Analysis of Urinary Lipid Biomarker Candidates from Tuberculosis Patients by Multiple Reaction Monitoring

Elizabeth Louise Waldron
(WLDELI001)

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN

In fulfilment for the degree:

MSc (Med) Chemical Biology

Faculty of Health Sciences
UNIVERSITY OF CAPE TOWN

February 2019

Supervisor: Professor Jonathan Blackburn
Department of Integrative Biomedical Sciences

Name: Elizabeth Louise Waldron

Student Number: WLDELI001

Course: IBS5003W

Declaration

I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

I have used the American Psychology Association convention for citation and referencing. Each contribution to, and quotation in, this project from the work(s) of other people has been attributed, and has been cited and referenced.

This project is my own work.

I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

Signature Signed by candidate

Date 25/07/2019

Acknowledgements

First and foremost, I would like to thank my husband, James Ross, who has been nothing but supportive of me achieving this goal. He has brought home cooked dinners to the office when I've had to be there until the early hours of the morning, he has delivered lunch to me on the weekends when I've been stuck in the laboratory and has been enthusiastic about listening to me ramble on about this project. His patient support has made this project strides easier. I would like to thank my parents, Howard and Miranda Waldron, who have encouraged me and believed in me through the difficulties of this project. I would especially like to thank my father, Dr Howard Waldron, for mentoring me through the writing up phases and for using his scientific writing expertise to finely tune this thesis.

This degree would not be possible if it were not for the opportunity granted to me by the Centre for Proteomic and Genomic Research (CPGR). I would like to thank them for generously funding this degree and for letting me work part-time for the duration of this degree. I would especially like to thank my manager, Dr Liam Bell, for his understanding when I was absorbed in this thesis, and for his practical advice. I would also like to thank him for his contributions in reviewing this report. I would further like to thank the rest of the team at the CPGR, namely Michelle du Plessis, who picked up my workload with no complaints. This was a great comfort and freed me up to focus on my masters.

The students of the Blackburn Laboratory have been instrumental in me completing this degree, whether it be help with the machines, or just a laugh over lunch. I would especially like to thank Alexander Giddey and Brandon Murugan who were always willing to help me rearrange equipment in the laboratory and gave me practical advice and guidance without which this project would not be half of what it is today. Thank you to Dr Zandile Mlamla who has mentored me and taught me the ways of lipids. She provided fundamental advice on chromatography and lipidomic sample preparation techniques. She was incredibly patient with my questions and made this work go smoothly. Thank you to Dr Bridget Calder who has provided much advice on multiple reaction monitoring and practical help by keeping the machines of the laboratory running and in good condition. I would also like to thank her for her contributions in reviewing this report. I would like to extend my gratitude to my supervisor, Professor Jonathan Blackburn, for being so encouraging throughout this degree. I had moments of panic where I did not believe in myself, and I always felt heaps better after a meeting with Prof. I would like to thank him for his patience and for his calm wisdom; his guidance kept me going and kept my eye on the target.

Finally, I would like to thank God for giving me faith to keep me grounded through the trying times. I am eternally grateful for my community of friends, particularly Cindy Adriaanse, who have kept my life balanced and inspired me to just stop and smell the roses.

Abstract

Background: Tuberculosis (TB) is an aggressive disease and is the leading cause of death by infectious disease in South Africa. With early diagnosis and correct treatment, almost all TB cases can be cured. The main diagnostic tests in South Africa are limited for people living in rural areas, require sputum which cannot be produced by very ill patients, and have low sensitivity in immune compromised individuals. There is an urgent need for a non-invasive and robust diagnostic test which uses an easily accessible biofluid and can be performed at the point-of-care. Urine has shown promise as a diagnostic biofluid for biomarker investigation. Full scan mass spectrometry is the gold standard for the unbiased discovery of biomarker candidates, and targeted multiple reaction monitoring (MRM) is the method of choice for subsequent validation of biomarker candidates. A list of candidate urinary biomarkers has previously been generated which can discriminate between latent and active TB infection using MS1 mass spectrometry, but these biomarkers have not yet been verified by targeted mass spectrometry.

Aims and Objectives: The aim of this project is to verify a list of biomarker candidates using MRM assays by: 1) developing MRM assays for known fatty acid standards, and 2) developing MRM assays for unidentified urinary lipid biomarker candidates *de novo*, which can be applied to clinical cohorts for future validation.

Methods: Fatty acid standards were initially assessed using direct infusion full-scan MS1 mass spectrometry on an orbitrap mass analyser. They were then optimised for fragmentation by compound optimisation on a triple-quadrupole mass analyser, the data from which was used to build MRM assays. Liquid chromatography was optimised for these lipids and the MRMs were validated by spiking the lipid standards into a complex mixture.

For the second part of the project, lipid extract (containing unidentified biomarker candidates) from patient derived urine samples were analysed by data-dependent acquisition with inclusion lists on an orbitrap mass analyser. From this experiment MS/MS data was acquired for biomarker candidates which were then compiled into MRM assays and verified using a triple-quadrupole mass analyser.

Results: From six fatty acid standards, reliable MRM assays were generated for five of them. The biomarker candidates formed a list of 70 molecules which were further refined to 10 molecules which were reproducibly measured by MRM assay.

Discussion and Conclusions: From this work the fatty acid standards can be used as internal retention time predictors for future lipidomic work and quality checks, as they eluted across a wide retention time range. The biomarker candidates have been verified using MRM assays and can be validated in larger clinical cohorts in the future. The end-goal is to use these biomarker candidates as part of a panel which represents a unique biosignature according to the disease state of the patient.

Table of Contents

List of Figures

List of Tables

Abbreviations and Symbols

ATB	Active TB disease
BSL3	Biosafety level three
CID	Collision induced dissociation
CV	Coefficient of variation
DDA	Data dependent acquisition
DIA	Data independent acquisition
^{o}C	Degrees Celsius
DNA	Deoxyribose nucleic acid
ESI	Electrospray ionisation
FWHM	Full width at half maximum
g	Gravitational force
HCD	Higher energy collisional dissociation
HIV	Human immunodeficiency virus
HPLC	High performance liquid chromatography
H	Hydrogen
IMS	Ion mobility spectrometry
iRT	Indexed retention time
IT	Injection time
IGRA	Interferon gamma release assay
LAM	Lipoarabinomannan
LAMn	LAM negative
LAMp	LAM positive
LC	Liquid chromatography
LCMS	Liquid chromatography mass spectrometry
LLOD	Lower limits of detection
LLOQ	Lower limits of quantification
LTBI	Latent TB infection
MALDI	Matrix assisted laser desorption/ionisation
µl	Microliter
µM	Micron
µM	Micromolar
ml	Millilitre
mm	Millimetre
ms	Milliseconds
min	Minute
M	Molar
MRM	Multiple reaction monitoring
MS1	Precursor mass spectra
MS/MS	Tandem mass spectra

Mtb	*Mycobacterium tuberculosis*
m/z	Mass-to-charge ratio
nmol	Nanomole
NTB	Non-TB lung infection
%	Percent
ppm	Parts per million
pmol	Picomole
RRHD	Rapid resolution high definition
RT	Retention time
s	Seconds
SRM	Selected reaction monitoring
TOF	Time of flight
TocAce	Tocopherol acetate
TSQ	Triple-stage Quadrupole Vantage
TB	Tuberculosis
V	Volts

Chapter 1 – Improved biomarkers for tuberculosis diagnosis

1.1 Tuberculosis diagnosis in South Africa

Tuberculosis (TB) is a disease caused by the bacterium *Mycobacterium tuberculosis* (Mtb) and is the leading cause of death by infectious disease in South Africa (StatsSA, 2015). In South Africa around 80% of the population is estimated to be infected with Mtb but do not display symptoms and are therefore classified as having a latent TB infection (National South African AIDS council, 2011). Around 454,000 South Africans progress to active TB infection every year and just over half of these patients are dually infected with human immunodeficiency virus (HIV) (World Health Organisation, 2015). With early diagnosis and correct treatment, almost all TB cases can be cured. Therefore, efforts to improve the diagnosis of TB are of utmost importance in tackling this epidemic.

In South Africa there are three main diagnostic TB tests that are currently in use. The most sensitive is the culture test whereby a patient sputum sample is sent to a laboratory where it is cultured to determine the presence of Mtb. A positive result can be confirmed after ten days and a negative result can only be confirmed after 42 days (Davies & Pai, 2008). The shortcomings of this test are that it can only be conducted in centralised laboratories, meaning patients must travel for the test, the test is time consuming, and it is costly. The second is the sputum microscopy test which is widely used in Africa. Patient derived sputum is acid stained and subsequently evaluated under a microscope for the presence of Mtb cells. The test is inexpensive; however, it is labour intensive which means a limited number of smear tests can be conducted in a single day. The test also has a low sensitivity rate of just 60% (Cattamanchi et al., 2009). Another diagnostic test is the Gene Xpert® MTB/RIF test (Cepheid, Sunnyvale, United States of America) which is a cartridge-based nucleic acid amplification test for Mtb in sputum and its resistance to the antibiotics (Steingart et al., 2014). The test does this by detecting specific sequences of bacterial deoxyribose nucleic acid (DNA) and it only takes two hours to produce a result. This test has limited availability to people living in rural areas as the instruments that conduct the test are centralised, and the cost per assay is still relatively high. All these tests are challenging for infants, the elderly, and people living with HIV as they are unable to produce a sputum sample. Additionally the tests have high false negative rates for people living with HIV as their sputum generally has low levels of bacteria (Bévilacqua, et al., 2002). The requirement of patient derived sputum places the health practitioners at risk of contracting a TB infection, and as a result all samples need to be processed in biosafety level three (BSL3) laboratories which are highly specialised and costly to set up.

1.2 TB biomarker research

There is an urgent need for a non-invasive and robust diagnostic test which can be routinely performed at the point-of-care. Because of the challenges associated with collecting sputum, such a test should ideally be conducted on more easily accessible samples such as blood, breath, saliva, or urine. There are several TB disease states and the ideal test would differentiate one from the other. The defined clinical groups are patients who are naturally immune to TB, patients who have become immune by vaccination, patients who have active TB, patients who are at risk of developing active TB, patients who have a latent TB infection, or patients who are responding positively or negatively to TB treatment (Goletti et al., 2018). In theory, each of these health-states possesses a unique biosignature which is yet to be discovered, tested and validated in well-characterised patient cohorts. Variables independent of disease state, *inter alia,* genetic and environmental factors, coinfections, exposure to

non-TB mycobacteria, nutritional status, metabolic status, and age must be considered when investigating biosignatures.

There are two approaches to studying biomarkers; either Mtb-specific antigens can be targeted, or host response markers can be monitored.

Urine has shown promise as a diagnostic biofluid with a clinically approved rapid urinary test that screens for lipoarabinomannan (LAM), a glycolipid found in the membrane of Mtb cells. LAM enters the urine through haematogenous TB dissemination, a condition which occurs primarily in HIV-positive patients with advanced immunodeficiency (Lawn & Gupta-Wright, 2015). Thus LAM detection in urine is a suitable biomarker for disease severity and mortality among HIV positive patients, but it is not useful as a routine diagnostic biomarker (Drain et al., 2017). Urine has shown further potential as a suitable biofluid through a shotgun proteomics experiment, where Mtb proteins and differentially expressed human proteins were discovered as potential biomarker candidates in the urine of patients with active and latent TB infections (Young et al., 2014).

Blood is the most researched biofluid and contains numerous human cytokines (Chegou et al., 2016) and has even been shown to contain detectable Mtb antigens (Liu et al., 2017). While the greatest need is for a first-line diagnostic test, many of the blood-based methods are suitable for assessing patient response to anti-TB treatment (Sweeney et al., 2017). Most diagnostic tests require centralised laboratory equipment, however some show promise of being conducted at the point of care for relatively low cost (Corstjens et al., 2016). Host response can be monitored in the blood to diagnose active TB infection (Jones et al., 2017) or to predict TB disease progression from latent to active (Zak et al., 2017; Khaliq et al., 2018).

Two research groups have tested a handheld, point-of-care electronic nose to diagnose TB in exhaled breath (Teixeira et al., 2017; Zetola et al., 2017). The electronic nose will not satisfactorily replace the gold standard diagnoses because Mtb culture is needed for determination and drug susceptibility testing; however, it is an example of how biomarkers can be used as a point-of-care screening tool which would lessen the burden on centralised facilities.

None of these methods have been approved for use in the public sector and there is an urgent need for current methods to be advanced so that they are field-ready, or for new methods to be developed.

1.3 Biomarker discovery and validation

A biomarker is defined as anything that can be measured and used to predict a biological or diseased state (Strimbu & Tavel, 2010). A biomarker can be anything from number of cigarettes smoked per day, to a genetic mutation, or a protein fold change measured in blood. The platforms for biomarker discovery and validation are broad and varied, depending on the candidate biomarker(s), the species, and the disease. This thesis describes chemical and immunological compounds for biomarker investigation and therefore mass spectrometry is the platform of choice given its greater analytical specificity in these fields. Mass spectrometers are rare in clinical settings because it is a difficult application to automate, they are costly, and they have relatively low throughput compared with other analysers (Crutchfield et al., 2016). At this stage however, mass spectrometers can be used at the research level as accurate screening tools for the discovery, verification and validation of

biomarkers. Ideally, biomarker development is a phased approach that gradually shifts from unbiased characterisation by mass spectrometry to candidate-driven assays (Rifai et al., 2006).

There are various methods for the discovery of biomarkers, however most methods require prior knowledge of the biomarker candidates, for example immunoaffinity capture. Mass spectrometry is currently the method of choice for the unbiased discovery of candidate biomarkers. Despite its superior sensitivity and specificity, mass spectrometry comes with a set of challenges that must be managed. First, the complexity and dynamic range of biofluids interfere with the low-abundant analytes which may result in them never being measured or detected in the discovery phase. It is unfortunately anticipated that disease biomarkers lie in this range of low-abundance analytes. This dynamic range issue can be addressed with multiple fractionation steps prior to mass spectrometry analysis and this introduces the second challenge which is the potential for high variation between experiments. This problem can be addressed with higher sample numbers and technical repeats (Cominetti et al., 2016).

For a sample to be analysed by mass spectrometry it must first be ionised; the two routinely used methods are matrix assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI). ESI is considered a “soft” ionisation method that disperses highly charged droplets near atmospheric pressure, followed by droplet evaporation allowing the charged analytes to enter the mass spectrometry source (Smith et al., 1990).

Mass accuracy largely depends on the resolution of the mass spectrometer. Resolution (or resolving power) is defined as the ability of a mass analyser to distinguish two separate ions with a small *m/z* difference. For example, high resolution mass spectrometers are able to resolve isotopes from each other, while low resolution machines will report an average mass measurement. Resolving power can be calculated using one or two peaks (de Hoffman & Stroobant, 2007):

If two peaks are measured, then resolving power $R = m/\Delta m$; where m is the mass of one peak and Δm is the smallest difference between the two peaks.

If just one peak is used, then the resolving power is the full width of the peak at half of the maximum height of the peak; R = FWHM.

Discovery experiments – also called shotgun experiments – aim to capture as many measurements as possible in a given sample. A common acquisition method is data-dependent acquisition (DDA) in which a mass spectrum is acquired and analysed by the machine and, based on rules set by the user, certain ions – usually the top ten most abundant – are selected from that precursor mass spectrum (MS1) pre-scan and fragmented and their tandem mass spectra (MS/MS) measured (Watson & Sparkman, 2007). Orbitrap mass analysers are well suited for discovery experiments due to their ability to resolve complex mixtures of ions for high resolution results (Zubarev & Makarov, 2013). The product of a discovery phase is a list of analytes that show differences between various disease states. This is a semi-quantitative list and is likely to contain a certain percentage of false negatives and false positives (Rifai et al., 2006): a false positive is a candidate that upon further testing is not differentially expressed, and a false negative is a candidate that is differentially expressed but is not detected as such in the discovery phase. This list then proceeds to the verification phase of the experiment where it is refined.

Data independent acquisition (DIA) combines the indiscriminate data collection of full scan methods with the accuracy of targeted methods, and for such an experiment a quadrupole-quadrupole-time of flight (TOF) mass spectrometer is best suited (Gillet et al., 2012). A window of precursors, for example 20 *m/z* units wide, is selected for fragmentation and a full scan is taken of all the fragment ions. The machine then moves onto the next window of precursors and cycles through in a systematic fashion until, theoretically, MS/MS data has been acquired for all species within a certain mass range. DIA experiments currently rely on spectral reference libraries generated on a shotgun mass spectrometry platform (Schubert et al., 2015). These libraries can also be used to validate targeted assays which are discussed below.

A robust platform for the verification and refinement of candidate biomarker lists is targeted mass spectrometry. Targeted mass spectrometry is conducted with *a priori* knowledge of the analytes in question and is routinely performed on a triple-quadrupole mass spectrometer which offers high selectivity, sensitivity and a wide dynamic range. Precursor ions, representing target analytes, are selected in the first quadrupole and fragmented in the second. Product ions that have been preselected by the user are then filtered in the third quadrupole and their signal intensities are reported. These precursor-product ion sets are called transitions and in modern mass spectrometers, a transition set can be measured in a few milliseconds (Gillette & Carr, 2013), thus allowing tens to hundreds of transitions to be multiplexed per analysis. Such an experiment is called a multiple reaction monitoring (MRM) assay or selected reaction monitoring (SRM) assay, depending on the vendor or instrument. In this review MRM will be utilised, but it is interchangeable with SRM.

1.3.1 MRM experiments

MRM experiments are the gold standard method for analysing small molecules and have been used extensively in pharmaceutical and drug research (Hoke et al., 2001). MRM is a hypothesis driven experiment and prior knowledge of the molecule is required. Because of advances in mass spectrometry technology over the last decade, MRM experiments have begun to be applied to complex biological samples such as proteomes and metabolomes.

When a triple-quadrupole mass spectrometer is operated in MRM mode, the first and third quadrupoles act as mass filters while the second quadrupole acts as a collision cell (Figure 1.1) (Gallien et al., 2011). An MRM experiment is ideal for measuring low level analytes in a complex biological mixture because of its ability to filter out much of the background noise. Fractionation, such as liquid chromatography (LC), usually precedes MRM experiments as it dramatically decreases the background signal and therefore increases the sensitivity of the measurements.

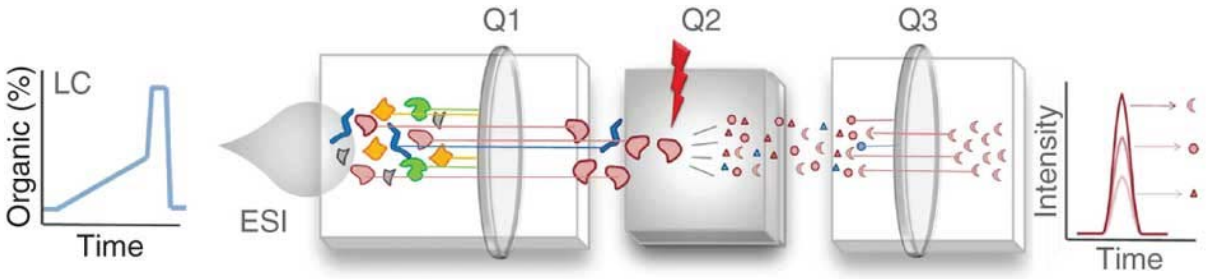


Figure 1.1: Overview of a typical liquid chromatography mass spectrometry (LCMS) MRM experiment.

