CHAPTER 6

A comparison of four different sputum pre-extraction preparations methods, prior to GCxGC-TOFMS metabolomics analysis, to characterise *M. tuberculosis* from patient collected sputum

1. INTRODUCTION

Due to the viscosity and uneven consistencies (or fluidity) of patient collected sputum, repeatable analyses of these samples are challenging, as the production of reproducible aliquots , prior to extraction, is impossible without either prior isolation of the bacterial cells or homogenisation of these viscous mixtures. These irregularities are due to the mucoproteins present in sputum. Considering this, we investigated and compared four different sputum preparation methods, prior to extraction, using the previously investigated total metabolome extraction procedure (Chapter 5), followed by GCxGC-TOFMS and statistical data analyses. Three of the four methods tested, using: 1) Sputolysin; 2) a combination of N-acetyl-L-cysteine and sodium hydroxide (NALC-NaOH); and 3) NaOH alone, are standard sputum *Mycobacterium* cell isolation procedures, used prior to culturing for diagnostic purposes (Alugupalli *et al.,* 1993; Petroff *et al.*, 1915; Kubica *et al.,* 1967). The fourth method investigated involved only a simple homogenization of the sputum samples with an ethanol solution. These pre-extraction preparation methods were once again compared on the basis of repeatability, extraction efficiency, and capacity to extract those compounds able to best differentiate between the *M. tuberculosis* spiked and control pooled sputum sample repeats. Those compounds best explaining the variation between the sputum sample groups were identified and once again validated using literature, in order to confirm whether or not the selected approach, does in fact allow the identification of metabolite markers of biological relevance, and not due to any other anomaly. Furthermore, as previously described, the effect of these sputum pre-extraction processing methods on the minimum number of cells required for analysis (i.e. the detection limit) was also investigated.

Aim: To comparatively investigate the most suited sputum pre-extraction approach for metabolomics investigations using patient collected sputum, prior to previously investigated total metabolome extraction procedure, followed by GCxGC-TOFMS analysis.

Note: During the course of this study, a comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometer (GCxGC-TOFMS) was purchased by the Centre for Human Metabonomics. Due to its superior metabolite resolution capabilities and increased analytical sensitivity through solute band re-concentration,

the GCxGC-TOFMS is far better suited for the analyses of complex biological samples, such as sputum, than the single dimensional triple quadrupole system previously used. The enhanced capacity of the GCxGC-TOFMS for generating comparatively more comprehensive biosignatures and potentially better metabolite marker identification, makes it more suitable for metabolomics investigations (Mondello *et al*., 2007), and therefore, this apparatus was used in the remaining investigations of this study. Although this research apparatus is relatively new to the larger science community, GCxGC-TOFMS metabolomics has already been used successfully to characterise a variety of complex biological samples (Hope *et al.*, 2005; Welthagen, *et al.*, 2005; Mohler *et al.*, 2006).

2. MATERIALS AND METHODS

2.1 Reagents and chemicals

NaOH, methoxyamine hydrochloride, Sputolysin, and 3-phenyl butyric acid were purchased from Sigma Aldrich (St. Louis, MO, USA). N-acetyl-L- cysteine (NALC), pyridine, TMCS and MSTFA, were purchased from Merck (Darmstadt, Germany). All organic solvents used were ultra pure Burdick & Jackson brands (Honewell International Inc., Muskegon, MI, USA).

2.2 Sputum samples

Sputum samples from 95 patients suspected of having TB, based on a medical assessment of the symptoms associated with the disease, were collected from various South African clinics, using standard sputum collection procedures, and sent to a centralised national laboratory (transported on ice) where standard diagnostic procedures, including both Ziehl-Neelsen staining and bacteriological culture, were performed. After the required amounts of sample material was removed for the completion of these diagnostic tests, the remaining proportion of these samples were immediately frozen (-80ºC) and transported to the North-West University (NWU), Centre for Human Metabonomics, for the metabolomics analyses pertaining to this study. Anonymity was ensured by the allocation of a unique code to each sample prior to this transport. Clinical information received with regards to these anonymous samples included: HIV status (only for those samples where the patient requested a

HIV test to be done); age, and gender. Collected sputum samples were kept at -80°C until further metabolomics analysis.

In order to accomplish the aims of this chapter, a sputum sample pool was made by combining equal volumes of only a small fraction of each of the TB-negative sputum samples, of which half the volume was spiked with cultured *M. tuberculosis* (1 X 10^8 bacteria per 250 µL sputum) and the other half was used as is, serving as a pooled TB-negative control. The individually collected TB-positive patient sputum samples, and the remaining fractions of each of the patient collected TB-negative sputum samples, were stored at -80°C for use in the analysis of Chapter 7.

In order to compare the four sputum pre-extraction preparation methods, 4 x 1.75 mL aliquots of both the spiked and control sputum pools were prepared and processed using each of the four methods, prior to extraction and GCxGC-TOFMS analyses.

For the purpose of investigating the detection limit or minimum amount of cells per volume sputum required for these metabolomics methods, 250 µL aliquots of the control sputum pool were spiked with various amounts of isolated *M. tuberculosis,* in order to create a sample concentration range of 1 x 10^8, 1 x 10^5, and 1 x 10^3 bacteria mL^{-1} (n = 6). Six un-spiked aliquots (1 x 10^0 bacteria mL^{-1}) of the control sputum pool served as a blank.

2.3 Bacterial cultures

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

2.4 Sputum pre-extraction preparation methods

2.4.1 NaOH sputum pre-extraction preparation method (Petroff *et al.*, 1915)

In this method, NaOH was added in a 2:1 v/v ratio to 1.75 mL of the spiked and control sputum pools respectively, followed by vortex mixing and incubation for 15 minutes at room temperature. Cells were then isolated from 7 x 750 µL sample repeat aliquots (each containing 250 µL of the original sputum) of the spiked and control pre-processed sputum pools via centrifugation for 15 min at 550 x g, after

which the supernatant was discarded and the pellet washed twice with 1 mL ddH_2O, prior to extraction via the total metabolome extraction procedure, as described in Chapter 5 section 2.4.

