Published in final edited form as:

J Mass Spectrom. 2012 October ; 47(10): 1269–1281. doi:10.1002/jms.3070.

Structural Determination of Glycopeptidolipids of *Mycobacterium smegmatis* by High Resolution Multiple-stage Linear Ion-trap Mass Spectrometry with Electrospray Ionization

Fong-Fu Hsu^{*,†}, Sophia Pacheco[§], John Turk[‡], and Georgiana Purdy[§]

[†]Mass Spectrometry Resource, Division of Endocrinology, Diabetes, Metabolism, and Lipid research, Department of Internal Medicine, Washington University School of Medicine, St. Louis, MO 63110

[§]Department of Molecular Microbiology & Immunology, Oregon Health & Science University, Portland, Oregon 97239

Abstract

Glycopeptidolipids (GPLs) are abundant in the cell walls of different species of mycobacteria and consist of tripeptide-amino-alcohol core of D-Phe-D-allo-Thr-D-Ala-L-alaninol linked to 3-hydroxy or 3-methoxy C_{26–34} fatty acyl chain at the N-terminal of D-Phe via amide linkage; and a 6-deoxytalose (6-dTal) and an *O*-methyl rhamnose residues respectively attach to D-allo-Thr and the terminal L-alaninol. They are important cell-surface antigens that are implicated in the pathogenesis of opportunistic mycobacteria belonging to the *Mycobacterium avium* complex. In this contribution, we described multiple-stage linear ion-trap in conjunction with high resolution mass spectrometry towards structural characterization of complex GPLs as [M + Na]⁺ ions isolated from *Mycobacterium smegmatis*, a fast-growing and non-pathogenic mycobacterial species. Following resonance excitation in an ion-trap, MSⁿ spectra of the [M + Na]⁺ ions of GPLs contained mainly b and y series ions that readily determine the peptide sequence. Fragment ions from MSⁿ also afford locating the 6-dTal and *O*-methyl rhamnose residues linked to the D-allo-Thr and terminal L-alaninol of the peptide core, respectively, as well as recognizing the modifications of the glycosides, including their acetylation and methylation states and the presence of succinyl group. The GPL families consisting of 3-hydroxy fatty acyl and of 3-methoxy fatty acyl substituents are readily distinguishable. The MS profiles of the GPLs from cells are dependant on the conditions they were grown and several isobaric isomers were identified for many of the molecular species. These multiple-stage mass spectrometric approaches give detailed structures of GPL in complex mixtures of which the isomeric structures are difficult to define using other analytical methods.

Introduction

Glycopeptidolipids (GPLs) (also called apolar C-mycosides) contain a tripeptide-amino alcohol core of D-phenylalanine-D-*allo*-threonine-D-alanine-L-alaninol (D-Phe-D-allo-Thr-D-Ala-L-alaninol), to which a 3-hydroxy or 3-methoxy C_{26–34} fatty acyl chain is attached at the N-terminal D-Phe through amide linkage ^[1,2]. The D-allo-Thr and the terminal L-alaninol are further linked with 6-deoxytalose (6-dTal) and *O*-methyl rhamnose, respectively, to form serovar non-specific GPL (nsGPL) that are found in all species of *M.*

*To whom the correspondence should be addressed: Dr. Fong-Fu Hsu, Box 8127, Washington University School of Medicine, 660 S Euclid, St. Louis, MO 63110. Tel: 314-362-0056; Fax: 314-362-7641; fhsuim.wustl.edu.

avium complex (MAC) [3,4,2]. The nsGPL is further glycosylated at 6-dTal to produce serovar-specific polar GPLs[5].

GPLs are abundant components of the cell walls of different species of non-tuberculous mycobacteria. They are important cell-surface antigens in *Mycobacterium avium* (*M. avium*), *Mycobacterium intracellulare* and *Mycobacterium scrofulaceum*, and have been implicated in the pathogenesis of the opportunistic mycobacteria belonging to the *M. avium* complex. GPLs from MAC and *M. smegmatis* were found to inhibit the process of phagocytosis by macrophages [6]. The antigenic polar GPL (serovar specific) has been associated with the change in colony morphology among natural isolates of *M. avium* [7]. GPLs also play important roles in biofilm formation, cell invasion, and immune modulation; and variations in GPL structure, for example, structural modifications of the carbohydrate moieties, may affect the pathogenicity of *M. avium* complex and other GPL-expressing mycobacteria [8]. The immunomodulation aspects of *M. avium* pathogenesis and the role of GPLs as virulence factors have been recently reviewed [9,10].

The *Mycobacterium smegmatis* (*M. smegmatis*) mc²155 strain is often used for biosynthetic studies as this strain is a "fast grower", non-pathogenic, easy to manipulate and produces simple range of nsGPL structures of which the *allo* Thr residue is glycosylated with a diacetylated or monoacetylated 6-deoxy talose (dTal); whereas the alaninol is glycosidically attached to 3,4 di-*O*-Methyl rhamnose (Rha) or 2,3,4 tri-*O*-Methyl Rha. Variations in the GPL composition can occur depending on culture conditions [6,11,12]. Four major species of GPLs occur in *M. smegmatis* mc²155 when cultivated aerobically on nutritionally complex media. Depending on the methylation state of Rha, the species contain a 3,4 di-*O*-Methyl Rha and a 2,3,4 tri-*O*-Methyl Rha, respectively, with a 3-methoxy C₂₆₋₃₄ long chain fatty acyl residue are named GPL-1 and GPL-2, respectively; while GPL-1a and GPL-2a contain a 3-hydroxy C₂₆₋₃₄ fatty acyl chain with 3,4 di-*O*-Methyl Rha and 2,3,4 tri-*O*-Methyl Rha, respectively. In addition, a minor GPLs III subfamily composed of a 6-deoxytalosyl unit and a 2-succinyl 3,4-di-*O*-CH₃-rhamnosyl group linked to the *allo*-threoninyl and terminal alaninol residues, respectively, is also present [6].

Mass spectrometric approaches employing various ionization methods including field ionization, FAB, MALDI, ESI and GC/MS in conjunction with chemical degradation and derivatization, together with nuclear magnetic resonance spectroscopy had play major roles in the structural characterization of GPLs [6,13,14,15,16,17,18]. However, these approaches are laborious and the detailed structural information of the minor isobaric isomers is not available.

We previously applied ESI multiple-stage mass spectrometry (MSⁿ) combined with high resolution mass spectrometry for structural elucidation of complex microbial lipids isolated from *Mycobacterium tuberculosis* [19,20,21], *Streptococcus pneumoniae* [22] and *Rhodococcus equi* [23,24]. The MSⁿ mass spectrometric approaches permit structures of the complex lipids to be unveiled in detail. Here, we describe the similar LIT ESI-MSⁿ approach in combination with high resolution mass spectrometry towards structural characterization of GPLs isolated from *M. smegmatis* grown in different conditions. Our results readily revealed the structures of this lipid family including many isobaric isomers.

