

CHAPTER 3

A fatty acid metabolomics approach to differentiate and characterise various *Mycobacterium* species

A part of this chapter has been published in the *Journal of microbiological methods*

OLIVIER, I., LOOTS, D.T. 2012. A metabolomics approach to characterise and identify various *Mycobacterium* species. *J. Microbiol. Meth.* 88:419-426.

1. INTRODUCTION

In 2008, there were an estimated 9.4 million new TB cases and 1.8 million TB related deaths reported globally, with the highest incidence rate recorded in sub-Saharan Africa (WHO, 2010). Lately, the new research field of metabolomics has been used successfully as a tool to characterise a variety of disease states in order to better understand the underlying mechanisms of these diseases (Schoeman & Loots, 2011).

In this chapter, we developed and investigated the capacity of a metabolomics research approach to identify unique metabolite markers which would differentiate and characterise TB and / or other related infectious diseases. Although *M. tuberculosis* is the primary cause of TB, other mycobacterial species, still encountered frequently in clinical samples, may also result in opportunistic infections in humans, resulting in similar symptoms to that caused by *M. tuberculosis* (Miguez-Burbano, 2006; Olivier, 1998). These species include: *M. avium*, which has the potential to cause pulmonary infection in patients with chronic lung disease; *M. kansasii*, which may cause skin and soft tissue infections, skeletal infections, lung infections, surgical site infections, and disseminated disease (Davis, 2007), and; *M. bovis*, the organism responsible for causing TB in cattle (Todar, 2005). It is important to note that cultured sample repeats of the four above mentioned *Mycobacterium* species were used in all the method development chapters (Chapters 3 and 4), for a minimum inter-sample variation, in order to determine the variation which potentially exists between the metabolomics methods investigated. In the later application of these developed methods for the identification of metabolite markers pertaining to various research questions, however, inter-sample variation was considered and included in the research design (Chapters 4 and 7). Furthermore, since these bacterial samples are all related to the same genus, a developed method capable of identifying unique metabolite makers, which are able to differentiate and characterise these on the species level, would be considered optimal. Therefore, this method would have an increased likelihood of success when applying the approach to individually cultured *M. tuberculosis* or patient collected sputum samples, where inter-sample variation is a given.

The fact that the lipid composition of these mycobacterial species are well known (Lambert *et al.*, 1986; Mosca *et al.*, 2006) and that their fatty acid composition sets them apart from other infectious organisms, three fatty acid extraction methods were compared in this chapter for their possible application to metabolomics TB research. The success of these methods however, resulted in their further refinement and the development of other research approaches (Chapters 5 and 6), for more specific research applications (Chapters 4 and 7). The three different fatty acid extraction methods investigated in this chapter were compared by considering their respective repeatability and extraction capacities. Furthermore, using

multivariate statistical data analyses, we identified those metabolites contributing most to the variation between the tested bacterial species, as extracted by the best performing extraction procedure. These metabolites were then evaluated by comparing them to previous literature, in order to validate the capacity of this extraction method, and the metabolomics research approach used, to identify markers of biological relevance, considering the later applications of this method for better disease characterisation. Finally, the potential detection limit (minimum sample requirements) for this approach was also determined in order to establish its analytical capacity and practicality for research applications.

Aim: To develop a fatty acid metabolomics approach for differentiating and characterising various *Mycobacterium* species isolated from culture.

2. MATERIALS AND METHODS

2.1. Reagents and chemicals

Sodium hydroxide (NaOH), potassium hydroxide (KOH), *N*-methyl-*N*-(*tert*-butyldimethylsilyl)trifluoroacetamide (MTBSTFA), and nonadecanoic acid (C19:0) were obtained from Merck (Darmstadt, Germany). All organic solvents used were ultra pure Burdick & Jackson brands (Honeywell International Inc., Muskegon, MI, USA) and were used without further purification.

2.2. Sample collection and preparation

All organisms used in this study were prepared by the Royal Tropical Institute, Amsterdam, The Netherlands. All bacteria (*M. tuberculosis*, *M. avium*, *M. bovis* and *M. kansasii*) were cultured (originally from patient sputum samples) in Middlebrook 7H9 medium with oleic acid-albumin-dextrose-catalase enrichment. The bacteria were incubated at 37°C while shaking at 120 – 150 rpm until an optical density (420 nm) of 0.30 ($\approx 2 \times 10^8$ bacteria mL⁻¹) was reached. The cells were washed once with PBS, collected via centrifugation, heat killed in a water bath at 80°C and stored at -80°C until extraction and GC-MS metabolomics analysis.

In order to compare the repeatability and extraction capacity of the 3 fatty acid extraction procedures investigated for later metabolomics applications, each extraction procedure was repeated four times using 4 x 1 mL of the isolated *M. tuberculosis* sample repeats, re-suspended in ddH₂O at a concentration of 1×10^8 bacteria mL⁻¹, followed by GC-MS, and statistical data analyses, as described below.

For the purpose of investigating the best performing extraction method's capacity for differentiating the various bacterial groups, 10 x 1 mL samples of each of the 4 isolated *Mycobacterium* species (*M. tuberculosis*, *M. avium*, *M. bovis* and *M. kansasii*), were prepared at a concentration of 1×10^8 bacteria mL^{-1} in ddH₂O, prior to fatty acid extraction, GC-MS analyses and multivariate statistical analysis of the generated data. In order to test the validity of this approach and the integrity of the potential metabolite markers used for differentiating the groups, two separately grown cultures of each of the four *Mycobacterium* species were cultured and analysed in the same manner, six months later.