When designing an MRM assay, the experimental question and target analytes are defined. Where possible, purified standards can be purchased for MRM development and used as a positive control.

If previous experiments have been done on the analyte, a library can be built and used to validate the MRM. For each analyte, the charge state and measured mass-to-charge ratio (*m/z*) is defined and then the fragment ions for each analyte are defined. Transitions are then optimised by tuning the machine to these values. If standards are being used, then the transitions can be validated by a dilution series where the measured signal is expected to decrease proportionally. This validation step also allows for the lower limits of detection (LLOD) and lower limits of quantification (LLOQ) to be determined for each analyte. Transitions can also be validated by assessing the retention time across multiple experiments. If standards are being used, it is ideal to validate transitions by spiking the standards into a complex biological matrix. This allows interference and noise to be assessed and the transition set refined accordingly.

When a triple-quadrupole mass spectrometer is operated in MRM mode, all the machine's analysis time is devoted to a small number of measurements. MRM assays will therefore provide a far more accurate quantitative measurement for each analyte than would be the case for a full scan or shotgun experiment, however, as stated, fewer analytes can be monitored per analysis. For quantification, the experiment must be optimised so that the mass spectrometer is able to measure between eight and ten points per chromatographic peak (Figure 1.2). Fewer measurements per peak can result in an inaccurate peak area reading (Agilent, 2007). Quantification by MRM assays is based on the signal intensity of specific transitions and can be done relatively or absolutely (Lange et al., 2008). In relative quantification experiments, the signal intensities for transitions are compared across multiple samples and results are reported as fold-changes rather than exact quantities. For relative quantification sample processing must be well controlled and sample matrices very similar. In absolute quantification experiments a known amount of isotopically labelled standard is spiked into the sample. The absolute amount of target analyte is then determined by its relative intensity to the heavy-labelled standard (Ankney et al., 2018).

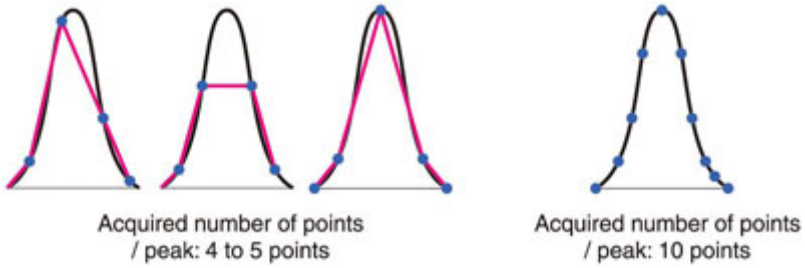


Figure 1.2: A demonstration of why eight to ten points per peak are required for accurate quantification. The three chromatograms on the left picture the real peak in black and the measured peak in red with too few points per peak. The far-right peak pictures the correct peak shape being measured due to sufficient number of points per peak.

To increase the number of analytes measured per analysis, scheduled MRM can be employed (Fillatre et al., 2010). In a scheduled MRM the transitions are only monitored when necessary according to prior knowledge about their LC retention times. Scheduled MRM decreases the number of transitions measured at any point in time, allowing more total measurements across a single experiment without compromising sensitivity.

1.3.2 An introduction to high performance liquid chromatography (HPLC)

HPLC makes the analysis of complex biological samples possible and therefore a mass spectrometry experiment is only as good as its preceding chromatography.

Reversed-phase HPLC is a method used to separate molecules based on their hydrophobicity. There are two phases involved; the mobile phase and the stationary phase. The stationary phase, or column, is made up of hydrophobic ligands to which molecules bind. The mobile phase is a solution that flows over the column and ranges from hydrophilic to hydrophobic. The separation of molecules depends on the binding of solute molecules from the mobile phase to the column. In reversed-phase HPLC, this binding happens in an aqueous mobile phase and the solutes are eluted by the addition of organic solvent to the mobile phase. The solutes therefore elute in order of increasing hydrophobicity, making it a powerful tool for the separation of complex biological samples (Aguilar, 2004). HPLC is a flexible method in which a wide range of chromatographic conditions can be employed and manipulated with relative ease.

The HPLC system employed in this thesis consists of a C18-based stationary phase from which molecules are eluted by a mobile phase of increasing acetonitrile concentration. The mobile phase also contains an ion-pairing agent such as trifluoroacetic acid or formic acid, which assists with the retention of charged molecules on the stationary phase (Bidlingmeyer, 1980). The column is packed with microparticles of porous silica which are chemically modified to support a hydrophobic ligand. The most commonly used ligand is a C18 chain, however C4 and C8 chains are also employed and can provide alternative selectivity and separation (Zhou et al., 1991).

Column dimensions depend on the level of efficiency that is required and generally resolution is proportional to column length (Jilge et al., 1987). The internal diameter is selected based on sample volumes and detection sensitivity. For most analytical applications an internal diameter of 4mm is employed, however if sample volumes are limited then an internal diameter of 1-2mm can be used to allow lower elution volumes. For further sensitivity nanoflow columns can be used which have internal diameters in the range of 75µm. Separation using nanoflow columns requires a longer gradient which subsequently increases analysis time (Hsieh et al., 2013), however the payoff is increased sensitivity and lower sample volumes required.

HPLC is a versatile application because of the ease at which the mobile phase can be manipulated to optimise solute retention and resolution. The most routinely used organic solvents – usually mixed with water to different ratios – are acetonitrile, methanol and isopropanol (Aguilar, 2004) as they are nonreactive and do not interfere with the detection of the analytes. Acetonitrile has the lowest viscosity, which allows higher flow rates through the column which in turn decreases the experimental time, while isopropanol is the strongest eluent and is often used to clean columns between experiments.

1.4 Introduction to aims and methodology

Metabolite changes occur before the onset of disease symptoms and these changes are often reflected in biofluids (Van et al., 2011). Urine has long been used as a non-invasive biofluid to detect abnormal metabolomic by-products, cells, bacteria, and hormones (Simerville et al., 2005). It is easy to collect, rich in metabolites, and is relatively stable *ex vivo*. One difficulty with urinary lipidomics, however, is that the expected concentration of complex lipids is relatively low (Rockwell et al., 2016), so extracts require additional concentrating steps. There are also multiple extraction methods and chromatographic conditions that favour specific lipid classes, some examples of which are shown in Table 1.1. A prior doctorate project within this laboratory has been done on the urinary lipidome to generate a list of candidate biomarkers to discriminate between latent and active TB infection

(Mlamla, 2018). Due to limiting scan times of the mass spectrometer used in that study, this discovery list was based on MS1 spectra only and few of the candidates have unambiguously assigned identities. For verification and quantification by mass spectrometry, it is important to transfer this list of candidates to a target-driven platform such as MRM assays. Chromatography must be identical to that of the previous experiment so that retention times can be used as a metric to match the MS1 values across experiments (Pabst et al., 2007).

Table 1.1: Examples of the various lipid extraction and chromatographic methods in lipidomics.

Extraction solvent	Butanol: Methanol (1:1)
Chromatography conditions	C18 column, tetrahydrofuran, methanol, water, 10 mM ammonium formate
Lipid class	All major lipid classes.
Reference	Alshehry et al., 2015
Extraction solvent	Chloroform: Methanol (2:1)
Chromatography conditions	HILIC column, acetonitrile, methanol, water, 5 mM ammonium formate, 0.05% ammonium hydroxide
Lipid class	Phospholipids, sphingolipids, glycerides, cholesterol esters, triglycerides, ceramides, glycosphingolipids and glycerophospholipids.
Reference	Bang et al., 2014
Extraction solvent	Methyl-*tert*-butyl ether: Methanol (2:1)
Chromatography conditions	C18 column, water, acetonitrile, methanol, isopropanol, 5 mM ammonium formate, 0.05% ammonium hydroxide
Lipid class	Phospholipids, sphingolipids, glycerides, triglycerides and glycerophospholipids.
Reference	Bang et al., 2014
Extraction solvent	Butanol: Methanol (3:1) and Heptane: Ethyl acetate (3:1), 1% acetic acid
Chromatography conditions	NA – direct infusion experiment
Lipid class	Cholesterol esters, free cholesterols, triacylglycerol, phosphatidylcholines, sphingomyelins, ceramides, diacylglycerols, and lyso-phospholipids.
Reference	Löfgren et al., 2012
Extraction solvent	Chloroform: Methanol (1:1) and Hexane: Isopropanol (3:2)
Chromatography conditions	NA – MALDI used
Lipid class	Phospholipids, sphingolipids, glycerides, triglycerides and glycerophospholipids.
Reference	Tipthara & Thongboonkerd, 2016

The research reported in this thesis addresses the development of MRM assays to verify a list of biomarker candidates. This can be done with prior knowledge of the candidate molecules and thus synthetic standards can be used to create an MRM assay. Alternatively, it can be done *de novo* with no knowledge of the candidates' identities. The aim of this project is to demonstrate two approaches in the generation of an MRM assay: namely

1. Developing MRM assays for known lipid standards which can be employed in a complex mixture, and
2. Developing MRM assays *de novo* to verify unidentified lipid biomarker candidates which can be applied in clinical cohorts for further validation. See Figure 1.3 for experimental overview.

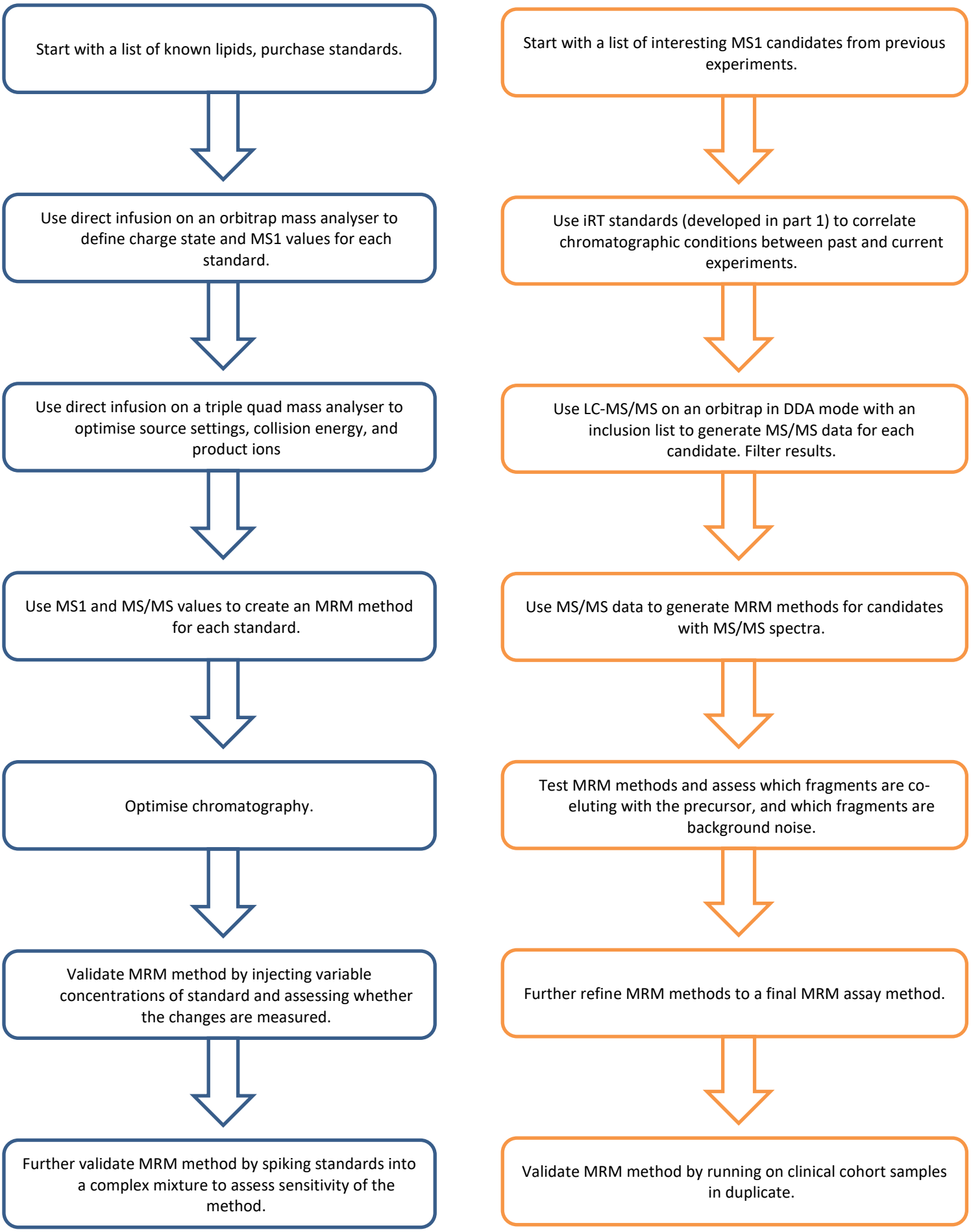


Figure 1.3: Experimental outline for 1) MRM development using standards (left in blue) and 2) MRM development of candidate biomarkers de novo (right in orange).

Chapter 2 – Development of MRM assays for lipids using standards

2.1 Introduction

MRM assays are most commonly developed using purified standards, before proceeding to a complex mixture. This chapter describes the development of MRM assays for heavy labelled fatty acid standards, which are also to be used as retention time predictors in subsequent experiments.

Heavy labelled fatty acid standards were selected because they are the basic building blocks of lipids and represent the stable carbon-chain associated with lipids (Sethi & Brietzke, 2017). Tocopherol acetate, more commonly known as vitamin E, was selected as it is a naturally occurring lipid with a complex head group and thus shares many fragmentation properties with biological lipids. See Figure 2.1 for lipid standard structure and identity.

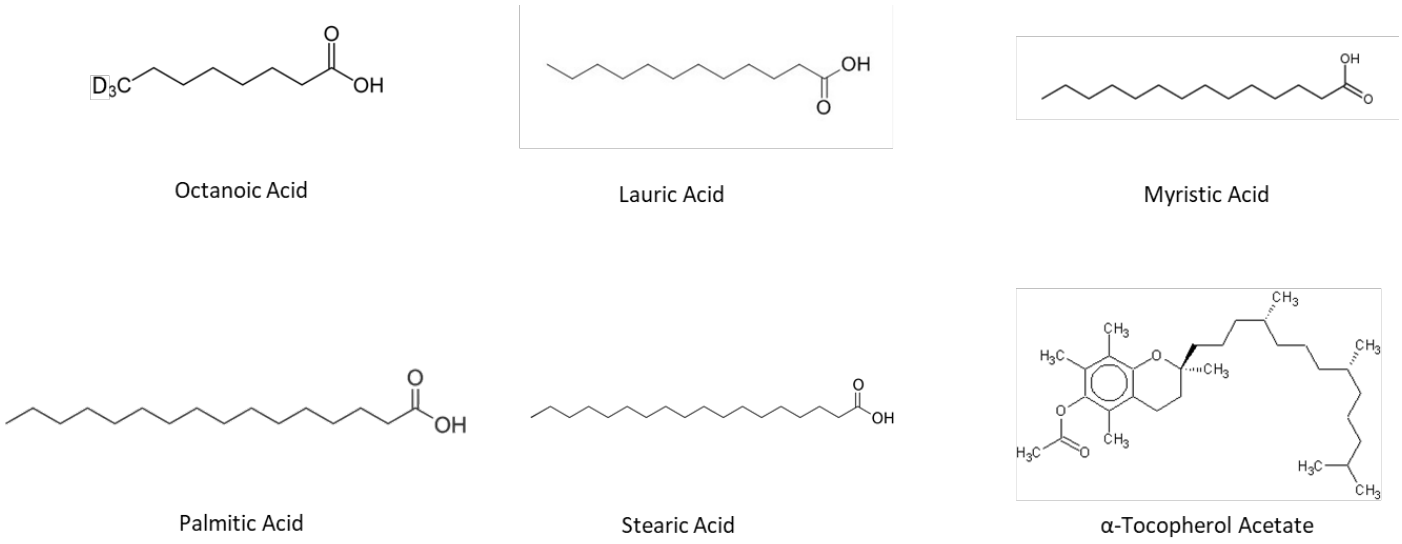


Figure 2.1: Structure of lipid standards. Hydrogens in the fatty acid carbon chains (except for Stearic acid) are replaced with deuterium, while ester group hydrogens are unchanged. Stearic acid carbons are replaced with 13C isotopes while the hydrogens remain in their light form. Tocopherol acetate is in its light form, i.e. unlabelled.

2.2 Aims

The aim of this chapter was:

1. To develop an MRM method for the analysis of lipid standards on a triple-quadrupole mass spectrometer, including optimising HPLC conditions and mass spectrometry settings for the measurement of the standards.
2. To optimise conditions for both negative mode and positive mode mass spectrometry.
3. To generate a work flow that can be applied to other future experiments using different standards.

2.3 Experimental Procedures

2.3.1 Preparation and quality control of standards

The following standards were purchased at 98% purity:

Octanoic-d_{15} Acid (Sigma 55471)
Lauric-d_{23} Acid (Sigma 451401)
Myristic-d_{27} Acid (Sigma 68698)
Palmitic-d_{31} Acid (Sigma 68277)
Stearic $^{13}C_{18}$ Acid (Sigma 89065)
+/- α-Tocopherol Acetate (Sigma T3376)

Standards were split into two samples, dried down with nitrogen gas, and resolubilised in each solvent system: chloroform : methanol (1:1) and isopropanol : hexane (70:30) to 1nmol/µl and stored at -20°C as the stock solutions.

Standards were combined for each solvent system, with each standard at a final concentration of 167pmol/µl in the mixture. This mixture was acidified with formic acid to a concentration of 0.1% and samples were then infused directly into a Q-Exactive quadrupole-Orbitrap mass spectrometer (Thermo-Fisher Scientific, USA) with a Hamilton syringe needle and electrospray ionisation (ESI) source at 5µl/min. Full scan MS1 data were acquired in both positive and negative mode using Xcalibur 3.0.63 (Thermo-Fisher Scientific, USA). Fatty acids were observed in negative mode, and tocopherol acetate was observed in positive mode.

2.3.2 Compound optimisation on a triple-quadrupole mass spectrometer

All standards were prepared individually in both chloroform : methanol (1:1) and isopropanol : hexane (70:30) solvent systems. Dilutions were made so that each standard was at 167pmol/µl, except for lauric acid which was prepared to 65nmol/µl, as it was found that an increased concentration was needed to allow visualisation of the compound. Samples were all acidified to 0.1% formic acid and directly infused into a Triple-stage Quadrupole Vantage mass spectrometer (TSQ; Thermo-Fisher Scientific, USA) using a Hamilton syringe and ESI source. Compound optimisation was done using TSQ Tune Master 2.5.0.1305 (Thermo-Fisher Scientific, USA) and MS/MS data acquisition was optimised using collision energy ramps from 0 to 80 eV.