Methods

Sample preparation

M. smegmatis strain mc²155 [25] were grown to mid-logarithmic phase at 37°C shaking in 7H9 liquid medium (Difco) containing oleate-albumin-dextrose-catalase (OADC) supplements (Becton Dickinson). A 50 mL culture was autoclaved 20 minutes, 121°C, and

bacteria harvested by centrifugation. The bacterial pellet was resuspended in 5 mL chloroform:methanol (2:1, v/v). The resuspended pellet was sonicated for 5 minutes and incubated for an additional 1 h to extract total lipids. Following a second extraction with 5 mL chloroform:methanol (2:1, v/v), lipid samples were pooled and dried under N₂ gas. The dry weight yield of lipids was ~3 mg. Total lipids were also isolated from *M. smegmatis* mc²155 grown on LB agar (Difco) plates. An aliquot of *M. smegmatis* (approximately 1×10⁴ cfu) was spread evenly on an agar plate. Three-days post-innoculation, the resulting lawn was harvested using a cell scraper, autoclaved and resuspended in 5 mL chloroform:methanol (2:1, v/v). Total lipids were isolated as above, yielding ~2 mg of total lipids. The crude lipids in 300 µL CHCl₃/CH₃OH (2:1; vol/vol) were loaded to a 3 mL/200 mg Macherey-Nagel amino Chromabond Sep-Pak column (Duren, Germany). The column was first washed with 2 mL EtOAc:Hexane (15:85; vol/vol), followed by 1.5 mL di-isopropyl ether:HOAc (98:2; vol/vol), and then eluted with 2 mL acetone/methanol (9:1.35; vol/vol) (by gravity) to a vial. The eluant containing GPLs was dried under a stream of nitrogen. The dried sample was re-dissolved in CHCl₃/CH₃OH (1/2) before ESI-MS analysis

Mass spectrometry

Both high-resolution (R=100,000 at m/z 400) and low-energy CAD tandem mass spectrometry experiments were conducted on a Thermo Scientific (San Jose, CA) LTQ Orbitrap Velos mass spectrometer (MS) with Xcalibur operating system. GPLs in chloroform/methanol (1/2) were infused (1.5 µL/min) to the ESI source, where the skimmer was set at ground potential, the electrospray needle was set at 4.0 kV, and temperature of the heated capillary was 300°C. The automatic gain control of the ion trap was set to 5×10⁴, with a maximum injection time of 100 ms. Helium was used as the buffer and collision gas at a pressure of 1×10⁻³ mbar (0.75 mTorr). The MSⁿ experiments were carried out with an optimized relative collision energy ranging from 35–40% and with an activation q value at 0.25, and the activation time at 10 ms to leave a minimal residual abundance of precursor ion (around 20%). The mass selection window for the precursor ions was set at 1 Da wide to admit the monoisotopic ion to the ion-trap for collision-induced dissociation (CID) for unit resolution detection in the ion-trap or high resolution accurate mass detection in the Orbitrap mass analyzer. Mass spectra were accumulated in the profile mode, typically for 3–10 min for MSⁿ-spectra (n=2,3,4).

Nomenclature

To simplify data interpretation, we adopt the rules recommended by IUPAC with modification for abbreviation. The *O*-acetyl 6-deoxytalose or di-*O*-acetyl 6-deoxytalose residue attached to the D-allo-Thr is abbreviated as Ac-dTal or (Ac)₂-dTal, respectively. The 3,4-di-*O*-methyl rhamnose, 2,3,4-tri-*O*-methyl rhamnose or 2-succinyl 3,4-di-*O*-methyl rhamnose residue attached to the terminal L-alaninol is designated as (CH₃)₂-Rha, (CH₃)₃-Rha, or HO₂C(CH₂)₂CO-(CH₃)₂-Rha, respectively. The 3-hydroxy or 3-methoxy C_{26–34} fatty acyl chain attached to the N-terminal D-Phe, such as 3-hydroxy-octaeicosanoyl or 3-methoxy-triacosenoyl is abbreviated as hC28:0 or mC30:1, respectively. Thus, GPL that consists of a common D-Phe-D-allo-Thr-D-Ala-L-alaninol core structure linked to a 3-hydroxy-octaeicosanoyl group, a di-*O*-acetyl 6-deoxytalose and a 2,3,4-tri-*O*-methyl rhamnose residues is designated as hC28:0/(Ac)₂-dTal/(CH₃)₃-Rha GPL; while an analogous GPL with 3-methoxy fatty acyl chain is designated as mC28:0/(Ac)₂-dTal/(CH₃)₃-Rha GPL. The convention for description of peptide fragments proposed by Roepstorff and Fohlman [26] and later modified by Biemann [27] was used for the observed peptide ions in the MSⁿ spectra.

Results and Discussion

When subjected to ESI, GPLs isolated from *M. smegmatis* readily formed abundant $[M + Na]^+$ ions in the positive-ion mode. As shown in Fig. 1, the spectra of GPLs from *M. smegmatis* grown on plates (Panel a) and in broth (Panel b) contained an array of homologous $[M + Na]^+$ ions with an intermittence of 14 Da. The profiles of the two spectra are different, demonstrating that the compositions of GPL varied dependent on culture conditions [6,11,12]. In the negative-ion mode, ions in the forms of $[M - H]^-$, and of $[M + X]^-$ ($X = CH_3CO_2$, HCO_2 , Cl) were observed dependent on the presence of X^- ions in the solution. The profile of the spectrum of the $[M - H]^-$ ions (Fig. 1c) of GPLs from *M. smegmatis* grown in broth is similar to that seen as the $[M + Na]^+$ ions (Fig. 1b). However, the spectrum also contained the homologous $[M + Cl]^-$ ions (the ions marked with “*” in Panel C) that complicate the analysis. MS^n on the $[M + Na]^+$ ions of GPLs yielded rich structural information applicable for nearly complete characterization. In contrast, structural information derived from MS^n on the $[M - H]^-$ ions of GPL is incomplete (data not shown). Thus, the LIT MS^n with high-resolution mass spectrometric approach towards structural characterization of various GPLs as described below will focus on the $[M + Na]^+$ ions desorbed by ESI.

High-resolution mass spectrometry on the $[M + Na]^+$ and $[M - H]^-$ ions of GPLs

High resolution mass measurements ($R=100,000$ at m/z 400) on the $[M + Na]^+$ (Table 1 & 2) and the corresponding $[M - H]^-$ ions (see supplemental material Table S1) of GPLs from *M. smegmatis* gave the elemental compositions that are consistent with the GPL structures reported previously [2]. High resolution mass spectrometry also readily resolved the GPL species that have same nominal masses but belong to different GPL subfamilies varied by the acetylation state of 6-deoxy-Tal, the methylation state and succinyl substitution of the terminal rhamnose residues. For example, the $[M + Na]^+$ ions with nominal mass of m/z 1241 comprise ions at m/z 1241.7771 and 1241.8127, corresponding to elemental compositions of $C_{65}H_{110}O_{17}N_4Na$ (calculated mass: 1241.7758) and $C_{66}H_{114}O_{16}N_4Na$ (calculated: 1241.8122), respectively, reflecting the presence of the GPLs containing a diacetyl-6-deoxy-talose and an acetyl-6-deoxy-talose residues (Table 1). The results are consistent with high-resolution mass measurements on the corresponding $[M - H]^-$ ions that resulted in the ion pair with m/z 1217.7789 and 1217.8150 corresponding to elemental compositions of $C_{65}H_{109}O_{17}N_4$ (calculated mass: 1217.7793) and $C_{66}H_{113}O_{16}N_4$ (calculated mass: 1217.8157), respectively (Supplemental material Table S1). The elemental composition of $C_{69}H_{116}O_{20}N_4Na$ (calculated mass: 1343.8075) deduced from the measured mass of m/z 1343.8081 and the analogous $[M - H]^-$ ion of m/z 1319.8110, which corresponds to elemental composition of $C_{69}H_{115}O_{20}N_4$ (calculated mass: 1319.8110), indicates the presence of a GPL with 2-succinyl 3,4-di-*O*-CH₃-rhamnosyl substituent. High resolution mass spectrometry also confirms the state of unsaturation of the fatty acyl chains. For example, the ion of m/z 1243.7911 ($C_{65}H_{112}O_{17}N_4Na$, calculated mass: 1243.7915) represent GPLs with saturated fatty acyl chain, while ion of m/z 1241.7771 ($C_{65}H_{110}O_{17}N_4Na$, calculated mass: 1241.7758) contains two less hydrogen atoms due to the presence of a double bond in the fatty acyl substituent. Although accurate mass measurements do not differentiate a GPL possessing a 3-hydroxy fatty acyl group from a GPL possessing a 3-methoxy fatty acyl group, MS^n on the $[M + Na]^+$ ions of GPLs yield fragment ions informative for distinction of these two GPL families. The MS^n mass spectrometric approaches leading to the elucidation of the GPL structures are described below.

