CHAPTER 5

Further development and comparison of a simpler fatty acid and total metabolome extraction procedure for differentiating various *Mycobacterium* species and *P. aeruginosa*

A part of this chapter has been published in the *African journal of microbiology research*.

OLIVIER, I., LOOTS, D.T. 2012. A comparison of two extraction methods for differentiating and characterising various *Mycobacterium* species and *Pseudomonas aeruginosa* using GC-MS metabolomics. *Afr. J. Microbiol. Res*. 6(13): 3159-3172.

1. INTRODUCTION

The fatty acid GC-MS metabolomics approach developed in Chapter 3 was successfully applied in Chapter 4 for the purpose of identifying new metabolite markers better characterising two rifampicin-resistant conferring *rpoB* mutations in *M. tuberculosis*. Not only was this study the first of its kind to link mutations in the *rpoB* gene to alterations in the fatty acid metabolism of these organisms, but the new markers identified also propose alterations to specific metabolic pathways, explaining a number of phenotypical characteristics of these mutants and additionally implying metabolic adaptations essential for *in vivo* survival. However, despite the excellent capacity of this developed methodology for metabolomics investigations, it may have a number of limitations when considering its complexity (number of steps required), the time taken from sample collection to obtaining an analytical result (16 hours) and the large volumes of reagents required (costs). Hence, further refinements to this approach is necessary, especially when considering its use for metabolomics investigations requiring the extraction of larger sample cohorts, such as in Chapter 7 of this thesis. Taking these limitations into account, the current chapter is dedicated to investigating a simpler, faster and more efficient way of extracting the fatty acid metabolome. An additional limitation to this fatty acid metabolomics approach however may be its targeted analysis of only the fatty acids. Considering that metabolomics as per definition, investigates "all metabolites", we additionally developed a second approach, using a similar extraction principal, capable of extracting and analysing a much wider variety of compounds, i.e. the "total metabolome" of the sample. The two extraction procedures developed in this chapter were compared in a related manner to those developed in Chapter 3, with regards to their respective repeatability, differentiation capacities, detection limits, and validation of the metabolite markers identified via the metabolomics data analysis using available literature. Furthermore, for this investigation, the same *Mycobacterium* species sample repeats used in Chapter 3, were analysed. However, just for interest, *Pseudomonas aeruginosa* was also included as an additional infectious organism species in this investigation. *P. aeruginosa*, a genus from the phylum Proteobacteria, is a common cause of iatrogenic infections, and infection of this organism to the respiratory passages of patients with immuno-compromising illnesses, results in symptoms similar to that of pulmonary TB (Todar, 2005), hence its inclusion here.

Aim: To investigate a simpler, faster and more efficient way of extracting the fatty acid metabolome, and the comparison of this method to a similar approach extracting the "total metabolome" of a sample, for later metabolomics research applications to analysing TB-positive and TB-negative patient collected sputa (Chapter 7).

2. MATERIALS AND METHODS

2.1 Reagents and chemicals

N-methyl-N-(tert.-butyldimethylsilyl) trifluoroacetamide (MSTFA) with 1% trimethylchlorosilane (TMCS), pyridine, KOH, methyl-nonadecanoic acid (Me-C19), glacial acetic acid, and trimethylpentane were purchased from Merck (Darmstadt, Germany). Methoxyamine hydrochloride and 3-phenyl butyric acid were purchased from Sigma-Aldrich (St. Louis, Mo., USA). All organic solvents used were ultra purity Burdick & Jackson brands (Honeywell International Inc., Muskegon, MI, USA) and were used without any further purification.

2.2 Bacterial cultures

All cultures used in this study were supplied by the Royal Tropical Institute, Amsterdam, The Netherlands, and were cultured as described in Chapter 3 section 2.2.

Similar to the approach used in Chapter 3, in order to compare the repeatability and extraction capacity of the two extraction procedures investigated for later metabolomics applications, each extraction procedure was repeated 12 times using 12 x 1 mL of isolated *M. tuberculosis* sample repeats re-suspended in ddH_2O 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 each extraction method's capacity for differentiating the various bacterial groups, as would typically be done during metabolomics studies, 12 x 1 mL sample repeats of each of the five re-suspended isolated infectious organism samples (*M. tuberculosis, M. avium, M. bovis, M. kansasii* and *P. aeruginosa*) were prepared at a concentration of 1×10^8 bacteria mL^{-1} in ddH_2O, prior to extraction (using the two methods investigated), GC-MS analyses and multivariate statistical analyses of the generated data.

In order to determine the potential detection limit of these extraction methods, 6 x 1 mL repeats of the isolated *M. tuberculosis* re-suspended in ddH_2O at a concentration gradient ranging from 1×10^1 - 1×10^6 bacteria mL^{-1}, including a blank (ddH_2O only), were extracted using the investigated methods, analysed using the GC-MS, and the collected data were processed using the multivariate statistics described below. In order to verify the differentiation capacity of these extraction methods at the determined detection limits, 6 x 1mL samples of each of the five re-suspended isolated bacterial species (*M. tuberculosis, M. avium, M. bovis, M. kansasii, and P. aeruginosa*) were prepared at the concentration determined by the detection limit experiments, in ddH_2O, prior to extraction, GC-MS analyses and multivariate statistical analysis.