2.3.3 HPLC trouble-shooting and optimisation

Standards were prepared in both chloroform : methanol (1:1), and isopropanol : hexane (70:30) to 1nmol/µl and all standards were acidified with formic acid to 0.1%. An EclipsePlus C18 Rapid Resolution High Definition (RRHD) 1.8µm 2.1 x 100mm (Agilent 959758-902) analytical column was used in conjunction with a C18 3mm guard column (Phenomenex AJ0-9000). Solvent A was water and 0.1% formic acid, Solvent B was acetonitrile and 0.1% formic acid. HPLC optimisation was performed on the Thermo TSQ mass spectrometer as described in Section 2.3.2 with a micro-flow ESI source set to paramaters which were optimised in compound optimisation (Table 2.1).

Table 2.1: Thermo TSQ source parameters.

Parameter	Value
Capillary temperature	320°C
Vaporiser temperature	60°C
Sheath gas pressure	10 bar
Aux gas pressure	10 bar
Positive polarity	3000V
Negative polarity	3500V

Each standard solution was analysed using a gradient that loaded at 0% Solvent B at 250µl/min and increased over seven minutes to 100% Solvent B where it remained for three minutes and then dropped back down to 0% Solvent B. Acidified standards were injected at 10nmol per analysis, however they were not observed using this chromatography. The loading conditions were changed to 2% Solvent B and the rest of the gradient unchanged, however this did not improve the results.

The HPLC configuration was altered so that the column was excluded but the loading conditions and source were not. The sample was collected by the autosampler and directed to the micro-flow source through empty tubing. The initial injection was done using an isocratic gradient of 2% Solvent B at 250µl/min over five minutes and none of the standards were observed. The column-free injections were repeated with an isocratic gradient of 100% Solvent B at 250µl/min over five minutes and the standards were all observed except for lauric acid. Data were collected in full-scan MS1 mode and to confirm that the observed peak was the standard, a 10nmol and 20nmol injection was done for each standard in each solvent system.

The HPLC configuration was returned to the standard method with the column in line with the autosampler and ESI source and the injections repeated with a 100% Solvent B isocratic gradient at 250µl/min for five minutes. Each standard was acidified and injected in duplicate (once each, in each solvent system) and the loading concentrations of Solvent B altered for each gradient according to Table 2.2.

Table 2.2: Gradients employed to determine binding point of standards to C18 column.

Gradient name	Column volumes	Time (mins)	% Solvent B	Flow (µl/min)
100% isocratic	3.5	0-5	100	250
90% gradient	1	0-1.4	90	250
	2	1.4-4.2	90-100	250
	1	4.2-5.6	100	250
	5	5.7-12.6	90	250
80% gradient	1	0-1.4	80	250
	3	1.4-5.6	80-100	250
	1	5.6-7.0	100	250
	5	7.1-14	80	250
70% gradient	1	0-1.4	70	250
	4	1.4-7.0	70-100	250
	1	7.0-8.4	100	250
	5	8.5-15.4	70	250
60% gradient	1	0-1.4	60	250
	5	1.4-8.4	60-100	250
	1	8.4-9.8	100	250
	5	9.9-16.8	60	250
50% gradient	1	0-1.4	50	250
	6	1.4-9.8	50-100	250
	1	9.8-11.2	100	250
	5	11.3-18.2	50	250
40% gradient	1	0-1.4	40	250
	7	1.4-11.2	40-100	250
	1	11.3-12.6	100	250
	5	12.7-19.6	40	250
30% gradient	1	0-1.4	30	250
	8	1.4-12.6	30-100	250
	1	12.6-14	100	250
	5	14.1-21	30	250
20% gradient	1	0-1.4	20	250
	9	1.4-14	20-100	250
	1	14-15.4	100	250
	5	15.5-22.4	20	250
10% gradient	1	0-1.4	10	250
	10	1.4-15.4	10-100	250
	1	15.4-16.8	100	250
	5	16.9-23.8	10	250
5% gradient	1	0-1.4	5	250
	11	1.4-16.8	5-100	250
	1	16.8-18.2	100	250
	5	18.3-25.2	5	250

Thermo raw files were imported into Skyline 3.7.0 10940 (University of Washington, USA) where the document had been set up for small molecule analysis and the target transitions set according to the compound optimisation done previously in this project (Section 2.3.2). Anything eluting before one minute was considered dead volume and thus not interacting with the column. Standards that eluted later than one minute were interacting with the column and the gradient was plotted as a linear graph

of % Solvent B versus time. A delay of one minute was considered when back-calculating the % Solvent B at the point of elution.

Based on the results from HPLC optimisation (Section 2.4.3), chloroform : methanol (1:1) was chosen as the preferred solvent system for solubilising fatty acid standards.

The final HPLC gradient was optimised to start and load at 8% Solvent B for 3.5 minutes at a flow rate of 250µl/min. Solvent B increased to 100% over 28 minutes and remained at 100% for 7 minutes. Solvent B was decreased to 8% for 6.5 minutes, the equivalent of 2.5 column volumes, to re-equilibrate the column between experiments.

2.3.4 MRM assessment and validation

Transitions were constructed for each standard based on the precursor-product ion pairs observed in the compound optimisation (see appendices, Table 5.1). The MRM method was created using Xcalibur 3.0.63 (Thermo Fisher Scientific, USA), combining all transitions into one method.

Standards were mixed into four different mixtures with altered concentrations as detailed in Table 2.3.

Table 2.3: Standard mixtures made up for injection.

Standard	Concentration Mix 1	Concentration Mix 2	Concentration Mix 3	Concentration Mix 4
Octanoic Acid	180 pmol/µl	400 pmol/µl	100 pmol/µl	100 pmol/µl
Lauric Acid	6.9 nmol/µl	13.8 nmol/µl	3.4 nmol/µl	3.4 nmol/µl
Myristic Acid	180 pmol/µl	100 pmol/µl	400 pmol/µl	100 pmol/µl
Palmitic Acid	180 pmol/µl	100 pmol/µl	400 pmol/µl	100 pmol/µl
Stearic Acid	180 pmol/µl	100 pmol/µl	100 pmol/µl	400 pmol/µl
Tocopherol Acetate	91 pmol/µl	50 pmol/µl	50 pmol/µl	200 pmol/µl

Each mixture was injected onto the column and measured by MRM using the Thermo TSQ mass spectrometer operated under the same conditions as described in section 2.3.3. Data was analysed in Skyline as previously described in section 2.3.3, and the transition sets were further refined to remove fragment ions that did not co-elute with the precursor ion.

To validate the MRM in a complex matrix, lipids were extracted from healthy volunteer urine by aliquoting 5ml of urine and filtering through a 0.45µM filter and then a 0.22µM filter. Water : ethylacetate (5:7) with 6M HCl was added and the mixture vortexed for 30min. The urine sample was then centrifuged at 3320 x *g* for 10min. The supernatant was collected into a clean 50ml amber vial that had been weighed. This process was repeated an additional two times and the supernatants combined each time. The aqueous residue was mixed with 1ml of chloroform by gentle vortexing for 30min. This was then centrifuged at 3320 x *g* for 10min and the supernatant combined with that from previous steps. This was repeated an additional two times and the supernatants combined each time. The aqueous residue was mixed with 1ml of methanol : chloroform (1:2) by gentle vortexing for 30min. This was then centrifuged at 3320 x *g* for 10min and the supernatant combined with that from previous steps. This was repeated an additional two times and the supernatants combined each time. The residue was mixed with 1ml of methanol : chloroform (1:1) by gentle vortexing for 30min. This was then centrifuged at 3320 x *g* for 10min and the supernatant combined with that from previous

steps. This was repeated an additional two times and the supernatants combined each time. Combined supernatants were dried down by vacuum centrifugation in the same pre-weighed amber vial as before. Dried extract was weighed, and the extraction yield determined by weight. Lipid extract was reconstituted in chloroform : methanol (1:1) and acidified to 0.1% formic acid for LCMS analysis.

Standards were spiked into the urine lipid extract at the following concentrations:

Octanoic acid – 250pmol/µl
Lauric acid – stock depleted by this stage
Myristic acid – 640pmol/µl
Palmitic acid – 540pmol/µl
Stearic acid – 250pmol/µl
Tocopherol acetate – 50pmol/µl
Urine lipid extract – 250ng/µl

This mixture was injected at 20µl onto the column and measured with the Thermo TSQ mass spectrometer using the refined MRM method and HPLC gradient as in section 2.3.3. All data was analysed in Skyline as previously described in section 2.3.3.

2.4 Results and Discussion

2.4.1 Direct infusion on the Q-Exactive

Standards were found to be soluble and ionizable in both solvent systems which allowed the charge state for each standard to be established. The fatty acids were observed with a charge state of -1 and tocopherol acetate was observed with a charge state of +1 (Table 2.4). Lauric acid was not detectable by direct infusion. The reasons for this may be that the ionisation conditions were not optimal for lauric acid, it was not in fact soluble in the solvent systems chosen, there may have been in-source fragmentation that was not detected, or the LLOD is significantly higher for lauric acid. Not too much time was spent on optimising conditions for lauric acid as it was not a critical component of this experiment. It did appear in later direct infusion experiments on the Thermo TSQ mass spectrometer where a higher concentration was infused. Formic acid was found to be a suitable ion-pairing agent in both positive and negative modes. While it is acknowledged that acidic conditions are not the best for negative mode small molecule work (Wu et al., 2004), formic acid was suitable for the purposes of these experiments, and it was the ion-pairing agent used in the study (Mlamla, 2018) of which this is a continuation, so no further ion-pairing systems were tested.

Table 2.4: Measured m/z values for each standard from the direct infusion experiment on the Q-Exactive.

Standard	Mass	Mass + H	Mass - H	Observed *m/z*	Charge state
Octanoic Acid	159.30	160.30	158.30	158.201	-1
Lauric Acid	223.46	224.46	222.46	222.30*	-1
Myristic Acid	255.54	256.54	254.54	254.371	-1
Palmitic Acid	287.62	288.62	286.62	286.428	-1
Stearic Acid	302.35	303.35	301.35	301.325	-1
Tocopherol Acetate	472.74	473.74	471.74	473.399	+1
*Lauric acid was observed at this *m/z* at a later stage during compound optimisation on the Thermo TSQ mass spectrometer.					

2.4.2 Compound optimisation on the Thermo TSQ

Masses from compound optimisation can be seen in Table 2.5 below and indicate a slight difference in precursor *m/z* from that measured on the Q-Exactive. Given that the data was generated using two different mass analysers, these slight differences in *m/z* are to be expected (Aebersold & Mann, 2003). The product ion measured for myristic acid was only 0.5 *m/z* smaller than the precursor. A mass shift of 0.5 *m/z* in MS1 spectra is commonly associated with doubly charged isotopes (Zhang & Marshall, 1998) however, when looking at the MS1 spectra from the Q-Exactive direct infusion experiments, this 253.9 *m/z* ion was not present and the isotopic envelope clearly showed a singly charged molecule (Figure 2.2). It may be that the precursor did not yield any measurable product ions in the compound optimisation experiment, and what is being measured here is an intact precursor, with a *m/z* shift caused by relatively poor resolution of the TQS mass spectrometer in full scan mode. If two peaks are separated by 1 *m/z* at high resolution, when the same molecule is measured at lower resolution the two peaks will merge and be measured as one peak. The apparent *m/z* measured on the low resolution mass spectrometer will be close to the average of the two isotope peaks.

Table 2.5: Results from compound optimisation on the Thermo TSQ mass spectrometer.

Compound	Precursor *m/z*	Voltage (V)	Cap. Temp. (°C)	Product Ions *m/z*	Collision Energy (V)
Octanoic Acid	158,229	3500	350	137,9 138,1	16 17
Lauric Acid	222,3	3500	350	149,5 164,9 151,3	6 24 45
Myristic Acid	254,4	3500	350	253,9	12
Palmitic Acid	286,5	3500	350	266,2	26
Stearic Acid	301,3	3500	350	283,4 283,0 284,7 212,8	23 24 21 61
Tocopherol Acetate	473,4	3200	320	165,1 207,1 147,1 149,1 137,1	24 19 30 28 48

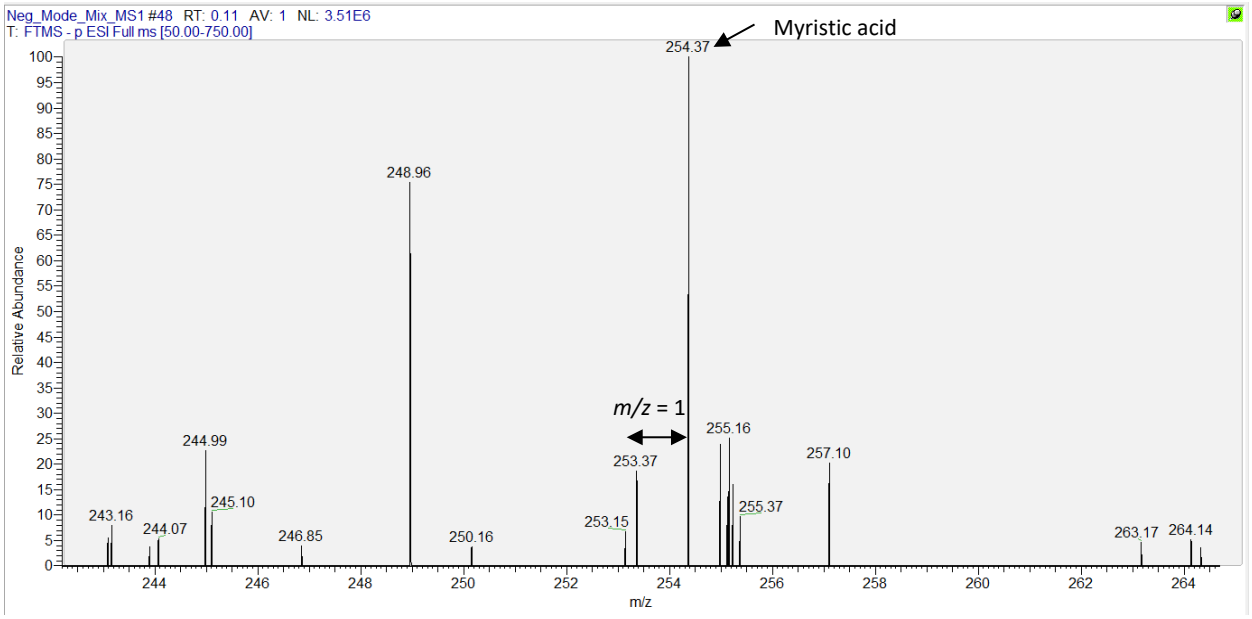


Figure 2.2: Spectra from the full scan MS data acquired on the Q-Exactive. Heavy labelled myristic acid is seen at 254.37 m/z with its light isotope one unit smaller at 253.37 m/z.

and that they ionised sufficiently to be observed by mass spectrometry. However, when HPLC was introduced the standards disappeared, so troubleshooting was done to address the following options: A) there was a problem with the column, B) there was a problem with the loading conditions, or C) there was a problem with the source at high flow as the direct infusion was done at 5µl/min and the HPLC was at 250µl/min. A column-free injection was used to test option A), and the standards were still not detected. To test option B), the loading conditions were changed from 0% Solvent B to 100% which resolved the issue as the standards were observed again at 100% Solvent B. Therefore, it was not necessary to test option C.

I hypothesise that the reason for this is that aqueous loading encourages micelle formation of the fatty acids (Blesic et al., 2007) thus making them undetectable by mass spectrometry. A micelle is an aggregate of fatty acids with the hydrophilic head regions in contact with the solution and the hydrophobic single-chain tail regions in the centre of the sphere (Figure 2.3). Micelles generally form in aqueous solutions, at moderate to high fatty acid concentrations, and are very difficult to detect with ESI mass spectrometry (Nohara et al., 1998).

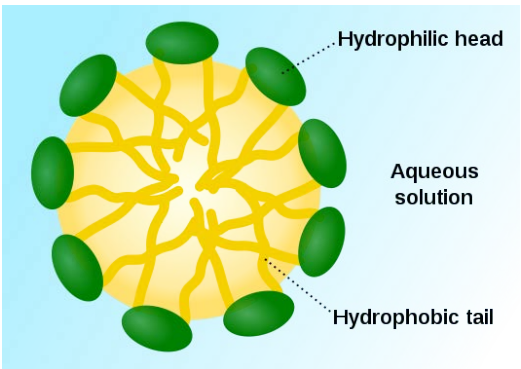


Figure 2.3: Schematic of a micelle formed by fatty acids in aqueous solution.

After the standards had been observed in the column-free system, the injection was doubled from 10nmol to 20nmol, and the peak intensities measured for each solvent system. It was found that the signal doubled from 10nmol to 20nmol, as expected, in the chloroform : methanol (1:1) solvent system but remained the same in the isopropanol : hexane (70:30) solvent system (Figure 2.4). This occurrence has not been recorded elsewhere in the literature, however it may be because the boiling points of chloroform and methanol are lower than the boiling points of hexane and isopropanol (Orgsoltab, 2017), which, in turn, means that the chloroform-methanol solvent mixture evaporates more readily in the ESI source than the isopropanol-hexane solvent mixture. Consequently, more sensitive measurements were observed in the former system than the latter. Because of this discrepancy, chloroform : methanol (1:1) was selected as the preferred solvent system for subsequent work.

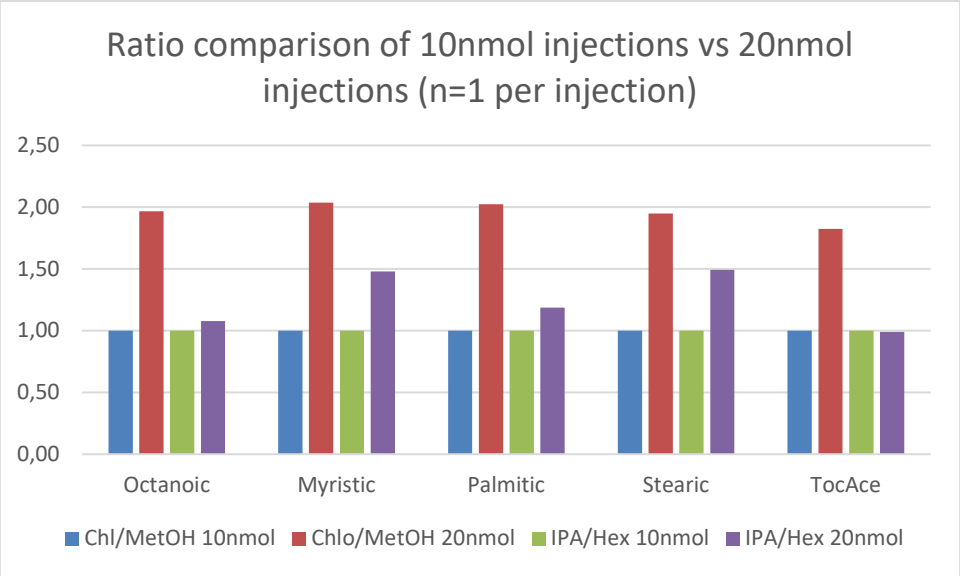


Figure 2.4: Signal intensity comparison of 10nmol injection versus 20nmol injection between chloroform: methanol (1:1) and isopropanol: hexane (70:30) solvent systems. The signal for each 20nmol injection was normalised to the corresponding 10nmol injection (made to =1 in this graph).

