Characterization of phthiocerol and phthiodiolone dimycocerosate esters of M. tuberculosis by multiple-stage linear ion-trap mass spectrometry.

2Kelly N. Flentie, 2Christina L. Stallings, 1John Turk, 3Adriaan J. Minnaard, *1Fong-Fu Hsu,

1Mass Spectrometry Resource, Division of Endocrinology, Diabetes, Metabolism, and Lipid research, Department of Internal Medicine, 2Department of Molecular Microbiology, Washington University School of Medicine St, Louis, MO USA, 3Stratingh Institute for Chemistry, University of Groningen, Groningen, The Netherlands

Running title: Characterization of PDIM by mass spectrometry

*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, e-mail: fhsu@im.wust.edu.

Abbreviations:

PDIM, phthiocerol/phthiodiolone dimycocerosate; PGL, phenolic glycolipids; DIM, dimycocerosate; LIT, linear ion-trap; MS, mass spectrometry; AMPP, N-(4-aminomethylphenyl) pyridinium; MALDI-TOF, matrix assisted laser desorption ionization time-of-flight; ESI-FTICR, electrospray ionization-fourier transform ion cyclotron resonance; HCD, higher collision energy dissociation; CID, collision induced dissociation; FA, fatty acid;
Abstract

Both phthiocerol/phthiodiolone dimycocerosate (PDIM) and phenolic glycolipids (PGL) are abundant virulent lipids in the cell wall of various pathogenic mycobacteria, which can synthesize a wide range of complex high molecular weight lipids. In this report, we describe LIT MS^n mass spectrometric approach for structural study of PDIMs, which were desorbed as the [M + Li]^+ and [M + NH₄]^+ ions by ESI. We also applied charge-switch strategy to convert the mycocerosic acid substituents to their N-(4-aminomethylphenyl) pyridinium (AMPP) derivatives and analyzed them as M^+ ions, following alkaline hydrolysis of the PDIM to release mycocerosic acids. The structural information from MS^n on the [M + Li]^+ and [M + NH₄]^+ molecular species and on the M^+ ions of mycocerosic acid-AMPP derivative afford realization of the complex structures of PDIMs in Mycobacterium tuberculosis biofilm, differentiation of phthiocerol and phthiodiolone lipid families; and complete structure identification, including the phthiocerol and phthiodiolone backbones, and the mycocerosic acid substituents including the locations of their multiple methyl side chains can be achieved.

Keywords: glycolipid; microbial lipid; lipidomics; PDIM; biofilm; Linear ion-trap mass spectrometry; ESI; HCD; M. Tuberculosis.
Introduction:

Both phthiocerol/phthiodiolone dimycocerosate (PDIM) esters and phenolic glycolipids (PGL) are dimycocerosate esters (DIMs) produced by pathogenic mycobacteria. PDIMs were originally isolated from *Mycobacterium tuberculosis* (*M. tuberculosis*) (1-4) and are specific tuberculosis biomarkers (5, 6). They are among the most abundant lipids in the cell wall of various pathogenic mycobacteria, including *M. tuberculosis, M. bovis, M. leprae, M. kansasii, M. microti* and *M. marinum* (7-10), which are known to synthesize a range of complex high molecular weight lipids (7). PGLs are produced by the same set of pathogenic mycobacteria species except that in *M. tuberculosis* only a subset of clinical isolates belonging to the W-Beijing family (11) produces PGLs.

Phthiocerol and phenolphthiocerol and other lipids such as mycolic acids and methyl-branched fatty acids in cell wall are among those that have been most extensively studied in terms of their biosynthesis and the role in *M. tuberculosis* virulence in vivo (12). In pathogenesis, the role of PDIMs of *M. tuberculosis* was recognized by the studies that identified mutants of *M. tuberculosis* were unable to either produce or properly localize PDIMs to the cell envelope; and demonstrated that PDIM-deficient strains were attenuated in animal models of infection (8, 11, 13-17). Recently, *M. tuberculosis* and its close pathogenic relative *M. marinum* were reported to manipulate macrophage recruitment through coordinated use of membrane PDIM and PGLs to initiate infections (18). However, the precise role of these molecules in the course of infection remains largely unknown; and its role in the multiplication of mycobacteria from the tuberculosis complex in organs other than the lungs is also unclear.

PDIM (see structure below) and PGL respectively consist of a long chain 3-methoxy, 4-methyl, 9,11-dihydroxy glycol (phthiocerol) and a p-glycosylated phenylglycol (glycosyl phenolphthiocerol) backbone diesterified with di-, tri-, and tetra-methyl-branched long-chain mycocerosic (mycoceranic) acids (10, 19, 20). Dependent on the species, the long-chain diol backbone ranges in size from C_{25} to C_{36} and the mycocerosic acid chain ranges from C_{23} or C_{24} to C_{32} (3, 4, 10, 21, 22). Other variants with 3-keto...
or 3-hydroxy diols also exist. The phthiocerol family consists of phthiocerol A, phthiocerol B, phthiodiolone A and phthiotriol A of which phthiocerol A and phthiodiolone A are the most commonly present. In *M. marinum*, a major analogous phenolic glycolipid family consisting of the long-chain β-diol backbone modified with a phenolic group at the terminal, to which a 3-O-methylrhamnose is β-O-linked to the phenol ring is also present (9, 23).

Mass spectrometry has played important roles in the elucidation of the structures of PDIMs and PGLs. For example, Minnikin et al (10, 24-26) and Daffe et al (9, 27-29) employed GC-MS analysis together with NMR spectroscopy for complete characterization of DIMs, following extraction steps, chromatographic separations, and chemical reactions. Recently, MALDI-TOF (16, 29, 30) and ESI-FTICR (31) mass spectrometry have also been applied for profiling PDIM and PGLs. However, ESI tandem mass spectrometric method useful for direct structural identification of these complex lipids has not been established. Here, we describe multiple stage (MS^n) linear ion-trap (LIT) with high resolution mass spectrometry toward characterization of the structures of PDIMs, which were desorbed as the [M + Alk]^+ (Alk = Li, Na, NH₄) ions by ESI. This mass spectrometric approach affords realization of the structures of this lipid family isolated from the biofilm formed by *M. tuberculosis*, including the identities of the mycocerosic acid side chains and the phthiocerol backbone.
Materials and Methods

Chemicals

All chemicals and solvents were purchased from Fisher Scientific (Waltham, MA USA). Standard C$_{30}$/C$_{30}$-mC$_{26:0}$ PDIM (abbreviation please refer to “nomenclature”) was synthesized as previously described (32).