To determine the potential detection limit of this metabolomics method, 6 x 1 mL repeats of the isolated *M. tuberculosis* re-suspended in ddH₂O at a concentration gradient ranging from 1×10^2 - 1×10^8 bacteria mL^{-1} , including a blank (ddH₂O only), were extracted using the selected method, analysed using the GC-MS, and the collected data processed using the multivariate statistics described below. In order to verify the differentiation capacity of this extraction method at the determined detection limit, 10 x 1mL samples of each of the 4 re-suspended isolated *Mycobacterium* species (*M. tuberculosis*, *M. avium*, *M. bovis* and *M. kansasii*) were prepared at the concentration determined by the detection limit experiment, in ddH₂O, prior to fatty acid extraction, GC-MS analyses and multivariate statistical analysis of the generated data.

2.3. Extraction procedures

Three extraction procedures, traditionally used for extracting fatty acids, were compared, prior to GC-MS analyses. Only one of these three methods was selected based on its comparatively superior repeatability and extraction capacity, for the further metabolomics method development investigations in this chapter.

The first extraction procedure tested was described by Vreken *et al.* (1998) for the extraction of very-long-chain fatty acids, and modified for our purposes. Briefly, 100 μL of nondecanoic acid (C19:0) in hexane ($14 \mu\text{g mL}^{-1}$), was added to 1 mL *M. tuberculosis* suspension (1×10^8 bacteria mL^{-1} in ddH₂O), as an internal standard. Acid hydrolysis was performed by adding 1.156 mL of 10 M HCl (37%) and 22.844 mL acetonitrile to the sample, followed by incubation at 110°C for 45 min. The sample was then cooled to room temperature and alkaline hydrolysis was performed by adding 2 mL of 1 M NaOH in methanol, followed by a second 45 min incubation period at 110°C. The sample was allowed to cool, followed by the addition of 462.5 μL of 10 M HCl. Fatty acids were subsequently extracted in 4 mL of hexane and the organic phase washed with 3.5 mL of 1 M KOH. The aqueous layer was acidified again by adding 578 μL of 10 M HCl (37%) and re-extracted with 4 mL of hexane. The combined collected solvent

was evaporated to dryness under nitrogen at 45°C. The sample was then derivatised with 50 µL of MTBSTFA and 50 µL of pyridine at 80°C for 30 min.

The second extraction method tested involved hexane extraction and HCl methylation as described by Van der Walt *et al.* (2008). Once again, 100 µL of the C19:0 internal standard (14 µg mL⁻¹) in hexane was added to 1 mL *M. tuberculosis* suspension (1 X 10⁸ bacteria mL⁻¹ in ddH₂O), followed by the addition of 2 mL methanolic HCl and a 4 hour incubation period at 90°C. Hexane (2 mL) was added, followed by a brief vortex step and then centrifuged for 5 min at 550 x g. The organic phase was collected, transferred to a new tube and the water phase re-extracted with hexane. The pooled organic phases were dried under nitrogen and the dried extract was re-suspended in 100 µL of hexane prior to injection.

Thirdly, a modified Bligh-Dyer extraction procedure (Rezwan *et al.*, 2007) was investigated. As internal standard, 100 µL of C19:0 (14 µg mL⁻¹) in hexane, was added to 1 mL of the *Mycobacterium* suspensions, followed by the addition of 2 mL chloroform, 4 mL methanol and 1.6 mL ddH₂O. The sample mixture was then sonicated for 2 min, vortexed for 30 sec, and allowed to stand for 2 hours at room temperature, in order to facilitate phase separation. Chloroform and ddH₂O (2 mL of each) were again added and the sample was centrifuged for 15 min at 550 x g. The organic phase was collected and the water phase was re-extracted with 2 mL of chloroform. The combined organic phase was washed with 5 mL of ddH₂O and once again left to stand for 2 hours at room temperature, in order to facilitate phase separation, followed by centrifugation for 20 min at 550 x g. The newly formed organic phase was collected and dried under nitrogen. Subsequently, 0.5 mL chloroform, 0.5 mL methanol, and 1 mL methanolic KOH (0.2 M) was added. The mixture was incubated for 30 min at 60°C followed by the addition of 2 mL hexane, 200 µL glacial acetic acid (1 M), and 2 mL ddH₂O. The organic phase was again collected and transferred to a new tube. The water phase was re-extracted 3 times with hexane. The combined organic phases from each extraction were dried under nitrogen and re-suspended in 100 µL hexane prior to injection.

2.4. GC-MS parameters

GC-MS analyses were done in the splitless mode by randomly injecting 1 µL of the extracts on an Agilent 7890A gas chromatograph coupled to an Agilent 5975 mass selective detector equipped with a 7683B injector, 7683 auto sampler, and VF1-MS capillary column (30 m x 250 µm i.d., 0.25 µm film thickness). The injector temperature was held constant at 250°C for the entire run time. The initial GC oven temperature was held at 50°C for 1 min after injection, followed by an increase of 10°C min⁻¹ to 240°C. The oven temperature was then increased at a rate of 20°C min⁻¹ to a final temperature of 300°C and maintained for 7 min. Helium was used

as the carrier gas and pressure programmed as such that the helium flow was kept constant at 1.2 mL min⁻¹. Detection was achieved by using MS detection in full scan mode (m/z 50-550).