2.3 Extraction procedure 1 – Fatty acid metabolome extraction

An extraction solvent mixture (1.25 mL), consisting of chloroform, methanol, and ddH_2O, in a ratio of 1:3:1 respectively, was added to 250 μL of the above mentioned sample suspensions in a microcentrifuge tube. The extraction was performed using an MM 400 vibration mill (Retsch GmbH & co. KG, Haan, Germany) at a frequency of 30 Hz for 5 min with a 3-mm tungsten carbide bead (Retsch GmbH & co. KG) for increased extraction efficiency. Chloroform and ddH_2O (250 μL of each) was subsequently added, followed by centrifugation for 5 min at 550 x g. The organic phase was collected, transferred to a kimax tube, and dried under a light stream of nitrogen. Thereafter, 0.5 mL of chloroform, 0.5 mL of methanol, and 1 mL of methanolic KOH (0.2 M) was added to the dried extract. The mixture was then incubated for 30 min at 60˚C followed by the addition of 2 mL of hexane, 200 μL of glacial acetic acid (1 M), and 2 mL of ddH_2O. After centrifugation at 550 x g for 5 min, the organic phase was once again collected and transferred to a new kimax tube. The remaining water phase was re-extracted 3 times with hexane and the combined organic phases from each of the previous extractions were dried under nitrogen and re-suspended in 50 μL of external standard (Me-C19) dissolved in trimethylpentane (15.6 ng mL^{-1}).

2.4 Extraction procedure 2 - Total metabolome extraction

As an internal standard, 50 μL of 3-phenyl butyric acid (26.25 mg mL^{-1}) was added to 250 μL of the above mentioned sample suspensions in a microcentrifuge tube, followed by the addition of 1.25 mL of an extraction solvent mixture containing chloroform, methanol, and ddH_2O, in a 1:3:1 ratio. The extraction was again performed using an MM 400 vibration mill at a frequency of 30 Hz, for 5 min, after adding a 3-mm tungsten carbide bead to each tube for increased extraction efficiency. After centrifugation, the supernatant (including the organic and water phases, but excluding the pellet) was collected, transferred to a GC-MS sample vial and dried under a light stream of nitrogen. The dried supernatant was subsequently derivatised using 50 μL of methoxyamine hydrochloride in pyridine (15 mg mL^{-1}) at 50˚C for 90 min, followed by 50 μL of MSTFA with 1% TMCS at 50˚C for 60 min.

2.5 GC-MS parameters

The above mentioned prepared extracts were analysed by randomly injecting 1 μL on an Agilent 7890A gas chromatograph (Agilent, Atlanta, GA) 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) in the splitless mode. Helium was used as the carrier

gas and the pressure was 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 for the fatty acid extracts and *m/z* 60-800 for the total metabolite extracts). For the fatty acid metabolome extracts, the most optimal analytical conditions were an injector temperature of 250˚C, for the total run time, and an initial GC oven temperature of 50˚C for 1 min, followed by an initial increase in oven temperature of 20˚C min^{-1} to 160˚C. This was then followed by an increase in oven temperature of 10˚C min^{-1} to 190˚C and then a further increase of 5˚C min^{-1}, to a final temperature of 300˚C, at which the temperature was maintained for 5 min. For the total metabolome extracts, the most optimal analytical conditions were an injector temperature of 270˚C, for the total run time, and an initial GC oven temperature of 70˚C for 2 min, followed by an increase in oven temperature of 4˚C min^{-1}, to a final temperature of 300˚C, at which the temperature was held for 2 min.

2.6 Data-acquisition

The raw GC-MS data were deconvoluted and analysed using AMDIS software (V2.65). Alignment of the detected compounds across the samples analysed was achieved by creating a new reference library in AMDIS, 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, as described in Chapter 3 section 2.5. Each analysed sample was subsequently processed using the aforementioned reference library, and the resulting output of each sample was combined in order to generate a data matrix containing the relative concentrations (normalised with the internal standard) for all compounds present or absent in each sample. This was done for each of the GC-MS data sets collected from each of the two extraction methods investigated.

2.7 Statistical data analysis

The repeatability of the two extraction methods investigated were compared using the distribution of the calculated CV values, of the relative concentrations, of all the compounds detected subsequent to GC-MS analysis of the extracted sample repeats.

The statistical packages, “R” (version 2.13.0) and Statistica (version 10) were used for all the remaining statistical data analyses. Data were normalised relative to the respective internal standards in order to compensate for sample loss during the extraction or chromatographic injection. To prevent variables with high concentrations from dominating the multivariate statistical analyses, data pre-treatment using a non-parametric transformation function (Koekemoer & Swanepoel, 2008) was used to scale the data prior to statistical data analyses, after which mean centering was applied.

In order to determine whether or not a natural grouping (differentiation) exists between the investigated sample groups, PCA was applied. PLS-DA was used to identify those compounds which best characterise the differentiated sample groups by ranking the metabolites according to the VIP parameter, as described in Chapter 3, section 2.6. 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.