2.4.2 Sputolysin sputum pre-extraction preparation method (Alugupalli *et al.,* 1993)

Sputolysin was added in a 1:1 v/v ratio to 1.75 mL of both the spiked and control sputum pools respectively, followed by vortex mixing and incubation for 15 minutes at room temperature. Cells were then isolated from 7 x 500 µL sample repeat aliquots (each containing 250 µL of the original sputum) of the spiked and control pre-processed sputum pools via centrifugation for 15 min at 550 x g, after which the supernatant was discarded and the pellet washed twice with 1 mL ddH_2O, prior to extraction via the total metabolome extraction procedure, as described in Chapter 5 section 2.4.

2.4.3 NALC-NaOH sputum pre-extraction preparation method (Kubica *et al.,* 1967)

In this method, 0.5 N NaOH and 20% w/v NALC was added in a 1:1 v/v ratio to 1.75 mL of the spiked and control sputum pools respectively, followed by vortex mixing and incubation for 15 minutes at room temperature. Cells were then isolated from 7 x 500 µL sample repeat aliquots (each containing 250 µL of the original sputum) of the spiked and control prepared sputum pools via centrifugation for 15 min at 550 x g, after which the supernatant was discarded and the pellet washed twice with 1 mL ddH_2O, prior to extraction via the total metabolome extraction procedure, as described in Chapter 5 section 2.4.

2.4.4 Ethanol homogenisation

In this method, samples were simply homogenised and extracted, without prior cell isolation. In this instance, 45% ethanol was added in a 2:1 v/v ratio to 1.75 mL of the spiked and control sputum pools respectively. These sample mixtures were then homogenised by shaking them in an MM 400 vibration mill (Retsch GmbH & co. KG, Haan, Germany) at 50 Hz for 2 min. Hereafter, 7 x 750 µL sample repeat aliquots (each containing 250 µL of the original sputum) of the homogenised spiked and control sputum pools were completely dried in a speedvac, prior to extraction via the total metabolome extraction procedure, as described in Chapter 5 section 2.4.

2.5 Extraction procedure

As internal standard, 150 µL of 3-phenyl butyric acid (0.525 mg mL^{-1}) was added to the above-mentioned processed samples, followed by analyses via the total metabolome extraction procedure, as described in Chapter 5 section 2.4.

2.6 GCxGC-TOFMS parameters

Chromatographic analyses of derivatised samples were done in a two-dimensional mode on an Agilent 7890A GCxGC (Agilent, Atlanta, GA) coupled to a time of flight mass spectrometer (TOFMS) (Leco Corporation, St. Joseph, MI, USA) equipped with a Gerstel Multi Purpose Sampler (MPS) (Gerstel GmbH & co. KG, Eberhard-Gerstel-Platz 1, D-45473 Mülheim an der Ruhr). Samples (1 µL) were randomly injected at split ratio of 15, and helium was used as the carrier gas at a constant flow rate of 1 mL min^{-1}. The injector temperature was held constant at 270°C for the entire run. A Restek Rxi-5Sil MS capillary column (30m, 0.25 mm i.d., 0.25 µm d.f.) served as the primary column and compound separation was achieved by programming the primary oven at an initial temperature of 70°C for 2 min, followed by an increase of 4°C min^{-1} to a final temperature of 300°C, at which it was maintained for a further 2 min. A Restek Rxi-17 (1 m, 100 µm i.d., 0.1 µm d.f.) column served as the column for the second dimensional separation of the compounds in the analysed samples. The secondary oven was programmed with an identical temperature gradient to that of the primary column, only with an offset of + 5°C. Cryomodulation and a hot pulse of nitrogen gas of 0.7 sec, every 3 seconds, was used to control the effluent emerging from the primary column onto the secondary column. No mass spectra were recorded for the first 550 sec of each run, as this period was considered a solvent delay. The transfer line was held at a constant 280°C and the ion source temperature at 200°C, for the entire run. The detector voltage was 1700 V and the filament bias -70 eV. Mass spectra were collected at an acquisition rate of 200 spectra's per second from 50-550 *m*/*z*.

2.7 Peak identification and alignment

Leco Corporation ChromaTOF software (version 4.32) was used for peak finding and mass spectral deconvolution at a S/N ratio of 300, with a minimum of 3 apexing peaks. Using the mass fragmentation patterns generated by the MS, together with their respective GC retention times, the identities of these peaks were determined using libraries generated from previously injected standards. In order to eliminate the effect of retention time shifts and create a data matrix containing the relative concentrations all the compounds present in all the samples investigated, peaks with similar mass spectra and retention times were aligned prior to statistical data analyses, using an optional function of the ChromaTOF software: Statistical Compare. Peak areas were normalised relative to the internal standard. As previously mentioned, this is done in order to eliminate any variation which may occur due to irregularities during the extraction process or injection of the sample onto the GCxGC-TOFMS.

2.8 Statistical data analysis

The repeatability of the four sputum pre-extraction preparation methods investigated were compared using the distribution of the CV values calculated from the relative concentrations of all the compounds detected subsequent to GCxGC-TOFMS analysis of the extracted sample repeats.

The remaining statistical data analysis was done using MetaboAnalyst, a web server for metabolomics data analysis and interpretation based on the statistical program "R" version 2.10.0 (Xia *et al.*, 2009). Data were pre-treated using the log transformation function, prior to mean centering, and those compounds which were not detected in at least 50% of the samples, in one or more of the sample groups, together with those compounds with no detected variation between the groups, were removed. All missing values, i.e. those compounds which were not detected in a specific sample, were replaced with small values (half of the minimum positive value in the original data), assuming that most missing values were a reflection of low

abundance metabolites (i.e. those below the detection limit). The log transformation function, as opposed to the non-parametric transformation function used in the previous chapters, was applied for data pre-treatment as it led to a better discrimination between the various sample groups when doing the PCA. In order to determine whether or not a natural differentiation exists between experimental groups, PCA was performed. To test the ability of this approach to identify biologically relevant metabolite markers for the characterisation of *M. tuberculosis* infection from complex sputum samples, the supervised multivariate method, PLS-DA, was used to rank the metabolites according to the VIP parameter, as previously described (Chapter 3 section 2.6).