Bacterial strains and growth conditions

*Mycobacterium tuberculosis* Erdman was cultured at 37°C in Middlebrook 7H9 or Middlebrook 7H10 agar plates supplemented with 60 μl/l oleic acid, 5 g/l bovine serum albumin (BSA), 2 g/l dextrose, 0.003 g/l catalase (OADC), 0.5% glycerol, and 0.05% Tween 80 (broth) or in Sauton’s liquid medium unless otherwise indicated. Bacterial biofilms were inoculated with stationary phase planktonic cultures into Sauton’s media at a 1:100 dilution. Culture vessels were closed tightly to restrict oxygen for 3 weeks, and then vented, as described by Ojha (33).

Bacterial lipid extraction

Bacterial biofilms were harvested at the indicated time, pelleted, and boiled for 30 minutes. Samples were then extracted twice by adding chloroform:methanol 2:1, sonicating for 5 minutes, incubating one hour, centrifuging and the organic phase from 2 extractions (twice) were pooled and dried under nitrogen gas. Lipids were dissolved in chloroform:methanol 2:1 before analysis.

LC-MS fractionation of PDIMs

Preparative HPLC experiments were carried out using a Thermo Scientific (San Jose, CA) Vantage TSQ mass spectrometer with Thermo Accela UPLC operated by Xcalibur software. Separation of lipid was achieved by a Supelco 100 x 2.1 mm (2.7 u particle size) Ascentis C-8 column at a flow rate of 260 μL/min. The mobile phase contained 10 mM ammonium formate (pH 5.0) in solvent A: acetonitrile:water (60:40, v:v); solvent B: 2-propanol:acetonitrile (90:10, v:v); and a gradient elution in the following
manner was applied: 68% A, 0-1.5 min; 68-55% A, 1.5-4 min; 55-48% A, 4-5 min; 48-42% A, 5-8 min; 42-34% A, 8-11 min; 34-30% A, 11-14 min; 30-25% A, 14-18 min; 25-3% A, 18-23 min; 3-0% A, 25-30 min; 0% A, 30-35 min; 35-40 min 68% A. During fractionation, around 95% of the lipid was sent to a fraction collector and a small percentage (5%) of the lipid was sent to the mass spectrometer via a tee to identify the structure. The PDIM fraction was eluted at 30.5-34.5 min; and family of phthiocerol dimycocerosate eluted earlier than that of phthiodiolone dimycocerosate (See supplemental data Figure s1, Panel b). PDIM fractions from several injections were collected and pooled, dried under a stream of nitrogen, and subjected to further structural analyses as described below.

*Alkaline hydrolysis and preparation of N-(4-aminomethylphenyl) pyridinium (AMPP) derivative with AMPP reagent.*

Alkaline hydrolysis to yield myroceranoic acids from PDIM was carried out following the protocol as previously described (34) with modification. To the tube containing dried PDIM, 400 μL methanol, 200 μL diethylether and 200 μL tetrabutylammonium hydroxide (40 wt% solution in water) were added. The tube was capped and heated at 100 °C overnight, cooled to room temperature and 50 μL HCl (37%) was added. Followed by addition of 1 mL H2O and 1 mL chloroform, vortexed for 1 min, and centrifuged at 1200xg for 2 min, the organic layer was transferred to another tube and was washed twice with 1 mL H2O. The final organic layer containing free fatty acids was dried under nitrogen, and AMPP derivative was made with the AMP+ Mass Spectrometry Kit, according to the user’s instruction. Briefly, the dried sample was resuspended in 20 μL ice-cold acetonitrile/DMF (4:1, v/v), and 20 ul of ice-cold 1 M EDCI (3-(dimethylamino)propyl)ethyl carbodiimide hydrochloride) in water was added. The vial was briefly mixed on a vortex mixer and placed on ice. To the sample tube, 10 μl of 5 mM N-hydroxybenzotriazole (HOBt) and 30 ul of 15 mM AMPP (in distilled acetonitrile) were added. After the solution was heated at 65°C for 30 min, cooled to room temperature, 1 ml of water and 1 ml of n-butanol were added. The final solution containing FA-AMPP derivative was vortexed for 1 min, centrifuged for 3 min at 1200 x g. The
organic layer was transferred to another vial, dried under a stream of nitrogen, and stored at –20°C until use.

**Mass spectrometry**

Both high-resolution (R=100,000 at m/z 400) higher energy collision activation dissociation (HCD) and low-energy CID tandem mass spectrometry experiments were conducted on a Thermo Scientific (San Jose, CA) LTQ Orbitrap Velos mass spectrometer (MS) with Xcalibur operating system. Lipid extracts in chloroform/methanol (2/1) 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 5x10⁴, with a maximum injection time of 50 ms. Helium was used as the buffer and collision gas at a pressure of 1x10⁻³ mbar (0.75 mTorr). The MSⁿ experiments were carried out with an optimized relative collision energy ranging from 30-45% 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 2-10 min for MSⁿ spectra (n=2,3,4).

**Nomenclature**

For simplicity, the terms phthiocerol dimycocerosate and phthiodiolone dimycocerosate, abbreviated as PDIM, will be used for all the families of the parent waxes without implying one particular stereochemistry for the component of the multimethyl-branched acids. The term phthioglycol, a modification of Stendal’s original nomenclature (35), be used to refer to the family of compounds and the term phthiocerol be reserved for the original 3-methoxy congener. Thus, the abbreviation of the phthiocerol backbone of a PDIM, for example, the 3-methoxy, 4-methyl, 9,11-dihydroxy hexaeicosane is
designated as mC\textsubscript{27:0} to reflect the C\textsubscript{27} chain length with a methoxy group attached at C-3; while glycol backbones with 3-keto, 4-methyl 9,11-dihydroxy hexacosane, and with 3-hydroxy, 4-methyl 9,11-dihydroxy hexacosane are designated as kC\textsubscript{27:0} and as hC\textsubscript{27:0}, respectively. The two mycocerosate residues consisting of two tetramethyl-branched C\textsubscript{30} -acyl chains located at 9 and 11, for example, are abbreviated as C\textsubscript{30}/C\textsubscript{30}. Therefore, a phthioglycol with a respective mC\textsubscript{27:0} and kC\textsubscript{27:0} backbone and two C\textsubscript{30} -acyl chains is designated as C\textsubscript{30}/C\textsubscript{30}-mC\textsubscript{27:0} PDIM and C\textsubscript{30}/C\textsubscript{30}-kC\textsubscript{27:0}-PDIM, respectively.
Results

CID MS\textsuperscript{n} on the 3-methoxy, 4-methyl, 9,11-di-(3,5,7,9-tetramethyloctaeicosanoyl) tetratriacontanediol (C\textsubscript{32:0}/C\textsubscript{32:0}-mC\textsubscript{35:0} PDIM) standard.

