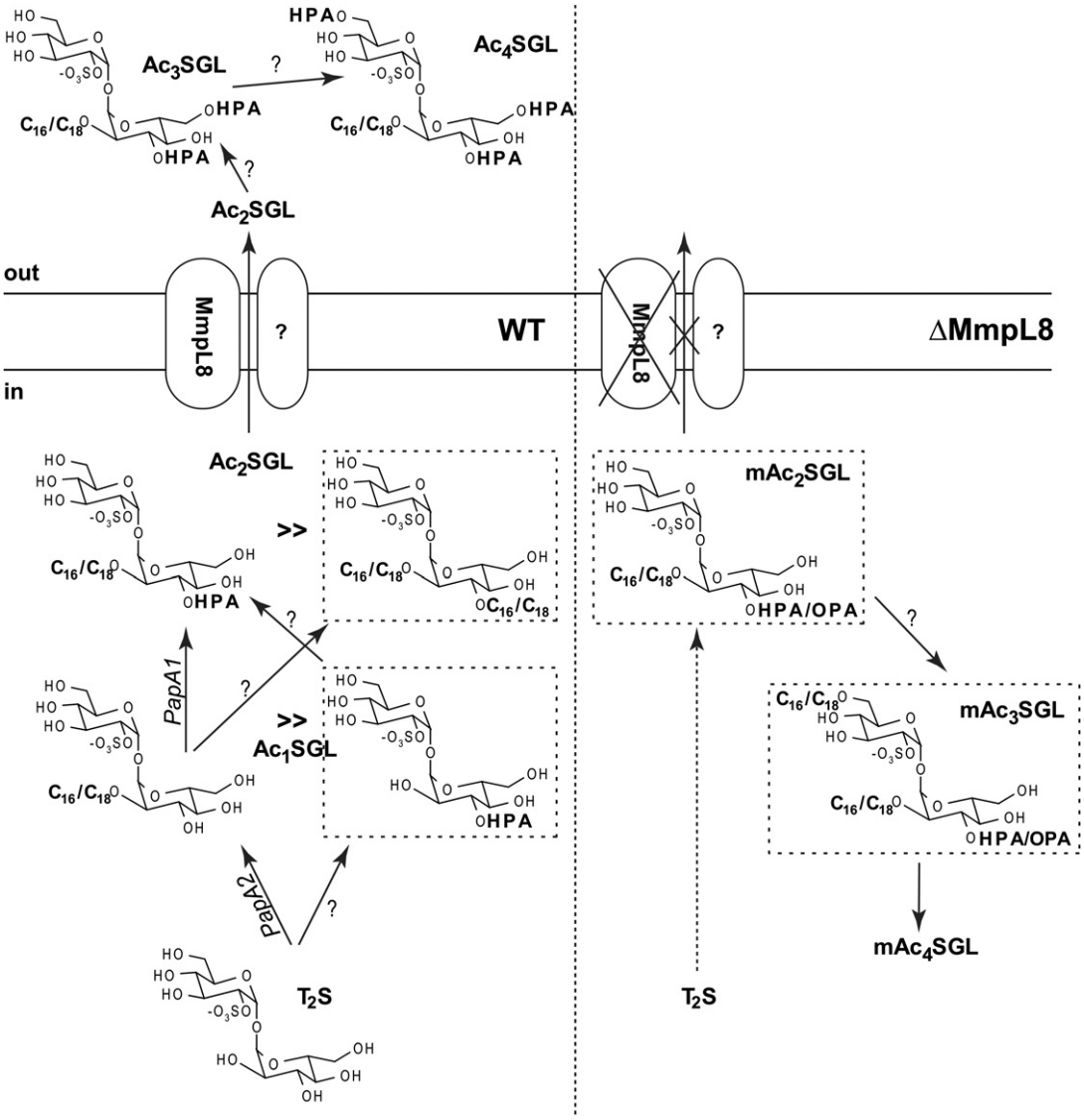


Fig. 7. Schematic representation of the pathway of completion of SGL biosynthesis in *M. tuberculosis* H37Rv (WT) and Δ*mmpL8* mutants of *M. tuberculosis* (Δ*mmpL8*) is shown [adapted from Bertozzi and Schelle (31) with permission]. The SGLs are not inserted in membrane, as they should be, for simplification and clarity purposes. Our results favor an extracellular acylation of Ac_2SGL, following the transport of Ac_2SGL by MmpL8. Indeed, the MmpL8 knockout strains also synthesize very weak amounts of Ac_3- and Ac_4SGL (mAc_3- and mAc_4SGL) but from a different structure. This model implies that HPA is transported by an unknown protein (annotated as "?") to generate Ac_3- and Ac_4SGL species. Dashed insets show the novel SGL species described for the first time in this study. T_2S, trehalose-2′-sulfate.