Structural Characterization of GPL as $[M + Na]^+$ adduct ions and the fragmentation processes

While fragment ions from MS^2 on the $[M + Na]^+$ ions of GPL provide information for nearly complete identification of the structure, the sequential MS^n ($n = 3, 4$) and high resolution mass spectra confirm the fragmentation processes and the structural identification. For example, the MS^2 spectrum of the $[M + Na]^+$ ion at m/z 1257 (Fig. 2a) is dominated by the ion at m/z 1009 arising from loss of 3,4-diacetyl 6-deoxy-talose (248 Da), together with the ion at m/z 1069, arising from loss of the terminal 2,3,4-trimethylrhamnose (188 Da). The ions at m/z 877 (f_2) arose from loss of the 3-hydroxy long-chain fatty acid residue as an aldehyde (380 Da) via cleavage of the C2–C3 bond and gave rise to 859 ($f_2 - H_2O$) by further loss of H_2O ; while the ion of m/z 835 (f_1) arose from elimination of the fatty acyl chain as a fatty acyl ketene (422 Da) (Scheme 1a). The results are consistent with the attachment of a 3-hydroxy-octaeicosanoyl group (hC28:0) to the D-Phe via an amide bond. The spectrum also contained the ions at m/z 966 (a_3), 923 (b_2), 895 (a_2), 592 (b_1), 357 (y_2) and 688 (y_3) arising from cleavages of peptide bonds. The observation of the ions at m/z 1197 and 1137 arising from consecutive losses of the acetyl groups (loss as acetic acid) are consistent with the presence of 3,4-diacetyl groups on the 6-deoxy talose residue. This notion is further supported by the MS^3 on the ions of m/z 1197 which yielded ions at m/z 1137, along with ions arising from the analogous losses (data not shown). Further dissociation of the ion of m/z 1009 ($1257 \rightarrow 1009$; Fig. 2b) gave rise to ions at m/z 821 arising from further loss of the terminal 2,3,4-trimethylrhamnose (Scheme 1b). The spectrum also contained ions at m/z 746 (b_3), 718 (a_3), 675 (b_2), 647 (a_2), 592 (b_1), 440 (y_3), 383 (x_2-2), 357 (y_2), and 286 (y_1) arising from similar cleavages of the peptide bonds, together with ions at m/z 629 (f_2) and 587 (f_1) arising from losses of the 3-hydroxy long-chain fatty acid residue as an aldehyde (380 Da) and as a fatty acyl ketene (422 Da), respectively, and the ion of m/z 611 ($629 - H_2O$). These later ions from cleavages of fatty acyl chain and the ions at m/z 440 (y_3), 675 (b_2), 647 (a_2), and 718 (a_3) are 248 Da lighter than the analogous ions of m/z 877, 835, 859, 688 (y_3), 923 (a_2), 966 (a_3) seen in Fig. 2(a), respectively; while ions at m/z 592 (b_1) and 357 (y_2) are identical. The results indicate that the diacyl-6-deoxy-talose residue is attached to the *allo*-threoninyl residue of the peptide core, consistent with the structure previously reported (Scheme 1b) [2]. In contrast, further dissociation of the ion of m/z 1069 ($1257 \rightarrow 1069$; Fig. 2c) yielded ions at m/z 966 (a_3), 923 (b_2), 895 (a_2), 592 (b_1) (Scheme 1c), identical to those seen in Fig. 2(a); while ions at m/z 689, 647, 671, and 500 (y_3) are 188 Da lighter than the analogous ions seen in Fig. 2(a). These mass shifts in the analogous ions readily locate the trimethylrhamnose residue attached to the terminal L-alaninol.

The ions at m/z 821 seen in Fig. 2(b) and 2(c) arose from losses of both of the glycoside substituents attached to the peptide core and are identical. This notion is supported by that the MS^4 spectra of the ion of m/z 821 originated from ion of m/z 1009 ($1257 \rightarrow 1009 \rightarrow 821$; Fig. 2d) and from ion of m/z 1069 ($1257 \rightarrow 1069 \rightarrow 821$; data not shown) are identical. The spectrum (Fig. 2d) contained the ions at 564 (a_1), 592 (b_1), 647 (a_2), 675 (b_2), 718 (a_3), 746 (b_3), 252 (y_3) arising from similar cleavages of the peptide chain (Scheme 1d), along with the ions at m/z 441 and 399 arising from elimination of the 3-hydroxy long chain fatty acyl chain as an aldehyde and as a ketene, respectively, and the ion at m/z 423 ($441 - H_2O$). The observed mass shifts of the analogous fragment ions arising from cleavages of the peptide bonds again, readily locate the 6-dTal and $(CH_3)_3Rha$ residues linked to the peptide core.

MS^4 on the ion of m/z 629 ($1257 \rightarrow 1009 \rightarrow 629$; Fig. 2e), gave rise to ions at m/z 267 (a_2), 295 (b_2), 338 (a_3), 366 (b_3), 440 (y_3), 357 (y_2), 383 (x_2-2) and 286 (y_1) (Scheme 1e). As compared to the analogous ions seen in Fig. 2(b), a mass shift of 380 Da was seen for the ions of 267 (a_2), 295 (b_2), 338 (a_3), 366 (b_3), while ions at 440 (y_3), 357 (y_2), 383 (x_2-2)

and 286 (y_1) remain unchanged. The results indicate that the 3-hydroxy fatty acyl group is located at N-terminal of D-Phe. The assignments of the fragment ions are consistent with high-resolution MS² mass spectrometry (See supplemental material Table S2).

Similarly, MS² on the $[M + Na]^+$ ion of m/z 1269 (Fig. 3a) arising from GPL isolated from cells grown in broth contained prominent ion at m/z 1021 arising from elimination of (Ac)₂-dTal. The spectrum also contained the ion at m/z 1095 arising from loss of 3,4-dimethylrhamnose ((CH₃)₂Rha) (174 Da), together with the ion of m/z 863 arising from cleavage of 3-hydroxy-triacosenoyl substituent as octaeicosenaldehyde. These results indicate that the molecule consists of an unsaturated hC30:1 fatty acyl chain, along with (Ac)₂-dTal and (CH₃)₂-Rha residues. Further dissociation of the ion of m/z 1021 ($1269 \rightarrow 1021$; Fig. 3b) yielded prominent ion of m/z 847 (loss of (CH₃)₂Rha), consistent with the presence of terminal (CH₃)₂Rha, together with ions at m/z 615 (f_2), 597 ($f_2 - H_2O$), and 573 (f_1) from cleavages of the hC30:1 fatty acyl substituent; and ions at m/z 343 (y_2), 426 (y_3), 673 (a_2), 701 (b_2), 744 (a_3), and 772 (b_3) arising from cleavages of the peptide bond. The results give assignment of the hC30:1/(Ac)dTal/(CH₃)₂-Rha GPL structure. The mass shifts of the observed fragment ions in Fig. 3(a) and 4(b) as compared to the analogous ions seen in Fig. 2(a) and 2(b) readily determine the fatty acyl chain length, and are consistent with the peptide sequence and the locations of the glycosyl groups as seen earlier. The fragmentation processes and the assignments of the fragment ions (Scheme S1) are also supported by high resolution tandem mass spectrometry (not shown).