PDIMs readily form [M + Alk]\textsuperscript{+} ions (Alk = Li, Na, NH\textsubscript{4}) when subjected to ESI in the presence of alkaline ion in positive-ion mode. Corresponding ions in the fashion of [M + X]\textsuperscript{-} (X = Cl, HCO\textsubscript{2}) were also observed in the negative-ion mode, however, CAD MS\textsuperscript{n} on these latter ions fail to provide sufficient information for structural identification, and will not be discussed further.

To gain insight into the fragmentation processes, we first performed MS\textsuperscript{n} on the [M + Li]\textsuperscript{+} and [M + Na]\textsuperscript{+} adduct ions of C\textsubscript{32:0}/C\textsubscript{32:0}-mC\textsubscript{35:0} PDIM standard to explore their utilities toward structural determination of PDIMs. As shown in Figure 1a, the MS\textsuperscript{2} spectrum of the [M + Li]\textsuperscript{+} ion at m/z 1486.5 contained ions at m/z 1006.0 arising from loss of 3,5,7,9-tetramethyloctaeicosanoic acid residue (32:0), and at m/z 525 (1006.0 - 480.5) arising from further loss of the remaining 32:0-fatty acid substituent (480.5 Da), along with ion at m/z 487, representing a lithiated 32:0-FA cation. The consecutive loss of the fatty acid substituent was further supported by the MS\textsuperscript{3} spectrum of the ions of m/z 1006 (1486 → 1006; Figure 1c), which contained major ions at m/z 525 and 487. Similarly, the MS\textsuperscript{2} spectrum of the corresponding [M+ Na]\textsuperscript{+} ions at m/z 1502.5 (Figure 1b) contained ions at m/z 1022.1, 541, and 503 that are 16 Da heavier, and the MS\textsuperscript{3} spectrum of the ions of m/z 1022 (1502 → 1022; data not shown) also contains the ions of m/z 541 and 503, which are also 16 Da heavier, consistent with the losses of the 32:0-fatty acid moieties.

In contrast, the MS\textsuperscript{2} spectrum of the [M+ NH\textsubscript{4}]\textsuperscript{+} ion at m/z 1497.5 (Figure 1d) is dominated by the ion at m/z 1000 arising from losses of the 3,5,7,9-tetramethyloctaeicosanoic acid and NH\textsubscript{3}, similar to that observed for the [M+ NH\textsubscript{4}]\textsuperscript{+} ion of triacylglycerol (TAG) (36). The spectrum also contained the ion at m/z 968, arising from further loss of CH\textsubscript{3}OH, along with the ions at m/z 519 from additional loss of the
32:0-FA substituent, and at m/z 487 (519 – CH$_3$OH) arising from further loss of the methoxy group. The above losses of the fatty acid substituents and NH$_3$, and methanol were supported by high resolution mass spectrometry (data not shown). The observation of the ions at m/z 968 (1000 – CH$_3$OH) and at 487 (519 – CH$_3$OH) corresponding to loss of CH$_3$OH residue, supports the notion that the phthiocerol possesses a methoxy side chain, and the molecule belongs to the 3-methoxy 4-methyl PDIM family.

It appeared that the 32:0-fatty acid substituent at C11 was first cleaved from the phthiocerol backbone via a charge-remote fragmentation process to yield a lithiated ion of m/z 1006.1 (Figure 1a), which possesses a double bond at C-11 (Scheme 1). This is followed by elimination of the remaining 32:0-FA at C9 to form the lithiated ion of m/z 525 that consists of a conjugated double bond at C9 and C11. This assumption is based on the findings that the MS$^4$ spectrum of the ion of m/z 525 (1486 → 1006→ 525; Figure 1e) contained the major ion at m/z 493, arising from loss of the methoxy group (loss as CH$_3$OH) attached to the phthiocerol chain, along with ions at m/z 231 arising from allylic cleavage of the C13-C14 bond and at m/z 161 arising from cleavages that may eliminate an alkene, acetylene and H$_2$ (Scheme 1).