2.5. Data-acquisition

Alignment of the detected compounds across the samples analysed was achieved by creating a new reference library in AMDIS (Automated Mass Spectral Deconvolution and Identification System, version 2.65), which contained the mass spectra of all the compounds detected above a threshold of 0.01% of the total signal, for all the samples analysed in each of the four *Mycobacterium* sample repeats. Each compound was given a code corresponding to the retention time at which it was first detected. This library was manually inspected and duplicate compound mass spectra identified by AMDIS due to GC-MS spectral shifting, were deleted. Each analysed sample was subsequently processed using the aforementioned reference library, and the resulting output of each sample was combined into a data matrix containing the areas of all compounds present or absent in each sample. Peak areas were normalised relative to the area of the internal standard, of which the concentration was known. This normalisation was done in order to eliminate any variation which may occur due to irregularities during the extraction process or injection of the sample onto the GC-MS. This data matrix was then subjected to multivariate statistical data analysis in order to determine if a separation of the groups can be achieved based on the variation in the detected metabolite profiles using this extraction method, as well as to identify those metabolite markers characterising these species.

2.6. Statistical data analysis

The distribution of the coefficient of variation (CV) values for all the detected compounds were used as a measure to compare the repeatability of the two extraction methods.

The statistical packages, “R” (version 2.13.0) and Statistica (version 10) were used for all the remaining statistical data analyses. To lessen the weight of the compounds present in drastically elevated concentrations and in order to bring those of lower concentrations into perspective, data were scaled using a non-parametric transformation function (Koekemoer & Swanepoel, 2008), after which mean centering was applied.

Principle component analysis (PCA) was performed in order to ascertain whether or not a natural grouping or differentiation exists between the various sample groups based on their detected metabolite profiles. PCA involves a mathematical procedure that transforms a number of possibly related variables (in this case metabolites) into a smaller number of unrelated variables known as principal components (PC's). PC 1 accounts for the most variance in the

data and each subsequent PC (PC 2, PC 3 etc.) accounts for the next highest variance of the remaining data. Using one PC per axis, the PCA can then be visualized as a scores plot. Using the data generated from the species differentiation analyses at a concentration of 1×10^8 bacteria mL^{-1} , the metabolites were ranked according to their respective PCA modelling powers.

Additionally, a partial least squares discriminant analysis (PLS-DA) model was built in order to identify those compounds that contribute most to the separation between the sample groups, by ranking the compounds according to the variable influence on the projection (VIP) parameter. VIP is a weighted sum of squares of the PLS-DA weights, indicating the importance of the metabolite to the total model.

To reduce the dataset into a set of potentially relevant metabolite markers, we created a combined list of the compounds with the highest modelling powers (PCA) and VIP values (PLS-DA). Using the mass fragmentation patterns generated by the MS, together with their respective GC retention times, the identities of these metabolite markers were determined using libraries generated from previously injected standards.

In order to examine the effect of inter-sample variation on this approach, the identified metabolite markers were then used to build a discriminant model based on Bayes' theorem, in conjunction with multivariate kernel density estimation, using the first 3 PCs as input. This model was developed for the purpose of estimating the class membership probabilities of an "unknown" bacterial sample based on the presence of the aforementioned metabolite markers. As a validation of this metabolomics approach and the metabolite markers selected, two new, "unknown" bacterial cultures, representing each of the above mentioned *Mycobacterium* species groups, were separately cultured from patient samples, extracted and analysed 6 months later, and the collected data were used to determine if the model would correctly identify or allocate these to their respective groups.

In order to determine the potential detection limit (minimum sample requirements) of this method, a PCA was performed on the data obtained after GC-MS analyses of the *M. tuberculosis* concentration gradient samples, as described in section 2.2. The sample group with the lowest bacterial concentration not overlapping with the blank sample was regarded as the potential detection limit, representing the minimum amount of sample material required for the detection of metabolite markers characterising the groups. To verify this detection limit, a second PCA was performed using data collected from the analyses of the four *Mycobacterium* species groups, prepared at the concentration determined by the previously described detection limit analyses.

3. RESULTS AND DISCUSSION

3.1 Comparative repeatability and extraction capacity of the three fatty acid extraction procedures

A comparison of the calculated CV values for the GC-MS determined internal standard areas for the three methods evaluated, indicates the modified Bligh-Dyer method (CV = 4.9%) to have a far better repeatability than the hexane extraction and HCl methylation method (CV = 31.58%) and the very-long-chain fatty acid extraction procedure (CV = 102.39%). Reasons for this hierarchy may be the fact that the latter method requires two incubation steps, at very high temperatures (110°C), followed by a third heating step during derivitisation of the sample at 80°C, potentially leading to a loss of temperature sensitive compounds. Additionally, Vreken *et al.* (1998) standardised this very-long-chain fatty acid extraction method for plasma and serum samples and not for bacterial samples, for which the Bligh-Dyer method is widely used. Similarly, the CV values calculated using the relative concentration of a selected fatty acid marker, hexadecanoic acid (C16:0), further confirms this result (1.83% vs. 7.21% vs. 117.21% for the modified Bligh-Dyer vs. the hexane extraction and HCl methylation vs. the very-long-chain fatty acid extraction method, respectively). However, as metabolomics is involved in determining “all” the compounds in a mixture in an unbiased manner, the repeatability of each of the three extraction methods was evaluated by calculating the CV values (using the relative concentrations) of all the compounds detected after GC-MS analysis of the extracted sample repeats (Figure 3.1). The distribution of these CV values indicates that, for example, only 34.4% of the compounds detected via the very-long-chain fatty acid extraction procedure had a CV value below 50%, as opposed to 66.7% for that of the hexane extraction and HCl methylation and 76.8% for the Bligh-Dyer method, respectively, once again indicating an overall better repeatability of the latter method, confirming the previous results.