The column was reconnected and loading conditions tested from 100% Solvent B to 5% Solvent B to determine at which point the standards interacted with the column and at which point they disappeared from the gradient altogether. Figure 2.5 below is a representation of the elution time of each standard from the column using multiple different gradients. Lauric acid was loaded at a much higher concentration (~760nmol) which allowed it to be visualised. As can be seen from the graph, each standard begins to elute at a later point as the starting concentration of Solvent B decreases. This indicates column interaction which begins at a different point for each standard.

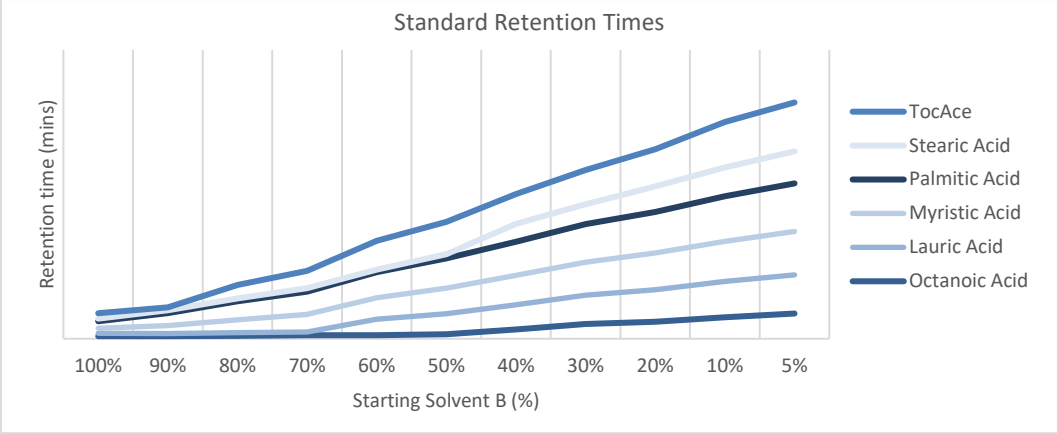


Figure 2.5: Stacked line graph of retention times for each standard on each gradient.

2.4.4 MRM assessment and validation

Four different mixtures with an increasing range of individual standards were injected onto the column and analysed by MRM (Section 2.3.4). The retention times of the standards were measured in each experiment and plotted below in Figure 2.6. The retention times are clearly reproducible and the relationship between carbon chain length and retention time is proportional. This was to be expected and is a well-described trend (Tanaka et al., 1980).

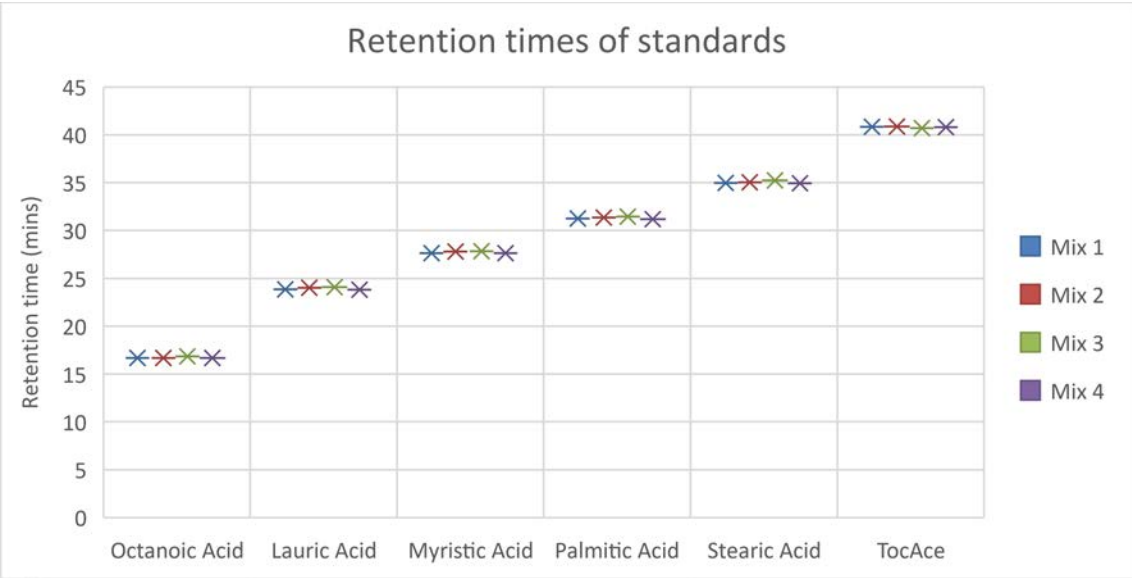
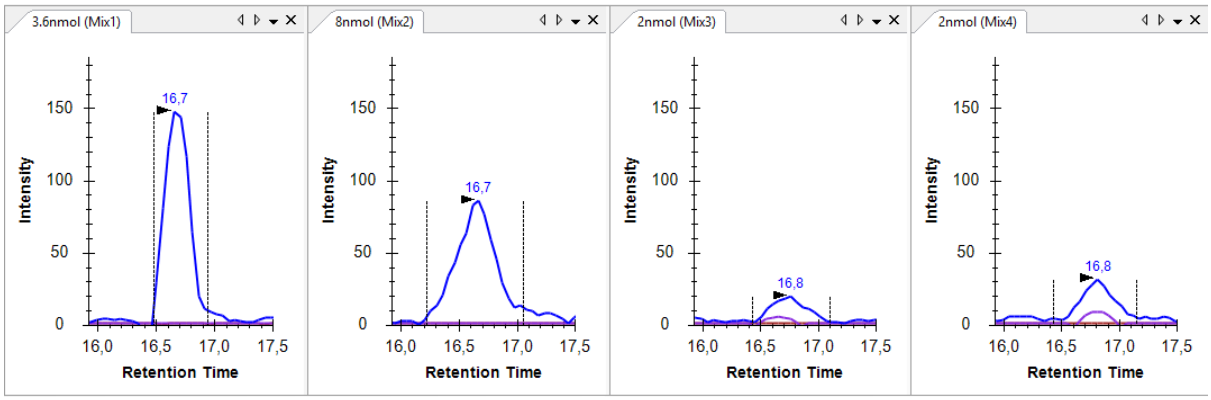


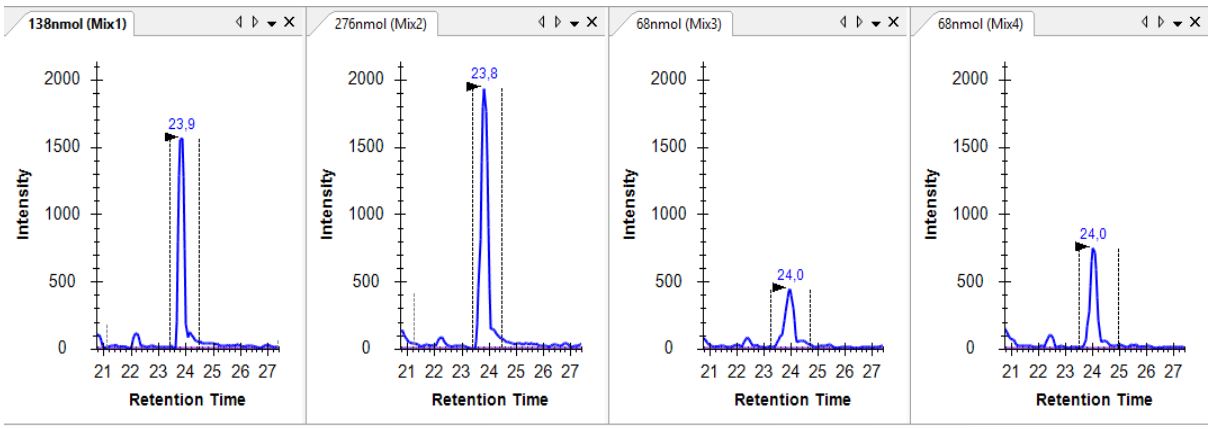
Figure 2.6: Measured retention time for each standard across four different injections.

This experiment, with different concentrations of each standard injected per mixture, was found useful in determining the correct peak annotation. Occasionally false positives can be measured in MRM experiments and a way to test which is the true peak is to increase/decrease the injected amount of the analyte and identify which peaks increase or decrease proportionally. Peaks where the signal remains unchanged are a consequence of matrix interference. Results are shown in Figure 2.7 below. Precursor ion (in blue) was used to compare signal changes between injections, and the fragment ion(s) (in colour) was/were used to confirm peak identity. The signal increases and decreases as expected for each standard, except for tocopherol acetate. According to the elution profile of the product ions, this is the correct peak as they are all measured and elute at the same time as the

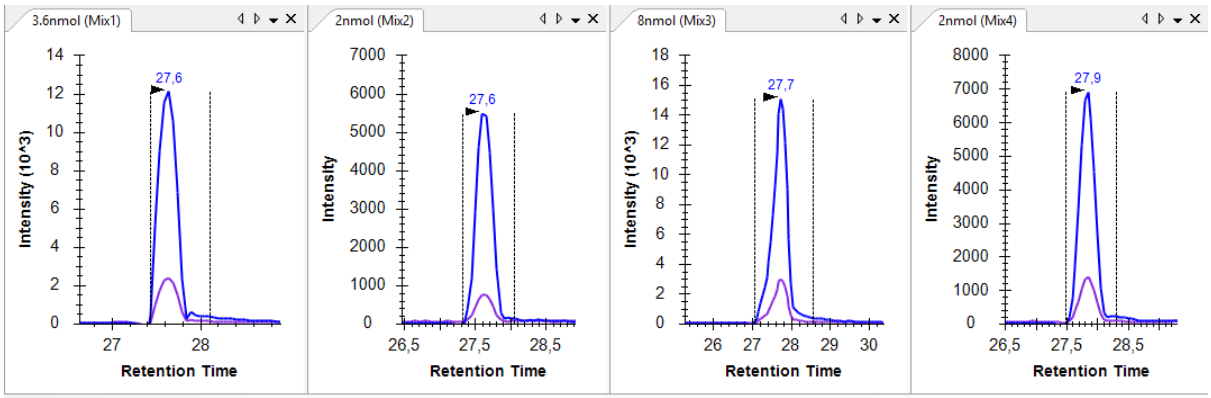
precursor, however the changes in signal are not proportional to the injected amount. It is presumed that this is tocopherol acetate being measured, however it is particularly difficult to elute from the column, so the erroneous signal changes may be due to carryover. Stronger organic solvents in the mobile phase could potentially resolve this issue. For octanoic and lauric acids the product ions were difficult to measure, and the peak selection was done based on the precursor ions. This may be because the shortened carbon chain length yields less stable product ions when compared with the longer carbon chains. Octanoic acid also did not show a large signal increase when the injection was increased from 3.6nmol to 8nmol, however the 2nmol injections showed the expected signal decrease and there may have been an experimental error when Mix 1 was prepared in experimental section 2.3.4.



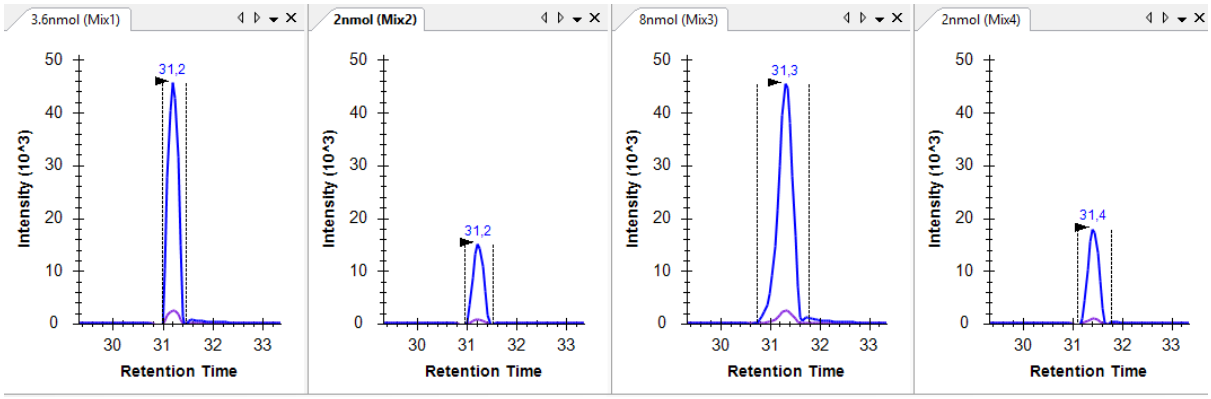
A: Octanoic acid



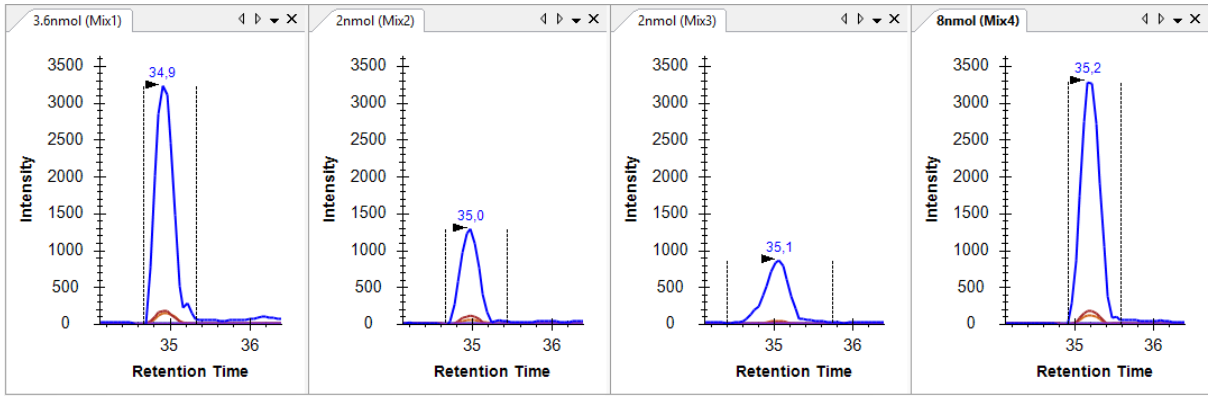
B: Lauric acid



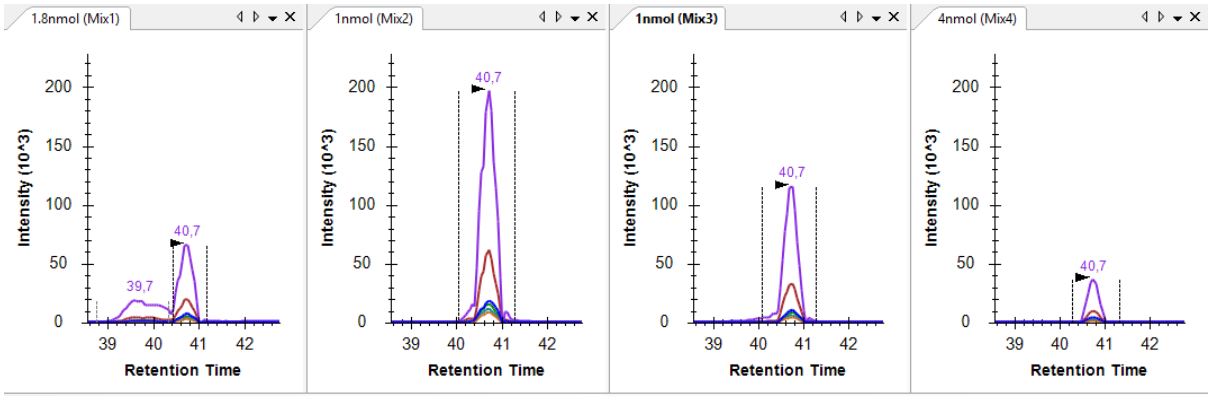
C: Myristic acid



D: Palmitic acid



E: Stearic acid

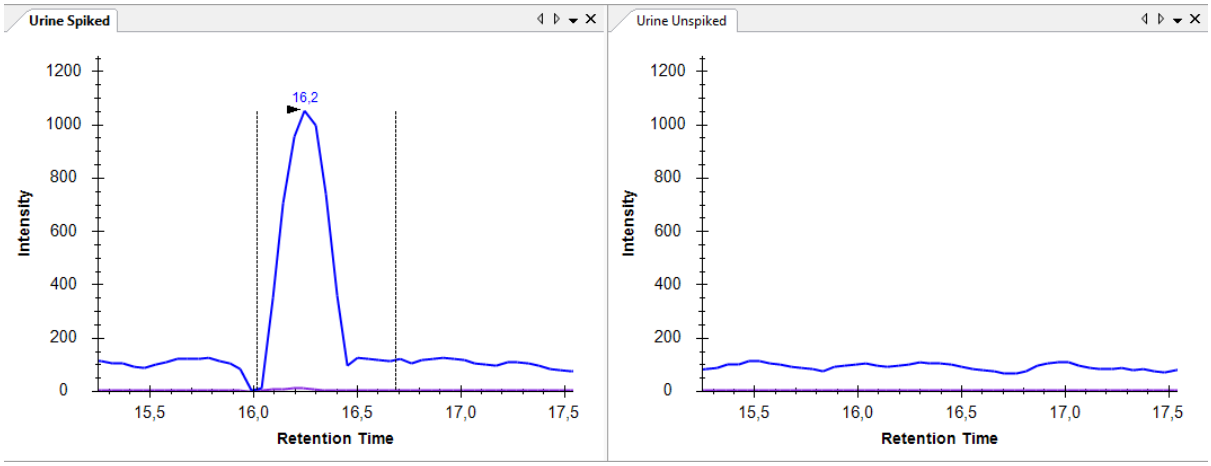


F: Tocopherol acetate

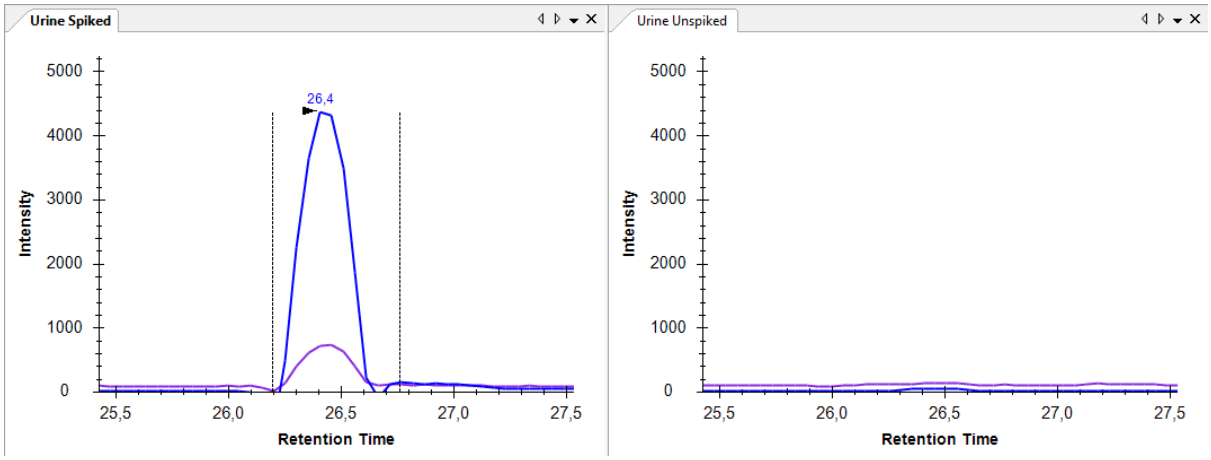
Figure 2.7A-F: MRM results for concentration changes of fatty acid standards across four different mixtures. Each figure is the integrated peak of precursor and product ions.