Similarly, the 32:0-FA residue at C11 was also preferentially cleaved (Figure 1d), resulting in the elimination of NH$_3$ and 32:0-FA residues to form the ion of m/z 1000 in which the proton may reattach to the carboxylate group of the remaining 32:0-FA (Scheme 2). Further dissociation of the ion of m/z 1000 (1497 → 1000; Figure 1f) eliminates the 32:0-FA substituent at C9 to give rise to the major ion of m/z 519, which may represent a protonated phthiocerol ion possessing a conjugated double bond with proton relocated at the methoxy side chain. This is followed by loss of a methanol molecule to yield the ion of m/z 487, in which the charge site may situate at C7 via hydrogen shift. These fragmentation processes leading to the ions of m/z 487 were supported by the MS$^4$ spectrum of the ion of m/z 519 (1497 → 1000 → 519; data not shown). The loss of the 32:0-FA substituent is also consistent with the observation of the ion at m/z 481, representing a protonated 32:0-FA ion, in Figure 1f.
The dissociation of the ion of m/z 487 (1497 → 1000 → 521 → 487; Figure 1g) gave rise to the ion series of m/z 445, 431, 417, 403, 389, 375, etc, along with the ion series of m/z 151, 165, 179, 193, 207, 221, etc (intensity in the descending order), arising from cleavages of the C-C bonds of phthiocerol chain. The observation of these two ion series is consistent with the notion that the charge may primarily reside at C7, and cleavages of these C-C bonds may be similar to the ‘mobile proton’ model (37, 38), in which more than one charge sites are energetically and/or kinetically favored due to proton rearrangement along the phthiocerol backbone. For example, C-C bond cleavages distal to the methyl side chain terminal from precursor ions possessing charge located at C7 gave rise to ions at m/z 193, 207, 221, 235, etc; while ions at m/z 445, 431, 417, 403, 389, etc arose from C-C cleavages from the methyl side chain terminal. Further dissociations of the ions of m/z 431, 417, and 403, also gave rise to ions of m/z 389, 375, and 361 by loss of propylene residue (39); while ions of m/z 151 (loss of C19-alkene), 137 (loss of C20-alkene), and 123 (loss of 21:0-alkene) can also arise from, e.g., m/z 417 by consecutive fragmentation processes that eliminate an alkene residue (Scheme 2). These fragmentation processes are supported by MS5 spectra of the ions of m/z 431 (1497 → 1000 → 521 → 487 → 431; not shown) 417 (1497 → 1000 → 521 → 487 → 417; Figure 1h) and 403 (1497 → 1000 → 521 → 487 → 403; not shown), which contained the whole array of ions arising from cleavages of C-C bonds. Similarly, fragmentations resulting from precursor ion holding charge at C-8 may also lead to the ion series in which the ion of m/z 375 formed by allylic cleavage becomes prominent. The cleavages from the various precursors with various locations of charge site due to proton rearrangement led to the formation of the whole array of ion series with CH2 interval (14 Da).

By contrast, the MS5 spectrum of m/z 485 arising from a keto PDIM specimen contained the ion series that are 2 Da lighter and are characteristic to the family (discuss below).

*CID MS^n on 3-keto, 4-methyl, 9,11-di-(3,5,7-trimethylhexaeicosanoyl) tetratriacontanediol (C29:0/C29:0-kC35:0 PDIM).*
The MS² spectra of the [M + Na⁺, [M + Li]⁺, and of the [M+ NH₄]⁺ ions of C₂₉₀/C₂₉₁₀-kC₃₅₀ PDIM at m/z 1402.5 (not shown), 1386.4 (Figure 2a), 1397.5 (Figure 2b), respectively, are similar to those observed for methoxy-PDIM family as shown in Figure 1. The former spectrum (Figure 2a) is dominated by the ions of m/z 948 arising from loss of the 29:0-FA substituent at C11, and of m/z 509 arising from further loss of the remaining 29:0-FA at C9 (Scheme 3). The speculation of the preferential loss of FA substituent at C11 is consistent with the observation of the ion of m/z 215 in the MS³ spectrum of the ion of m/z 509 (1386 → 509; Figure 2c). This ion is equivalent to the ion of m/z 231 in Figure 1e, arising from cleavage of allylic bond (Scheme 3) similar to that observed for the methoxy-PDIM family (Scheme 1). The spectrum (Figure 2c) also contained the ions of m/z 385 (another allylic cleavage), and of m/z 399, 481 (cleavage of C2-CO bond), 491 (509 – H₂O), and 423 (β-cleavage with γ-H rearrangement (Scheme 3), indicating that the ion of m/z 509 may represent a lithiated Δ⁹,¹¹ k₃₅₀ phthiocerol. Similar fragmentation processes (scheme 3) were also seen in the MS³ spectrum of the analogous ion of m/z 481 (Figure 2e), representing a lithiated Δ⁹,¹¹ k₃₃₀ phthiocerol arising from the [M + Li]⁺ ion of m/z 1400 consisting of the 29:0/32:0-k₃₃₀ major isomer together with 29:0/30:0-k₃₅₀ and 27:0/32:0-k₃₅₀ minor isomers (equivalent to the NH₄⁺ adduct ion of m/z 1411 in Table 1; the structural assignments of the 29:0/32:0-k₃₃₀, 29:0/30:0-k₃₅₀ and 27:0/32:0-k₃₅₀ isomers can be found in the supplemental material Figure s2).

The MS² spectrum of the [M+ NH₄]⁺ ions at m/z 1397 (Figure 2b), again, contained the prominent ion of m/z 941.9, arising from expulsion of 29:0-FA and NH₃; and further loss of the remaining 29:0-FA gave rise to the ion of m/z 503 (Figure 2d), in which the charge site is relocated at the carbonyl group (Scheme 4). The ion at m/z 485 arose from additional loss of a H₂O molecule (Figure 2d), probably involving a prior 1,6-H shift to yield a 3-OH carbonium ion, whose charge may reside at C-7; and elimination of a water molecule to form a stable allylic carbonium ion of m/z 485 with the participation of the adjacent hydrogen at C-4 is followed (Scheme 4). These fragmentation processes are
supported by the MS\(^3\) spectrum of the ion of m/z 941 (1397 → 941; Figure 2d), and the MS\(^4\) spectrum of the ion of m/z 503 (not shown).

Further dissociation of the ion of m/z 485 (1397 → 941 → 503 → 485; Figure 2f) gave rise to the ion series of m/z 457, 443, 429, 415 and 401, .., etc, in which the ion of m/z 457 is likely arising from loss of an ethane residue, in contrast to loss of a C\(_3\)-alkane observed for the ion of m/z 487 originated from a methoxy PDIM compound as shown in Figure 1g. The spectrum also contained the ion series of m/z 403, 389, 375, 361, etc, consistent with the notion that elimination of water molecule involves the participation of the hydrogen at C-4.

**Characterization of M. tuberculosis biofilm PDIMs as [M+ NH\(_4\)]\(^+\) ions.**

Previous report indicates that *M. tuberculosis* produced mainly diesters of phthiocerol A and phthiodiolone A, and phthiocerol B diesters were present in a very small amounts that their detection was unreliable (10). High resolution mass measurements (supplemental material Figure s3) on the NH\(_4\)\(^+\) adduct ion of these mycobacterial waxes clearly resolved phthiodiolone A dimycolates, which gave elemental compositions of C\(_{86}\)H\(_{168}\)(CH\(_2\))\(_n\)O\(_5\)NH\(_4\) (n=1, 2, .., 12), from phthiocerol A dimycolates having elemental compositions of C\(_{86}\)H\(_{170}\)(CH\(_2\))\(_n\)O\(_5\)NH\(_4\) (Table 1). More than one isomers were found in most of the species, and molecules in the phthiocerol B family were not detected. Examples for complete characterization of the entire PDIM family from biofilm of *M. tuberculosis* (Table 1) applying LIT MS\(^n\) mass spectrometry were given below.