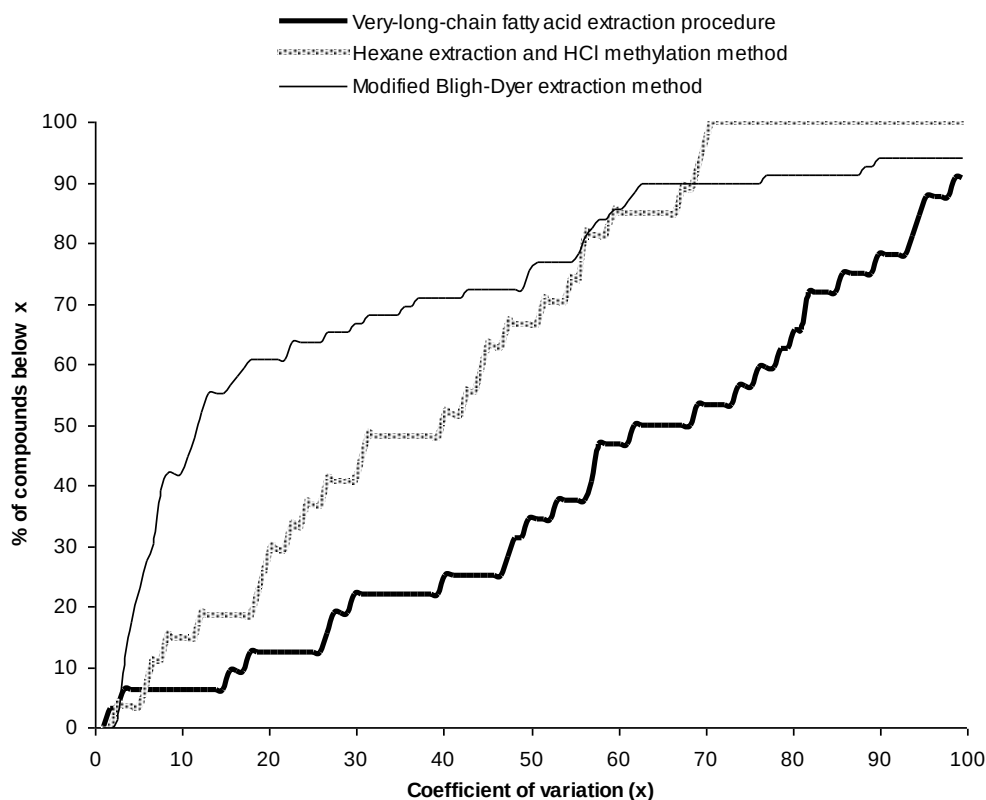


Figure 3.1: Distribution of the coefficient of variation (CV) values of all the compounds detected through GC-MS analyses of the *M. tuberculosis* sample repeats after extraction via each of the three compared extraction procedures. The best repeatability was obtained using the modified Bligh-Dyer extraction method.

A further comparison of the GC-MS results obtained from the latter two methods, after AMDIS deconvolution at a threshold setting of 0.01%, showed that considerably more peaks were identified when the samples were extracted using the modified Bligh-Dyer extraction method (141 ± 12) vs. the hexane extraction and HCl methylation extraction method (95 ± 2 compounds). Additionally, a comparatively higher compound abundance, when comparing the peak intensities of the internal standard, was obtained using the modified Bligh-Dyer extraction method vs. the hexane extraction and HCl methylation method (2116667 ± 104083 vs. 515000 ± 162634), respectively. Considering that metabolomics, per definition, requires the detection of the maximum number of metabolites when analysing the data in a repeatable manner, the modified Bligh-Dyer extraction method was chosen for the subsequent metabolomics investigations.

3.2 Bacterial fatty acid content

Considering the compounds detected via GC-MS analysis of the modified Bligh-Dyer method extracted sample repeats of each of the four *Mycobacterium* species, 347 mass spectra were

added to the library. The most abundant compounds detected in each of the species groups are given in Table 3.1.

Table 3.1: Mean relative concentrations ($\mu\text{g mg}^{-1}$ sample) of the most abundant fatty acids detected in the four *Mycobacterium* species, extracted using the modified Bligh-Dyer approach. Standard deviations are given in parenthesis.

Species	<i>M. tuberculosis</i>	<i>M. avium</i>	<i>M. bovis</i>	<i>M. kansasii</i>
Compound ^a	Concentration (SD)			
C16:0	234.8 (6.3)	457.0 (13.9)	178.7 (57.0)	262.5 (9.3)
C18:1 ω 9c	121.9 (3.7)	275.3 (91.6)	49.7 (9.2)	152.1 (48.3)
TBSA	134.6 (0.7)	203.5 (9.4)	69.6 (14.5)	52.0 (2.1)
C16:1 ω 7c	14.9 (0.6)	60.4 (2.4)	2.3 (0.3)	12.1 (0.7)
BPA	7.4 (0.1)	8.1 (3.5)	6.2 (4.9)	13.5 (3.3)
C17:0	23.0 (0.8)	8.6 (0.8)	0.2 (0.6)	T
Hexadecane	2.0 (0.6)	5.6 (0.6)	21.2 (7.4)	3.4 (1.0)
C26:0	85.6 (2.5)	12.3 (4.1)	103.2 (20.4)	3.9 (0.2)
C24:0	23.3 (0.8)	116.9 (4.1)	23.8 (3.5)	102.8 (6.0)
C20:0	23.0 (0.6)	7.3 (6.2)	20.8 (3.1)	17.7 (0.7)
C14:0	6.4 (2.2)	37.5 (2.4)	28.7 (5.8)	14.2 (2.0)
C22:0	9.3 (0.2)	34.9 (0.9)	9.9 (1.1)	36.1 (1.9)
1, ME-C15:0	0	23.0 (6.2)	85.1 (10.4)	0
C24:1 ω 9c	0	13.3 (1.6)	0	21.5 (1.1)
Squalene	0.3 (0.8)	1.7 (0.8)	32.2 (25.1)	1.4 (0.6)
C22:1 ω 9c	0	8.1 (0.3)	T	14.7 (0.7)

^aTBSA, tuberculostearic acid. BPA, benzenepropanoic acid. ME, methylethyl. T, trace amount.