3. RESULTS AND DISCUSSION

Although the purpose of this investigation wasn't aimed at directly comparing any of the two described extraction methods to those conventionally used for similar purposes, it is of value to mention that the use of a vibration mill as described in this study, significantly improves the extraction efficiency of any extraction procedure, especially when extracting compounds from more robust matrixes (such as cell or tissues samples). This increased extraction efficiency allows for faster extraction of 24 samples simultaneously, using far less solvent volumes, in a 2 mL centrifuge tube. This fatty acid extraction method, for instance, can be completed within 5 hours, including the drying and derivitisation steps, using only 1.75 mL of solvent, which is a vast improvement on the 16 hours required for the completion of the original Bligh-Dyer method, which also requires far greater volumes of solvent. Comparatively, the total metabolome extraction method is even simpler (as no wash step is required) and faster (completed in less than 4 hours), and requires even smaller volumes of solvent (1.25 mL).

3.1 Extraction capacity

The AMDIS generated reference library (a compilation all the compounds extracted from all five sample groups) for the fatty acid metabolome extraction procedure contained 270 compounds, predominantly fatty acids and various hydrocarbons, while the AMDIS generated reference library for the total metabolome extraction procedure contained a total 1194 compounds, belonging to a variety of compound classes including, amongst others: fatty acids, amino acids, alcohols, organic acids, monosaccharides, alkenes, alkanes, purines, pyrimidines etc. Although both of the tested methods extracted fatty acids, a more complete profile of these compounds were seen after the application of the fatty acid metabolome extraction procedure. The reduced number of fatty acids detected after applying the total metabolome extraction procedure can be ascribed to the complexity of the obtained chromatograms, overwhelming fatty acids in lower concentrations. Table 5.1 indicates the number of compounds extracted via each of the compared extraction methods, for each of the infectious disease sample groups. Considering

this, the total metabolome extraction procedure detected, on average, 358 more compounds per sample group, which could be expected, as the fatty acid metabolome extraction method is a more targeted approach, and is not aimed at extracting all compounds present in the sample matrix.

Table 5.1: Number of compounds in the AMDIS generated reference libraries, originating from each of the bacterial sample groups, for the two investigated extraction methods. The standard deviations, representing the variation in the number of compounds detected between individual samples, are given in parenthesis.

Bacterial sample group	Fatty acid extraction method	Total metabolome extraction method
M. tuberculosis	180 (10.09)	461 (10.93)
M. avium	178 (06.12)	501 (13.24)
M. bovis	174 (13.12)	388 (14.40)
M. kansasii	191 (05.87)	641 (14.94)
P. aeruginosa	161 (11.31)	679 (19.66)
Total number of compounds in library	270	1194

It should be noted that this comparison of the number of compounds is merely for descriptive purposes, and may not necessarily be an indication of which of the two described methods performs better for metabolomics applications and for the identification of makers characterising the above mentioned infectious organisms. Although, as per definition of metabolomics, the total metabolome extraction procedure should better fulfil the required criteria for identifying "all compounds in an unbiased manner", the increased dimensionality of the resulting extract may potentially overwhelm the dimensionality limits of the GC-MS. The chromatographic analyses of these complex extracts may consequently lead to compound co-elution and poor repeatability, and consequently larger variation or unwanted "noise" in the collected data, hampering this extraction method's application for metabolomics research purposes. Knowing this beforehand, however, GC-MS flow rates and temperature programming was optimised, in order to minimise the chance of this happening.

3.2 Repeatability

As previously described in Chapter 3, the repeatability of the 2 extraction methods investigated were compared by calculating the CV values of the relative concentrations of all the compounds detected subsequent to GC-MS analysis of the extracted sample repeats. The distribution of these CV values is shown in Figure 5.1.

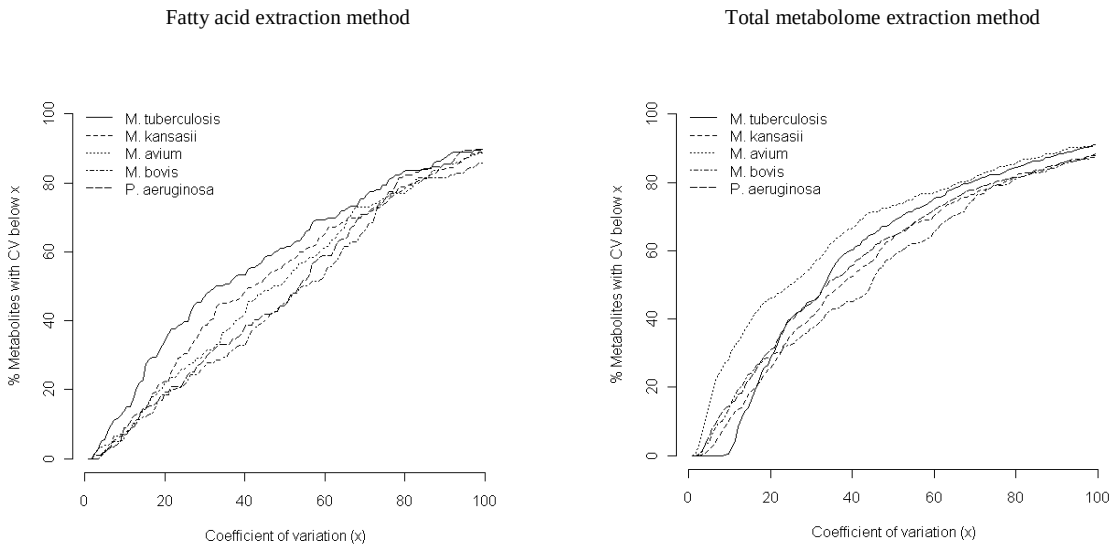

Figure 5.1: Distribution of the calculated coefficient of variation (CV) values for all the GC-MS detected compounds from each infectious organism group (12 sample repeats of each organism), extracted using the two extraction methods comparatively investigated.