3. RESULTS AND DISCUSSION

3.1 Repeatability

In order to compare the four sputum pre-extraction preparation methods, the repeatability of each was determined by calculating the CV values (using the relative concentrations) of all the compounds detected after GCxGC-TOFMS analysis of the extracted sample repeats. Figure 6.1 graphically depicts the distribution of these CV values for all the compounds detected, subsequent to chromatographic analyses and extraction of the samples prepared by the four sputum pre-extraction procedures.

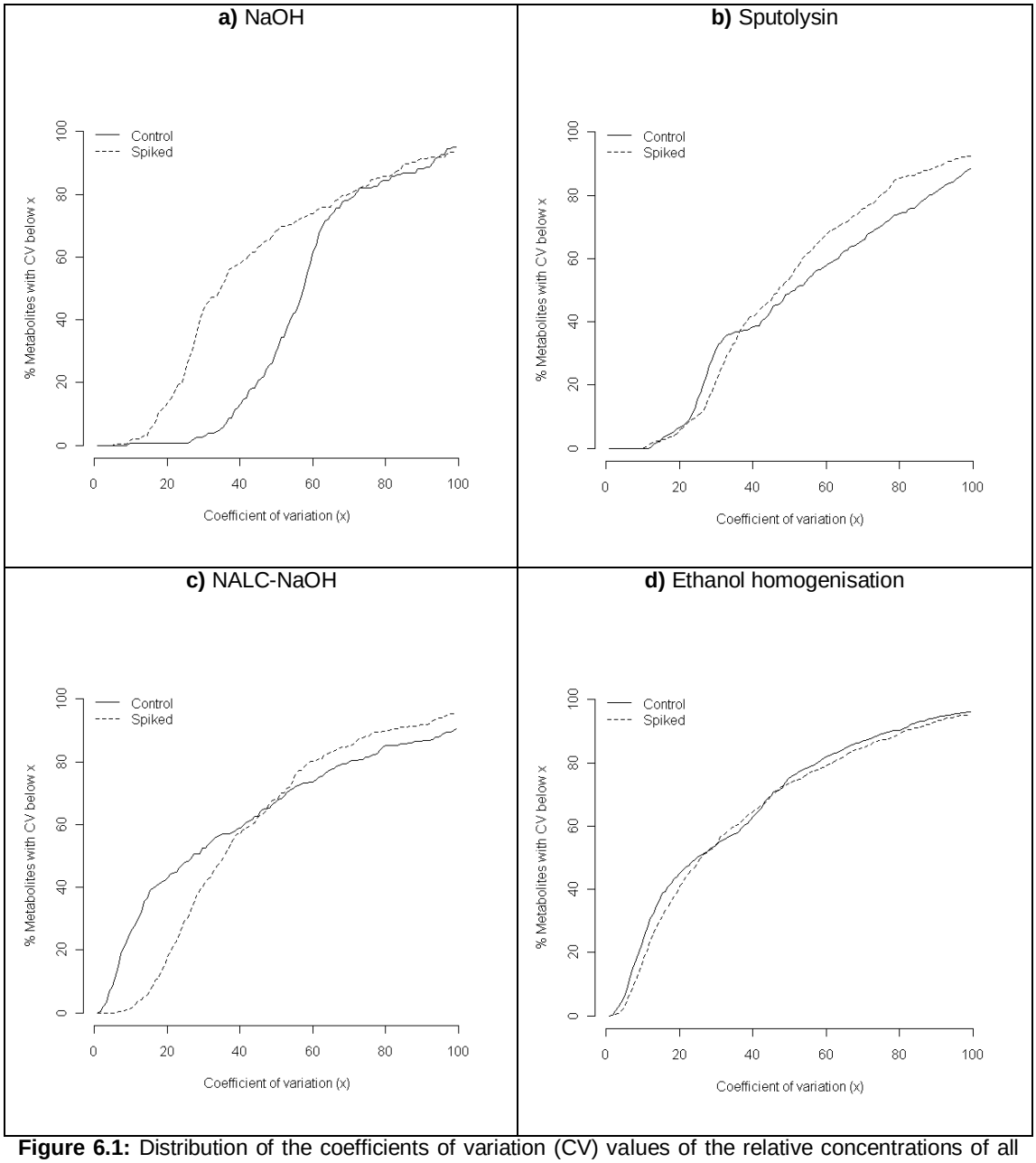


Figure 6.1: Distribution of the coefficients of variation (CV) values of the relative concentrations of all the compounds detected in the control and *M. tuberculosis* spiked sputum samples subsequent to sputum pre-extraction preparation using one of the four investigated methods, followed by extraction and GC x GC-TOFMS analysis.

As the concentrations of the detected compounds diminish and fall to levels closer to that of the detection limit of the analytical technique, extraction and analysis becomes less repeatable, consequently resulting in larger CV values (Rocke and Lorenzato,

1995). Comparatively, an average of 74.7%, 68.1%, 52% and 50%, of all the compounds detected from the spiked and control sputum sample repeats, using the ethanol homogenisation, NaLC-NaOH, Sputolysin and NaOH pre-extraction preparation methods respectively, fell below a 50% CV value (Table 6.1), indicating that the ethanol homogenisation is by far the most repeatable of these methods for the majority of the compounds detected. For further comparative purposes, using more conventional CV interpretations, the calculated CV values for ten compounds, representative of various compounds classes and detected at retention time intervals throughout the total chromatographic run, are also given in Table 6.1. The average CV values for these compounds confirm the previously determined hierarchical arrangements of the compared sputum pre-extraction preparation methods.

Furthermore, 4 of the 14 samples analysed were regarded as outliers for the NaOH pre-extraction preparation method, in comparison to only 1 outlier for the Sputolysin and NALC-NaOH methods and none using ethanol homogenisation (Table 6.1). Outlier samples were identified based on indiscretions in the GCxGC-TOFMS chromatograms, possibly due to irregularities which occurred during the sputum processing, making the data of these samples unusable.

Table 6.1: Comparative calculations including: coefficients of variation (CV) values, number of compounds detected and number of sample outliers identified, for the four sputum pre-extraction preparation methods.