To confirm that the MRM method designed in this section was applicable to a complex mixture, the standards were spiked into a lipid extract from urine prepared in section 2.3.4 and the results are shown in Figure 2.8 below. This extraction method (which is a modification of the well-known Folch method) was used so as to mimic the conditions of a previous study (Mlamla, 2018). An unspiked urine extract was also analysed with the same method to confirm the level of interference. Lauric acid was not spiked as the stock had depleted by this stage of the experiment. Octanoic, myristic, palmitic and stearic acids all gave convincing evidence that an MRM assay had successfully been generated with minimal interference in a biological matrix. Tocopherol acetate showed an increased signal in the unspiked sample versus the spiked sample which indicates that this may be due to a contaminant which shares the same mass as tocopherol acetate. This is also consistent with the previous evidence

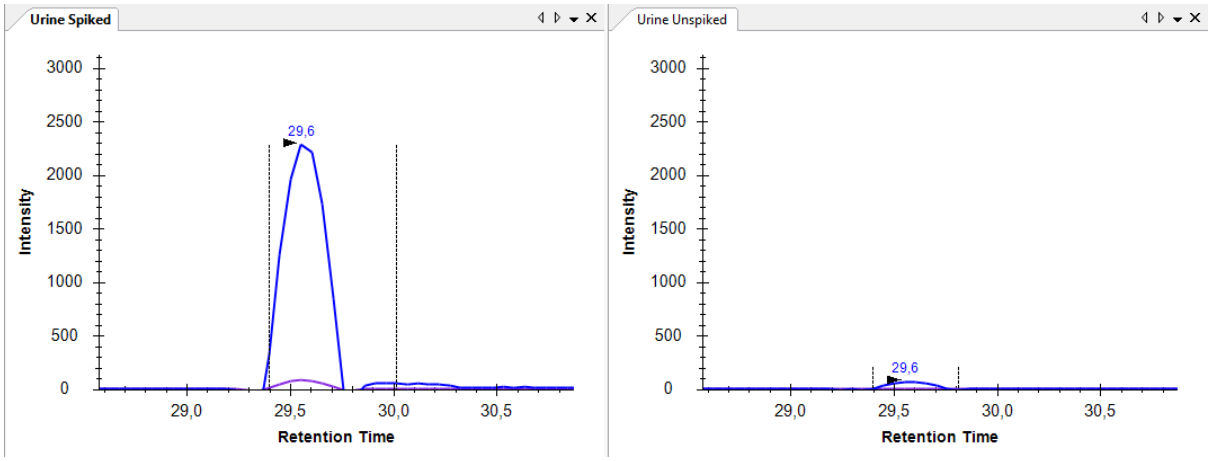
showing that tocopherol acetate MRM results were not proportional to the spiked concentrations, and hence the MRM assay for this analyte is considered unreliable. This finding further confirms the importance of validating an MRM assay before using it for quantitative experiments.



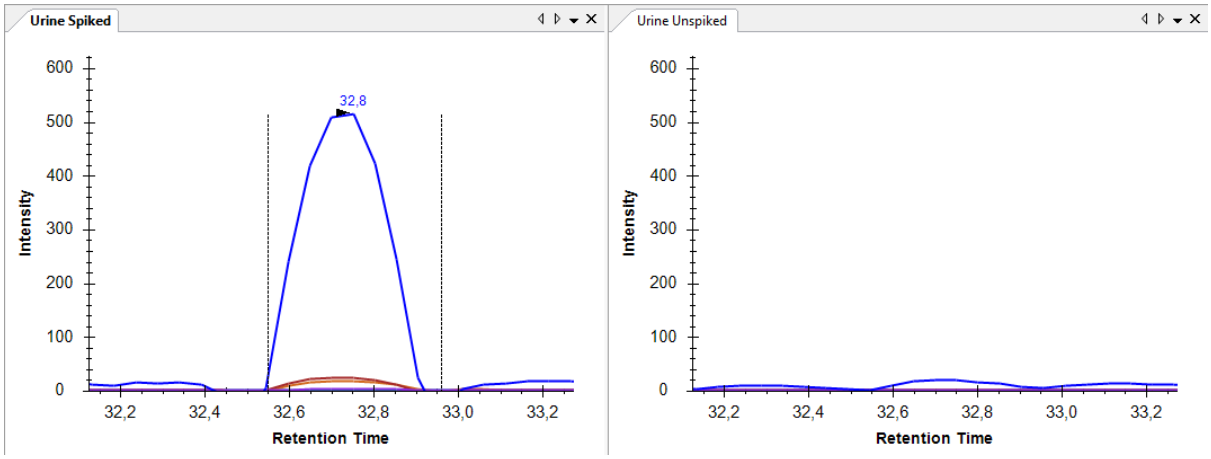
A: Octanoic acid



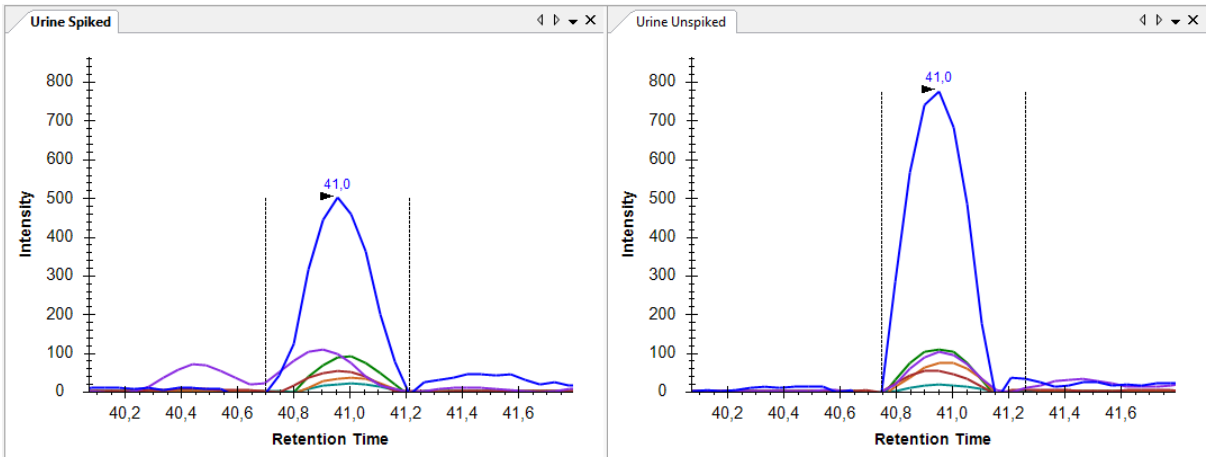
B: Myristic acid



C: Palmitic acid



D: Stearic acid



E: Tocopherol acetate

Figure 2.8A-E: MRM results from standards spiked in a complex mixture of lipids extracted from urine. The left panel represents the urine lipid extract which was spiked with standards, and the panel on the right is the urine extract unspiked to serve as a negative control.

2.5 Conclusions

This chapter demonstrated the methodology used to generate an MRM assay for an analyte for which a commercial standard is available. Given the broad retention time range, this specific set of MRM assays can further be used for cross-column calibration or as an internal retention time predictor. This Chapter also demonstrated the importance of all validation steps when generating an MRM assay as not all candidate analytes will be suitable for MRM analysis. These methods can be applied to any biomolecule for which there is a commercially available standard.

This Chapter also demonstrated that when working with fatty acids, one must be aware of micelle formation and therefore highly aqueous loading solutions should be avoided.

Chapter 3 – Verification of biomarker candidates

3.1 Introduction

A list of some 2000 lipid biomarker candidates have already been generated based on a urinary lipidomic discovery experiment (Mlamla, 2018). Shotgun mass spectrometry was used to acquire MS1 values for these biomarker candidates across three different patient groups namely:

1. active TB infection,
2. latent TB infection, and
3. non-TB lung infection as a control group.

Pairwise comparative analysis identified statistically significant lipids across these three groups, and these lipids created a metabolomic pattern that differentiated the three groups from each other. Hierarchical Cluster analysis showed marked differences in the expression levels of several molecular features between the patient groups. This list was refined to 70 of the most statistically significant biomarker candidates and this is the list that is verified in this Chapter. None of these biomarker candidates have their identities or structures confirmed due to no fragmentation data being acquired in the previous study; a limitation of the MS/MS cycle time on the Agilent qTOF 6530 used in that work.

These biomarker candidates are described at the MS1 level; however, MS/MS data is required to differentiate a biomarker candidate from similar but separate molecules. Compared with the uniform fragmentation of peptides along the amine backbone, metabolites and lipids do not fragment in a predictable way and they often yield common fragments (Herzog et al., 2011). Simple DDA experiments are not good enough to unambiguously match product ions to precursor as it is difficult to define what is noise and what is a product ion without being able to predict the fragmentation patterns, which requires a spectral library. Even with a library and an MS/MS spectrum it is often difficult to entirely characterise a small molecule. MRMs are required so that retention times can be used to link product ions to precursors using peak shape matching.

As discussed in the literature review (Section 1.3.1), MRM assays are a highly selective and sensitive mass spectrometry technique and an excellent method for verifying biomarker candidates. This Chapter describes the process of progressing from shotgun data to refined MRM assays which can then be applied to future biomarker validation.

3.2 Aims

The aim of this Chapter is to demonstrate a framework to generate MRMs for biomarker candidates *de novo* without standards or prior knowledge of the molecule. The secondary aim is to generate MS/MS data for a panel of biomarker candidates previously measured in the urine of TB patients (Mlamla, 2018). This chapter does not infer biological significance of the biomarker candidates, but rather creates refined MRM assays for as many candidates as possible.

3.3 Experimental Procedures

3.3.1 Cross-column calibration

Biomarker candidates were originally discovered in 2014 using a Phenomenex Aqua 3µm C18 100 x 3.0mm column coupled to an Agilent qTOF 6530 mass spectrometer, however the present body of work was conducted using an Agilent EclipsePlus 1.8µm C18 100 x 2.1mm column, hereafter referred

to as “Aqua” and “Plus” respectively. The reason for the column change is that the Aqua column was no longer performing optimally, and the closest column available was the Plus column. Given that MS/MS data had not been generated for the candidate biomarkers, retention time was used as a metric to support the premise that the MS1 measurement in the validation experiment corresponded to the MS1 measurement in the discovery experiment. A calibration was required between the two columns so that the retention times of the candidates could be predicted in the validation experiment.

A mixture of standards was created using octanoic acid, lauric acid, myristic acid, palmitic acid, stearic acid and tocopherol acetate, from the same source as those used in Chapter 2. The standards were solubilised in chloroform : methanol (1:1) and acidified to 0.1% formic acid.

The standard mix was injected onto each column and separated as described below. Both columns were heated to 40^{o}C and connected to a C18 3mm guard column (Phenomenex AJ0-9000). The gradient that was used to generate the original list of candidate biomarkers is detailed in Table 3.1 and the equivalent gradient which was used on the Plus column is detailed in Table 3.2. Solvent A was water acidified to 0.1% formic acid, and Solvent B was acetonitrile acidified to 0.1% formic acid. The second gradient was adapted so that the ratio between the column volumes and change in %B was the same in both gradients (Table 3.2).

Table 3.1: Original gradient used in biomarker discovery experiment with the Aqua column.

Column Volumes	Time (mins)	% Solvent B	Flow (µL/min)
2.5	0-5	8	350
10	5-25	8-61	350
5	25-35	61-80	350
5	35-45	80-100	350
5	45-55	100	350
2.5	55.1-60	8	350

Table 3.2: Adjusted gradient used on the Plus column.

Column Volumes	Time (mins)	% Solvent B	Flow (µL/min)
2.5	0-3.5	8	250
10	3.5-17.5	8-61	250
5	17.5-24.5	61-80	250
5	24.5-31.5	80-100	250
5	31.5-38.5	100	250
2.5	38.6-45.0	8	250

Analytes were ionised using a Thermo Ion Max Source operated in ESI mode, and data was acquired using a Thermo TSQ mass spectrometer in MRM mode, with a cycle time of 3s, a total of 22 transitions and an average peak width of 0.70 FWHM. Data were acquired using Xcalibur 3.0.63 (Thermo Fisher Scientific, USA).

Standards were analysed on both the previously used Aqua and the newer Plus columns on the gradients described in Table 3.1 and Table 3.2 respectively. The standards’ retention times from each experiment were plotted on X- and Y-axes as a linear graph, providing an equation that can be used to predict the retention time correlation between the two columns.

3.3.2 Clinical groups

Patient urine samples had previously been stratified according to disease status: active TB disease (ATB), latent TB infection (LTBI), and non-TB lung infection (NTB) (Figure 3.1). ATB disease was defined as patients with a positive result for the presence of Mtb in culture and smear tests, as well as the patient displaying characteristic clinical symptoms of TB infection. The ATB group was further split into those who tested positive for the presence of LAM in urine (LAMp) and those who tested negative for the presence of LAM in urine (LAMn). LTBI was defined as asymptomatic patients who had tested negative for Mtb in culture and smear tests but had tested positive on interferon gamma release assay (IGRA). NTB patients were defined as patients who tested negative in IGRA and tuberculin skin tests, as well as negative in TB smear and culture tests, but had a positive diagnosis for another respiratory disorder. Urine was collected from patients and stored at -80°C prior to extraction.

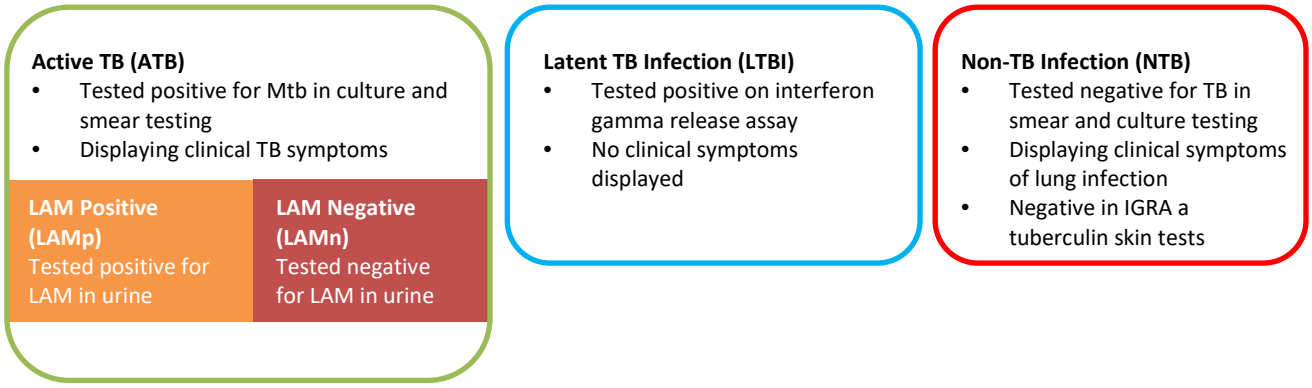


Figure 3.1: Patient stratification according to clinical groups.

3.3.3 Preparation of clinical urine extract

Lipids were extracted from urine by aliquoting 5ml of urine and passing through a 0.45µM filter and then a 0.22µM filter. Water : ethylacetate (5:7) with 6M hydrogen chloride was added and the mixture vortexed for 30 minutes. The urine sample was then centrifuged at 3320 x *g* for 10min. Supernatant was collected into a clean 50ml amber vial that had been pre-weighed. This process was repeated two more times and the supernatants combined each time. A diverse set of organic solvents were added to the residue from the previous step and after each addition the sample was vortexed for 30min, centrifuged at 3320 x *g* for 10min, and the supernatants combined with that from previous steps. This was performed three times for each organic solvent in the following order:

1. methanol : chloroform (2:1) with 14µM potassium chloride
2. chloroform;
3. methanol : chloroform (1:2);
4. methanol : chloroform (1:1).

Combined supernatants were dried down by vacuum centrifugation in the same pre-weighed amber vial as before. Dried extract was weighed, and the extraction yield determined by weight. Lipid extract was reconstituted in chloroform: methanol (1:1) and stored at -25°C until further use.

3.3.4 Preparation of pooled samples

Nine pools were created by aliquoting 80µg of lipid extract from each sample and combining them according to Table 3.3. All the pools were dried down and the concentration corrected to 4µg/µl. The pools were acidified to 0.1% formic acid for LCMS analysis.

Table 3.3: Pools created from urinary lipid extracts of clinical samples.

ATB pools	Samples	**LTBI pools**	Samples	**NTB pools**	Samples
LAMp1	TB_144 TB_167 TB_137 TB_145	LTBI1	LT_10 LT_12 LT_15µA LT_18	NonTB1	NT_LAMP157 NT_L257 NT_L280 NT_L283
LAMn2	TB_122 TB_133 TB_N340	LTBI2	LT_10µA LT_12µA LT_13µA LT_20 LT_21 LT_37	NonTB2	NT_1 NT_17 NT_5 NT_9
LAMn3	TB_N134 TB_N51 TB_TS121	LTBI3	LT_4 LT_5 LT_7 LT_8 LT_12	NonTB3	NT_L029 NT_L060 NT_L261 NT_L281 NT_L259

3.3.5 Data-dependent MS/MS data acquisition on the Q-Exactive

A total of 80µg lipid extract was injected onto the Plus column and the gradient employed as previously described in Table 3.2, section 3.3.1. Analytes were ionised using a Thermo Ion Max Source operated in negative or positive ESI mode and data was acquired using the Thermo Q-Exactive operated in data-dependent mode with inclusion lists. Full MS1 resolution was 70,000 and the maximum injection time (IT) was 100ms. The resolution for MS/MS was 17,500, the maximum injection time was 50ms, the loop count was 10, the isolation window was 4.0 *m/z* and the collision energy set to 35. Four inclusion lists were created based on the biomarker candidates originally observed in each sample subset: Active TB positive mode, Active TB negative mode, LTBI positive mode, and LTBI negative mode. Additionally, a method was used with no inclusion list for the NTB samples where no biomarker candidates were expected. Refer to Table 5.2 in the appendices for the full inclusion lists.

Each pool was analysed with the inclusion list of biomarker candidates that were previously seen in that cohort (Table 3.4). A separate injection was done for positive mode, and for negative mode.

Table 3.4: Sample pools analysed according to their corresponding inclusion lists. Please see Table 5.2 in the appendices for further information on the inclusion lists.