For all the organism groups, an averaged 50% ± 6.7% of all the compounds detected via the fatty acid extraction procedure, comparative to 62% ± 4.6% for that of the total metabolome extraction, fell below a 50% CV value. These numbers indicate that the total metabolome extraction procedure, despite extracting a larger number and variety of compounds with potentially larger variations in their concentrations, still showed a comparatively better repeatability. As the most optimal GC-MS conditions were selected for analysing the extracts derived from both extraction methods, the better repeatability of the total metabolome extraction procedure is most easily explained by its comparative simplicity, resulting in a far lower likelihood for analytical variation than the fatty acid metabolome extraction method, which requires one additional step.

3.3 Comparison of each extraction method's capacity for metabolomics species differentiation

When considering the application of these methods for comparative metabolomics, not only is the repeatability and extraction capacity as described above of importance, but also their capacity to extract those compounds most important for differentiating the groups being compared, in order to answer the biological question at hand. Subsequently, two multivariate statistical data analyses methods (PCA and PLS-DA) were used to summarise and interpret the collected metabolite data. Prior to these analyses, a 50% filter was applied to the data collected, in order to exclude those compounds which do not appear in at least 50% of the

samples, in one or more of the sample groups, as a data pre-processing step for eliminating unnecessary "noise" in the data.

Three PCs were extracted for the fatty acid metabolome extraction analyses, explaining a total of 64.4% of the variation in the data (R^2X cum), of which PC 1 explained 34.8%, PC 2 explained 15.7% and PC 3 explained 13.9%. Similarly, three PCs were extracted for the total metabolome extraction procedure's analysed data set, explaining a total of 48.8% of the variance in the data (R^2X cum), of which PC 1 explained 27.2%, PC 2 explained 14.5% and PC 3 explained 7.08%.

The PCA scores plot of the GC-MS data acquired using the fatty acid metabolome sample extracts, is shown in Figure 5.2, indicating a clear grouping of the individual samples of the respective infectious bacterial groups and a differentiation between these groups. Furthermore, all the related *Mycobacterium* species groups were grouped together, relative to the *P. aeruginosa* group, indicating that organisms with similar genetic characteristics group together due the similarities in their genomes and resultant metabolite profiles.

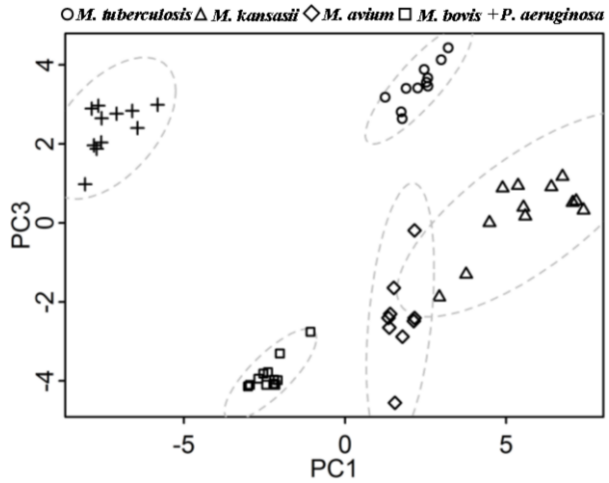

Figure 5.2: PCA scores plot (PC1 vs. PC3) of the GC-MS data acquired using the fatty acid metabolome extraction, indicating a grouping of the individual samples into their respective organism groups on the basis of the variation in the detected fatty acid metabolomes.

Figure 5.3 illustrates the PCA scores plot for the GC-MS data acquired using the total metabolome sample extracts. Similar to that of the fatty acid metabolome extraction method's PCA scores plot, the related *Mycobacterium* species groups were, once again, grouped together, relative to the *P. aeruginosa* group. This time however, a visibly smaller inter-sample, and hence, within group variation was detected, comparative to that of the fatty acid metabolome extraction procedure, resulting in a better differentiation of the infectious bacterial groups investigated. This smaller within group variation can be explained by the comparatively

better extraction repeatability of the total metabolome extraction procedure for those compounds which best describe the variation between the infectious organism groups.

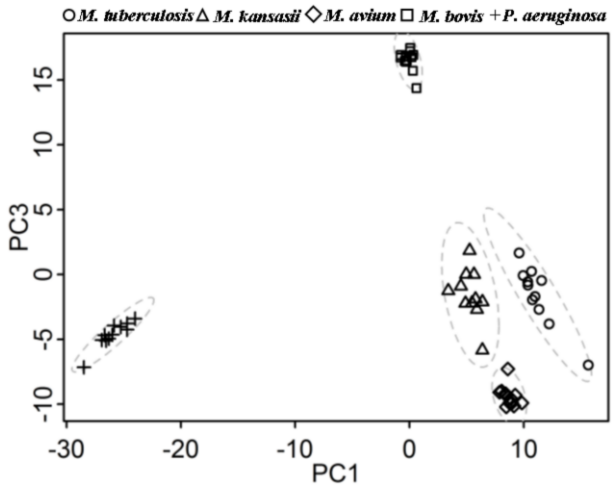

Figure 5.3: PCA scores plot (PC1 vs. PC3) of the GC-MS data acquired using the total metabolome extraction, indicating a grouping of the individual samples into their respective organism groups, with a comparatively smaller within group variation, and hence, bigger inter-group variation, compared to that of the fatty acid extraction procedure.