Characterization of GPLs with 3-methoxy fatty acyl chain

The ions arising from cleavages of C2-C3 bond of 3-hydroxy fatty acyl chain (Scheme 1; f ion series), e.g., the ions at m/z 877 in Fig. 2(a) and at m/z 863 in Fig. 3(a), are abundant in the MSⁿ spectra of the GPLs containing 3-hydroxy fatty acyl substituent. However, similar fragment ions were not observed in those arising from GPLs consisting of 3-methoxy fatty acyl chain. For instance, MS² on the $[M + Na]^+$ ions of m/z 1271 (Fig. 3c) yielded ions of m/z 1211 and 1151 arising from consecutive losses of the acetyl groups as acetic acids, indicating the presence of acetyl groups on the d-Tal group; and the prominent ion at m/z 1023 arising from loss of (Ac)₂-d-Tal (loss of 248 Da), together with ions at m/z 578 (a_1), 606 (b_1), 909 (a_2), 937 (b_2), 980 (a_3), 688 (y_3), 383 (x_2-2), 357 (y_2) arising from cleavages of peptide bonds (Scheme 2a). However, analogous ions expected at m/z 877 (f_2) and 835 (f_1) arising from cleavage of the fatty acyl chain is absent; while ion at m/z 1239 arising from loss of CH₃OH was observed. The profile of the MS³ spectrum of the ion of m/z 1023 ($1271 \rightarrow 1023$; Fig. 3d) is readily distinguishable from that of Scheme 1(b), arising from similar loss of (Ac)₂-dTal residue attached to D-allo-Thr. These results demonstrate the utility of MSⁿ in the differentiation of the GPL families consisting of 3-hydroxy fatty acyl and of 3-methoxy fatty acid substituents. The spectrum (Fig. 3d) contained ions at 578 (a_1), 606 (b_1), 661 (a_2), 689 (b_2), 732 (a_3), 760 (b_3), 440 (y_3), 383 (x_2-2), 357 (y_2), and 286 (y_1) from cleavages of the peptide bonds (Scheme 2b), consistent with the sequence of the peptide chain as seen earlier. The ions at m/z 991 arose from loss of 3-methoxy group on the long-chain fatty acyl residue as methanol, and gave rise to the prominent ion of m/z 973 by further loss of a H₂O molecule. These fragmentation processes are further supported by the MS⁴ spectrum of the ion of m/z 991 ($1271 \rightarrow 1023 \rightarrow 991$; Fig. 3e), which is dominated by the ion of m/z 973 (loss of H₂O), along with ions at m/z 629 (a_2), 657 (b_2), 700 (a_3), 728 (b_3), 440 (y_3), 383 (x_2-2), 357 (y_2) and 286 (y_1) from cleavages of the peptide bonds (Scheme 2c). The ions at m/z 629 (a_2), 657 (b_2), 700 (a_3), 728 (b_3) are 32 Da lighter than the analogous ions seen in Fig. 3(d), while ions at m/z 440 (y_3), 383 (x_2-2), 357 (y_2) and 286 (y_1) are unchanged, supporting that the loss of CH₃OH residue indeed arises from elimination of the methoxy group on the fatty acyl residue. This loss of CH₃OH unique to

the GPLs containing 3-methoxy-fatty acyl group also permits their distinction from the GPLs containing 3-hydroxy-fatty acyl group.

The MS⁴ spectrum of the ion of m/z 835 ($1271 \rightarrow 1023 \rightarrow 835$; Fig. 3f), arising from losses of 6-deoxy-talose and the terminal rhamnose residues also confirm their location attached to the peptide core. As compared to its MS² (Fig. 3d) and MS³ (Fig. 3e) spectra, the observed mass shifts of the analogous b and y ions (Scheme 2d) readily confirms the above described structure.

Characterization of GPLs containing a terminal 2- succinyl 3,4-dimethylrhamnose

The presence of a succinyl group in the rhamnose residue drastically altered the MS² spectra of GPLs. For example, the $[M + Na]^+$ ion of m/z 1369 (Fig. 4a) is dominated by the ion of m/z 1095 arising from loss of the terminal 2-succinyl 3,4-dimethyl rhamnose ($HO_2C(CH_2)_2CO-(CH_3)_2-Rha$; 274 Da); whereas the ion at m/z 1121 arising from loss of diacetyl-dTal is less abundant. This reversal in the appearance of these fragment ions that arise from loss of the sugar residues due to modification of the terminal rhamnose readily afford its distinction from other GPL families possessing a terminal trimethylrhamnose residue (Fig. 2a and Fig. 3a). The spectrum also contained the ion at m/z 963 (f₂) arising from loss of a hC30:1 fatty acyl chain as aldehyde (loss of 406 Da), and the ion of m/z 689 arising from further loss of the same aldehyde from m/z 1095, indicating that the ion represents an hC30:1/(Ac)₂-dTal/ $HO_2C(CH_2)_2CO-(CH_3)_2-Rha$ GPL. The presence of the 2-succinyl group in the molecule is readily recognized by the observation of the ion of m/z 1269, arising from loss of succinic acid (loss as dehydrated succinic acid; 100 Da) (Scheme 3a). Further dissociation of the ion of m/z 1269 ($1369 \rightarrow 1269$; data not shown) yielded MS³ spectrum identical to that arising from hC30:1/(Ac)₂-dTal/(CH₃)₂-Rha GPL shown in Fig. 3(a). The MS⁴ spectrum of the ion of m/z 1021 ($1369 \rightarrow 1269 \rightarrow 1021$; data not shown) is also identical to that shown in Fig. 3(b). The results demonstrated that the ion of m/z 1369 first eliminates a 2-succinyl residue to yield the ion of m/z 1269, which is equivalent to a sodiated hC30:1/(Ac)₂-dTal/(CH₃)₂-Rha GPL ion. MS³ on the ion of m/z 1095 ($1369 \rightarrow 1095$; Fig. 4b) yielded prominent ions at m/z 847 and 689 (f₂) arising from further losses of (Ac)₂-dTal (248 Da) and of hC30:1- fatty acyl group as an aldehyde, respectively; together with ions at m/z 500 (y₃), 618 (b₁), 921 (a₂), 949 (b₂), 992 (a₃), and 1020 (b₃) arising from cleavages of the peptide bonds (Scheme 3b). The profile of the spectrum (Fig. 4b) is similar to that of the MS³ spectrum of the ion of m/z 1069 (Fig. 2c), consistent with the notion that both the ions of m/z 1095 and 1069 arose from the GPLs of which the terminal rhamnose has been cleaved. The structural assignments are consistent with the previous report by Villeneuve et al [25]. The presence of this GPL class consisting a succinyl residue is also supported by high resolution mass measurements of the $[M + Na]^+$ ions as shown earlier (Table 1 & 2), in which ion series of m/z 1341, 1355, ... and 1411 possessing elemental compositions of C₆₉H₁₁₄(CH₂)_nO₂₀N₄Na (n=0, 1, ..5) arise from the GPLs consisting of a unsaturated fatty acyl chain, together with the ions of m/z 1343 and 1357 (C₆₉H₁₁₆(CH₂)_nO₂₀N₄Na (n=1,2)) bearing a saturated fatty acyl chain.