		Sputum pre-extraction preparation method					
		Retention time (sec)		NaOH	Sputolysin	NALC-NaOH	Ethanol homogenisation
		1st dim	2nd dim				
CV values (%) calculated for the relative concentrations of compounds detected in the control samples	3-Penten-2-one	443	0.92	42.78	48.64	35.56	28.14
	Propanoic acid	992	1.22	59.47	81.46	25.36	36.45
	L-Proline	1613	1.70	73.73	106.88	49.23	36.52
	1,4-Butanediamine	1979	1.10	64.02	115.66	66.96	20.84
	Cadaverine	2132	1.09	64.07	114.4	21.86	12.42
	Tetradecanoic acid	2153	1.25	61.29	48.15	32.16	21.35
	Hexadecanoic acid	2435	1.32	82.16	47.78	54.99	22.08
	Myo-Inositol	2492	1.12	99.29	111.92	88.24	32.54
	9,12-Octadecadienoic acid	2651	1.44	ND	55.40	81.42	32.04
	2-Monopalmitin	3056	1.35	ND	80.89	110.37	31.66
% of all compounds with CV values below 50%	Control samples			31.3	49.6	67.9	75.6
	Spiked samples			68.7	54.4	68.3	73.7
	All samples			50.0	52.0	68.1	75.6
Number of peaks identified (standard deviation)	Control samples			138 (9)	395 (21)	210 (18)	731 (56)
	Spiked samples			168 (7)	465 (20)	290 (23)	833 (24)
Number of samples identified as outliers (n=14)				4 (28%)	1 (7%)	1 (7%)	0 (0%)
Average internal standard area for all analysed samples (standard deviation)				43859740 (39138750)	31484056 (21512867)	39560266 (14061124)	70693517 (17581883)

3.2 Extraction efficiency

Although repeatability for metabolomics investigations is a major consideration for comparative method selection, extraction efficiency (the number of compounds and their intensity of detection), also plays a significant role in this selection.

Consequently, each of the four sputum pre-extraction preparation procedures were compared with reference to the number and intensity of the peaks detected after extraction and GCxGC-TOFMS analysis, using the pooled control sample repeats as an example. Comparatively, the ethanol homogenisation procedure resulted in the extraction and detection of the most compounds (731 ± 56), followed by the Sputolysin (395 ± 21), NALC-NaOH (210 ± 18), and NaOH procedure (138 ± 9), respectively (Table 6.1). A similar comparison of the *M. tuberculosis* spiked sputum pool samples indicated that, apart from the expected larger number of peaks due to the presence of *M. tuberculosis*, an identical ranking of these methods' extraction capacities to that determined using in the control pooled sample repeats occurred, confirming the previous hierarchy of these methods. The ethanol homogenisation would be expected to result in comparatively more extracted compounds, as it allows for the analysis of all compounds in the sputum, including the bacterial cells present, the metabolites excreted by these bacteria, and those excreted by the host, which, from a metabolomics perspective, is an advantage.

In order to compare the intensities at which compounds are extracted and detected after sputum processing using the four pre-extraction preparation methods, the average area of the internal standard, detected via each method, is given in Table 6.1. These areas for each of the four methods clearly indicate that, although the same amount of internal standard (150 µL of 3-phenyl butyric acid (0.525 mg mL^{-1})) was added to all the analysed samples, when using the ethanol homogenisation sputum pre-extraction preparation method, it was detected at almost double the intensity, in comparison to the other sputum pre-extraction preparation methods.

3.3 Capacity to extract those compounds differentiating the investigated sample groups

When considering the application of the investigated methods to comparative metabolomics, not only is the repeatability and extraction capacity, as described above, of importance, but also the methods ability to extract those compounds most important for the differentiation of the groups being compared, in order to answer the biological question at hand. To determine the capacity of the above described sputum pre-extraction procedures for isolating those compounds most important for differentiating the control from the *M. tuberculosis* spiked sputum, a PCA was

performed on the data matrixes generated after peak alignment with ChromaTOF Statistical Compare, for each method (Figure 6.2).