Corresponding to that previously reported (Lambert *et al.*, 1986; Mosca *et al.*, 2006), the most abundant fatty acids detected in the *Mycobacterium* samples were hexadecanoic (C16:0) and oleic acid (C18:1 ω 9c). Additionally, as was expected, we detected the well-known *Mycobacterium* biomarker, tuberculostearic acid (TBSA) (Larsson *et al.*, 1987; Stopforth *et al.*, 2004) as one of the major compounds in all *Mycobacterium* species groups. Other shorter-chain fatty acids detected in relatively high concentrations included: tetradecanoic acid (C14:0), palmitoleic (C16:1 ω 7c) and heptadecanoic acid (C17:0), which is also consistent with previous findings (Lambert *et al.*, 1986; Luquin *et al.*, 1991). Long-chain fatty acids including: eicosanoic acid (C20:0), docosanoic acid (C22:0), tetracosanoic acid (C24:0) and hexacosanoic acid (C26:0) were also identified as part of the most abundant fatty acids using this approach. According to Guerrant *et al.* (1981), these fatty acids are formed as a result of heat cleavage of the mycolic acids in the injection port of the gas chromatograph at injector temperatures exceeding that of 235°C, as was the case in our study. Mycolic acids are unique, high-molecular-weight, α -alkyl, β -hydroxy, very-long fatty acids (C70-C90), representing 40 – 60% of the dry cell weight of *Mycobacterium* and *Corynebacterium* species (Salman *et al.*, 1999). The concentrations of these mycolic acid derived products are known to vary consistently among different *Mycobacterium* species (Knisley *et al.*, 1985; Lambert *et al.*, 1986; Luquin *et al.*, 1991).

Tetracosanoic acid (C24:0), for instance, was detected in higher amounts in *M. avium* and *M. kansasii*, and hexacosanoic acid (C26:0) was detected in comparatively higher amounts in *M. tuberculosis* and *M. bovis*. Using this approach, we additionally detected minor concentrations of C26:0 in *M. kansasii*, and isopropyl tetradecanoic acid (1, ME-C14:0) in elevated amounts in *M. bovis* and in minor concentrations in *M. avium*, which to date have not yet been reported. Lastly, alkanes (hexadecane and squalene) were also detected in major concentrations, reflecting the ability of this method to additionally extract this compound class.

3.3 Differentiation capacity

Prior to multivariate statistical data analyses, a 50% filter was applied to the GC-MS obtained data, excluding those compounds which do not appear in at least 50% of the samples, in one or more of the sample groups. This filter, together with the removal of compounds with no detected variation between the groups, led to a reduction of the total number of compounds from 347 to 151. Using these 151 variables, a clear differentiation of all of the sample groups, at a concentration of 1×10^8 bacteria mL^{-1} , was achieved when using the first three PCs of the PCA (Figure 3.2). The total amount of variance explained by the first three PCs (R^2X cum) was 71%, of which PC 1 contributed 35%, PC 2 contributed 21.3%, and PC 3 contributed 14.7%.

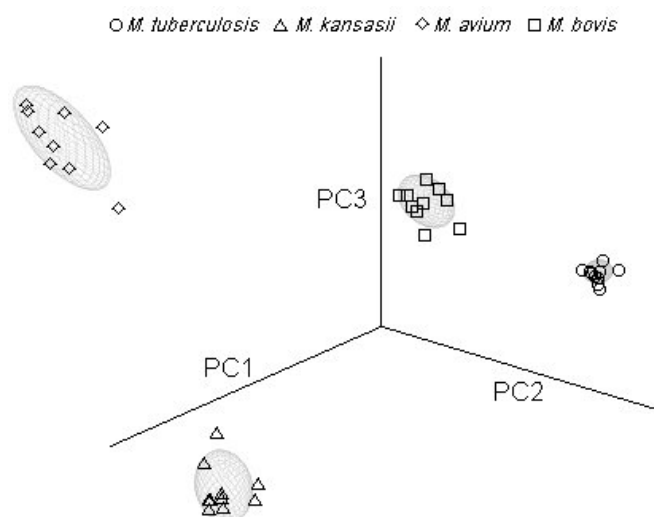


Figure 3.2: Three-dimensional PCA scores plot of the GC-MS generated metabolite data after extraction of the bacterial samples using the modified Bligh-Dyer method, showing a clear differentiation between the four *Mycobacterium* sample groups at a concentration of 1×10^8 bacteria mL^{-1} .

The PLS-DA model also used 3 components, with the modelling parameter R^2Y (cum) being 98.4%, indicative of the total explained variation of the response Y. Q^2 (cum), the cross-validated variation explained by the response Y, was 97.8%. This high degree of validation may be attributed to the good analytical repeatability of the extraction method, previously

determined, in addition to the capacity of the method to extract and detect the unique characteristic metabolite profiles from the four *Mycobacterium* species used for this investigation

By selecting those metabolites with the highest modelling powers for the PCA and the highest VIP values for the PLS-DA, the 12 compounds contributing most to the differentiation of these bacterial species, were identified as potential metabolite markers. These metabolite markers, together with their relative concentrations (normalised using the internal standard) are listed in Table 3.2.

Table 3.2: Mean relative concentrations ($\mu\text{g mg}^{-1}$ sample) of the metabolite markers identified for the various *Mycobacterium* species. Standard deviations are given in parenthesis.