3.4 Detection limits

As described in the methods section, 6 repeats of each of the six different concentrations (1 X 10^1 to 1 X 10^6 bacteria mL^{-1}) of the *M. tuberculosis* concentration gradient samples, together with 6 blank samples (only ddH_2O), were used for the detection limit analyses. Once again, PCA models were built using the AMDIS generated data matrixes, consisting of the relative concentrations of all metabolites present in each of the analysed samples. The detection limit for each extraction procedure was defined as the group with the lowest concentration not overlapping with the blank samples (Fend *et al.*, 2006). Figure 5.4 represents the PCA output obtained using the fatty acid metabolome extraction procedure on the *M. tuberculosis* concentration series. The lowest concentration group not overlapping with the blank, and thus the detection limit for this extraction method, is 1 X 10^4 bacteria mL^{-1}. Considering that only 250 μL of the sample suspension was used for these extractions, the minimum amount of sample material required for detecting metabolite markers differentiating *M. tuberculosis* from a sample blank would be 2500 cells.

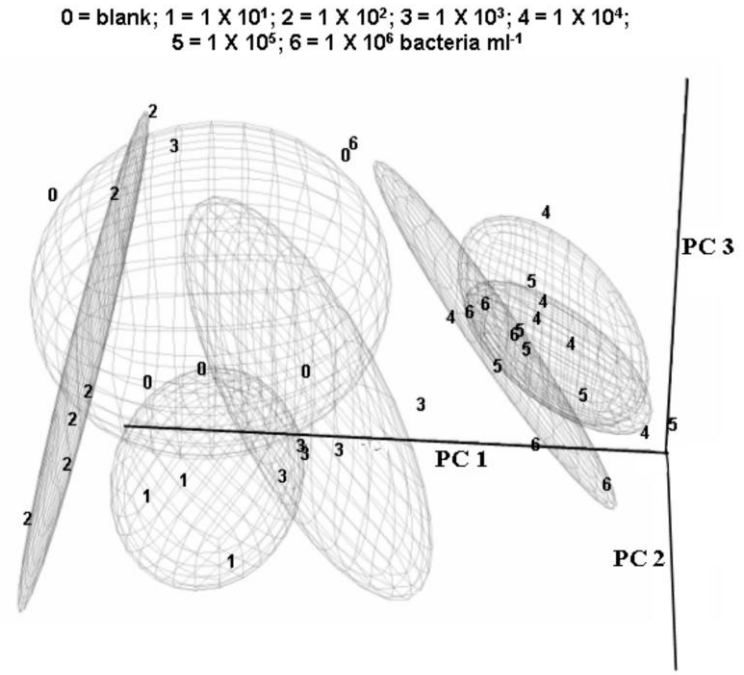

Figure 5.4: PCA scores plot (PC 1 vs. PC 2 vs. PC 3) of the GC-MS generated data of the concentration gradient samples of *M. tuberculosis* using the fatty acid metabolome extraction method. The lowest bacterial concentration not overlapping with the blank, and hence the detection limit for this method, is 1 x 10^4 bacteria mL^{-1}.

Comparatively, considering the results from the total metabolome extraction procedure (Figure 5.5), the lowest concentration not overlapping with the blank was 1 X 10^3 bacteria mL^{-1}. Since this method also requires only 250 µL of the sample suspension, the amount of sample material required for detecting metabolite markers differentiating *M. tuberculosis* from a sample blank would be 250 cells. This indicates that the latter extraction approach allows for the detection of those compounds which inherently occur in higher concentrations in these organisms.