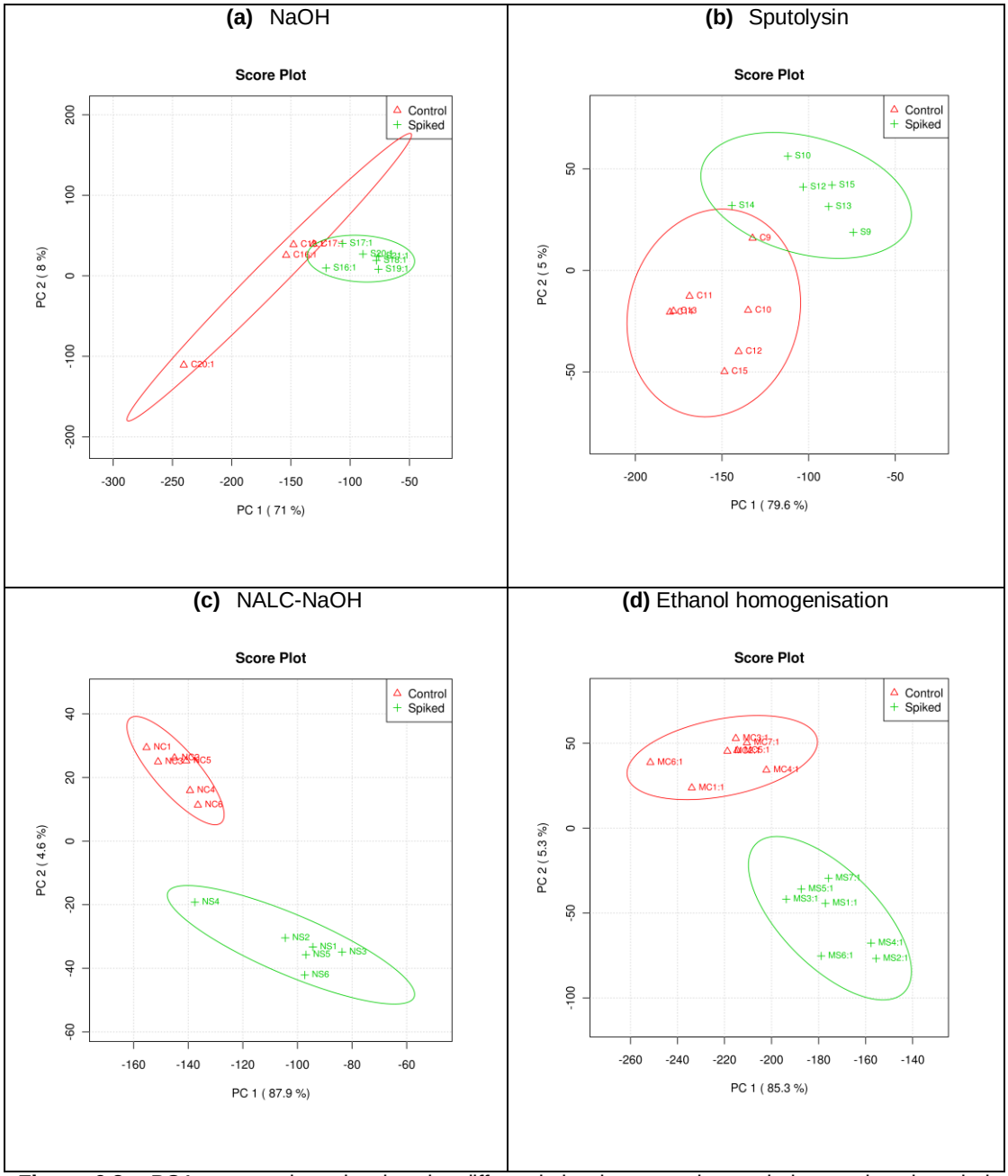


Figure 6.2: PCA scores plots showing the differentiation between the pooled control and pooled spiked sputum sample repeats using the GCxGC-TOFMS generated data of the analysed extracts obtained via the four sputum preparation methods. In each case, three PCs were extracted. The cumulative variance explained by each PC is indicated in parenthesis.

As can be seen in Figure 6.2a, the PCA scores plot generated from the data obtained after NaOH sputum preparation, extraction and GCxGC-TOFMS analyses, resulted

in an overlap of the pooled control and *M. tuberculosis* spiked sample groups. This overlap may be ascribed to: 1) the comparatively poor repeatability of the method previously described, and; 2) the fact that the high concentration of this decontaminant may cause bacterial cells lyses, leading to inconsistent cell loss during the cell isolation step. The PCA scores plot using the data obtained after Sputolysin sputum pre-extraction preparation, extraction and GCxGC-TOFMS analyses (Figure 6.2b), also showed a marginal overlap between the 2 sample groups, most probably due to similar reasons as described above. The data obtained after NALC-NaOH sputum pre-extraction preparation (Figure 6.2c) and ethanol homogenisation (Figure 6.2d), prior to extraction and GCxGC-TOFMS analyses, however, resulted in a complete differentiation of the control and *M. tuberculosis* spiked sputum sample repeats into their respective groups.

Considering these results, the NALC-NaOH and ethanol homogenisation pre-extraction preparation methods showed the best repeatability, extraction capacity (considering the number and intensity of the extracted compounds), and ability to extract those compounds which best differentiate the control from the *M. tuberculosi*s spiked sputum, using PCA. Therefore, the evaluation of the extracted compounds from a biological perspective and a comparison of the detection limits (minimal sample requirements) of only these two methods were further investigated.

3.4 Metabolite marker identification

PLS-DA was performed using the generated GCXGC-TOFMS data of the extracted sample repeats, after sputum pre-extraction preparation using the NALC-NaOH and ethanol homogenisation methods. This was done purely for the purpose of identifying those compounds which differ most between the two groups, in order to validate these from a biological perspective, evaluating these methodologies for extracting those compounds which not only differentiate the groups, but also characterise *M. tuberculosis* infection. These metabolites were subsequently ranked according to their VIP values, hence describing their importance in differentiating the control sputum from the *M. tuberculosis* spiked sputum samples. It is important to note that these markers are representative of the physical presence of *M. tuberculosis* cells spiked into the TB-negative sputum samples and does not reflect changes in the host metabolism due to a TB disease state. The identification of these compounds may, however, later assist when applying this methodology to the

collected TB-positive and TB-negative patient sputum, in order to determine whether the identified metabolite markers are a reflection of the direct presence of the *M. tuberculosis* in these samples, or if they represent TB induced alterations to the normal host metabolism.

Table 6.2 indicates the top 20 ranked metabolite markers, as identified using the NALC-NaOH pre-extraction preparation method.

Table 6.2: The top 20 ranked VIPs identified via PLS-DA, using the data obtained after NALC-NaOH sputum preparation, extraction and GCxGC-TOFMS analysis of the control and *M. tuberculosis* spiked sputum samples.

Rank	Compound	Control Sputum		Spiked sputum	
		Concentration (mg mL^{-1} sputum) ± SD	Samples detected (n = 7)	Concentration (mg mL^{-1} sputum) ±SD	Samples detected (n = 6)
1	C19:0	0	0 (0%)	1.737 ± 0.812	6 (100%)
2	TBSA	0	0 (0%)	0.797 ± 0.39	6 (100%)
3	C26:0	0	0 (0%)	0.175 ± 0.087	6 (100%)
4	Acetohydroxamic acid	0	0 (0%)	0.474 ± 0.161	5 (83.3%)
5	Myo-inositol	0.02 ± 0.004	3 (43%)	0.21 ± 0.116	6 (100%)
6	á-D-glucopyranoside	0.83 ± 0.087	2 (29%)	0.741 ± 0.368	6 (100%)
7	C20:0	0	0 (0%)	0.103 ± 0.038	6 (100%)
8	D-glycero-L-manno-heptonic acid	0	0 (0%)	0.099 ± 0.043	6 (100%)
9	C16:1 ω6c	0.017 ± 0.003	3 (43%)	0.153 ± 0.092	6 (100%)
10	Furan	0	0 (0%)	0.096 ± 0.062	6 (100%)
11	C24:0	0	0 (0%)	0.08 ± 0.035	6 (100%)
12	C18:0	4.182 ± 1.507	7 (100%)	5.287 ± 2.087	6 (100%)
13	C17:0	0.056 ± 0.157	7 (100%)	0.489 ± 0.207	6 (100%)
14	C22:0	0	0 (0%)	0.039 ± 0.018	6 (100%)
15	C20:4 ω6c	0	0 (0%)	0.078 ± 0.023	5 (83.3%)
16	D-glucose	0	0 (0%)	0.033 ± 0.016	6 (100%)
17	C18:1 ω9c	0.136 ± 0.003	2 (29%)	0.257 ± 0.13	6 (100%)
18	D-mannose	0	0 (0%)	0.03 ± 0.014	6 (100%)
19	D-galactose	0	0 (0%)	0.056 ± 0.019	5 (83.3%)
20	Propane	0	0 (0%)	0.169 ± 0.062	4 (66.6%)