Characterization of GPL species with different fatty acyl chain length and modification in the glycoside substituents

The structure diversity of GPLs isolated from cells grown in different conditions is seen not only in the variation of the methylation state of rhamnose and acetylation state of 6-dTal residues linked to the peptide core, but also in the different fatty acyl chain lengths (Table 1). For example, the MS² spectrum of the $[M + Na]^+$ ion of m/z 1199 (Fig. 5a) yielded ions at m/z 993 and 1025, arising from losses of acetyl-dTal (206 Da) and dimethyl-Rha (174 Da), respectively, indicating the presence of Ac-dTal and (CH₃)₂-Rha substituents. These structural information combined with the ions seen at m/z 821 (f₂), 803 (821 – H₂O) and

779 (f1) arising from cleavages of hC28:1 fatty acyl substituent, indicated the presence of hC28:1/Ac-dTal/(CH₃)₂-Rha GPL isomer. This structure assignment is further supported by the MS³ spectrum of m/z 993 (1199 → 993, Fig. 5b), which contained ions at m/z 819 arising from loss of (CH₃)₂-Rha residue, together with ions at m/z 615 (f2), 597 (615 – H₂O), 571 (f1), 343 (y₂), 360 (x₂-2), 426 (y₃), 645 (a₁), 673 (b₂), 716 (a₃), 744 (b₃), consistent with the assigned structure. The spectrum (Fig. 5a) also contained ions at m/z 1039 and 951 arising from losses of methyl-Rha (160 Da) and diacetyl-dTal residues, respectively, together with ions at m/z 849 (f2) and 807 (f1) arising from cleavages of the hC26:1-fatty acyl chain, indicating the presence of minor isomer of hC26:1/(Ac)₂-dTal/CH₃-Rha. The assignment of this latter structure is confirmed by MS³ on the ion of m/z 951 (1199 → 951; Fig. 5c). The profile of the spectrum is similar to those of Fig. 2(b) and Fig. 5(b) and contained abundant ions at m/z 791 arising from loss of CH₃-Rha, and ions at m/z 601 (f2), 583 (601 – H₂O), and 559 (f1) deriving from cleavages of the hC26:1-fatty acyl chain, along with ions at m/z 329 (y₂), 355 (x₂-2), 412 (y₃), 617 (a₂), 645 (b₂), 688 (a₃), 716 (b₃). The results showed the presence of GPLs with lesser acetylation of dTal and methylation of Rha. The identification of hC26:1-fatty acyl substituent also represents the shortest fatty acyl chain found in GPLs. The presence of the isobaric isomers is consistent with high resolution mass measurements (Table 1), which showed two ions at m/z 1199.7293 and 1199.7661, corresponding to elemental compositions of C₆₂H₁₀₄O₁₇N₄Na (calculated mass; 1199.7289) and of C₆₃H₁₀₈O₁₆N₄Na (calculated mass; 1199.7653), respectively, arising from the two isomeric structures.

The MS² spectrum of the ion of m/z 1325 seen in the GPL mixture isolated from *M. smegmatis* grown on LB agar plates (Fig. 5d) is dominant by the ion of m/z 1077 arising from loss of diacyl-dTal, along with ion at m/z 1137 arising from loss of trimethyl-Rha. However, ions arising from cleavages of fatty acyl chain were absent, suggesting the molecule contains a 3-methoxy fatty acyl rather than a 3-hydroxy fatty acyl substituent. The profile of the MS³ spectrum of the ion of m/z 1077 (1325 → 1077; Fig. 5e) is similar to that shown in Fig. 3(d), consistent with that the molecule consist of a 3-methoxy fatty acyl substituent. The ions at 357 (y₂) 383 (x₂ – 2), 440 (y₃) are identical to those seen in Fig. 3(d), but ions at 660 (b₁), 715 (a₂), 743 (b₂), 786 (a₃), 814 (b₃) are 54 Da (C4:1) heavier indicating the molecule represents a mC32:1/(Ac)₂-dTal/(CH₃)₃-Rha GPL. The identification of mC32:1-fatty acyl chain is agreement with the report of the longest fatty acyl substituent found for *M. smegmatis* GPLs [25].

Characterization of GPL species consisting of isobaric isomers

Nearly all the GPL species isolated from *M. smegmatis* grown on LB agar contained more than one isomer (Table 1). For example, the MS² spectrum of the ion of m/z 1269 (Fig. 6a) contained ions at m/z 1095, and 1081 arising from losses of the terminal rhamnose with 3 and with 2 methyl groups, respectively, together with ions at m/z 1063 and 1021, arising from losses of the 6-deoxytalose with one and with two acetyl groups, respectively. Further dissociation of the ion of m/z 1021 (1269 → 1021; Fig. 6b) yielded prominent ions at m/z 1003 and 971 arising from consecutive losses of H₂O and CH₃OH residues, and the ion at m/z 833 arising from loss of trimethylrhamnose, together with ions at m/z 659 (a₂), 687 (b₂), 730 (a₃), 758 (b₃), 440 (y₃), 383 (x₂-2), 357 (y₂), and 286 (y₁). The profile of these ions is similar to that of Fig. 3(d), suggesting the presence of mC28:1-fatty acyl chain. The presence of a double bond on the fatty acyl chain is supported by observation of the ions at m/z 659 (a₂), 687 (b₂), 730 (a₃), and 758 (b₃), which are 2 Da lighter than the analogous ions seen in Fig. 3(d). The results gave assignment of mC28:1/(Ac)₂-dTal/(CH₃)₃-Rha GPL. The spectrum (Fig. 6b) also contained additional set of ions at m/z 847 (loss of dimethylrhamnose) and prominent ions at m/z 615 (f2), 597 (f2 – H₂O) and 573 (f1) arising from cleavages of h30:1-fatty acid substituent, along with ions at m/z 343 (y₂), 426 (y₃), 369

($x_2 - 2$), 673 (a_2), 701 (b_2), 744 (a_3), and 772 (b_3) from cleavages of the peptide bonds. The profile of the ion set is similar to that of Fig. 3(b), indicating the presence of h30:1-fatty acid residue and a h30:1/(Ac)₂-dTal/(CH₃)₂-Rha GPL isomer as shown earlier (Fig. 3a & 4b).

Further dissociation of the ion of m/z 1063 ($1269 \rightarrow 1063$; Fig. 6c) yielded prominent ion at m/z 629 (f_2), 601 ($f_2 - H_2O$) and 587 (f_1) arising from cleavages of h32:1-fatty acyl chain, consistent with the presence of a 3-hydroxy rather than a 3-methoxy fatty acyl group attached to the phenylalanine of the peptide core. The spectrum also contained the ion at m/z 875 arising from loss of the terminal trimethyl rhamnose, together with ions at m/z 646 (b_1), 701 (a_2), 729 (b_2), 772 (a_3), 800 (b_3) 357 (y_2) and 440 (y_2), indicating the presence of hC32:1/(Ac)-dTal/(CH₃)₃-Rha GPL isomer.

The above structural assignments were further supported by the MS⁴ spectra of the ion of m/z 847 ($1269 \rightarrow 1021 \rightarrow 847$; Fig. 6d) and of m/z 833 ($1269 \rightarrow 1021 \rightarrow 833$; Fig. 6e). The profile of the former spectrum (Fig. 6d) is similar to that of Fig. 2(d) arising from GPLs consisting of 3-hydroxy fatty acyl group. The observation of the ions at m/z 441 (f_2) and 399 (f_1) from cleavages of h30:1-fatty acid substituent together with ions at 590 (a_1), 618 (b_1), 673 (a_2), 701 (b_2), 744 (a_3), 772 (b_3), further supports the assignment of the structure of hC30:1/(Ac)₂-Tal/(CH₃)₂-Rha GPL. In contrast, the profile of the MS⁴ spectrum of m/z 833 (Fig. 6e) is similar to that of Fig. 3(f), arising from GPL containing a 3-methoxy-fatty acid substituent. The spectrum also contained ions at m/z 252 (x_3), 280 (y_3), 576 (a_1), 604 (b_1), 659 (a_2), 687 (b_2), 730 (a_3), 758 (b_3), consistent with the assignment of mC28:1/(Ac)₂-dTal/(CH₃)₃-Rha GPL. The MS⁴ spectrum profile of the ion of m/z 875 ($1269 \rightarrow 1063 \rightarrow 875$; data not shown) is similar to that of Fig. 6(d) and contained prominent ions of m/z 441 (f_2), 423 ($f_2 - H_2O$) and 399 (f_1), from cleavages of hC32:1-fatty acyl substituent, along with the analogous a, b, and y ion series arising from cleavages of peptide bonds. The results confirm the assignment of hC32:1/(Ac)-dTal/(CH₃)₃-Rha GPL isomer.