Species	<i>M. tuberculosis</i>	<i>M. avium</i>	<i>M. bovis</i>	<i>M. kansasii</i>
Compound ^a	Concentration (SD)			
1, ME-C15	0	23.0 (6.2)	85.1 (10.4)	0
C16:1 ω 7c	14.9 (0.6)	60.4 (2.4)	2.3 (0.3)	12.1 (0.7)
C17:1 ω 7c	3.5 (0.2)	4.5 (0.4)	0	1.7 (0.5)
C17:0	23.0 (0.8)	8.6 (0.8)	0.2 (0.6)	T
TBSA	134.6 (0.7)	203.5 (9.4)	69.6 (14.5)	52.0 (2.1)
C20:1 ω 9c	0	4.7 (0.7)	0	3.5 (0.3)
C22:1 ω 9c	0	8.1 (0.3)	T	14.7 (0.7)
C24:1 ω 9c	0	13.3 (1.6)	0	21.5 (1.1)
Unknown 422	4.2 (0.2)	0	5.6 (1.1)	0
Unknown 408	0	11.6 (1.5)	0	2.0 (0.5)
Tricosane	0	3.3 (0.8)	0	T
Unknown 494	2.7 (0.5)	0	0	1.8 (0.4)

^aTBSA, tuberculostearic acid. ME, methylethyl, T, trace amounts.

As seen in Table 3.2, the majority of these metabolite markers are not necessarily novel to a particular bacterial group, but were identified due to the constant differences in their concentrations between the groups. Of the 12 identified metabolite markers described here, 6 were also detected as part of the most abundant compounds in the individual species groups, and their occurrence have therefore already been discussed in section 3.2. Using this approach, the unsaturated fatty acids, 10-heptadecenoic acid (C17:1 ω 7c) and gondoic acid (C20:1 ω 9c), were determined to be potential metabolite markers despite the fact that they were not detected in the most abundant concentrations. Consistent with our results, Mosca *et al.*, (2007) reported 10-heptadecenoic acid (C17:1 ω 7c) to be present in comparatively higher concentrations in *M. avium* than in the other *Mycobacterium* species tested, however, gondoic acid has not yet been described as a characteristic marker for *M. tuberculosis* or *M. kansasii*. Other potential metabolite markers detected included 3 as yet unknown compounds, of molecular masses 408, 422, and 494 respectively, which were classified as unknown when using the in house and commercial NIST mass spectral reference libraries.

Using the 12 identified metabolite makers, a discriminant model based on Bayes' theorem, in conjunction with multivariate kernel density estimation, was built in order to predict whether this approach and the markers identified using this extraction method, could be used to predict the class membership of the newly extracted “unknown” test samples (2 cultures of each species extracted and analysed 6 months later). This prediction would validate these markers and subsequently the developed metabolomics approach. This model can be defined as:

$$P_i^{(x,y,z)} = \hat{f}_{XYZ}^{(i)}(x,y,z) / \sum_{j=1}^G \hat{f}_{XYZ}^{(j)}, \text{ where } G \text{ is the number of groups (in this case 4) and } \hat{f}_{XYZ}^{(i)} \text{ is a}$$

multivariate kernel density estimate of the first 3 PCs (labelled X, Y, and Z) for group i . This model assigns a group membership to the “unknown” sample using probability estimates ($P_i^{(x,y,z)}$), based on the analysed metabolite profile of the “unknown” sample, comparative to those previously used for building the model. Figure 3.3 is a scatter plot of the predicted bacterial group probabilities as determined by the aforementioned discriminant model. All the previously “unknown” samples could be correctly assigned to their respective groups with probabilities of 72 – 100%.

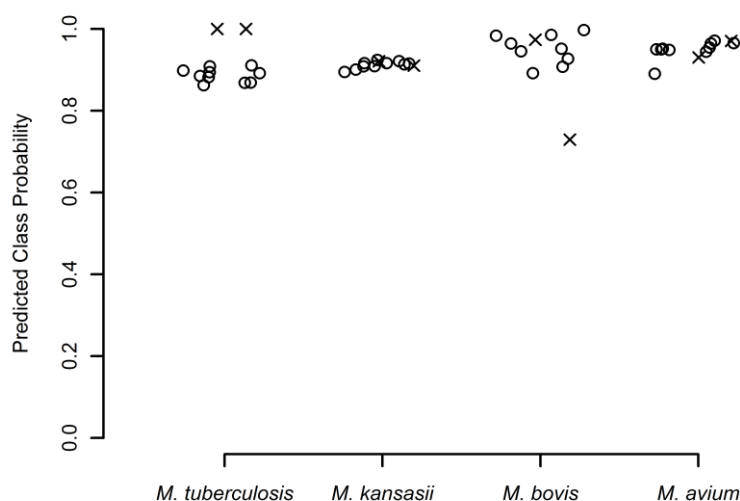


Figure 3.3: Probabilities for the “unknown” test samples (2 for each group) correctly identified for: *M. tuberculosis* are 100% in both cases; for *M. kansasii*: 92% and 91%; for *M. bovis*: 97% and 72%; and for *M. avium*: 97% and 93%. (o = original samples used to build the model, x = “unknown” test samples)

From these results it is evident that the potential sample variation and / or analytical variation of this metabolomics research approach, using the identified metabolite markers previously determined and again extracted from the individually cultured four *Mycobacterium* species prepared and analysed 6 months later, is negligible. This prediction additionally indicates that

this metabolomics research approach has the capacity to identify reliable metabolite markers for differentiating and characterising these species and potentially other infectious organism samples.