Figure 3a showed the MS\(^2\) spectrum of the [M+ NH\(_4\)]\(^+\) ions of m/z 1385.4, which is dominated by m/z 929.9 arising from loss of 29:0-FA acid substituent as NH\(_4\)\(^+\) salt (455.5 Da). Further dissociation of the ion of m/z 929 (1385 → 929; Figure 3b) gave rise to ions at m/z 491 arising from loss of the remaining 29:0-FA substituent (438.5 Da), along with m/z 459 (491 – CH\(_3\)OH) and 897.9 (929.9 - CH\(_3\)OH) arising from further loss of CH\(_3\)OH. The results indicate that the molecule belongs to the PDIM
family consisting of a major 29:0/29:0-m33:0 species. The spectrum also contains the 519/487 ion pair, signifying the presence of 27:0-Fatty acid/m33:0 phthiocerol substituents, and the presence of 29:0/27:0-m35:0-PDIM isomer. The recognition of this latter isomer is further supported by observation of the ion of m/z 958.1 arising from loss of 27:0-FA in Figure 3a. The MS\(^3\) spectrum of m/z 958 (1385 → 958; Figure 3c) contained the m/z 519/487 ion pairs indicating the presence of 29:0-Fatty acid/m35:0 phthiocerol substituents, consistent with the assignment of 29:0/27:0-m35:0-PDIM.

In Figure 3a, ions at m/z 915, arising from loss of 30:0-FA acid substitutent were also observed. MS\(^3\) on the ions of m/z 915 (1385 → 915; Figure 3d) yielded ion pairs of 519/487, 505/473, 491/459, 477/445, reflecting the presence of 26:0/m35:0, 27:0/m34:0, 28:0/m33:0, 29:0/m32:0 substituents, respectively. The results indicate that the ion of m/z 1385 also represent 30:0/26:0-m35:0, 30:0/27:0-m34:0, 30:0/28:0-m33:0, and 30:0/29:0-m32:0 minor isomers. The assignment of 30:0/26:0-m35:0 isomer is consistent with the observation of the ion of m/z 972.1, arising from loss of 26:0-FA substituent in Figure 3a. The MS\(^3\) spectrum of m/z 972 (1385 → 972; Figure 3e) contained the ion pairs of m/z 519/487 and of 491/459, revealing the 30:0/m35:0, and 32:0/m33:0 substituents of the molecules, respectively, and the assignment of 26:0/30:0-m35:0 and 26:0/32:0-m33:0 isomers. The assignment of 26:0/30:0-m35:0 isomer is consistent with the earlier identification of 30:0/26:0-m35:0 isomer (in this assignment, fatty acyl groups at C9 and C11 are not specified); and the identification of the 26:0/32:0-m33:0 isomer is also consistent with the observation of the ion of m/z 897 in Figure 3a, arising from loss of 32:0-FA substituent; and is further supported by the MS\(^3\) spectrum of the ion of m/z 897 (1385 → 897; data not shown), which contains the m/z 491/459 and 477/445 ion pairs.

The profiles of the MS\(^4\) spectra of the ions of m/z 487 (1385 → 958 → 487; not shown), 473 (1385 → 958 → 473; not shown), 459 (1385 → 930 → 459; Figure 3g), and of 445 (1385 → 915 → 445; not shown) are similar to that shown in Figure 1g, supporting the notion that they represent the de-methoxylated C35-, C34-, C33-, and C32-phthiocerol ions with conjugated double bond (Scheme 1).
Similar approaches were also applied to reveal the structural complexity of phthiodiolone dimycerosate family of the extract (Table 1). For example, the MS$^2$ spectrum of the [M+ NH$_4$]$^+$ ion at m/z 1411.5 (Figure 4a) contained major ions at m/z 956 and 914 arising from losses of NH$_3$ and 29:0- and 32:0-mycosanoic acids respectively, indicating the linkage of 29:0- and 32:0-mycosanoic acids to 9-, and 11- of the phthiodiolone backbone. Further dissociation of the ion of m/z 956 (1411 → 956; Figure 4b) gave rise to ions at m/z 475, arising from loss of the remaining 32:0-acid, consistent with the presence of the protonated 32:0-mycosanoic acid ion at m/z 481. The spectrum also consisted of the ion of m/z 457 arising from further loss of H$_2$O from m/z 475. This loss of H$_2$O is supported by the MS$^4$ spectrum of the ion of m/z 475 (1411 → 956 → 475; data not shown) which is dominated by the ion of m/z 457, indicating that the molecule belongs to the phthiodiolone family. The above structural information led to assignment of the major 29:0/32:0-k33:0 structure. Since the spectrum (Figure 4b) also contained the ions of m/z 503 (loss of 30:0-FA) and 485 (503 – H$_2$O), an 29:0/30:0-k35:0 isomeric structure can be defined. MS$^3$ on the ion of m/z 914 (1411 → 914; Figure 4c) gave rise to the major ions of m/z 475 and 457 (475 – H$_2$O), arising from loss of 29:0-FA substituent, consistent with the structural assignment of the 29:0/32:0-k33:0 isomer. The spectrum also contained the minor ions of m/z 503 and 485 (503 – H$_2$O), arising from loss of 27:0-FA substituent, pointing to the presence of 27:0/32:0-k35:0 minor isomer.