Sample Pool	Inclusion lists
LamP1 LamN2 LamN3	1. Active TB (positive mode) 2. Active TB (negative mode)
LTBI1 LTBI2 LTBI3	1. LTBI (positive mode) 2. LTBI (negative mode)
NonTB1 NonTB2 NonTB3	DDA mode only; no inclusion list

Data was processed in Xcalibur 2.2 (Thermo Fisher, 2011). Using the Chromatogram Ranges function, full MS/MS spectra for each biomarker candidate precursor were extracted and the top ten intense *m/z* values selected as a set of preliminary MRM transitions. Of the 70 biomarker candidates, 13 were measured with MS/MS spectra and thus proceeded to the preliminary MRMs. In many cases the precursor that was measured was not exactly the mass that was expected and a maximum deviation of 0.09 *m/z* units from the original mass was allowed. These minor deviations are expected as the Agilent qTOF 6530 has a mass accuracy of 5ppm while the Q-Exactive has a mass accuracy of 1ppm. For example, the biomarker candidate 146.0927 was in the inclusion list, however an ion of *m/z* 146.1185 was observed in one pool, and an ion of *m/z* 146.0865 observed in another. Both ions proceeded to preliminary MRMs as separate transition sets. Please see Table 5.3 in the appendices for the transition sets applicable to each sample pool.

3.3.6 Preliminary MRM development on the Thermo TSQ

Transition sets were assembled into MRM assays for each sample pool and these were refined using the same sample pools as the DDA experiment. Pools were prepared as previously described in section 3.3.4, with the exception that 100µg of lipid extract was injected onto the Plus column and the gradient was employed as previously described in Table 3.2, section 3.3.1.

Analytes were ionised using a Thermo Ion Max Source operated in ESI mode with polarity switching, and data acquired using the Thermo TSQ mass spectrometer in MRM mode with a cycle time of 3s, a maximum of 35 transitions per method, collision energy of 35, and an average peak width of 0.70 FWHM. Where there were more than 35 transitions per pooled sample, the transition set was divided over two or three methods (i.e. two or three separate injections) to maximise the points per peak as these were unscheduled experiments. LAMp1 had a total of 71 transitions targeted over two methods; LAMn2 had a total of 69 transitions targeted over two methods; LAMn3 had a total of 33 transitions targeted using a single method; LTBI1 had a total of 22 transitions targeted using a single method; LTBI2 had a total of 88 transitions targeted over three methods; LTBI3 had a total of 33 transitions targeted using a single method; NTB1 had a total of 88 transitions targeted over three methods; and NTB3 had a total of 33 transitions targeted over a single method. These transition sets are detailed in Table 5.3 in the appendices.

Data were analysed in Skyline 3.7.0.10940 (University of Washington, 2015). Transition sets were manually generated in Skyline using the small molecule manual insert function, and the raw MRM data from the Thermo TSQ were imported. Peaks were integrated according to the expected retention

time, which was calculated using the column calibration equation generated in section 3.3.1. Product ions that did not co-elute and form a uniform peak were excluded, thus refining each MRM to just a few transitions. Further, where there had been two slightly different MS1 *m/z* values for the same candidate biomarker, retention time and fragment ions were compared. For example, candidate biomarker 817.329 was observed with *m/z* 817.331 and 817.336 in the DDA experiment. These were both treated as separate transition sets in the preliminary MRM, but the results showed that 817.331 and 817.336 had a common retention time of 16.9 minutes and shared common product ions, thus suggesting that they were in fact the same molecule. Therefore, these two separate transition sets were merged into one with a precursor *m/z* of 817.3.

The output for this experiment was a final set of 68 transitions for 13 biomarker candidates. This was compiled into one scheduled MRM method.

3.3.7 MRM validation using clinical samples

The final MRM method was validated using the following individual clinical urine extracts:

LTBI group (n=13)	**NTB group (n=12)**	**ATB group (n=8)**
LT_12	NT_LAMP157	TB_TS121
LT_4	NT_L257	TB_167
LT_10	NT_1	TB_137
LT_8	NT_L283	TB_145
LT_5	NT_17	TB_122
LT_10µA	NT_5	TB_133
LT_13µA	NT_L029	TB_N340
LT_21	NT_L060	TB_N134
LT_37	NT_L261	
LT_7	NT_L281	
LT_18	NT_L259	
LT_15µA	NT_9	
LT_20		

The extracts were prepared for LCMS analysis by correcting the concentrations to 5µg/µl in chloroform : methanol (1:1) and acidifying to 0.1% formic acid. Once prepared, 100µg of lipid extract was injected onto the Plus column and the gradient employed as previously described in Table 3.2, section 3.3.1.

Analytes were ionised using a Thermo Ion Max Source operated in ESI mode with polarity switching, and data were acquired using the Thermo TSQ mass spectrometer in scheduled MRM mode with a cycle time of 2s, a total of 68 transitions scheduled over the 45-minute gradient, with a collision energy of 35, and an average peak width of 0.40 FWHM. This experiment was repeated under the same conditions two weeks later. In both analyses the sample order was randomised.

Data were analysed in Skyline 3.7.0.10940 using the previously refined document. Peaks were integrated based on precursor and product ion alignment and where there was uncertainty, predicted retention time was used to integrate the correct peak.

3.3.8 Quality checks

For all MRM experiments on the Thermo TSQ, machine performance was monitored using the standard mixture and MRM method described in the “Cross-column calibration” experiment (Section 3.3.1). A standard mixture was analysed at the beginning of the sequence, after every 10 samples, and

once again after the last sample. Data were processed using Skyline where they were evaluated for peak area and retention time shifts.

3.4 Results and Discussion

3.4.1 Cross-column calibration

Two columns were compared, and a cross-column calibration performed by drawing a line through all the retention time points which produced a linear equation (Figure 3.2). This regression equation can be used to predict the retention time of the biomarker candidates from the Plus column, based on their measured retention times from the Aqua column. This is an adaptation of the proteomic indexed retention time (iRT) method which uses peptides of known elution times to compare and calibrate different columns (S. J. Parker et al., 2015).

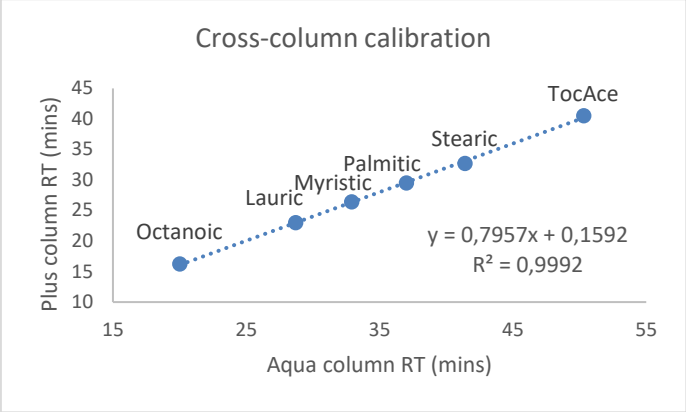


Figure 3.2: Cross-column calibration results. A trendline (dotted line) was drawn through the points with the R-squared value and linear equation displayed on the graph.

3.4.2 Data dependent acquisition on the Q-Exactive

The total number of biomarker candidates being screened was 70. There were 30 positive mode candidates and 40 negative mode candidates. The biomarker candidates were split into three groups based on their MS1 observations in the original discovery experiment: LTBI (21 candidates), Active TB and/or LTBI (24 candidates), and Active TB (25 candidates). Based on the DDA experiments, the LTBI group was refined to five possible candidates, the Active TB/LTBI group was refined to three possible candidates and the Active TB group was refined to five possible candidates. These candidates were considered possible based on the generation of MS/MS spectra and therefore their potential for MRM targeting. Biomarker candidates for which no MS/MS measurements were made were excluded from subsequent analyses. The possible reasons for MS/MS data not being acquired for 57 biomarker candidates are: 1) very low abundant molecules were not selected for MS/MS measurement due to compromised machine performance, i.e. a false negative results, 2) the mass error on the Agilent qTOF 6530 generated MS1 measurements, which were used in the inclusion lists, but may not have accurately represented the mass of the molecule, 3) biomarker candidates present in few samples would fall below the LLOD in a pool of samples where they are significantly diluted, or 4) the larger lipids, likely to be glycolipids, degraded over the four years of sample storage. Table 5.3 in the appendices displays the masses observed in each group across all the pools.

The observed MS1 *m/z* for each biomarker candidate was marginally different from the expected MS1 *m/z* and this is to be expected, as the candidate list was identified using an Agilent qTOF 6530 whereas in this experiment a Thermo Q-Exactive Orbitrap machine was used. In some cases, the same

candidate was observed across different pools but with a marginally different MS1 *m/z*. When looking at the MS/MS spectra from these candidates, some of them had common spectra, suggesting that they may be the same molecule. They were however all treated as separate molecules until further validation by MRM.

It appears that positive mode ions and ions with smaller mass were favoured. The lipid samples that were used were extracted in 2014, and while they were stored at -25°C multiple freeze-thaw cycles would have occurred in that time. Literature suggests that freeze-thaw cycles have little effect on lipids in a sample (Zivkovic et al., 2009), but duration of storage leads to degradation and changes in the lipidome of a sample (Soyer, et al., 2010). Given that these samples were analysed four years after the lipids were extracted, some changes are to be expected, particularly in the larger more complex lipids. The tendency towards positive mode ions may have been because formic acid was used as the ion-pairing agent. Formic acid is a proton donor and is therefore more suited for positive mode ESI than negative mode ESI (Cech & Enke, 2001).

The output from this experiment was a list of observed MS1 *m/z* values for 13 biomarker candidates, each with a subset of MS/MS values which will be referred to as transition sets. Please see Table 5.3 in the appendices for these values. The transition sets which arose from each pool were compiled into an MRM experiment. Therefore, this experiment resulted in nine separate MRM methods; one for each of the pooled samples.

3.4.3 Preliminary MRM experiment on the Thermo TQS mass spectrometer

The MRM methods generated in the DDA experiment needed to be verified to assess which biomarker candidates produced co-eluting product ions and if these molecules eluted at the expected retention times.

3.4.3.1 LTBI biomarker candidates

In the DDA experiment the LTBI biomarker candidate 146.0927 (+1) was measured in the LAMn3 pool as 146.1185 *m/z* and in the LTBI2 pool as 146.0865 *m/z*. Based on the discovery experiments and column calibration equation, this molecule was expected to elute in the dead volume and not interact with the column. The transition sets for these two observations were quite different, already suggesting that 146.1185 and 146.0865 are different molecules. The preliminary MRM supports that they are likely not the same molecule as 146.1185 eluted at 8.0min with four co-eluting product ions; and 146.0865 eluted in the dead volume with only one product ion being measured (Figure 3.3). Given the lack of evidence in the dead volume, the next earliest eluting molecule with fragmentation was considered. Therefore 146.1185 at retention time 8.0min with four product ions was selected as the MRM for this biomarker candidate.

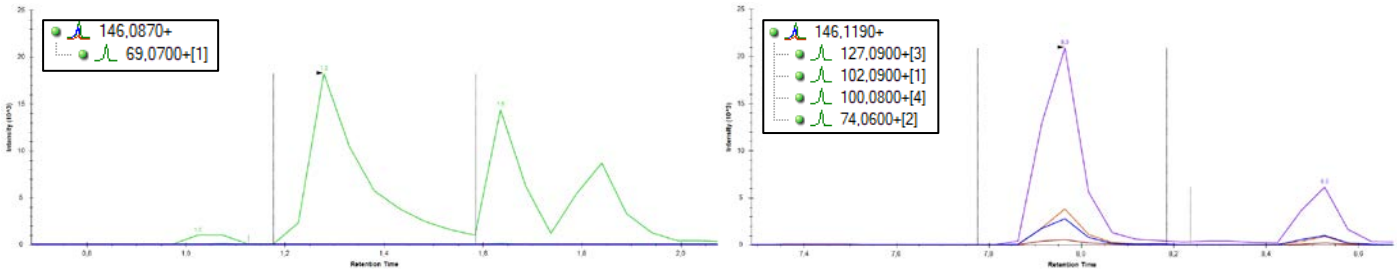


Figure 3.3: Chromatograms of the fragments from 146.0870 in the LTBI2 pool (left) and from 146.1185 in the LAMn3 pool (right).

In the DDA experiment the LTBI biomarker candidate 188.1999 (+1) was measured in the LTBI1 and LTBI3 pools as 188.1758 *m/z*; in the LTBI2 pool as 188.1759 *m/z*; in the NonTB2 pool as 188.1756 *m/z*; and in the NonTB3 pool as 188.1760 *m/z*. Based on the discovery experiments and column calibration equation this molecule was expected to elute at 9.8min. Peaks were observed at around 4min and 10min for this molecule. The fragmentation pattern was the same at both time points suggesting that this is the same molecule however the differences in retention time suggest not (Figure 3.4). One explanation could be that these are structural isomers of each other and therefore interact with the column in different ways, but the literature suggests that a C18 matrix, as used here, does not achieve isomer separation and further separation methods would have been required (Damen, et al., 2014; Kozlowski, et al., 2015; Kyle et al., 2016). It is therefore feasible that these are different molecules, but are too similar to be filtered out by the machine's mass error window.

Looking at Figure 3.4 below, it seems the same molecule is being measured at 10min in LTBI1, LTBI2 and NonTB3, while a similar chromatogram for a different molecule is being measured in LTBI3 and NonTB2 at ~4min. It seems this is not the same molecule in LTBI3 and NonTB2 because of the differences in fragmentation. Because of the uncertainty of the molecule eluting at ~4min, the peak at 10min was integrated with three product ions and selected as the MRM for this biomarker candidate.

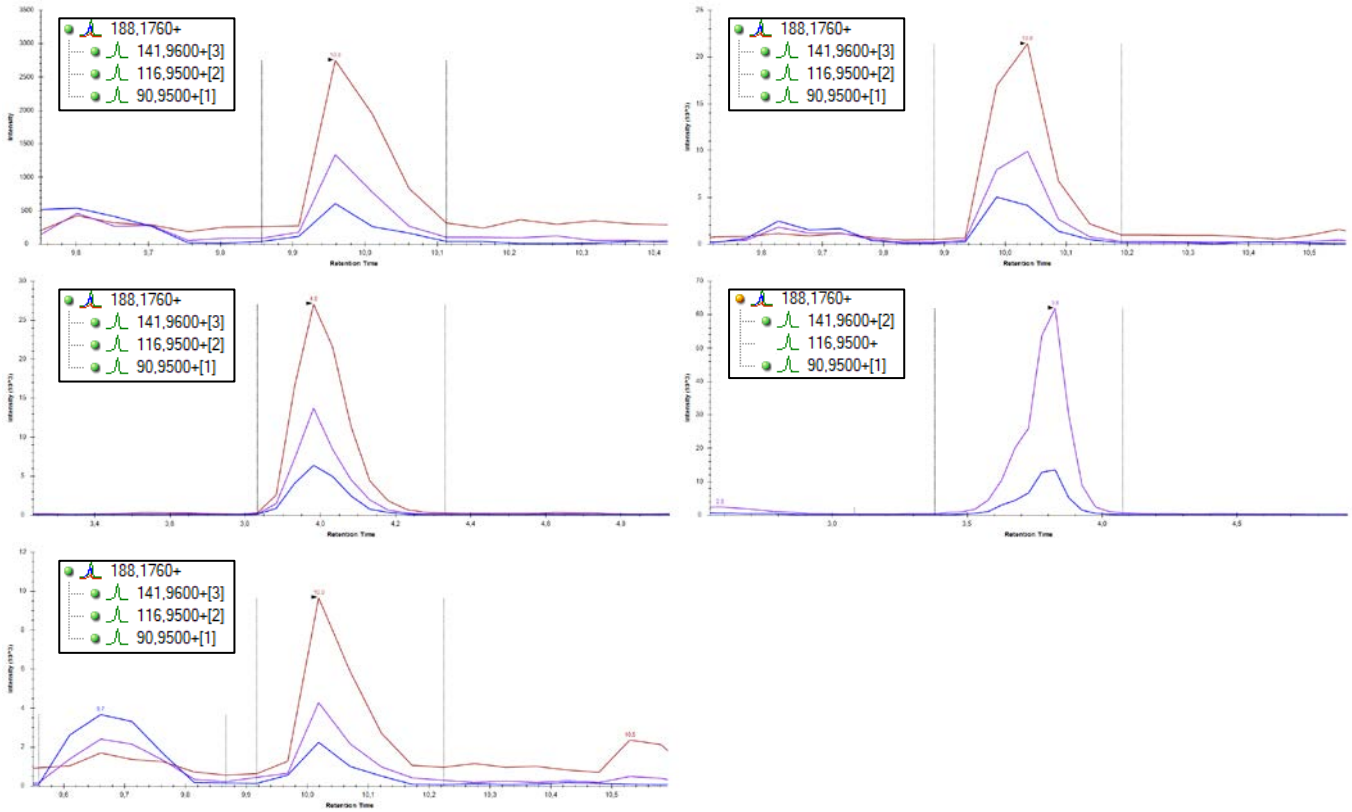


Figure 3.4: Chromatograms of the product ions for 188.1760 as measured in the LTBI1 pool (top left), LTBI2 pool (top right), LTBI3 pool (middle left), NonTB2 pool (middle right), and NonTB2 pool (bottom left).

The LTBI biomarker candidate 189.1348 (+1) had been observed in the DDA experiment in the LAMn2 pool as 189.0654 *m/z*; the LTBI3 pool as 189.1598 *m/z*; the NonTB1 pool as 189.0655 *m/z*; the NonTB2 pool as 189.0657 *m/z*; and the NonTB3 pool as 189.0658 *m/z*. Based on the discovery experiments and column calibration equation this molecule was expected to elute in the dead volume and have no column interaction. In the preliminary MRM, co-eluting product ions were found in the LAMn2 pool at 6.2min, in the LTBI3 and NonTB2 pools at 41.1min, and no co-eluting product ions were observed in the NonTB1 and NonTB3 pools. In Figure 3.5 below it appears that the same molecule was measured in LTBI3 and NonTB2 pools, however given the retention time it is improbable that this is the biomarker candidate of interest. Therefore, the fragment ions measured at 6.2min in the LAMn2 pool were integrated and selected as the transition set for this biomarker candidate.