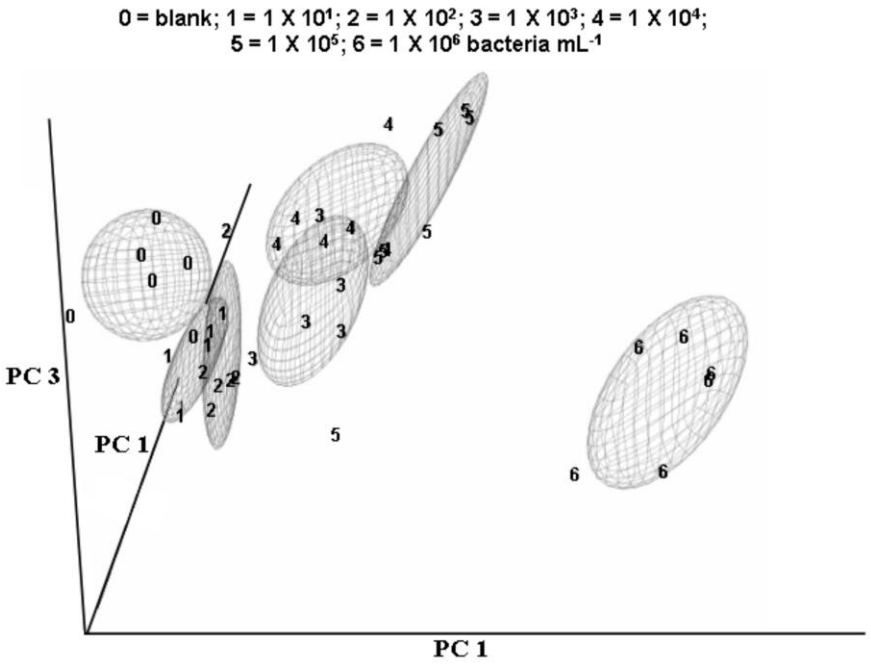

Figure 5.5: PCA scores plot (PC 1 vs. PC 2 vs. PC 3) of the GC-MS generated data of the concentration gradient samples of *M. tuberculosis* using the total metabolome extraction method. The lowest bacterial concentration not overlapping with the blank, and hence the detection limit for this method, is 1 x 10^3 bacteria mL^{-1}.

In order to confirm these detection limits 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 five bacterial species previously described, prepared at a concentration of 1 X 10^4 bacteria mL^{-1} for the fatty acid extraction method and 1 X 10^3 bacteria mL^{-1} for the total metabolome extraction method. Once again, in both instances, 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) using the fatty acid extraction method was 65.4%, of which PC 1 explained 35.6%, PC 2 explained 19.7%, and PC three explained 10.06%. When using the total metabolome extraction method, the total amount of variance explained by the first three PCs (R^2X cum) was 78.5%, of which PC 1 explained 50.7%, PC 2 explained 20.9%, and PC three explained 6.9%. The PCA scores plots of the processed data (Figure 5.6) indicates that all the infectious species groups differentiated from one another on the basis of the extracted metabolite profiles, confirming the detection limits for this approach, using both the investigated extraction methods.

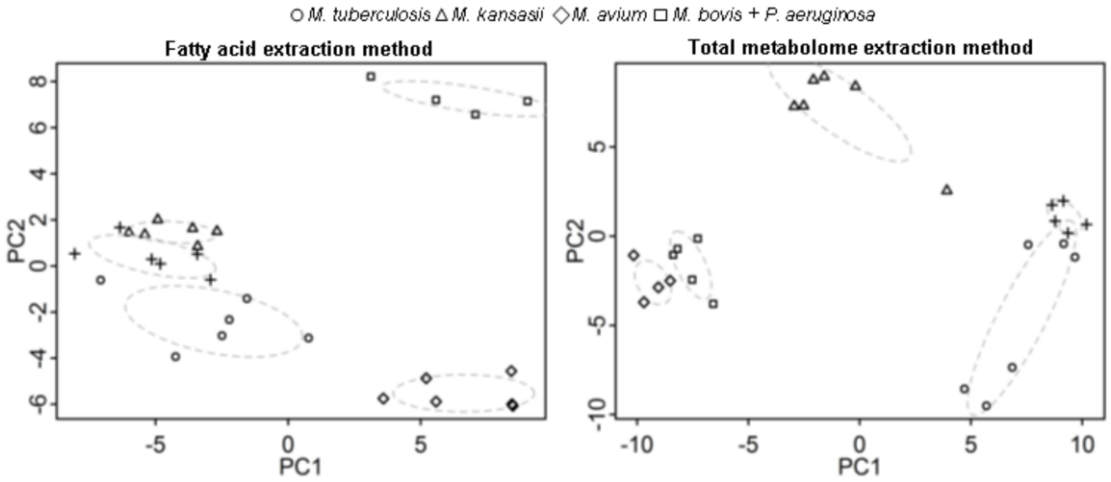

Figure 5.6: Three-dimensional PCA scores plot of the GC-MS generated data after extraction of the five bacterial sample repeats, showing a clear differentiation between all the sample groups at a concentration of 1 x 10^4 bacteria mL^{-1} using the fatty acid extraction method and 1 x 10^3 bacteria mL^{-1} using the total metabolome extraction method.

3.5 Metabolite maker identification

One of the primary functions of a metabolomics research approach is not only determining if a differentiation of various sample groups exists on the basis of varying metabolite profiles, but also to identify those metabolite markers which vary the most between these groups. These markers, in turn, may be used to better characterise the sample groups investigated, and in so doing answer a biological question. Considering this, those metabolites with the highest VIP values (identified via PLS-DA), extracted by each of the investigated methods, were identified and evaluated from a functional biology perspective. This was done in order to ensure that the variation detected is in fact due to the extraction of those compounds of biological relevance. In Table 5.2, the mean relative concentrations of those metabolite markers identified using the fatty acid metabolome extraction approach are presented. The PLS-DA model used for identifying these had a modelling parameter R^2Y (cum) of 93.7%, indicative of the total explained variation of the response Y. Q^2 (cum), the cross-validated variation explained by the response Y, was 89.1%.

Table 5.2: Mean relative concentrations (ug mg^{-1} sample) of the metabolite markers identified for the various infectious organism groups using the fatty acid metabolome extraction procedure. Standard deviations are given in parenthesis.