to the exterior of the cell by MmpL8. In the second hypothesis, an extracellular acyl-transferase would convert Ac_2SGL to Ac_4SGL after transport of Ac_2SGL to the exterior of the cell by MmpL8. In conclusion, our results favor the extracellular acylation of both C6 and C6′ positions. This process seems to parallel what is observed for the generation of cord factor, where acylation of both C6 and C6′ positions of trehalose is catalyzed by a mycoloyl-transferase found to be associated with the bacterial cell wall surface (33).

The generation of HPA is not yet completely understood. HPA could result from *i*) the introduction of the hydroxyl group into preformed PA or *ii*) an incomplete synthesis by Pks2. In the second hypothesis, the keto-acid generated by condensation of methylmalonyl CoA with the *n*-fatty acyl moiety during the first cycle could be reduced to the hydroxyl acid but not dehydrated, leaving the hydroxyl group in the final multi-methyl-branched product (27). Our data favor this last hypothesis. The occurrence of 50% OPA in Ac_2SGL of the *mmpL8* mutant suggests that only half of the keto-acid generated by condensation of methylmalonyl-CoA with the *n*-fatty acyl moiety would be reduced to the hydroxyl acid.

The authors thank Dr. Clifton E. Barry III and Pr. Jeffery Cox for providing *M. tuberculosis* Δ*mmpL8* mutants and Amélie Vax for technical assistance. We are also grateful to Drs. P. Domenech and C. Chalut for discussions, Dr. L. Sweet for reading the manuscript, and Dr. J. Nigou for critical comments and helpful suggestions.

REFERENCES

1. Middlebrook, G., C. Coleman, and W. B. Schaefer. 1959. Sulfolipid from virulent tubercle bacilli. *Proc. Natl. Acad. Sci. U S A.* **45:** 1801–1804.
2. Goren, M. B., O. Brokl, and W. B. Schaefer. 1974. Lipids of putative relevance to virulence in *Mycobacterium tuberculosis*: correlation of virulence with elaboration of sulfatides and strongly acidic lipids. *Infect. Immun.* **9:** 142–149.
3. Goren, M. B., O. Brokl, B. C. Das, and E. Lederer. 1971. Sulfolipid I of *Mycobacterium tuberculosis*, strain H37Rv. Nature of the acyl substituents. *Biochemistry.* **10:** 72–81.
4. Goren, M. B. 1970. Sulfolipid I of *Mycobacterium tuberculosis*, strain H37Rv. I. Purification and properties. *Biochim. Biophys. Acta.* **210:** 116–126.
5. Goren, M. B. 1970. Sulfolipid I of *Mycobacterium tuberculosis*, strain H37Rv. II. Structural studies. *Biochim. Biophys. Acta.* **210:** 127–138.
6. Goren, M. B. 1990. Mycobacterial fatty acid esters of sugars and sulfosugars. In Handbook of Lipid Research, vol. 6. D. J. Hanahan, editor. Plenum Press, NY. 363–461.
7. Daffe, M., F. Papa, A. Laszlo, and H. L. David. 1989. Glycolipids of recent clinical isolates of Mycobacterium tuberculosis: chemical characterization and immunoreactivity. *J. Gen. Microbiol.* **135:** 2759–2766.
8. Goren, M. B., B. C. Das, and O. Brokl. 1978. Sulfatide III of *Mycobacterium tuberculosis*, Strain H37Rv. *New J. Chem.* **2:** 379–384.