3.4 Detection limit

Figure 3.4 shows the PCA scores plot of the GC-MS data generated after the modified Bligh-Dyer extraction of the *M. tuberculosis* samples prepared at the mentioned concentration gradient (1×10^2 to 1×10^8 bacteria mL^{-1}). Although differentiation of the various concentrations of the *M. tuberculosis* sample repeat groups was attained, it can be seen that inter-group differentiation is less evident as the sample concentrations of the groups diminish. Despite this however, from Figure 3.4, it is clear that the sample concentration of 1×10^2 bacteria mL^{-1} overlaps with the blank (0 bacteria mL^{-1}), whilst 1×10^3 bacteria mL^{-1} is clearly separated, indicating that the potential detection limit of this approach for identifying the differentiating markers from these samples, would be in the range of 1×10^3 bacteria mL^{-1} .

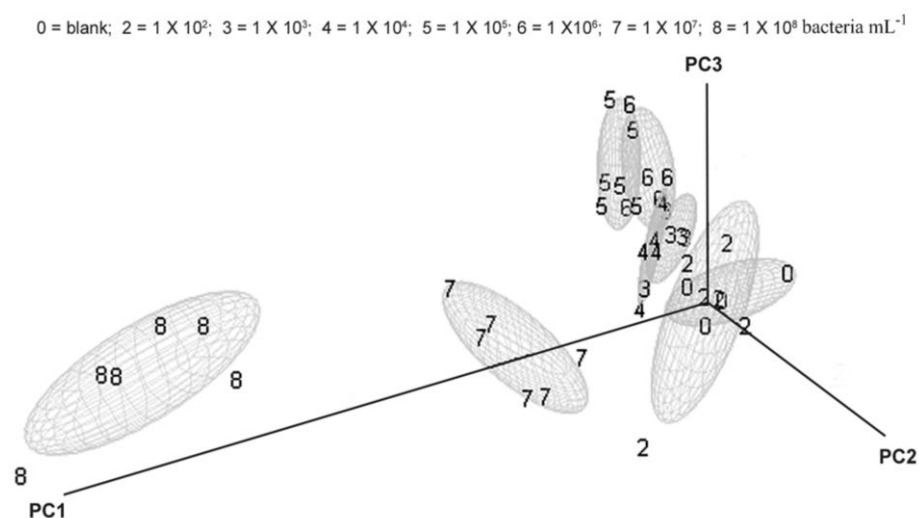


Figure 3.4: Three-dimensional PCA scores plot of the GC-MS generated data (after extraction using the modified Bligh-Dyer method) of the various prepared *M. tuberculosis* dilutions. All the groups, except that containing 1×10^2 bacteria mL^{-1} , could be distinguished from the blank group, determining a potential detection limit of 1×10^3 bacteria mL^{-1} .

In order to confirm this detection limit, determined for *M. tuberculosis*, for extracting those compounds capable of differentiating between the various TB-causing *Mycobacterium* species, we repeated the procedure using sample repeats of the four *Mycobacterium* species previously described, prepared at a concentration of 1×10^3 bacteria mL^{-1} . Once again, three PCs were used for the PCA of the GC-MS collected metabolite data, and the total amount of variance

explained by the first three PCs (R^2X cum) was 68.7%, of which PC 1 explained 27.7%, PC 2 explained 24.6%, and PC 3 explained 16.4%. The PCA scores plot of the processed data (Figure 3.5) indicates that all the infectious species groups, as expected, differentiated from one another on the basis of the extracted metabolite profiles, confirming the detection limit of 1000 cells per sample for this approach, using the modified Bligh-Dyer extraction method.

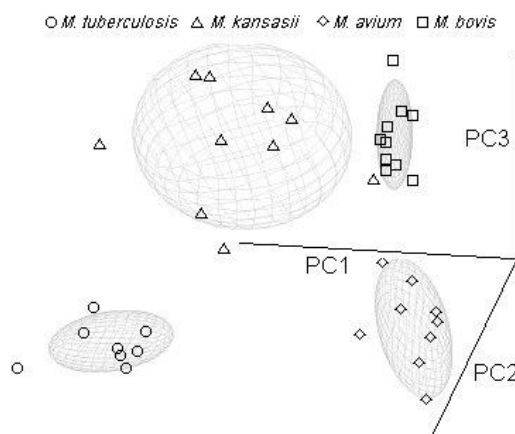


Figure 3.5: Three-dimensional PCA scores plot of the GC-MS generated data after extraction of the four *Mycobacterium* sample repeats using the modified Bligh-Dyer method, showing a clear differentiation between *Mycobacterium* sample groups at a concentration of 1×10^3 bacteria mL^{-1} .

It should however be mentioned that, although only 1000 cells are required per analysis for this differentiation, this is not necessarily the optimal concentration for detecting the most metabolite information for characterising these species using metabolomics. This determined detection limit simply indicates the smallest amount of sample required for extracting the minimum metabolite information for differentiating the groups.

4. CONCLUSIONS

According to Dunn *et al.* (2005), metabolomics is “the non-biased identification and quantification of *all* the metabolites in a biological system”, using highly sensitive analytical procedures. Using specialised software and multivariate statistical analyses, metabolite profiles can be compared in order to differentiate between the various biological systems investigated, such as the *Mycobacterium* species used in this study, for the purpose of identifying potential metabolite markers characteristic of these systems. Currently, however, no single analytical technique has the capacity to analyse *all* metabolites in a biological system simultaneously, and therefore we chose to analyse the lipid fraction of these mycobacterial species as these organisms are known to synthesise characteristic lipid / fatty acid components. Although this analyses added a certain degree of bias to the extraction method used, all data analyses where,

nevertheless, done in a non-biased manner, in order to comply with the requirement of a true metabolomics study.