VIP ranking	Species	*M. tuberculosis*	*M. kansasii*	*M. avium*	*M. bovis*	*P. aeruginosa*
	Compound	**Concentration (SD)**				
1	TBSA	273.9 (28.2)	171.3 (50.4)	139.8 (27.3)	32.7 (8.7)	0
2	C22:0	14.1 (2.7)	63 (26.1)	17.7 (5.4)	2.4 (0.6)	T
3	2-octyl-CPOA	0	0	0	0	195.9 (25.8)
4	C17:0	43.2 (8.4)	8.7 (3)	6.3 (0.9)	4.5 (1.2)	3.6 (0.6)
5	C22:1 ω9c	0	21.3 (9)	1.5 (0.3)	0	0
6	C24:1 ω9c	0	25.2 (11.7)	2.4 (0.06)	0	0
7	2,4-DM C14:0	0	3.3 (1.8)	0	T	0
8	Unknown 466	0.6 (0.3)	4.8 (2.7)	0.3 (0.6)	0	0
9	C16:1 ω7c	18 (4.5)	20.7 (8.7)	25.5 (8.4)	0.3 (0.3)	19.5 (6.9)
10	C20:0	36.9 (6.9)	29.4 (11.1)	20.7 (3)	5.4 (1.2)	1.2 (0.3)
11	C14:0	4.8 (2.7)	17.4 (10.8)	8.7 (4.2)	0.3 (0.3)	2.1 (1.5)
12	C28:0	3.3 (0.9)	2.7 (1.2)	T	0	0
13	Unknown 494	1.5 (0.3)	1.5 (0.6)	0.9 (0.3)	0	0
14	Unknown 422	5.7 (1.2)	0	0	0.6 (0.3)	0
15	C25:0	4.2 (0.9)	0	0	0.3 (0.03)	0
16	C18:1 ω9c	192.6 (32.1)	309 (86.1)	193.5 (24.9)	10.5 (2.7)	0
17	2-hexyl-CPOA	0	0	0	0	50.1 (8.4)
18	C20:1 ω9c	0.9 (0.3)	5.4 (2.4)	0.9 (0.6)	0	0
19	Unknown 340	1.8 (0.3)	0	0	0	0

TBSA, tuberculostearic acid; CPOA, cyclopropaneoctanoic acid; DM, dimethyl; T, trace amounts

From the data in Table 5.2, it is clear that it is not only those compounds totally novel to a particular group which are considered important for differentiating these infectious organisms, but also those compounds which are common to two or more sample groups, showing constant differences in their concentrations.

Of the two methods investigated, the fatty acid extraction procedure was selected based on the prior knowledge that *Mycobacterium* species produce rather unique fatty acids, which are involved in a number of structural and functional processes (Knisley *et al.,* 1985). Consistent with previous reports, we identified the well known characteristic fatty acid, tuberculostearic acid (TBSA) (Larsson *et al.*, 1987; Stopforth *et al.*, 2004), as well as oleic acid (C18:1 ω9c), which is known to be one of the most abundant fatty acids in these bacteria (Lambert *et al*., 1986, Mosca *et al*., 2006, Chapter 3), as novel metabolite markers for all the *Mycobacterium* species groups investigated comparative to *P. aeruginosa.* Additionally, the mycolic acid cleavage products (MACPs), eicosanoic acid (C20:0) and docosanoic acid (C22:0), formed as a result of chromatographic heat cleavage of the characteristic *Mycobacterium* cell wall mycolic acids (Guerrant *et al.* 1981), were also detected in relatively abundant concentrations in all the *Mycobacterium* species, as compared to only trace amounts in the *P. aeruginosa* sample group. Another MACP, octacosanoic acid (C28:0), together with gondoic acid (C20:1 ω9c) and two unknown compounds with molecular masses of 466 and 494 respectively, may potentially also characterise *Mycobacterium* species. Two unsaturated long-chain fatty acids, erucic acid

(C22:1 ω9c) and nervonic acid (C22:1 ω9c), were detected only in the *M. kansassi* and *M. avium* sample groups, whereas pentacosanoic acid (C25:0) was detected exclusively in the *M. tuberculosis* and *M. bovis* sample groups. Furthermore, as previously identified, 2,4 dimethyl-tetradecanoic (2,4-DM C14:0) was once again detected as a novel metabolite for *M. kansasii* (Lambert *et al.*, 1986; Mosca *et al.,* 2006). We additionally detected another unknown compound, with a molecular mass of 340, specifically characterising *M. tuberculosis* and 2-octyl-cyclopropaneoctanoic acid (2-octyl-CPOA) and 2-hexyl-cyclopropaneoctanoic acid (2-hexyl-CPOA) as unique characteristic metabolites of *P. aeruginosa.*

The PLS-DA model used for identifying the characteristic marker metabolites extracted using the total metabolome extraction procedure, had a modelling parameter R^2Y (cum) of 97.8%, with a Q^2 (cum) of 98.8%. This extraction method was developed for the purpose of extracting the entire metabolome and consequently the metabolites identified included a variety of fatty acids, amino acids, alcohols, organic acids, monosaccharides, alkenes, alkanes, purines, pyrimidines etc. These metabolite markers identified are listed in Table 5.3.

Table 5.3: Mean relative concentrations (ug mg^{-1} sample) of the metabolite markers identified for the various infectious organism groups investigated using the total metabolome extraction procedure. Standard deviations are given in parenthesis.