In Figure 4a, ions at m/z 984 and 942, arising from loss of 27:0-, and 30:0-fatty acid substituents, respectively, were also observed. Further dissociation of the ion of m/z 984 (1411 → 984; Figure 4d) gave rise to ions of m/z 503 and 485 (503 – H$_2$O), defining the the 32:0-fatty acid substituent. These results further support the earlier assignment of the 27:0/32:0-k35:0 isomer. MS$^3$ on the ion of m/z 942 (1411 → 942; Figure 4e) also yielded ions of m/z 503 and 485 (503 – H$_2$O), indicating the presence of 29:0-fatty acid substituent, and the k35:0-phthiodiolone chain; and confirming the assigned 29:0/30:0-k35:0 structure as described earlier.
The MS^5 spectra of the ions of m/z 485 stemmed from m/z 942 (1411 → 942 → 503 → 485), 984 (1411 → 984 → 503 → 485), and from 956 ((1411 → 956 → 503 → 485) are identical to that shown in Figure 2d, and the profiles of the MS^5 spectra of the ions of m/z 457 originated from m/z 914 (1411 → 914 → 475 → 457; Figure 4f), and from 956 (1411 → 956 → 475 → 457; not shown) are also identical. These spectra comprise the ion series defining the k35-, and k33:0-phthiodiolone chains, respectively. These results led to the notion that the ion of m/z 1411 consists of a major 29:0/32:0-k33:0 isomer together with minor isomers of 27:0/32:0-k35:0 and 29:0/30:0-k35:0 (Table 1).

Characterization of multiple-methyl-branched long-chain mycocerosic (mycoceranic) acid substituents

To assign the structure of mycoceranic acid substituents, released free acids by alkaline hydrolysis were converted to the AMPP derivatives and subjected to ESI MS^a analysis in the positive-ion mode. Three major species were observed at m/z 647, 619 and 605, corresponding to C32-, C30-, and C29-FA AMPP derivatives, respectively. The MS^2 spectra of m/z 649 (Figure 5a), 619 (Figure 5b) and of 605 (Figure 5c) all contained abundant ions at m/z 183 and 169 which are characteristic to FA-AMPP derivatives (40-44).

The MS^2 spectrum of the ion of m/z 649 (Figure 5a) also contained the ion series at m/z 365, 323, 281, 239, along with the ion series of m/z 351, 295 and 253, pointing to the position of the methyl side chains at C-2, 4, 6, and 8 of 32:0-mycocerosic acid (see insets for fragmentation pathways), and gave assignment of 2, 4, 6, 8-tetramethyl-octaeicosanoic acid structure. The MS^2 spectrum of the ion of m/z 619 (Figure 5b) contained the similar ion series, indicating the presence of 2, 4, 6, 8-tetramethyl-hexaeicosanoic acid (30:0-mycocerosic acid). In contrast, the MS^2 spectrum of the ion of m/z 605 (Figure 5c) contained the ion series of m/z 323, 281, and 239 along with the ion pair of m/z 295 and 253, indicating the presence of a 2, 4, 6-trimethyl-hexaeicosanoic acid (29:0-mycocerosic acid) structure.
Discussion

PDIM formed alkaline metal adduct ions ([M+ Alk]+; Alk=Li, Na, K,) and [M + NH₄]+ ions when subjected to ESI in the presence of Li⁺, Na⁺, K⁺ or NH₄⁺. ESI LIT MSⁿ with high resolution mass spectrometry on these adduct ions affords a near complete structural determination of PDIMs, revealing the presence of many homologous and isomeric structures (Table 1). Elemental compositions derived from high resolution mass spectrometry readily distinguish methoxy-PDIM and keto-PDIM molecules, while MSⁿ on the [M+ Alk]+ adduct ions (Alk=Li, Na) and [M + NH₄]+ provide spectroscopic evidence for unambiguous distinction of this two families via the specific neutral losses (i.e., methanol vs. water loss). The more facile cleavages of the mycoceranic acid substituent at C11 than that at C9 upon CID as observed in this study may be applicable for specifically defining the mycoceranic acid substituents on the phthiocerol backbone (at C9 or C11). However, more studies with standard compounds with two different mycoceranic acid chains are required to confirm this finding. In the negative-ion mode, in contrast, the MSⁿ spectra obtained from the [M + X]- adduct ions do not provide sufficient informative ions applicable for structure identification, despite that ions in the fashions of [M + X]- (X=Cl, HCO₂) are readily formed, and the elemental compositions derived from high resolution mass spectrometry are also distinguishable among the methoxy-PDIM and keto-PDIM families (data not shown).

Fragment ions arising from cleavages of C-C bonds along the phthiocerol backbone following loss of the mycoceranic acid substituents may involve the precursors in various resonance forms differed by the charge sites due to proton delocalization (37, 38). Thus, charge-remote fragmentation processes may not be invoked (45, 46) and other fragmentation processes become available (47). The stark differences between the MS³ spectra of m/z 485 (Figure 2d) and of m/z 487 (Figure 1g) lies on the notion that the former spectrum contains abundant ions at m/z 191, 177, 163, 149, 135 etc; while the analogous ions seen at m/z 193, 179, 165, 151, 137 etc in the latter spectrum are less prominent (Figure 2d). The prominence of the ion series of m/z 191, 177, 163, 149, etc, arising from m/z 485 (Figure 2d) may be
attributable to the fact these ions consist of one more double bond than the analogous ions of m/z 193, 179, 165, etc arising from m/z 487 and are more conjugated and more stable. Similar results were also observed for the MS$^5$ spectra of the ions of m/z 459 arising from m34:0 (Figure 3g) and of 457 arising from k33:0 backbones (Figure 4e). These differences in the profiles of the MS$^n$ spectra between those arising from phthiocerol and from phthiodiolone families also provide useful information for their structural differentiation by mass spectrometry.

The advantage of characterization of mycoceranic acid substituents using charge-switch formation of the FA-AMPP derivatives is that feature ions of m/z 183 and 169 are readily recognizable (Figure 5) (40-44), and ions from charge-remote fragmentations can locate the methyl side chains (43) and the functional groups unambiguously (Figure 5, inset). By contrast, the traditional GC/MS method for identification of mycocerosic acids is laborious, requiring acid methanolysis to form the methyl ester, separated by TLC, followed by reductive degradation to mycocerosic alcohol and another two dimensional TLC purification, before formation of the final t-butyldimethylsilyl ether derivatives (23-25). The superb sensitivity gained for the acid detected as the AMPP derivative (as compared to its underivatized form) also facilitate structural identification and quantitation (40-44). This aspect may deserve further investigation.