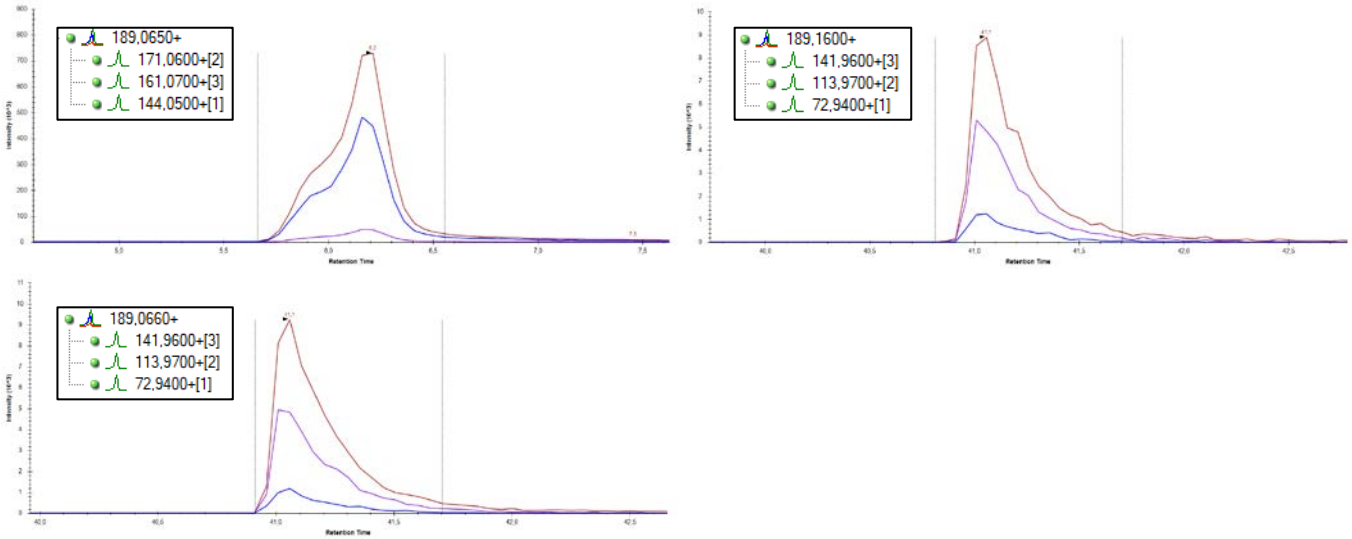


Figure 3.5: Chromatograms for the product ions for 189.0650 as observed in LAMn2 (top left), LTBI3 (top right), and NonTB2 (bottom left)

The LTBI biomarker candidate 233.0793 (+1) had been observed in the DDA experiment in the LTBI2 pool as 233.1212 *m/z*. Based on the discovery experiments and column calibration equation this molecule was expected to elute at 3.5min. In this preliminary MRM, the fragments were observed at 1.8min. As can be seen in Figure 3.6 these peaks are not well aligned, and this is not enough evidence to say they come from the same molecule. Despite the uncertainty, it was the best measurement obtained for this biomarker candidate and with caution this transition set was selected to be used as an MRM for further verification.

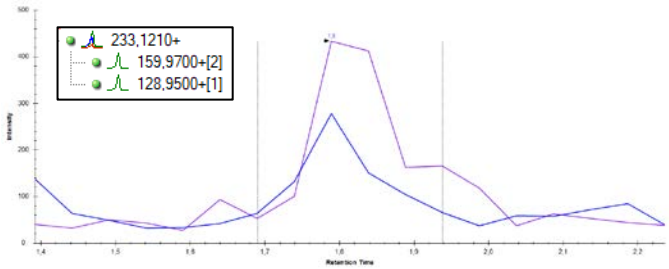


Figure 3.6: Chromatogram of the fragment ions for 233.1210 in the LTBI2 pool.

The LTBI biomarker candidate 332.3304 (+1) had been observed in the DDA experiment in the LAMn3 pool as 332.2811 *m/z*. Based on discovery experiments and the column calibration calculations this molecule was expected to elute at 19.2min. In this preliminary MRM five co-eluting fragments were observed at 15.6min which was within the acceptable retention time range, so this transition set was accepted as the MRM for this biomarker candidate (Figure 3.7).

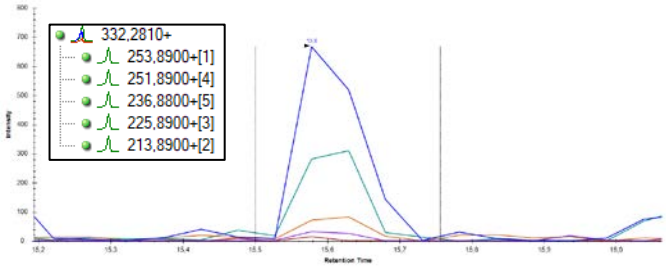


Figure 3.7: Chromatogram of the fragment ions for 332.2811 in the LAMn3 pool.

3.4.3.2 Active TB/LTBI biomarker candidates

The first biomarker candidate in this group is 284.3311 (+1) which was observed in the DDA experiment as 284.3318 *m/z* in the LAMp1 pool, as 284.2956 *m/z* in the LTBI2 pool, and as 284.2947 *m/z* in NonTB1 pool. Based on the discovery experiment and the column calibration equation, this molecule was expected to elute at the very end of the gradient at around 40-45min. A very weak signal was detected in the preliminary MRM for co-eluting product ions in LAMp1 and NonTB1 at 41min (Figure 3.8). Even though the signal was so low and different product ions were observed in each sample pool, these were combined to form one MRM assay for further verification. The cause of the very weak signal may be that this molecule is present in only a few of the samples that make up each pool, meaning that it has been diluted significantly in the pooling process. This can only be confirmed by analysing individual samples with the same MRM transition set.

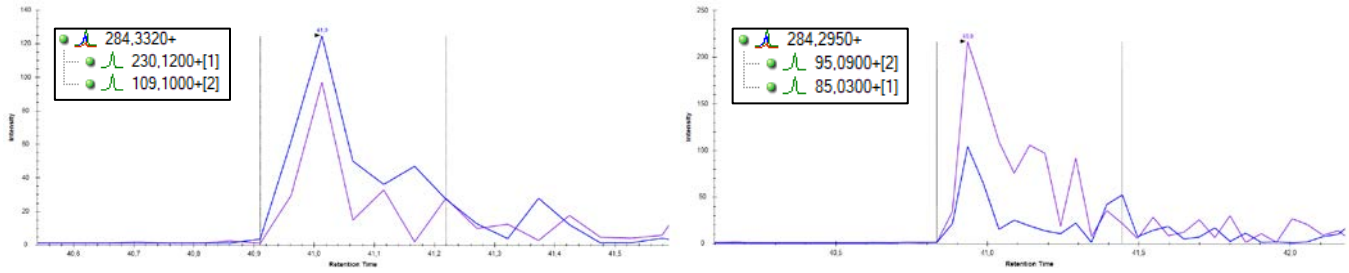


Figure 3.8: Transition sets observed for biomarker candidate 284.3311 as measured in the LAMp1 pool (left) and the NonTB1 pool (right).

The next biomarker candidate in the group is 403.2348 (+1) which was observed as 403.1794 *m/z* in the LAMn2 pool, as 403.1749 *m/z* in the NonTB1 pool, and as 403.1748 *m/z* in the NonTB3 pool in the DDA experiments. Based on the discovery experiment and the column calibration calculations, this molecule was expected to elute at 27.9min. In the preliminary MRM experiment a peak was detected at 28.6min in all three pooled samples (Figure 3.9). As can be seen in this figure, four product ions were co-eluting in the NonTB1 pool, while only two of these were co-eluting in the LAMn2 and NonTB3

pools, where the signal was also very weak. The product ions measured in the NonTB1 pool were used to create a transition list for this biomarker candidate's MRM assay.

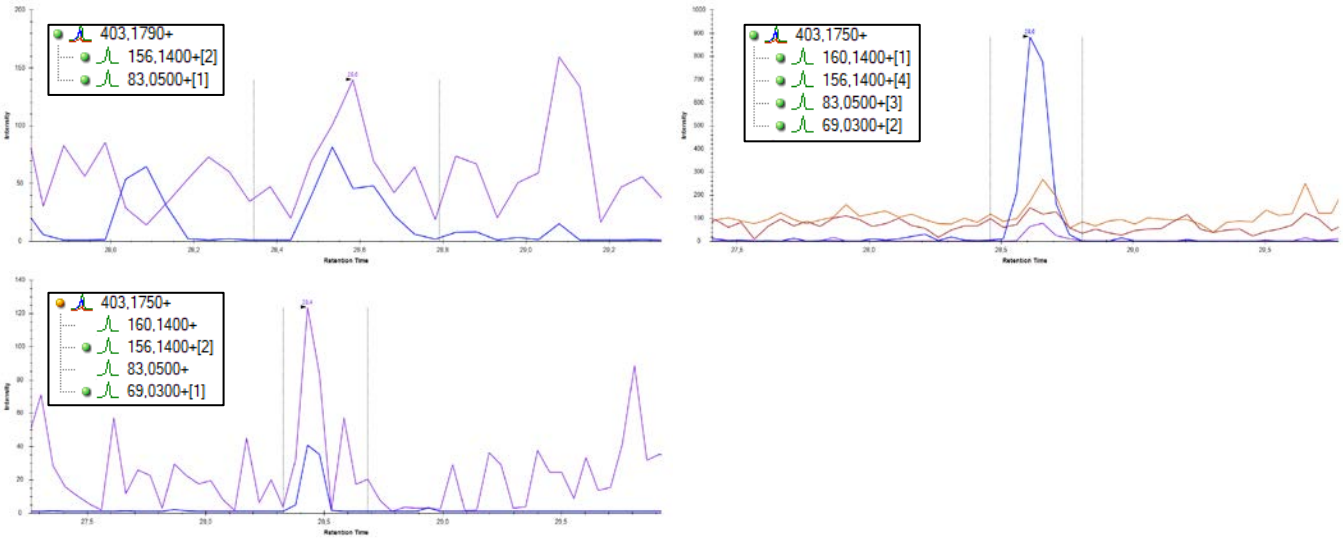


Figure 3.9: Chromatograms for the biomarker candidate 403.2348 which was measured as 403.1790 in the LAMn2 pool (top left), and 403.1750 in the NonTB1 (top right) and NonTB3 (bottom right) pools.

The next candidate in this group is 445.3637 (+1) which was observed in the DDA experiment in pool LAMp1 as 445.3675 *m/z*. The molecule was expected to elute at the end of the gradient at around 40-45 minutes. No co-eluting peaks were seen in the preliminary MRM experiment, so this biomarker candidate was removed from the list of those eligible for MRM assay development.

3.4.3.3 Active TB biomarker candidates

Biomarker candidate 284.1939 (-1) was observed in most of the sample pools in the DDA experiment as follows: in LAMp1 as 284.1692 *m/z*, in LAMn2 as 284.1501 *m/z*, in LAMn3 as 284.1500 *m/z*, in LTBI1 and LTBI2 as 284.1700 *m/z*, in NonTB1 as 284.1504 *m/z*, and in NonTB3 as 284.1512 *m/z*. The expected elution time of this molecule was 14.5min based on the discovery experiment and column calibration calculations. In the chromatograms (Figure 3.10), co-eluting fragments were measured in LAMp1, LAMn3 and LTBI2 at 12.4min. This is possibly the same molecule being measured in each pool as the transition sets overlap and the retention times are the same. The transition lists across the sample pools were combined to make one MRM assay for this candidate with a precursor mass of 284.1.

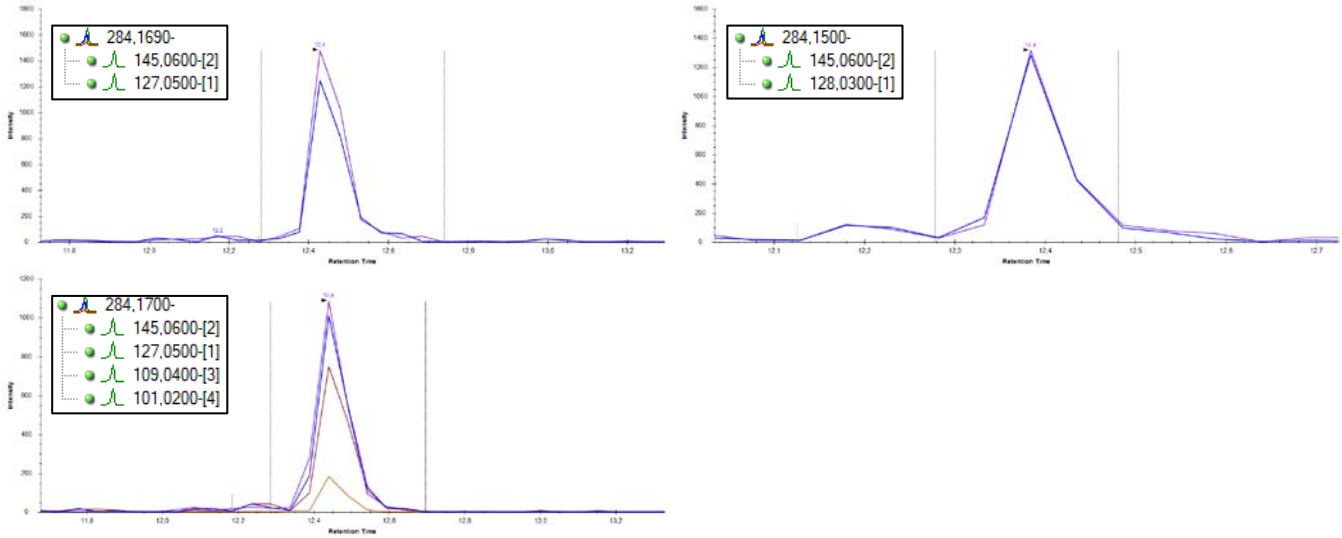


Figure 3.10: Chromatograms for biomarker candidate 284.1939 measured as 284.1690 in the LAMp1 pool (top left), as 284.15 in the LAMn3 pool (top right), and as 284.17 in the LTBI2 pool (bottom left).

The next biomarker candidate in this group is 432.2241 (-1) which was observed in the LAMn2 pool as 432.2025 *m/z*, the LTBI2 pool as 432.2033 *m/z* , the NonTB1 pool as 432.202 *m/z*, and in the NonTB3 pool as 432.2039 *m/z*. Based on the column calibration calculations and the discovery experiment this molecule was expected to elute at 12.9min. Co-eluting fragment ions for this biomarker candidate were observed in the LAMn2 pool, the LTBI2 pool, and the NonTB1 pool at 10.6min (Figure 3.11). There was an overlap in the transition sets which suggests this is the same molecule, and the transition sets were combined to generate the final MRM for this candidate with a precursor mass of 432.2 (-1).

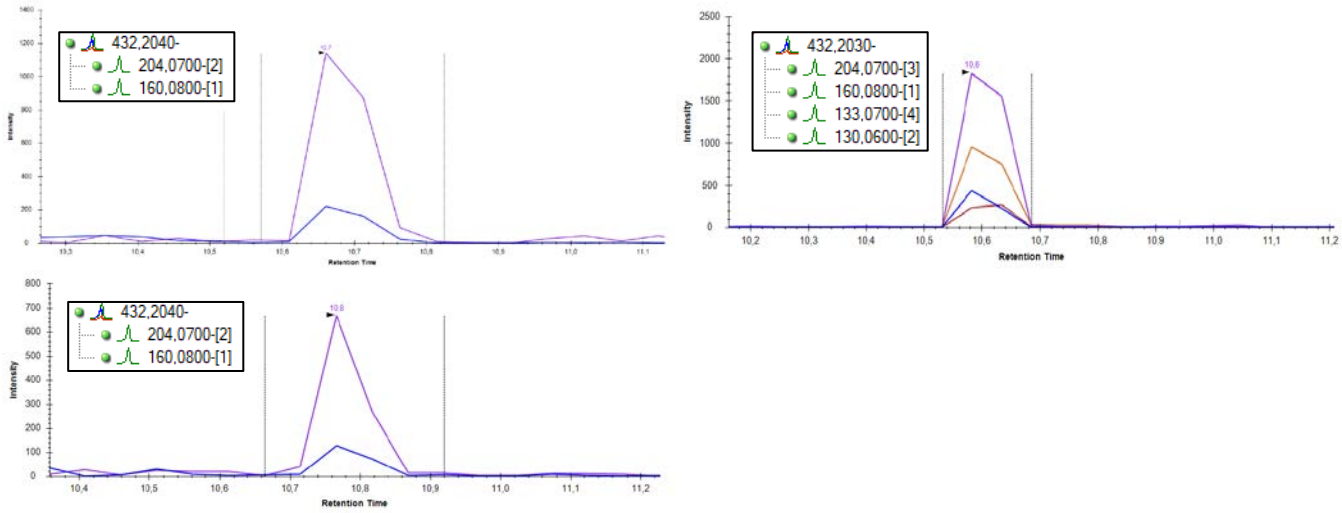


Figure 3.11: Chromatograms for biomarker candidate 432.2241 measured as 432.204 in the LAMn2 pool (top left), as 432.203 in the LTBI2 pool (top right), and as 432.204 in the NonTB1 pool (bottom left).

The next biomarker candidate in this group is 685.4386 (+1) which was measured in the DDA experiment as 685.4362 *m/z* in the LTBI2 and LTBI3 pools, each with a slightly different subset of product ions. This molecule was expected to elute at the end of the gradient at around 40-45 minutes. In the preliminary MRM, co-eluting product ions were measured at 40.9min in the LTBI2 pool only and at a very low signal. It is possible that this signal is coming from a small number of samples in the pool,

which would explain the low signal intensity. The chromatogram and transition set that was selected for the MRM assay is shown in Figure 3.12.

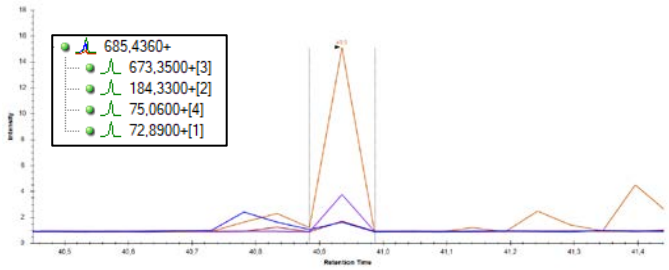


Figure 3.12: Chromatogram for biomarker candidate 685.4386 measured as 685.436 in the LTBI2 pool.

The next biomarker candidate was 785,3224 (-1) which was expected to elute at 26.1min according to the column calibration calculations and discovery experiments. In the DDA experiment it was observed as 785.3507 *m/z* in the NonTB1 pool. In the preliminary MRM experiment, no co-eluting product ions were measured in the NonTB1 pool at the expected retention time, however two co-eluting fragments were observed at 10.7min (Figure 3.13) and these were selected for future MRM validation, with the intention of excluding them later if their inclusion proved unhelpful.

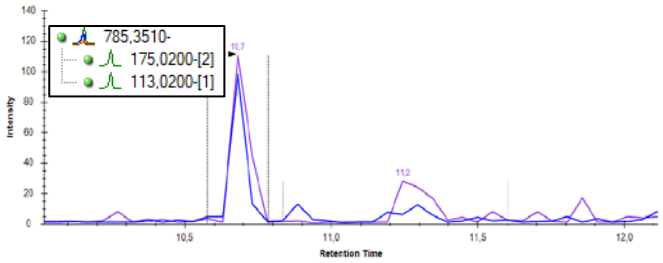


Figure 3.13: Co-eluting product ions for biomarker candidate 785.3224 measured as 785.351 in NonTB1.