In contrast, the ion of m/z 1269 isolated from *M. smegmatis* grown in broth (Fig. 3a & 4b) represents a major hC30:1/(Ac)₂-dTal/(CH₃)₂-Rha GPL isomer. The spectrum of Fig. 3(b) is readily distinguishable from that shown in Fig. 6(b), which is a composite spectrum from hC30:1/(Ac)₂-dTal/(CH₃)₂-Rha GPL and mC28:1/(Ac)₂-dTal/(CH₃)₃-Rha GPL isomers. The variation in the GPL structures among the same m/z isolated from lipid extracts grown in different media was also seen for many other ion species (Table 1&2).

Conclusions

The multiple stage mass spectrometric approach as described here is simple and provides structural information of GPLs including their chain length (from C26 to C32 with 0 or 1 double bond) of the 3-hydroxy or 3-methoxy fatty acyl substituents, the acetylation state of the 6-deoxy talose, the methylation and succinylation states of the rhamnose residues. MSⁿ on the $[M + Na]^+$ ions also confirms the attachment of the 6-deoxy talose and the rhamnose residues to the D-allo-Thr and L-alaninol of the peptide backbone, respectively. This mass spectrometric approach also affords distinction among GPL families consisting of 3-hydroxy-, 3-methoxy fatty acyl chain, and of succinyl group attached to the terminal rhamnose. GPL glycosylation and modification play important roles in GPL-mediated functions, and variations in GPL structure may also affect the pathogenicity of *Mycobacterium avium* complex and other GPL-expressing mycobacteria. Thus, the present method will be useful for study of GPL, leading to a better understanding of the intricate biosynthetic pathway and the role of GPL in mycobacterial pathogenesis and immunogenicity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by US Public Health Service Grants P41-RR-00954, P60-DK-20579, P30-DK56341, P01-HL-57278 (mass spectrometry facility); and AI087840 (GEP).

Abbreviations

ESI-MS	electrospray ionization-MS
HRMS	high resolution mass spectrometry
LIT	linear ion-trap
GPLs	glycopeptidolipids

References

1. Laneelle G, Asselineau J. Structure of a peptidolipid glycoside isolated from a Mycobacterium. *Eur. J. Biochem.* 1968; 5:487–491. [PubMed: 5698610]
2. Brennan PJ, Goren MB. Structural studies on the type-specific antigens and lipids of the mycobacterium avium. Mycobacterium intracellulare. Mycobacterium scrofulaceum serocomplex. Mycobacterium intracellulare serotype 9. *J. Biol. Chem.* 1979; 254:4205–4211. [PubMed: 438184]
3. Aspinall GO, Chatterjee D, Brennan PJ. The variable surface glycolipids of mycobacteria: structures, synthesis of epitopes, and biological properties. *Adv. Carbohydr. Chem. Biochem.* 1995; 51:169–242. [PubMed: 7484363]
4. Belisle JT, Brennan PJ. Chemical basis of rough and smooth variation in mycobacteria. *J. Bacteriol.* 1989; 171:3465–3470. [PubMed: 2722755]
5. Brennan PJ, Nikaido H. The Envelope of Mycobacteria. *Annu. Rev. Biochem.* 1995; 64:29–63. [PubMed: 7574484]
6. Villeneuve C, Etienne G, Abadie V, Montrozier H, Bordier C, Laval F, Daffe M, Maridonneau-Parini I, Astarie-Dequeker C. Surface-exposed glycopeptidolipids of Mycobacterium smegmatis specifically inhibit the phagocytosis of mycobacteria by human macrophages. Identification of a novel family of glycopeptidolipids. *J. Biol. Chem.* 2003; 278:51291–51300. [PubMed: 14534313]
7. Barrow WW, Ullom BP, Brennan PJ. Peptidoglycolipid nature of the superficial cell wall sheath of smooth-colony-forming mycobacteria. *J. Bacteriol.* 1980; 144:814–822. [PubMed: 7430072]
8. Schorey JS, Sweet L. The mycobacterial glycopeptidolipids: structure, function, and their role in pathogenesis. *Glycobiology.* 2008; 18:832–841. [PubMed: 18723691]
9. Mukherjee R, Chatterji D. Glycopeptidolipids: Immuno-modulators in greasy mycobacterial cell envelope. *IUBMB Life.* 2012; 64:215–225. [PubMed: 22252955]
10. Rocco JM, Irani VR. Mycobacterium avium and modulation of the host macrophage immune mechanisms. *Int. J. Tuberc. Lung Dis.* 2011; 15:447–452. [PubMed: 21396201]
11. Patterson JH, McConville MJ, Haites RE, Coppel RL, Billman-Jacobe H. Identification of a methyltransferase from Mycobacterium smegmatis involved in glycopeptidolipid synthesis. *J. Biol. Chem.* 2000; 275:24900–24906. [PubMed: 10801784]
12. Ripoll F, Deshayes C, Pasek S, Laval F, Beretti J-L, Biet F, Risler J-L, Daffe M, Etienne G, Gaillard J-L, Reyat J-M. Genomics of glycopeptidolipid biosynthesis in Mycobacterium abscessus and M. chelonae. *BMC Genomics.* 2007; 8:114. [PubMed: 17490474]
13. Chatterjee D, Khoo K-H. The surface glycopeptidolipids of mycobacteria: structures and biological properties. *Cell. Mol. Life Sci.* 2001; 58:2018–2042. [PubMed: 11814054]
14. Lanéeelle M-A, Silve G, López Marín LM, Daffé M. Structures of the Glycolipid Antigens of Members of the third Biovariant Complex of Mycobacterium Fortuitum. *European Journal of Biochemistry.* 1996; 238:270–279. [PubMed: 8665946]