The insight into the detailed structures of the PDIMs (Table 1) in this study also permits the realization of the structure similarities among the methoxy- and keto-PDIM families. For examples, the major structures of the ions of m/z 1369 (29:0/29:0-k30:0) and 1385 (29:0/29:0-m30:0) all contained 29:0/29:0 substituents; and the fatty acid substituents in the species of m/z 1411 (29:0/32:0-k30:0) of keto family and of m/z 1427 (29:0/32:0-m30:0) in the methoxy family are all 29:0/32:0. The ESI/MS profiles of these two families are also similar (see supplemental material Figure s1 Panels c and d). The results are consistent with the notion that methoxy-PDIM is formed from keto-PDIM (14, 30, 48). It is also notable
that keto-PDIMs are more prominent than the methoxy-PDIM species; while phthiodiolone
dimycocerosates were the minor components previously reported in the *M. tuberculosis* cells (30).
Acknowledgments

This research was supported by US Public Health Service Grants P41-GM103422, P60-DK-20579, P30-DK56341, and 1R21HL120760-01. C.L.S. is supported by a Beckman Young Investigator Award from the Arnold and Mabel Beckman Foundation and an Interdisciplinary Research Initiative grant from the Children's Discovery Institute of Washington University and St. Louis Children's Hospital. K.N.F is supported by a pilot award from the Center for Women’s Infectious Disease Research (cWIDR) at Washington University School of Medicine.
References:


Figure legends

Figure 1. The MS² spectra of the [M+ Li]⁺ ion of C₃₂:₀/C₃₂:₀⁻mC₃₄:₀ PDIM at m/z 1486.5 (a), of the corresponding [M+ Na]⁺ ions at m/z 1502.5 (b), and the consecutive MS³ spectrum of the ion of m/z 1006.1 (1486.5 → 1006.1) (d); the MS⁴ spectrum of the ion of 525 (1486.5 → 1006.1 → 525) (e) from the M+ Li⁺ ion. Panel d shows the MS² spectrum of the corresponding [M+ NH₄]⁺ ion at m/z 1497.5, and the MS³ spectrum of the ion of m/z 1000.1 (1497.5 → 1000.1) (f), MS⁵ spectrum of the ion of m/z 487 (1497.5 → 1000.1 → 519 → 487) (g), MS⁶ spectrum of the ions of m/z 417 (1497.5 → 1000.1 → 521 → 487 → 417) (h) support the fragmentation processes depicted in Scheme 2. Please also note that the mass discrepancy of the labelings in all the figures is due to mass defect.

Figure 2. The MS² spectra of the [M+ Li]⁺ ion of C₂₉:₀/C₂₉:₀⁻kC₃₄:₀ PDIM at m/z 1386.4 (a), MS³ spectra of the ions of m/z 509 (1386 → 509) (b), and of m/z 481 (1400.5 → 481) (from C₂₉:₀/C₃₂:₀⁻kC₃₂:₀) (c). Also shown are the MS² spectra of the corresponding [M+ NH₄]⁺ ion at m/z 1397.5 (d), its MS³ spectrum of the ion of m/z 941 (1397 → 941) (e), and MS⁵ spectrum of the ion of m/z 485 (1397 → 941 → 503 → 485) (f).

Figure 3. The MS² spectrum of the [M+ NH₄]⁺ ion at m/z 1385.4 (a), its MS³ spectra of the ions of m/z 930 (1385 → 930) (b), of m/z 958 (1385 → 958) (c), of m/z 916 (1385 → 916) (d), of m/z 972 (1385 → 972) (e), and its MS⁴ spectrum of the ions of m/z 459 (1385 → 930 → 459) (f).

Figure 4. The MS² spectrum of the [M+ NH₄]⁺ ion at m/z 1411.5 (a), its MS³ spectra of the ion of m/z 956 (1411 → 956) (b), of m/z 914 (1411 → 914) (c), of m/z 984 (1411 → 984) (d), of m/z 942 (1411 → 942) (e), and of the MS⁵ spectrum of the ion of m/z 457 (1411 → 914 → 475 → 457) (f).

Figure 5. The MS² spectra of the PDIM alkaline hydrolysate-AMPP derivative of the ions of m/z 649 (a), of m/z 619 (b), and of m/z 605 (c). The fragmentation pathways leading to locating the position of the methyl side chains of the fatty acid substituents (subset) are also shown.
### Table 1. The molecular species of PDIMs ([M + NH₄]⁺) from M. tuberculosis identified by LIT MSn with high resolution mass spectrometry