All of the identified markers were detected in elevated concentrations in the spiked sputum samples, highlighting the ability of this approach to detect changes in the sputum metabolome due to the presence of *M. tuberculosis*. Furthermore, 11 of

these top 20 markers are fatty acids, of which 7, including the well known *Mycobacterium* biomarker TBSA (Stopforth *et al.*, 2004), nonadecanoic acid (C19:0) (Larsson *et al*., 1987) and several mycolic acid cleavage products (MACPs), were exclusively detected in the *M. tuberculosis* spiked sputum samples. The identification of these compounds confirms our findings in Chapters 3 and 5, further proving the consistency and repeatability of this metabolomics approach. As explained in the previous chapters, mycolic acids (long-chain, high-molecular-weight α-alkyl, β-hydroxyl fatty acids) are major components of the mycolylarabinogalactan (mAG) cell wall skeleton of *M. tuberculosis*. During GC analyses, these mycolic acids are heat cleaved in the injection port of the GC at injector temperatures exceeding 235°C, leading to the detection of a variety of these MACPs including: hexacosanoic acid (C26:0); eicosanoic acid (C20:0); tetracosanoic acid (C24:0) and; docosanoic acid (C22:0) (Guerrant *et al*., 1981; Lambert *et al*, 1986), exclusively in the *M. tuberculosis* spiked sputum samples. Furthermore, although palmitoleic acid (C16:1 ω7c), heptadecanoic acid (C17:0), and oleic acid (C18:1 ω9c), were also detected in the control samples, these fatty acids were detected in relatively elevated concentrations in the spiked samples, confirming previous studies indicating that these compounds form part of the most abundant fatty acids in *M. tuberculosis* (Lambert *et al*, 1986; Chapter 3). Additionally, myo-inositol was detected exclusively in the *M. tuberculosis* spiked sputum samples, and is known to play an important role in the synthesis of phosphatidyl inositol, a structural component of the *M. tuberculosis* cell wall (Brown *et al.,* 2007).

Table 6.3 represents the top 20 ranked metabolite markers, as identified by PLS-DA, using the GCxGC-TOFMS analysed data after extraction of the ethanol homogenised sputum samples. Due to the lack of a cell isolation step prior to extraction, the metabolite markers identified using this method, are different to those detected using the NALC-NaOH method. Once again, MACPs (C26:0, C22:0, and C24:0) were detected as markers for the *M. tuberculosis* spiked sputum sample group. The MACPs and fatty acid markers detected using this methodology also correspond to those metabolite markers identified previously for the *M. tuberculosis* culture isolates analysed in Chapter 3 and 5. The identification of these markers further validates the capacity of this metabolomics approach for identifying *M. tuberculosis* metabolites, despite potential background "noise" extracted from the complex sputum matrix. The majority of identified markers however, when using the ethanol homogenisation pre-extraction preparation method, were monosaccharides, in either their chain or ring

configurations, in comparison to only 4 monosaccharides (glucopyranoside, d-glucose, d-mannose, and d-galactose) detected in elevated concentrations in the *M. tuberculosis* spiked sputum samples using the NALC-NaOH method.

Table 6.3: The top 20 ranked VIPs identified via PLS-DA, using the data obtained after ethanol homogenisation, extraction and GCxGC-TOFMS analysis of the control and *M. tuberculosis* spiked sputum samples.

Rank	Compound	Control Sputum		Spiked sputum	
		Concentration (mg ml^{-1} sputum) ± SD	samples detected (n=7)	concentration (mg ml^{-1} sputum) ± SD	Samples detected (n=7)
1	D-glucosamine	0	0	1.109 ± 0.476	7 (100%)
2	Glycerol	12.134 ± 1.403	7 (100%)	39.635 ± 8.444	7 (100%)
3	Uridine	0.306 ± 0.164	7 (100%)	1.302 ± 0.567	7 (100%)
4	2-O-glycerol-à-d-galactopyranoside	0.158 ± 0.096	7 (100%)	0.988 ± 0.242	7 (100%)
5	C26:0	0	0	0.344 ± 0.106	7 (100%)
6	D-galactose	0	0	0.339 ± 0.187	7 (100%)
7	á-D-galactopyranoside	0.054 ± 0.007	7 (100%)	0.345 ± 0.097	7 (100%)
8	á-D-xylopyranose	0	0	0.192 ± 0.131	7 (100%)
9	Arabinofuranose	0.319 ± 0.064	7 (100%)	0.51 ± 0.068	7 (100%)
10	D-erythro-pentitol	0	0	0.138 ± 0.079	7 (100%)
11	á-D-galactofuranose	0.008 ± 0.013	7 (100%)	0.229 ± 0.116	7 (100%)
12	D-glucose	1.02 ± 0.334	7 (100%)	5.289 ± 1.329	7 (100%)
13	Phenylethanolamine	0.658 ± 0.046	7 (100%)	0.726 ± 0.138	7 (100%)
14	D-galactose,	2.112 ± 3.43	7 (100%)	4.756 ± 1.682	7 (100%)
15	C22:0	0	0	0.062 ± 0.027	7 (100%)
16	Cadaverine	10.955 ± 0.981	7 (100%)	13.071 ± 1.689	7 (100%)
17	C24:0	0.124 ± 0.001	1 (14.3%)	0.149 ± 0.030	7 (100%)
18	L-threonine	2.428 ± 0.279	7 (100%)	3.323 ± 0.629	7 (100%)
19	D-fructose	0.417 ± 0.188	7 (100%)	0.714 ± 0.117	7 (100%)
20	Myo-inositol	1.368 ± 0.186	7 (100%)	2.441 ± 0.531	7 (100%)

The increased amounts of monosaccharides detected in the *M. tuberculosis* spiked sputum samples may be ascribed to several characteristic components of the cell wall of the *M. tuberculosis* present in these samples, including: mycolylarabinogalactan (mAG), comprising of a mycolic acid cluster, arabinan, and D-galactan; and lipoarabinomannan (LAM), a complex lipoglycan with the main monosaccharide components being arabinose and mannose. Several other glycolipids (Aspinall *et al.,* 1995) and polysaccharides including: mannan, α-1,4-glucan, and arabinomannan (Chatterjee, 1997), have also been identified as

elements of the mycobacterial cell wall. Furthermore, the peptidoglycan found in these organisms consists of *N*-glycolylmuramic (a d-glucosamine polymer) and *N*-acetylglucosamine (Chatterjee, 1997).