9. Lemassu, A., M. A. Laneelle, and M. Daffe. 1991. Revised structure of a trehalose-containing immunoreactive glycolipid of Mycobacterium tuberculosis. *FEMS Microbiol. Lett.* **62:** 171–175.
10. Baer, H. H. 1993. The structure of an antigenic glycolipid (SL-IV) from *Mycobacterium tuberculosis. Carbohydr. Res.* **240:** 1–22.
11. Cruaud, P., J. T. Yamashita, N. M. Casabona, F. Papa, and H. L. David. 1990. Evaluation of a novel 2,3-diacyl-trehalose-2'-sulphate (SL-IV) antigen for case finding and diagnosis of leprosy and tuberculosis. *Res. Microbiol.* **141:** 679–694.
12. Mougous, J. D., M. D. Leavell, R. H. Senaratne, C. D. Leigh, S. J. Williams, L. W. Riley, J. A. Leary, and C. R. Bertozzi. 2002. Discovery of sulfated metabolites in mycobacteria with a genetic and mass spectrometric approach. *Proc. Natl. Acad. Sci. U S A.* **99:** 17037–17042.
13. Domenech, P., M. B. Reed, C. S. Dowd, C. Manca, G. Kaplan, and C. E. Barry III. 2004. The role of MmpL8 in sulfatide biogenesis and virulence of Mycobacterium tuberculosis. *J. Biol. Chem.* **279:** 21257–21265.
14. Gilleron, M., S. Stenger, Z. Mazorra, F. Wittke, S. Mariotti, G. Bohmer, J. Prandi, L. Mori, G. Puzo, and G. De Libero. 2004. Diacylated sulfoglycolipids are novel mycobacterial antigens stimulating CD1-restricted T cells during infection with *Mycobacterium tuberculosis. J. Exp. Med.* **199:** 649–659.
15. Converse, S. E., J. D. Mougous, M. D. Leavell, J. A. Leary, C. R. Bertozzi, and J. S. Cox. 2003. MmpL8 is required for sulfolipid-1 biosynthesis and *Mycobacterium tuberculosis* virulence. *Proc. Natl. Acad. Sci. U S A.* **100:** 6121–6126.
16. Filliol, I., A. S. Motiwala, M. Cavatore, W. Qi, M. H. Hazbon, M. Bobadilla Del Valle, J. Fyfe, L. Garcia-Garcia, N. Rastogi, C. Sola, et al. 2006. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J. Bacteriol.* **188:** 759–772.
17. Gilleron, M., C. Ronet, M. Mempel, B. Monsarrat, G. Gachelin, and G. Puzo. 2001. Acylation state of the phosphatidylinositol mannosides from *Mycobacterium bovis* bacillus Calmette Guerin and ability to induce granuloma and recruit natural killer T cells. *J. Biol. Chem.* **276:** 34896–34904.
18. Besra, G. S., R. C. Bolton, M. R. Mcneil, M. Ridell, K. E. Simpson, J. Glushka, H. Van Halbeek, P. J. Brennan, and D. E. Minnikin. 1992. Structural elucidation of a novel family of acyltrehaloses from Mycobacterium tuberculosis. *Biochemistry.* **31:** 9832–9837.
19. Lee, J. S., R. Krause, J. Schreiber, H. J. Mollenkopf, J. Kowall, R. Stein, B. Y. Jeon, J. Y. Kwak, M. K. Song, J. P. Patron, et al. 2008. Mutation in the transcriptional regulator PhoP contributes to avirulence of *Mycobacterium tuberculosis* H37Ra strain. *Cell Host Microbe.* **3:** 97–103.
20. Chesne-Seck, M. L., N. Barilone, F. Boudou, J. Gonzalo Asensio, P. E. Kolattukudy, C. Martin, S. T. Cole, B. Gicquel, D. N. Gopaul, and M. Jackson. 2008. A point mutation in the two-component regulator PhoP-PhoR accounts for the absence of polyketide-derived acyltrehaloses but not that of phthiocerol dimycocerosates in *Mycobacterium tuberculosis* H37Ra. *J. Bacteriol.* **190:** 1329–1334.