Using such a metabolomics approach, we were able to prove that a modified Bligh-Dyer extraction method, followed by GC-MS fatty acid metabolome analysis and multivariate statistical data processing, is capable of differentiating between various TB and non-TB causing *Mycobacterium* species on the basis of their characteristic metabolite profiles, and that this approach is additionally able to identify the biologically relevant metabolite markers repeatedly over time. However, in order to explore the full potential of this approach, this methodology should be tested on a broader spectrum of species, as opposed to the limited number of species investigated in this study. Further refinement of this approach using other extraction approaches and / or more sensitive analytical techniques (such as GCxGC/TOFMS), may lead to even faster turnaround times, and the identification of even more metabolite markers. Therefore, these preliminary investigations strongly prove the capacity of this metabolomics methodology for metabolite marker identification, as well as the potential of such an approach to address various research topics, such as the mechanisms of drug-resistance, by identifying the metabolite markers associated with various strains of drug-resistant *M. tuberculosis*.

5. REFERENCES

- DAVIS, R. 2007. Tuberculosis and the intensivist. *Curr. Anaesthesia Crit. Care*. 18:76-85.
- DUNN, W.B., BAILEY., N.J.C., JOHNSON, H.E. 2005. Measuring the metabolome: current analytical technologies. *Analyst*. 130:606-625.
- GUERRANT, G.O., LAMBERT, M.A., MOSS, C.W. 1981. Gas-chromatographic analysis of mycolic acid cleavage products in mycobacteria. *J. Clin Microbiol.* 13:899-907.
- KNISLEY, C.V., DAMATO, J.J., MCCLATCHY, K., BRENNEN, P.J. 1985. Rapid and sensitive identification of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 22(5):761-767.
- KOEKEMOER, G., SWANEPOEL, J.W.H. 2008. A semi-parametric method for transforming data to normality. *Stat. Comput.* 8:241-257.
- LAMBERT, M.A., MOSS, C.W., SILCOX, V.A., GOOD, R.C. 1986. Analysis of mycolic acid cleavage products and cellular fatty acids of *Mycobacterium* species by capillary gas chromatography. *J. Clin. Microbiol.* 23:731-736.
- LARSSON, L., ODHAM, G., WESTERDAHL, G., OLSSON, B. 1987. Diagnosis of pulmonary tuberculosis by selected-ion monitoring: improved analysis of tuberculostearate in sputum using negative-ion mass spectrometry. *J. Clin. Microbiol.* 25(5):893–896.

- LUQUIN, M., AUSINA, V., LÓPEZ CALAHORRA, F., BELDA, F., GARCÍA BARCELÓ, M., CELMA, C., PRATS, G. 1991. Evaluation of Practical Chromatographic Procedures for Identification of Clinical Isolates of Mycobacteria. *J. Clin. Microbiol.* 29:120-130.
- MIGUEZ-BURBANO, M.J., FLORES, M., ASHKIN, D., RODRIGUEZ, A., GRANADA, A.M., QUINTERO, N., PITCHENIK, A. 2006. Non-tuberculous mycobacteria disease as a cause of hospitalization in HIV-infected subjects. *Int. J. Infect. Dis.* 10:47-55.
- MOSCA, A., RUSSO, F., MIRAGLIOTTA, L., IODICE, M.A., MIRAGLIOTTA, G. 2006. Utility of gas chromatography for rapid identification of mycobacterial species frequently encountered in clinical laboratory. *J. Microbiol. Methods.* 86:392-395.
- OLIVIER, KN. 1998. Nontuberculous mycobacterial pulmonary disease. *Curr. Opin. Pulm.* 4(3):148-53.
- REZWAN, M., LANÉELLE, M., SANDER, P., DAFFÉ, M. 2007. Breaking down the wall: Fractionation of mycobacteria. *J. Microbiol. Methods.* 68:32-39.
- SALMAN, M., BRENNAN, P.J. LONSDALE, J.T. 1999. Synthesis of mycolic acids of mycobacteria: an assessment of the cell-free system in light of the whole genome. *Biochim. et Biophys. Acta.* 1437:325-332.
- SCHOEMAN, J.C., LOOTS, D.T. 2011. Improved disease characterisation and diagnostics using metabolomics: a review. *J. Cell Tissue Res.* 11(1):2673-2683.
- STOPFORTH, A., TREDoux, A., CROUCH, A., VAN HELDEN, P., SANDRA, P. 2004. A rapid method of diagnosing pulmonary tuberculosis using stir bar sorptive extraction-thermal desorption-gas chromatography-mass spectrometry. *J. Chromatogr. A.* 1071:135-139.
- TODAR, K. 2005. *Todar's Online Textbook of Bacteriology*. University of Wisconsin-Madison Department of Bacteriology. <http://www.textbookofbacteriology.net/tuberculosis.html>
- VAN DER WALT, A.M., IBRAHIM, M.I.M., BEZUIDENHOUT, C.C., LOOTS, D.T. 2008. Linolenic acid and folate in wild-growing African dark leafy vegetables (morogo). *Public Health Nutr.* 10:1017.
- VREKEN, P., VAN LINT, A.E.M., BOOTSMA, A.H., OVERMARS, H., WANDERS, R.J.A., VAN GENNIP, A.H. 1998. Rapid stable isotope dilution analysis of very-long-chain fatty acids, pristanic acid and phytanic acid using gas chromatography-electron impact mass spectrometry. *J. Chromatogr. B.* 713:281-287.
- WHO. 2010. The Global Tuberculosis Epidemic: Fact sheet. June 2010. <http://www.kff.org/globalhealth/7883.cfm>