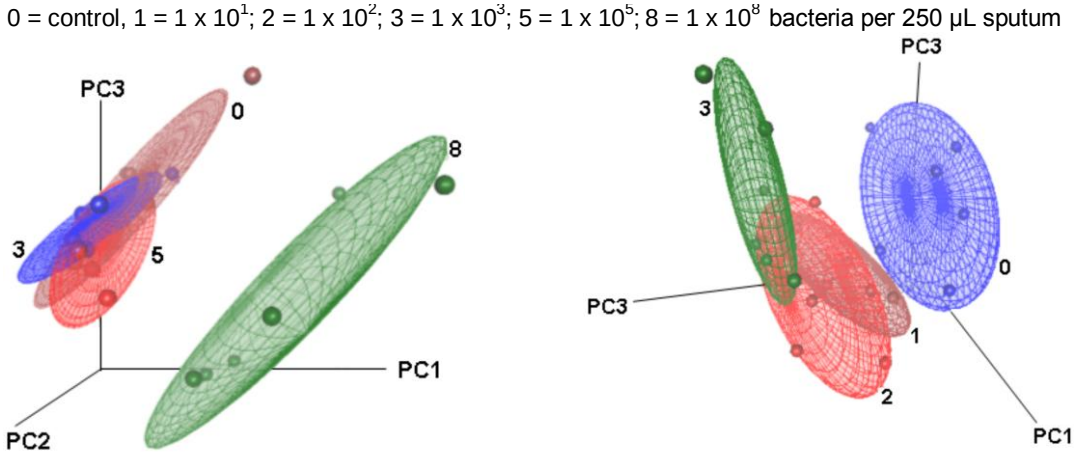
Although all the above mentioned factors may contribute to the increased amounts of monosaccharides detected in the *M. tuberculosis* spiked sputum samples, these compounds were also detected in lower amounts in the control sputum samples. In order to interpret this occurrence, one should note that the TB-negative samples used to make up the control sputum pool were not obtained from healthy controls, but from TB-negative patients with a suspicion of TB (due to the manifestation of similar symptoms), most likely suffering from a pulmonary infection caused by other bacteria. Thornton *et al.* (1991) reported relatively large amounts of various carbohydrates including: N-acetylgalactosamine, N-acetylglucosamine, galactose, fucose, and N-acetylneuraminic, to be present in the sputum of cystic fibroses (CF) patients. Due to the fact that sputum from patients with other lung diseases such as; chronic bronchitis, asthma and bronchiectasis, have similar properties to that of CF patients, the presence of these carbohydrates in the TB-negative samples analysed, could be expected. Therefore, we feel it safe to hypothesise that the elevated monosacharides detected in the *M. tuberculosis* spiked sputum samples, with respect to the control samples, may be due to the higher number of carbohydrate containing *M. tuberculosis* cell wall products present in these samples, in addition to the monosaccharides naturally occurring in the control sputa. Furthermore, confirming previous results obtained from the NALC-NaOH preparation method, myo-inositol was once again detected in comparatively elevated concentrations in the *M. tuberculosis* spiked sputum samples.

From these results, it is clear that both the NALC-NaOH sputum pre-extraction preparation method and the ethanol homogenisation pre-extraction preparation method, although different to one another, each have the capacity to allow for the extraction and detection of biologically relevant metabolite markers. However, it is once again important to mention that this marker identification step was done exclusively for method validation purposes, and the variation of a human population and the consequent variation in the content of sputum collected from these individuals, may have an impact on the outcome of these markers, and consequently, their validation using patient collected TB-positive and TB-negative sputum, is required. By doing so, we may further verify the markers detected in this chapter, due to the presence of *M. tuberculosis*, but may also identify markers due to

alterations in the human metabolome as a result of TB infection, consequently better characterising the disease state, and not only the disease causing bacteria. The latter investigation will subsequently be done in Chapter 7, using the developed methodology.

3.5 Detection limits and minimum sample volumes required

When developing an analytical approach, the verification of the minimum sample amounts required for the methods investigated, is also a necessity. Consequently, the detection limit for the NALC-NaOH and ethanol homogenisation sputum pre-extraction preparation methods was determined. For these analyses, the previously prepared control sputum pool was spiked with varying amounts of the isolated *M. tuberculosis* cells, in order to make up a concentration gradient of a 1×10^0 (control sputum without any *M. tuberculosis* added), 1×10^3, 1×10^5, and 1×10^8 bacteria per 250 µL sputum (6 samples of each). As described above, PCA analysis of the GCxGC-TOFMS analysed data of the extracted samples, following sputum pre-extraction preparation using the NALC-NaOH and ethanol homogenisation methods, was done in order to determine the smallest amount of cells required by each method for differentiating the *M. tuberculosis* spiked sputum samples from the controls. Considering this, the sample group containing the lowest amount of spiked *M. tuberculosis*, not overlapping with the control samples on a PCA scores plot, was considered as the detection limit for the specific method for metabolomics applications. Figure 6.3a represents the PCA scores plot for the GCxGC-TOFMS data of the samples prepared using the NALC-NaOH method, prior to extraction. From this result it is evident that a minimum amount of 1×10^8 cells is required in a sample in order to extract enough biologically relevant differentiating metabolites to differentiate those samples containing *M. tuberculosis* from those that don't. Therefore, considering that the vast majority of sputum samples collected from TB-positive patients in South Africa has concentrations of *M. tuberculosis* well below 1×10^8 cells mL^{-1} sputum, sputum volumes larger than 250 µL would be required when applying this methodology.



a) NALC-NaOH **b)** Ethanol homogenisation

Figure 6.3: PCA scores plot (PC 1 vs. PC 2 vs. PC 3) showing the concentration gradient analysis of the *M. tuberculosis* spiked sputum, comparing the (a) NALC-NaOH and (b) ethanol homogenisation sputum pre-extraction preparation methods. The lowest bacterial concentration not overlapping with the control was considered the detection limit or minimum amount of *M. tuberculosis* required to be present in a sputum sample for extraction and GCxGC-TOFMS analysis of the characteristic metabolites required for differentiation of TB-positive and TB-negative sputa, using this metabolomics research approach.