21. Dhariwal, K. R., G. Dhariwal, and M. B. Goren. 1984. Observations on the ubiquity of the *Mycobacterium tuberculosis* sulfatides in mycobacteria. *Am. Rev. Respir. Dis.* **130:** 641–646.
22. Soto, C. Y., M. Cama, I. Gibert, and M. Luquin. 2000. Application of an easy and reliable method for sulfolipid-I detection in the study of its distribution in *Mycobacterium tuberculosis* strains. *FEMS Microbiol. Lett.* **187:** 103–107.
23. Gilleron, M., J. Vercauteren, and G. Puzo. 1994. Lipo-oligosaccharidic antigen from *Mycobacterium gastri.* Complete structure of a novel C4-branched 3,6-dideoxy-alpha-xylo-hexopyranose. *Biochemistry.* **33:** 1930–1937.
24. Jain, M., C. J. Petzold, M. W. Schelle, M. D. Leavell, J. D. Mougous, C. R. Bertozzi, J. A. Leary, and J. S. Cox. 2007. Lipidomics reveals control of *Mycobacterium tuberculosis* virulence lipids via metabolic coupling. *Proc. Natl. Acad. Sci. U S A.* **104:** 5133–5138.
25. Kumar, P., M. W. Schelle, M. Jain, F. L. Lin, C. J. Petzold, M. D. Leavell, J. A. Leary, J. S. Cox, and C. R. Bertozzi. 2007. PapA1 and PapA2 are acyltransferases essential for the biosynthesis of the *Mycobacterium tuberculosis* virulence factor sulfolipid-1. *Proc. Natl. Acad. Sci. U S A.* **104:** 11221–11226.
26. Mougous, J. D., C. J. Petzold, R. H. Senaratne, D. H. Lee, D. L. Akey, F. L. Lin, S. E. Munchel, M. R. Pratt, L. W. Riley, J. A. Leary, et al. 2004. Identification, function and structure of the mycobacterial sulfotransferase that initiates sulfolipid-1 biosynthesis. *Nat. Struct. Mol. Biol.* **11:** 721–729.
27. Sirakova, T. D., A. K. Thirumala, V. S. Dubey, H. Sprecher, and P. E. Kolattukudy. 2001. The *Mycobacterium tuberculosis* pks2 gene encodes the synthase for the hepta- and octamethyl-branched fatty acids required for sulfolipid synthesis. *J. Biol. Chem.* **276:** 16833–16839.
28. Trivedi, O. A., P. Arora, V. Sridharan, R. Tickoo, D. Mohanty, and R. S. Gokhale. 2004. Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria. *Nature.* **428:** 441–445.
29. Lynett, J., and R. W. Stokes. 2007. Selection of transposon mutants of *Mycobacterium tuberculosis* with increased macrophage infectivity identifies fadD23 to be involved in sulfolipid production and association with macrophages. *Microbiology.* **153:** 3133–3140.
30. Bhatt, K., S. S. Gurcha, A. Bhatt, G. S. Besra, and W. R. Jacobs, Jr. 2007. Two polyketide-synthase-associated acyltransferases are required for sulfolipid biosynthesis in *Mycobacterium tuberculosis. Microbiology.* **153:** 513–520.
31. Bertozzi, C. R., and M. W. Schelle. 2008. Sulfated metabolites from *Mycobacterium tuberculosis*: sulfolipid-1 and beyond. In The Mycobacterial Cell Envelope, M. Daffe and J. M. Reyrat, editors. ASM Press, Washington, DC. 291–304.
32. Jain, M., E. D. Chow, and J. S. Cox. 2008. The MmpL protein family, In The Mycobacterial Cell Envelope, M. Daffe and J. M. Reyrat, editors. ASM Press, Washington, DC. 201–210.
33. Wiker, H. G., and M. Harboe. 1992. The antigen 85 complex: a major secretion product of *Mycobacterium tuberculosis. Microbiol. Rev.* **56:** 648–661.