<table>
<thead>
<tr>
<th>measured Da</th>
<th>Theo. Mass Da</th>
<th>Deviat. mDa</th>
<th>Rel. Intensity %</th>
<th>Composition</th>
<th><strong>major structures</strong></th>
<th><strong>minor structures</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1245.2760</td>
<td>1245.2759</td>
<td>0.11</td>
<td>0.27</td>
<td>C82 H166 O5 N</td>
<td>27.0/27.0-m27.0</td>
<td>*</td>
</tr>
<tr>
<td>1257.2762</td>
<td>1257.2759</td>
<td>0.32</td>
<td>0.32</td>
<td>C83 H166 O5 N</td>
<td>27.0/27.0-m29.0</td>
<td>*</td>
</tr>
<tr>
<td>1273.3074</td>
<td>1273.3072</td>
<td>0.23</td>
<td>0.41</td>
<td>C84 H170 O5 N</td>
<td>27.0/27.0-m29.0</td>
<td>*</td>
</tr>
<tr>
<td>1285.3074</td>
<td>1285.3074</td>
<td>0</td>
<td>0.67</td>
<td>C85 H170 O5 N</td>
<td>26.0/29.0-m31.0</td>
<td>*</td>
</tr>
<tr>
<td>1299.3235</td>
<td>1299.323</td>
<td>0.52</td>
<td>0.71</td>
<td>C86 H172 O5 N</td>
<td>27.0/27.0-m33.0</td>
<td>*</td>
</tr>
<tr>
<td>1313.3397</td>
<td>1313.3387</td>
<td>0.41</td>
<td>1.52</td>
<td>C87 H174 O5 N</td>
<td>29.0/29.0-m32.0</td>
<td>*</td>
</tr>
<tr>
<td>1327.3543</td>
<td>1327.3543</td>
<td>-0.04</td>
<td>5.66</td>
<td>C88 H176 O5 N</td>
<td>29.0/29.0-m33.0</td>
<td>*</td>
</tr>
<tr>
<td>1341.3700</td>
<td>1341.3700</td>
<td>0.06</td>
<td>8.5</td>
<td>C89 H178 O5 N</td>
<td>29.0/29.0-m33.0</td>
<td>*</td>
</tr>
<tr>
<td>1355.3856</td>
<td>1355.3865</td>
<td>-0.03</td>
<td>11.82</td>
<td>C90 H180 O5 N</td>
<td>29.0/29.0-m33.0</td>
<td>*</td>
</tr>
<tr>
<td>1371.4176</td>
<td>1371.4169</td>
<td>0.69</td>
<td>14.47</td>
<td>C91 H182 O5 N</td>
<td>29.0/29.0-m33.0</td>
<td>*</td>
</tr>
<tr>
<td>1383.4170</td>
<td>1383.4169</td>
<td>0.07</td>
<td>23.97</td>
<td>C92 H184 O5 N</td>
<td>29.0/30.0-m33.0</td>
<td>*</td>
</tr>
<tr>
<td>1385.4327</td>
<td>1385.4326</td>
<td>0.12</td>
<td>61.93</td>
<td>C92 H186 O5 N</td>
<td>29.0/30.0-m33.0</td>
<td>*</td>
</tr>
<tr>
<td>1397.4322</td>
<td>1397.4326</td>
<td>-0.37</td>
<td>55.45</td>
<td>C93 H186 O5 N</td>
<td>29.0/30.0-m33.0</td>
<td>*</td>
</tr>
<tr>
<td>1399.4488</td>
<td>1399.4482</td>
<td>0.63</td>
<td>27.54</td>
<td>C93 H188 O5 N</td>
<td>29.0/30.0-m33.0</td>
<td>*</td>
</tr>
<tr>
<td>1411.4489</td>
<td>1411.4482</td>
<td>0.67</td>
<td>58.04</td>
<td>C94 H188 O5 N</td>
<td>29.0/30.0-m33.0</td>
<td>*</td>
</tr>
<tr>
<td>1413.4639</td>
<td>1413.4639</td>
<td>0.27</td>
<td>96.37</td>
<td>C94 H196 O5 N</td>
<td>29.0/30.0-m33.0</td>
<td>*</td>
</tr>
<tr>
<td>1425.4647</td>
<td>1425.4639</td>
<td>0.84</td>
<td>21.92</td>
<td>C95 H190 O5 N</td>
<td>30.0/32.0-m33.0; 29.0/32.0-m34.0</td>
<td>*</td>
</tr>
<tr>
<td>1427.4806</td>
<td>1427.4795</td>
<td>1.13</td>
<td>52.4</td>
<td>C95 H192 O5 N</td>
<td>29.0/32.0-m33.0</td>
<td>*</td>
</tr>
<tr>
<td>1439.4799</td>
<td>1439.4795</td>
<td>0.4</td>
<td>61.38</td>
<td>C96 H192 O5 N</td>
<td>29.0/32.0-m35.0</td>
<td>*</td>
</tr>
<tr>
<td>1453.4952</td>
<td>1453.4952</td>
<td>0.53</td>
<td>34.67</td>
<td>C97 H194 O5 N</td>
<td>32.0/32.0-m33.0; 30.0/32.0-m35.0</td>
<td>*</td>
</tr>
<tr>
<td>1455.5114</td>
<td>1455.5108</td>
<td>0.6</td>
<td>57.3</td>
<td>C97 H196 O5 N</td>
<td>29.0/32.0-m35.0</td>
<td>*</td>
</tr>
<tr>
<td>1467.5113</td>
<td>1467.5108</td>
<td>0.49</td>
<td>9.6</td>
<td>C98 H196 O5 N</td>
<td>32.0/32.0-m34.0; 31.0/32.0-m35.0</td>
<td>*</td>
</tr>
<tr>
<td>1469.5273</td>
<td>1469.5265</td>
<td>0.86</td>
<td>29.83</td>
<td>C98 H198 O5 N</td>
<td>30.0/32.0-m35.0</td>
<td>*</td>
</tr>
<tr>
<td>1481.5268</td>
<td>1481.5265</td>
<td>0.32</td>
<td>35.06</td>
<td>C99 H198 O5 N</td>
<td>32.0/32.0-m35.0</td>
<td>*</td>
</tr>
<tr>
<td>1495.5421</td>
<td>1495.5421</td>
<td>0.21</td>
<td>3.99</td>
<td>C100 H200 O5 N</td>
<td>32.0/33.0-m35.0</td>
<td>*</td>
</tr>
<tr>
<td>1497.5582</td>
<td>1497.5578</td>
<td>0.4</td>
<td>34.11</td>
<td>C101 H202 O5 N</td>
<td>32.0/32.0-m35.0</td>
<td>*</td>
</tr>
<tr>
<td>1509.5583</td>
<td>1509.5578</td>
<td>0.58</td>
<td>1.41</td>
<td>C101 H202 O5 N</td>
<td>32.0/34.0-m35.0</td>
<td>*</td>
</tr>
<tr>
<td>1511.5755</td>
<td>1511.5734</td>
<td>2.07</td>
<td>2.74</td>
<td>C101 H204 O5 N</td>
<td>32.0/33.0-m35.0</td>
<td>*</td>
</tr>
<tr>
<td>1523.5742</td>
<td>1523.5734</td>
<td>0.78</td>
<td>1.84</td>
<td>C102 H204 O5 N</td>
<td>33.0/34.0-m35.0</td>
<td>*</td>
</tr>
<tr>
<td>1539.6051</td>
<td>1539.6047</td>
<td>0.44</td>
<td>1.62</td>
<td>C103 H208 O5 N</td>
<td>32.0/35.0-m35.0</td>
<td>*</td>
</tr>
</tbody>
</table>

*minor isomers not defined; **abundances in the descending order.
Scheme 1. The fragmentation processes proposed for the [M + Li]$^+$ ion of 32:0/32:0-m35:0-PDIM at m/z 1486.
Scheme 2. The fragmentation processes proposed for the [M + NH₄]⁺ ion of 32:0/32:0-m35:0-PDIM at m/z 1497.
Scheme 3. The fragmentation processes proposed for the [M + Li]+ ion of 29:0/29:0-k35:0-PDIM at m/z 1386.
Scheme 4. The fragmentation processes proposed for the [M + NH₄]⁺ ion of 29:0/29:0-k34:0-PDIM at m/z 1397.