Considering the results obtained using the GCxGC-TOFMS data collected from the ethanol homogenisation prepared sputum extracts, none of the spiked sputum sample concentration gradient groups overlapped with the controls (results not shown). Consequently, two additional concentration gradient sample groups ($1x10^1$ and $1x10^2$ cells per 250 µL sputum) were prepared. From the subsequent PCA scores plot (Figure 6.3b), it is clear that even the sample group consisting of $1x10^1$ bacterial per 250 µL sputum, can be differentiated from the control sample group, i.e. only 10 cells are required for detecting metabolites potentially differentiating TB-positive and TB-negative sputum sample groups. It is important, however, not to over interpret this data. This is the detection limit under ideal conditions, without any variation, and simply gives a comparative estimation of the minimum amount of cells required for successful metabolomics applications. As described in Chapter 3, this doesn't necessarily mean that this would be considered the most optimal number of cells required for detecting the most metabolite information for characterising these species using metabolomics. This detection limit is simply an indication of the minimum sample material required for extracting the minimum metabolite information for differentiating the groups. When used on patient collected sputum samples for

characterisation purposes (Chapter 7), one would inherently use as much sample material as analytically viable, in order to extract the maximum amount of metabolite information for characterising the disease state.

4. CONCLUSIONS

According to Lee *et al.* (2001), no one dimensional GC separation technique has the ability to resolve more peaks in a single analytical run than GCxGC. This separation, together with the high acquisition rate of the TOFMS, makes this analytical technique ideal for analysing samples as complex as sputum, using an extraction method as described in this study.

When comparing the four different sputum pre-extraction preparation methods for application to metabolomics, the NALC-NaOH method proved to have the best extraction efficiency, repeatability and capacity to extract those compounds best characterising and differentiating the *M. tuberculosis* spiked samples from the controls, when evaluating it in comparison to the two other cell isolation methods used (Sputolysin and NaOH). The ethanol homogenisation sputum pre-extraction preparation method, however, outperformed all the tested cell isolation methods, considering the aforementioned criteria, and was additionally able to differentiation between the *M. tuberculosis* spiked sputum and control sputum sample groups using only 10 cells. Furthermore, the latter method does not only have the capacity to identify characteristic metabolite markers due to the physical presence of *M. tuberculosis* in the patient collected sputum, but also those metabolites characteristic of an altered TB induced host metabolism, which when interpreted using a metabolomics research approach, may give additional clues to the impact of this disease on the human metabolome.

Considering this, the ethanol homogenisation pre-extraction preparation method evaluated in this chapter, in conjunction with the previously investigated total metabolome extraction method, followed by GCxGC-TOFMS analysis, and univariate and multivariate statistical processing of the generated data, was applied as the metabolomics approach for metabolite marker identification using the 95 patient sputum samples collected for this purpose, in Chapter 7.

5. REFERENCES

ALUGUPALLI, S., OLSSON, B., LARSSON, L. 1993. Detection of 2-Eicosanol by gas chromatography-mass spectrometry in sputa from patients with pulmonary mycobacterial infections. *J. Clin. Microbiol.* 31:1575-8.

ASPINALL, G.O., CHATTERJEE, D., BRENNAN, P.J. 1995. The variable surface glycolipids of mycobacteria: structures, synthesis of epitopes, and biological properties. *Adv. Carbohydr. Chem. Biochem.* 51:169-242.

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 Mg^{2+} *BMC Struct. Biol.* 7:55.

CHATTERJEE, D. 1997. The mycobacterial cell wall: structure, biosynthesis and sites of drug action. *Curr. Opin. Chem. Biol*. 1:579-568.

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.

HOPE, J.L., PRAZEN, B.J., NILSSON, E.J., LIDSTROM, M.E., SYNOVEC, R.E. 2005. Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry detection: analysis of amino acid and organic acid trimethylsilyl derivatives, with application to the analysis of metabolites in rye grass samples. *Talanta.* 65:380–388.

KUBICA, G.P. 1967. Sputum digestion and decontamination with N-Acetyl-L-Cysteine-Sodium Hydroxide for culture of Mycobacteria. *Am. Lung. Assoc*. 87:775-779.

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.

LEE, A.L., BARTLE, K.D., LEWIS, A.C. 2001. A Model of Peak Amplitude Enhancement in Orthogonal Two-Dimensional Gas Chromatography. *Anal. Chem*. 73:1330-1335.

MOHLER, R.E., DOMBEK, K.M., HOGGARD, J.C., YOUNG, E.T., SYNOVEC, R.E. 2006. Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Analysis of Metabolites in Fermenting and Respiring Yeast Cells. *Anal. Chem*., 78:2700-2709.

MONDELLO, L., TRANCHIDA, P.Q., DUGO, P. AND DUGO, G. 2007. Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry: A Review. *Mass Spectrometry Reviews.* 27:101– 124.

PETROFF, S. A. 1915. A new and rapid method for the isolation and cultivation of the tubercle bacillus directly from sputum and feces. *J. Exp. Med*. 21(1):38-42.

ROCKE, D.M., LORENZA, S. 1995. Two-component model for measurement error in analytical chemistry. *Technometrics*. 37(2):176-184.

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.

THORNTON, D.J., SHEEHAN, J.K., LINDGREN, H., CARLSTEDT, I. 1991. Mucus glycoproteins from cystic fibrotic sputum: Macromolecular properties and structural 'architecture'. *Biochem. J.* 276:667-67.

WELTHAGEN, W., SHELLIE, R.A., SPRANGER, J., RISTOW, M., ZIMMERMANN, R., FIEHN, O. 2005. Comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry (GCxGC-TOF) for high resolution metabolomics: biomarker discovery on spleen tissue extracts of obese NZO compared to lean C57BL/6 mice. *Metabolomics*, 1(1):65-73.

XIA, J., PSYCHOGIOS, N., YOUNG, N., WISHART, D.S. 2009. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Research*, 37(Web Server issue):W652-W660.