15. Besra GS, McNeil MR, Rivoire B, Khoo KH, Morris HR, Dell A, Brennan PJ. Further structural definition of a new family of glycopeptidolipids from *Mycobacterium xenopi*. *Biochemistry*. 1993; 32:347–355. [PubMed: 7678195]
16. Mederos L, Valdivia JA, Valero-Guillén PL. New variants of polar glycopeptidolipids detected in *Mycobacterium simiae*, including 'habana' strains, as evidenced by electrospray ionization-ion trap-mass spectrometry. *J. Appl. Microbiol.* 2008; 105:602–614. [PubMed: 18397261]
17. Daffe M, Laneelle MA, Puzo G. Structural elucidation by field desorption and electron-impact mass spectrometry of the C-mycosides isolated from *Mycobacterium smegmatis*. *Biochim. Biophys. Acta*. 1983; 751:439–443. [PubMed: 6849953]
18. Fujiwara N, Nakata N, Maeda S, Naka T, Doe M, Yano I, Kobayashi K. Structural Characterization of a Specific Glycopeptidolipid Containing a Novel N-Acyl-Deoxy Sugar from *Mycobacterium intracellulare* Serotype 7 and Genetic Analysis of Its Glycosylation Pathway. *J. Bacteriol.* 2007; 189:1099–1108. [PubMed: 17122347]
19. Hsu FF, Turk J, Owens RM, Rhoades ER, Russell DG. Structural Characterization of Phosphatidyl-myo-inositol Mannosides from *Mycobacterium bovis* Bacillus Calmette Guerin by Multiple-Stage Quadrupole Ion-Trap Mass Spectrometry with Electrospray Ionization. I. PIMs and Lyso-PIMs. *J. Am. Soc. Mass Spectrom.* 2007; 18:466–478. [PubMed: 17141526]
20. Hsu FF, Turk J, Owens RM, Rhoades ER, Russell DG. Structural Characterization of Phosphatidyl-myo-Inositol Mannosides from *Mycobacterium bovis* Bacillus Calmette Guerin by Multiple-Stage Quadrupole Ion-Trap Mass Spectrometry with Electrospray Ionization. II. Monoacyl- and Diacyl-PIMs. *J. Am. Soc. Mass Spectrom.* 2007; 18:479–492. [PubMed: 17141525]
21. Rhoades ER, Streeter C, Turk J, Hsu F-F. Characterization of Sulfolipids of *Mycobacterium tuberculosis* H37Rv by Multiple-Stage Linear Ion-Trap High-Resolution Mass Spectrometry with Electrospray Ionization Reveals That the Family of Sulfolipid II Predominates. *Biochemistry*. 2011; 50:9135–9147. [PubMed: 21919534]
22. Tatituri RV, Brenner MB, Turk J, Hsu FF. Structural elucidation of diglycosyl diacylglycerol and monoglycosyl diacylglycerol from *Streptococcus pneumoniae* by multiple-stage linear ion-trap mass spectrometry with electrospray ionization. *J Mass Spectrom.* 2012; 47:115–123. [PubMed: 22282097]
23. Hsu FF, Soehl K, Turk J, Haas A. Characterization of mycolic acids from the pathogen *Rhodococcus equi* by tandem mass spectrometry with electrospray ionization. *Anal. Biochem.* 2011; 409:112–122. [PubMed: 20946862]
24. Hsu F-F, Wohlmann J, Turk J, Haas A. Structural Definition of Trehalose 6-Monomycolates and Trehalose 6,6'-Dimycolates from the Pathogen *Rhodococcus equi* by Multiple-Stage Linear Ion-Trap Mass Spectrometry with Electrospray Ionization. *J. Am. Soc. Mass Spectrom.* 2011; 22:2160–2170. [PubMed: 21972013]
25. Snapper SB, Melton RE, Mustafa S, Kieser T, Jacobs WR Jr. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol. Microbiol.* 1990; 4:1911–1919. [PubMed: 2082148]
26. Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* 1984; 11:601. [PubMed: 6525415]
27. Biemann K. Contributions of mass spectrometry to peptide and protein structure. *Biomed. Environ. Mass Spectrom.* 1988; 16:99–111. [PubMed: 3072035]

Fig. 1.

The ESI mass spectra of the $[M + Na]^+$ ions of *M. smegmatis* GPLs isolated from cells grown in broth (a), in plate (b), and corresponding $[M - H]^-$ ions of GPL from broth (c). Ions marked with "*" (panel c) are $[M + Cl]^-$ ions.

Fig. 2.
The MS² spectrum of the ion of [M + Na]⁺ ions of m/z 1257 (a), its MS³ spectra of the ions of m/z 1009 (1257 → 1009) (b) of the ions of 1069 (1257 → 1069) (c), its MS⁴ spectra of m/z 821 (1257 → 1009 → 821) (d) and of m/z 629 (1257 → 1009 → 629) (e).

Fig. 3.

The MS^2 spectrum of the $[M + Na]^+$ ion of m/z 1269 (a), its MS^3 spectrum of the ion of m/z 1021 (1269 $\rightarrow$ 1021) (b); the MS^2 spectrum of the $[M + Na]^+$ ions of m/z 1271 (c), its MS^3 spectrum of the ion of m/z 1023 (1271 $\rightarrow$ 1023) (d), MS^4 spectra of the ion of m/z 991 (1271 $\rightarrow$ 1023 $\rightarrow$ 991) (e), and of the ion of m/z 835 (1271 $\rightarrow$ 1023 $\rightarrow$ 835) (f).

Fig. 4. The MS² spectrum of the $[M + Na]^+$ ion of m/z 1369 (a), and its MS³ spectrum of the ion of m/z 1095 ($1369 \rightarrow 1095$) (b).

Fig. 5. The MS² spectrum of the [M + Na]⁺ ions of m/z 1199 (a), its MS³ spectra of m/z 993 (1199 → 993) (b), and of m/z 951 (1199 → 951) (c); and the MS² spectrum of the [M + Na]⁺ ions of m/z 1325 (d), and its MS³ spectrum of m/z 1077 (1325 → 1077) (e).

Fig. 6.
The MS² spectrum of the ion of m/z 1269 (a), its MS³ spectra of the ion of m/z 1021 (1269 $\rightarrow$ 1021) (b), of m/z 1063 (1269 $\rightarrow$ 1063) (c), its MS⁴ spectra of the ion of m/z 847 (1269 $\rightarrow$ 1021 $\rightarrow$ 847) (d), and of m/z 833 (1269 $\rightarrow$ 1021 $\rightarrow$ 833) (e).

Scheme 1.

The fragmentation processes proposed for the $[M + Na]^+$ ion of h28:0/(Ac)₂dTal/(CH₃)₃Rha at m/z 1257 seen in MS² (a), MS³ (1257→1009) (b), MS³ (1257→1069)(c), MS⁴ (1257→1069→821) (d), MS⁴ (1257→1009→629)(e). Fragment ions shown are all sodiated ions; to simplify, “Na⁺” is not drawn.

Scheme 2.

The fragmentation processes proposed for the $[M + Na]^+$ ion of $m28:0/(Ac)_2dTal/(CH_3)_3Rha$ at m/z 1271 seen in MS^2 (a), MS^3 ($1271 \rightarrow 1023$) (b), MS^3 , MS^4 ($1271 \rightarrow 1023 \rightarrow 991$) (c), MS^4 ($1271 \rightarrow 1023 \rightarrow 835$) (d). Fragment ions shown are all sodiated ions and “ Na^+ ” is not drawn.

Scheme 3.

The fragmentation processes proposed for the $[M + Na]^+$ ion of h30:0/(Ac)₂dTal/
HO₂C(CH₂)₂CO (CH₃)₂-Rha at m/z 1369 seen in MS² (a), MS³ (1369 → 1095) (b).
Fragment ions shown are all sodiated ions and “Na⁺” is not drawn.

Table 1

Elemental compositions (deduced from HRMS) and structures of GPLs from *M. smegmatis* MC²155 grown in broth

measured m/z	Calculated m/z	elemental composition		Deviation	fatty acyl chain	dTal	Rhamnose methylation state	succinyl group	Rel. Int.
Da	Da	[M +Na] ⁺ ions	mmu			acetylation state		%	
1103.7087	1103.7077	C57 H100 O15 N4 Na	0.96	hC26:0	0	0	0	0.08	
1117.7242	1117.7234	C58 H102 O15 N4 Na	0.81	hC26:0	0	1	1	0.25	
1131.7401	1131.7390	C59 H104 O15 N4 Na	1.04	hC26:0	0	2	2	0.9	
				hC28:0	0	0	0		
1143.7398	1143.7390	C60 H104 O15 N4 Na	0.8	hC28:1	0	2	2	0.9	
1145.7559	1145.7547	C60 H106 O15 N4 Na	1.23	hC26:0	0	3	3	1.2	
				hC28:0	0	1	1		
1157.7554	1157.7547	C61 H106 O15 N4 Na	0.71	hC28:1	0	2	2	2.4	
1159.7350	1159.7340	C60 H104 O16 N4 Na	1.03	hC26:0	1	1	1	0.4	
1159.7717	1159.7703	C61 H108 O15 N4 Na	1.37	hC28:0	0	2	2	1.8	
1171.7348	1171.7340	C61 H104 O16 N4 Na	0.8	hC26:1	1	2	2	0.25	
1171.7711	1171.7703	C62 H108 O15 N4 Na	0.77	hC28:1	0	3	3	2.1	
				hC30:1	0	1	1		
1185.7506	1185.7496	C62 H106 O16 N4 Na	1.02	hC28:1	1	1	1	1.5	
1185.7867	1185.7860	C63 H110 O15 N4 Na	0.75	hC30:1	0	2	2	2.2	
1199.7293	1199.7289	C62 H104 O17 N4 Na	0.46	hC26:1	2	1	1	0.9	
1199.7661	1199.7653	C63 H108 O16 N4 Na	0.86	hC28:1	1	2	2	2.39	
1201.7457	1201.7445	C62 H106 O17 N4 Na	1.21	hC26:0	2	0	0	1.55	
1201.7821	1201.7809	C63 H110 O16 N4 Na	1.2	hC28:0	1	2	2	4.62	
1215.7614	1215.7602	C63 H108 O17 N4 Na	1.18	hC26:0	2	2	2	12.85	
1215.7948	1215.7966	C64 H112 O16 N4 Na	0.93	hC28:0	1	3	3	6.89	
1227.7617	1227.7602	C64 H108 O17 N4 Na	1.55	hC28:1	2	1	1	4.72	
1227.7979	1227.7966	C65 H112 O16 N4 Na	1.3	hC30:1	1	2	2	4.23	
1229.7771	1229.7758	C64 H110 O17 N4 Na	1.28	hC26:0	2	3	3	25.07	
				hC28:0	2	1	1		
1241.7771	1241.7758	C65 H110 O17 N4 Na	1.28	hC28:1	2	2	2	28.04	