VIP ranking	Species	*M. tuberculosis*	*M. kansasii*	*M. avium*	*M. bovis*	*P. aeruginosa*
	Compound	**Concentration (SD)**				
1	Indole-acetic acid	0	0	0	0	123.6 (0.6)
2	Unknown 343	0	0	0	0	216.6 (36.6)
3	Citric acid	0	213.3 (55.2)	0	4.5 (0.3)	0
4	Cadaverine	0	0	0	0	84.3 (15.6)
5	Purine	0	0	0	0	399 (67.5)
6	Inositol	262.8 (138.6)	6.9 (6.0)	168.6 (20.1)	0.3 (0.6)	0
7	5'-Adenylic acid	0	0	55.2 (7.8)	0	735 (111.3)
8	C16:1 ω7t	0.3 (0.6)	0.9 (1.2)	62.1 (5.7)	0	0.3 (1.2)
9	Myo-Inositol	227.1 (149.1)	27.9 (18.0)	348 (50.4)	1.5 (1.8)	0
10	Succinic acid	0	119.4 (19.8)	0	0	0
11	Putrescine	0	0	0	0	191.1 (42.9)
12	C10:0	0	T	19.8 (1.5)	0	53.1 (13.5)
13	Unknown 268	0	22.5 (0)	0	0	0
14	C16:1ω7c	6.6 (2.2)	7.2 (4.5)	303 (32.1)	0	53.1 (9.9)
15	Erythritol	2.4 (1.2)	19.2 (5.4)	12 (0.9)	0	0
16	Unknown 367	0	0.6 (0.9)	15.9 (1.8)	0	39 (7.2)
17	Unknown 373	0	0	0	0	60.6 (14.7)
18	Unknown 541	84 (42)	0	0	0	0
19	Valeric acid	0	0	0	0	114.9 (15)

The highest ranked metabolite marker extracted using the total metabolome extraction procedure, and novel to *P. aeruginosa,* was indole-acetic acid. This metabolite has previously been identified in this organism and in other related plant growth-promoting rhizobacteria (Karnwal, 2009). Other novel metabolite markers identified for *P. aeruginosa* using the total metabolome extraction procedure included: cadaverine, an intermediate in the alternative decarboxylation and catabolism of L-lysine (Fothergill & Guest, 1977; Steward, 1970); putrescine, an intermediate used in its L-arginine decarboxylase catabolic pathway (Mercenier *et al.*, 1980); purine; valeric acid; and two unknown compounds with molecular masses of 343 and 373. For the *Mycobacterium* species groups analysed, inositol and its stereoisomer, myo-inositol, were identified as novel comparative to *P. aeruginosa*. As summarized by Brown *et al.* (2007), the myo-inositol supply in mycobacteria is believed to be sustained by *de novo* synthesis initiated by the conversion of glucose-6-phosphate to inositol-1-phosphate, followed by dephosphorylation, and this is a novel characteristic of these organisms, substantiating our results. Succinic acid and an unknown compound of molecular mass of 268, were identified as novel *M. kansasii* markers, in addition to elevated concentrations of citric acid, which has never before been documented. It is also important to note that an unknown compound of mass 541

was uniquely detected in *M. tuberculosis* using this extraction approach, which could be investigated for potential diagnostic applications.

4. CONCLUSIONS

Considering these results, both the fatty acid metabolome and total metabolome extraction procedures developed and investigated in this chapter, in conjunction with GC-MS metabolomics analyses, show promise for not only differentiating various TB and non-TB causing *Mycobacterium* species from one another and from other lung pathogens, but also for identifying those markers best characterising these infectious organisms. However, the total metabolome extraction procedure has advantages over the fatty acid metabolome extraction procedure, as it is: 1) simpler, 2) faster (taking less than 4 hours), 3) more repeatable 4) extracts metabolites from all compound classes (as opposed to only fatty acids), and 5) has a 10 times better detection limit.

Considering this, these preliminary results suggest that the total metabolome extraction procedure would be the method of choice for differentiating and identifying biologically relevant metabolite markers, from all compound classes, should it be applied to the metabolomics analysis of patient collected TB-positive vs. TB-negative sputum samples.

5. REFERENCES

BROWN, A.K., MENG, G., GHADBANE, H., SCOTT, D.J., DOVER, L.G., NIGOU, J., BESRA, G.S., FÜTTERER, K. 2007. Dimerization of inositol monophosphatase *Mycobacterium tuberculosis* SuhB is not constitutive, but induced by binding of the activator Mg2+. *BMC Struct. Biol.* 7:55.

FEND, R., KOLK, A.H., BESSANT, C., BUIJTELS, P., KLATSER, P.R., WOODMAN, A.C. 2006. Prospects for clinical application of electronic-nose technology to early detection of *Mycobacterium tuberculosis* in culture and sputum. *J Clin Microbiol*. 44:2039-2045.

FOTHERGILL, J.C., GUEST, J.R. 1977. Catabolism of L-Lysine by *Pseudomonas aeruginosa*. *J. Genl. Microbiol.* 99:139-155.

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.

KARNWAL, A. 2009. Production of indole acetic acid by fluorescent *Pseudomonas* in the presence of L-Tryptophan and rice root exudates. *J. Plant Pathol*. 91(1):61-63.

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.

MERCENIER, A., SIMON, J., HAAS, D., STALON, V. 1980. Catabolism of L-Arginine by *Pseudomonas aeruginosa*. *J Genl Microbiol.* 116:381-389.

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.

STEWART, D.J. 1970. The Decarboxylation of L-lysine by *Pseudomonas aeruginosa*. *J. appl. Bact.* 33:501-504.

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