measured m/z	Calculated m/z	elemental composition	Deviation	fatty acyl chain	dTol acetylation state	Rhamnose methylation state	succinyl group	Rel. Int. %
Da	Da	[M+Na] ⁺ ions	mmu					
1241.8127	1241.8122	C66 H114 O16 N4 Na	0.54	hC30:1	1	3		4.58
1243.7911	1243.7915	C65 H112 O17 N4 Na	-0.33	hC28:0	2	2		64.59
1255.7922	1255.7915	C66 H112 O17 N4 Na	0.77	hC30:1	2	1		
				hC28:1	2	3		35
				mC28:1	2	2		
1255.8285	1255.8279	C67 H116 O16 N4 Na	0.64	hC32:1	1	2		1.48
1257.8075	1257.8071	C66 H114 O17 N4 Na	0.41	hC28:0	2	3		100
1269.8074	1269.8071	C67 H114 O17 N4 Na	0.25	hC30:1	2	2		56.19
	1269.8435	C68 H118 O16 N4 Na	-0.49	hC32:1	1	3		1.74
1283.8225	1283.8228	C68 H116 O17 N4 Na	-0.26	hC30:1	2	3		63.31
1297.8383	1297.8384	C69 H118 O17 N4 Na	-0.16	hC32:1	2	2		22.32
1311.8538	1311.8541	C70 H120 O17 N4 Na	-0.31	hC32:1	2	3		28.57
1325.8698	1325.8697	C71 H122 O17 N4 Na	0.03	mC32:1	2	3		3.22
1341.7920	1341.7919	C69 H114 O20 N4 Na	0.13	hC28:1	2	2	1	0.53
1343.8081	1343.8075	C69 H116 O20 N4 Na	0.59	hC28:0	2	2	1	1.48
1355.8080	1355.8075	C70 H116 O20 N4 Na	0.47	hC30:1	2	1	1	0.5
				mC28:1	2	2	1	
1369.8235	1369.8232	C71 H118 O20 N4 Na	0.32	hC30:1	2	2	1	1
1383.8394	1383.8388	C72 H120 O20 N4 Na	0.54	hC32:1	2	1	1	0.37
				mC30:1	2	2	1	
1397.8551	1397.8545	C73 H122 O20 N4 Na	0.64	hC32:1	2	2	1	0.35
1411.8715	1411.8701	C74 H124 O20 N4 Na	1.36	mC32:1	2	2	1	0.2

Elemental compositions (deduced from HRMS) and structural assignments of GPLs isolated from *M. smegmatis* MC²155 grown in agar plate

Table 2

measured m/z	Calculated m/z	elemental composition	Deviation	fatty acyl chain	dTal	Rhamnose		Rel. Intensity
(Da)	(Da)	[M +Na] ⁺ ions	mmu		acetylation state	methylation state	Succinyl group	%
1199.7652	1199.7653	C63 H108 O16 N4 Na	-0.06	hC28:1	1	2		0.49
1199.8020	1199.8016	C64 H112 O15 N4 Na	0.36	hC30:1	0	3		0.31
1201.7822	1201.7809	C63 H110 O16 N4 Na	1.32	hC28:0	1	2		1.98
1215.7615	1215.7602	C63 H108 O17 N4 Na	1.31	hC26:0	2	2		2.43
1215.7979	1215.7966	C64 H112 O16 N4 Na	1.3	hC28:0	1	3		8.13
1227.7618	1227.7602	C64 H108 O17 N4 Na	1.67	hC28:1	2	1		1.19
1227.7981	1227.7966	C65 H112 O16 N4 Na	1.54	hC30:1	1	2		2.51
1229.7776	1229.7758	C64 H110 O17 N4 Na	1.77	hC26:0	2	3		18.83
				hC28:0	2	1		
1241.7772	1241.7758	C65 H110 O17 N4 Na	1.4	hC28:1	2	2		4.95
1241.8137	1241.8122	C66 H114 O16 N4 Na	1.52	hC30:1	1	3		7.45
1243.7931	1243.7915	C65 H112 O17 N4 Na	1.62	hC28:0	2	2		18.49
				mC26:0	2	3		
1255.7929	1255.7915	C66 H112 O17 N4 Na	1.38	hC30:1	2	1		26.11
				hC28:1	2	3		
1255.8285	1255.8279	C67 H116 O16 N4 Na	0.64	mC30:1	1	3		3.55
1257.8073	1257.8071	C66 H114 O17 N4 Na	0.13	hC28:0	2	3		100
1269.8080	1269.8071	C67 H114 O17 N4 Na	0.86	hC30:1	2	2		25.34
				mC28:1	2	3		
1269.8436	1269.8435	C68 H118 O16 N4 Na	0.12	hC30:1	2	2		2.3
				mC28:1	2	3		
				hC32:1	1	3		
1271.8230	1271.8228	C67 H116 O17 N4 Na	0.23	mC28:0	2	3		36.08
1283.8231	1283.8228	C68 H116 O17 N4 Na	0.35	hC30:1	2	3		83.81

measured m/z	Calculated m/z	elemental composition	Deviation	fatty acyl chain	dTal acylation state	Rhamnose		Rel. Intensity
						methylation state	Succinyl group	
(Da)	(Da)	[M +Na] ⁺ ions	mmu					%
1297.8385	1297.8384	C69 H118 O17 N4 Na	0.08	mC30:1	2	3		38.86
1311.8540	1311.8541	C70 H120 O17 N4 Na	-0.07	hC32:1	2	3		28.94
1325.8698	1325.8697	C71 H122 O17 N4 Na	0.03	mC32:1	2	3		9.73
1341.7926	1341.7919	C69 H114 O20 N4 Na	0.74	hC28:1	2	2	1	1.71
1343.8081	1343.8075	C69 H116 O20 N4 Na	0.59	hC28:0	2	2	1	5.46
1355.8079	1355.8075	C70 H116 O20 N4 Na	0.35	mC28:1	2	2	1	1.29
1357.8245	1357.8232	C70 H118 O20 N4 Na	1.3	mC28:0	2	2	1	2.29
1369.8234	1369.8232	C71 H118 O20 N4 Na	0.2	hC30:1	2	2	1	5.79
1383.8390	1383.8388	C72 H120 O20 N4 Na	0.18	mC30:1	2	2	1	3.06
1397.8547	1397.8545	C73 H122 O20 N4 Na	0.27	hC32:1	2	2	1	1.62
1411.8721	1411.8701	C74 H124 O20 N4 Na	1.96	mC32:1	2	2	1	0.62