The next biomarker candidate in this group is 817.3290 (+1) which was expected to elute at 18.6min. In the DDA experiment two possible masses were observed in the LAMp1 pool; 817.3309 *m/z* and 817.3313 *m/z*. In the LAMn2 pool the *m/z* 817.3361 was observed. The two precursor masses observed in the LAMp1 pool in the DDA experiment are likely the same molecule as these *m/z* values are 0.5ppm apart and the mass error of the Q-Exactive is ~1ppm (Strupat, et al., 2013). They were treated as such in the preliminary MRM and their MS/MS spectra from the DDA experiment were converged into the same transition list. Figure 3.14 below displays the chromatograms with the transition lists for each preliminary MRM experiment in LAMp1 and LAMn2 respectively. A peak was observed at 16.9min in both pools and some of the product ions overlapped which suggests that these are the same molecules. The product ion lists were combined to be used as the final MRM assay for this biomarker candidate.

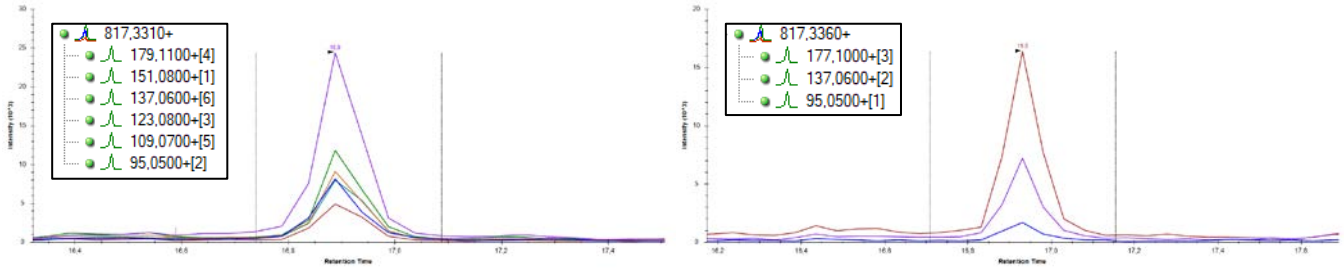


Figure 3.14: Chromatogram for biomarker candidate 817.3290 measured as 817.3310 in the LAMp1 pool (left) and as 817.3360 in the LAMn2 pool (right).

The final biomarker candidate in this group is 1273.6095 (+1) which was observed as 1273.613 *m/z* in the LAMp1 pool and as 1273.601 *m/z* in the NonTB3 pool. Based on the column calibration equation and results from the discovery experiment, this molecule was expected to elute at 31.5min. Three co-eluting product ions were measured at 29.2min in the LAMp1 pool and were not observed in the NonTB3 pool. Figure 3.15 below pictures the chromatogram and transition set.

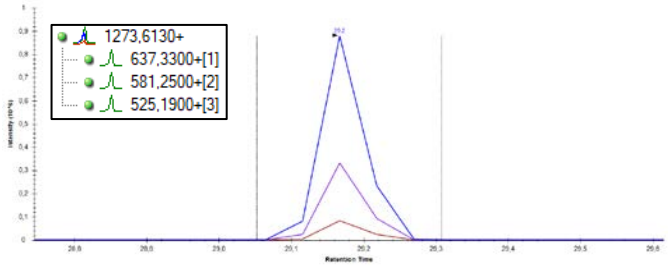


Figure 3.15: Chromatogram for biomarker candidate 1273.6095 measured as 1273.6130 in the LAMp1 pool.

The result from this preliminary MRM experiment was a tentative MRM assay for 13 biomarker candidates which needed to be put through further validation on individual samples, rather than pooled samples. This experiment also helped to filter out product ions that had been observed on the Q-Exactive orbitrap but were not observed on the Thermo TSQ mass spectrometer. The reason for the differences in MS/MS spectra may be because these two machines have different collision energy cells: the Q-Exactive uses a higher energy collisional dissociation (HCD) cell while the Thermo TSQ operates with a collision induced dissociation (CID) cell. HCD and CID both induce fragmentation by the same mechanism of raising the energy of the ions and forcing them to collide with neutral atoms. The difference is that HCD operates at a slightly higher energy than CID, and HCD occurs separate from the mass analyser cell. While both collision methods are similar, it is expected to see some differences in the fragmentation patterns (Demarque, et al., 2016). These results do show, however, that the fragmentation patterns are sufficiently consistent between the two cells to use them interchangeably.

Another reason for the differences are that the product ions measured in the Q-Exactive were in fact a result of chemical noise, and not part of the biological molecule. Therefore it is important to validate any MRM that was generated using a full scan mass spectrometer, as it allows noise to be filtered out.

3.4.4 MRM verification using clinical samples

At this stage, the MRM assays for each candidate biomarker had been generated using pooled samples that theoretically represented each patient group. This was a prudent way to minimise sample loss and machine analysis time while developing and refining the assays, however it is important to validate these MRM assays on individual samples as the pools may not be representative of the individual samples. The aim of this section of the study was to establish 1) how reproducible each MRM is by doing repeat injections, 2) how feasible it is to use each MRM for quantitation, 3) how variable the retention times are for each MRM, and 4) how prevalent the biomarker candidates are across this sample set. On the fourth point, it is difficult to predict the diagnostic ability of this set of biomarker candidates because of the limitations in the clinical samples: These samples were four years old at the time of measurement, so it is expected that the lipids have degraded or been modified over time. These samples were analysed in duplicate with the second batch undergoing LCMS analysis two weeks later than the first. It was found that from batch 1 to batch 2, the peaks consistently shifted to an earlier retention time. The type of variability between experiments is acceptable as retention time is not considered a fixed property and can vary slightly between experiments due to day-to-day fluctuations in room temperature or solvent batches.

3.4.4.1 LTBI biomarker candidates

The first biomarker candidate in this group is 146.119 (+1) which was targeted with the following MRM transition set:

Precursor	>	Product ions
146.119 (+1)	>	74.06
146.119 (+1)	>	100.08
146.119 (+1)	>	102.09
146.119 (+1)	>	127.09

The results for this MRM showed two clear peaks with the same fragmentation approximately 30 seconds apart from each other. Each peak was present in some samples and not others, and some samples contained both peaks (Figure 3.16). It was decided that the later eluting peak at 8.9min be integrated as it was closer to the previously observed retention time. Where the integrated peak was present in one sample it was always present in the repeat injection showing that this MRM assay is reproducible. See Table 3.5 for a summary of results for this biomarker candidate. This molecule was observed in 7/8 ATB samples, in 11/13 LTBI samples, and in 12/12 NTB samples.

Table 3.5: Results summary for 146.119 (+1) MRM assay.

No. of samples observed in 30/33	No. of repeat observations 30/30	RT batch 1 ~8.9 mins	RT batch 2 ~8.3 mins	RT CV 8.5%

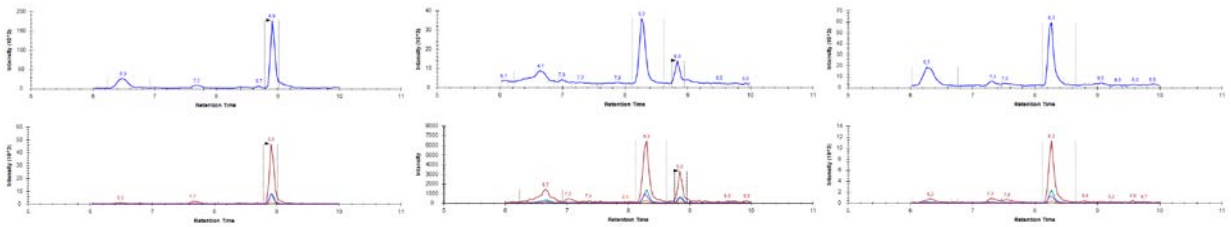


Figure 3.16: Peaks eluting for biomarker candidate 146.119. A peak was observed at either ~8.9 minutes (left), both ~8.3 and ~8.9 minutes (middle), or at just ~8.3 minutes (right). Blue peaks in the upper panels represent the precursor, while the bottom panels represent the product ions.

The next biomarker candidate in this group is 188.176 (+1) which was measured using the following MRM transition set:

Precursor	>	Product ions
188.176 (+1)	>	90.95
188.176 (+1)	>	116.95
188.176 (+1)	>	141.96

The results for this MRM showed one clear peak with co-eluting fragments at ~10.2min in the first batch, and then at ~9.7min in the second batch (Figure 3.17). Table 3.6 summarises the results for this candidate biomarker's MRM assay which was observed in 8/8 ATB samples, 8/13 LTBI samples, and 7/12 being in the NTB samples.

Table 3.6: Results summary for 188.176 (+1) MRM assay.

No. of samples observed in 23/33	No. of repeat observations 23/23	RT batch 1 ~10.2 mins	RT batch 2 ~9.7 mins	RT CV 10%

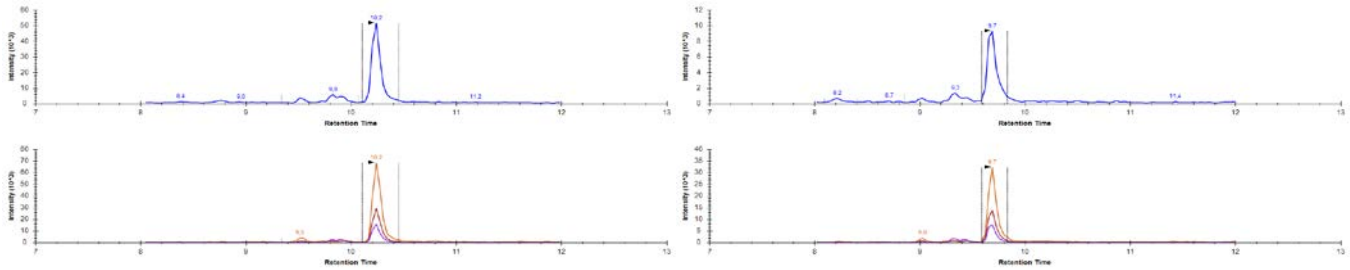


Figure 3.17: An example of a peak for precursor 188.176 (top panel in blue) and its coeluting product ions (bottom panel in multicolours) appearing at 10.2 minutes in batch 1 (left), and at 9.7 minutes in the same experiment repeated two weeks later (batch 2; right).

The next biomarker candidate in this group is 189.0 (+1) which was analysed using the following MRM transition set:

Precursor	>	Product ions
189.0 (+1)	>	130.09
189.0 (+1)	>	131.98
189.0 (+1)	>	144.05
189.0 (+1)	>	161.07
189.0 (+1)	>	171.06

The results for this MRM showed one clear peak with co-eluting product ions at ~6.6min in the first batch, and then at ~5.6min in the second batch (Figure 3.18). Where the integrated peak was present in one sample it was present in the repeat injection except for in sample NT_L157 where the first injection showed no peaks for this biomarker candidate, and the repeat injection showed a convincing peak. Each batch was freshly prepared for LCMS and sample handling in the first batch may have resulted in this analyte being degraded. Other MRMs were measured repeatedly in this sample, indicating that this specific analyte may be sensitive to perturbations while others are not. This biomarker candidate was observed in 3/8 ATB samples and in 2/13 LTBI samples with one irreproducible observation being made in one NTB sample. A summary of results for this candidate biomarker is detailed in Table 3.7. In previous work (Mlamla, 2018) this molecule was predicted to have the molecular formula $C_7H_{16}N_4O_2$. The most abundant product ion in this MRM is *m/z* 144.05

which is consistent with the fragmentation of homoarginine and targinine (Mass Bank of North America, 2007), both of which share this chemical formula.

Table 3.7: Results summary for 189.0 (+1) MRM assay.

No. of samples observed in 6/33	No. of repeat observations 5/6	RT batch 1 ~6.6 mins	RT batch 2 ~5.6 mins	RT CV 6%

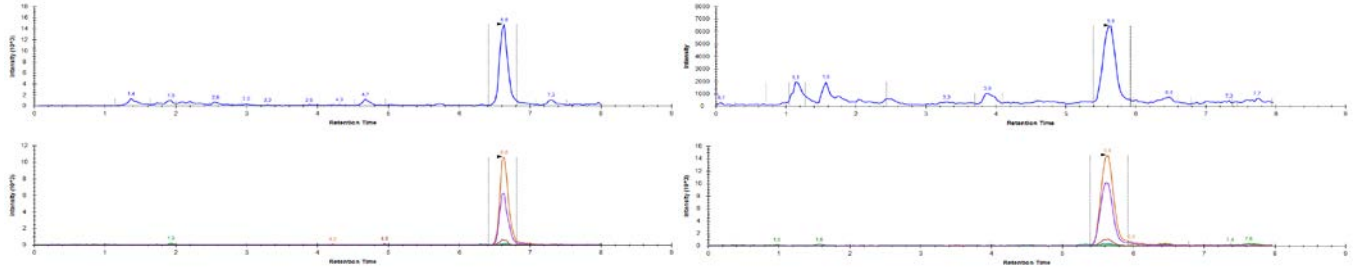


Figure 3.18: An example of a peak for precursor 189.0 (top panel in blue) and its coeluting product ions (bottom panel in multicolours) appearing at 6.6 minutes in batch 1 (left), and at 5.6 minutes in the same sample repeated two weeks later (batch 2; right).

The next biomarker candidate in this group is 233.12 (+1) which was targeted with the following MRM transition set:

Precursor	>	Product ions
233.12 (+1)	>	128.95
233.12 (+1)	>	159.97

The results for this MRM showed one clear precursor peak with co-eluting fragments at ~2.0min in the first batch, and then at ~1.7min in the second batch (Figure 3.19). The peak shapes for the fragments were not very well matched indicating that they may not originate from the same precursor. Where the integrated peak was present in one sample it was present in the repeat injection indicating that this MRM is reproducible. This biomarker candidate was observed in 1/8 ATB samples, in 7/13 LTBI samples, and in 3/12 NTB samples. The data for this MRM show a reliable measurement that is favoured in the LTBI group, however the shapes of the product ion peaks are a cause for caution. It would be advised that just one product ion (128.95 *m/z*) be used to measure this analyte as its shape closely matches the precursor ion shape. A summary of the results for this biomarker candidate are in Table 3.8.

Table 3.8: Results summary for 233.12 (+1) MRM assay.

No. of samples observed in 11/33	No. of repeat observations 5/6	RT batch 1 ~2.0 mins	RT batch 2 ~1.7 mins	RT CV 2%

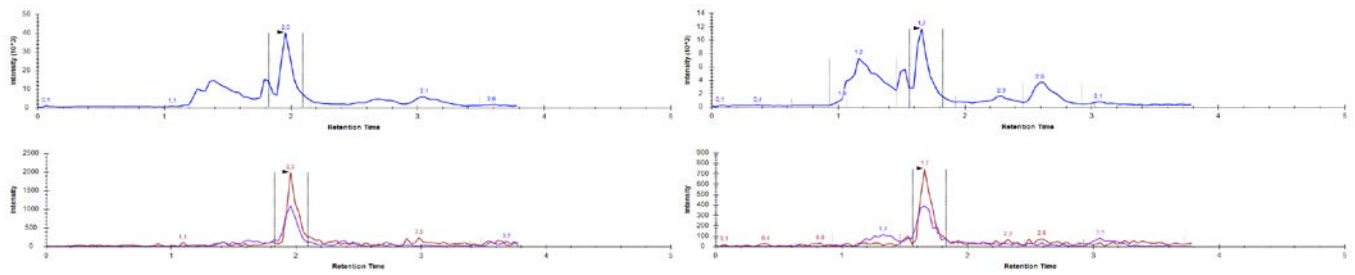


Figure 3.19: An example of a peak for precursor 233.12 (top panel in blue) and its coeluting product ions (bottom panel in multicolours) appearing at 2.0 minutes in batch 1 (left), and at 1.7 minutes in the same sample repeated two weeks later (batch 2; right).

The next biomarker candidate in this group is 332.281 (+1) which was measured using the following MRM transition set:

Precursor	>	Product ions
332.281 (+1)	>	213.89
332.281 (+1)	>	225.89
332.281 (+1)	>	236.89
332.281 (+1)	>	251.89
332.281 (+1)	>	253.0

The results for this MRM showed two peaks with co-eluting fragments at ~16.4min and at ~15.8min. While the earlier peak was closer to the expected retention time of 15.6min, its appearance was erroneous, and it was rarely present in both the sample and its repeat. The fragmentation of the peak at 15.8min was also inconsistent between samples, so the peak at 16.4min was selected for this MRM. Where the integrated peak was present in one sample it was present in the repeat injection indicating that this MRM is reproducible, however in sample TB_N134 the peak and its product ions were observed in the repeat, but not in the first injection. In the first injection for this sample, only the precursor ion was measured (Figure 3.20) while in the repeat inject the product ions co-eluted and formed clear peaks. While the MS1 intensity is similar in both experiments, the MS/MS intensity in the first repeat appears to fall into the noise signal. This may be why the product ion peaks are unclear and could have been a batch-specific machine error. This biomarker candidate was observed in just three of the 33 clinical extracts, with one observation in each clinical group. Table 3.9 summarises the results from this MRM assay.

Table 3.9: Results summary for 332.281 (+1) MRM assay.

No. of samples observed in 3/33	No. of repeat observations 2/3	RT batch 1 ~16.4 mins	RT batch 2 ~15.8 mins	RT CV 15%

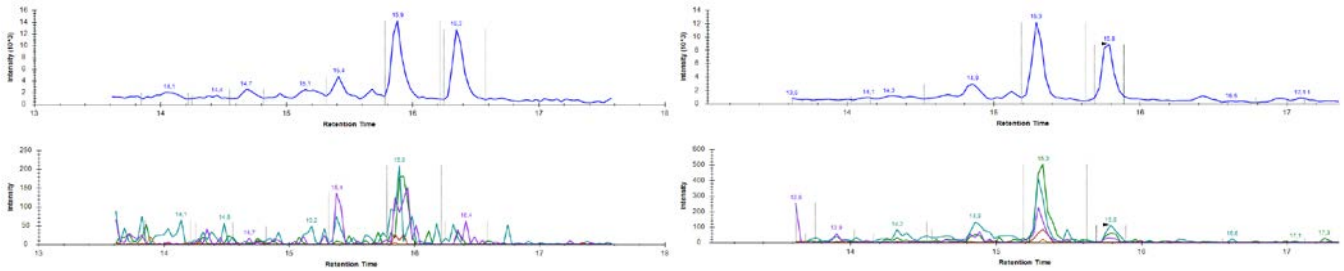


Figure 3.20: An example of a peak for precursor 332.281 (top panel in blue) and its coeluting product ions (bottom panel in multicolours) appearing in sample TB_N134 (left panel) and its repeat injection two weeks later (right panel).

3.4.4.2 Active TB/LTBI biomarker candidates

The first biomarker candidate in this group is 284.3 (+1) which was analysed using the following MRM transition set:

Precursor	>	Product ions
284.3 (+1)	>	85.03
284.3 (+1)	>	95.09
284.3 (+1)	>	109.1
284.3 (+1)	>	230.12

The results for this MRM showed one peak at ~40.8min with the product ions present but not co-eluting (Figure 3.21). This peak was not reproducible between the sample and its repeat, indicating that it is not a reliable MRM and should not be used to measure this biomarker candidate.

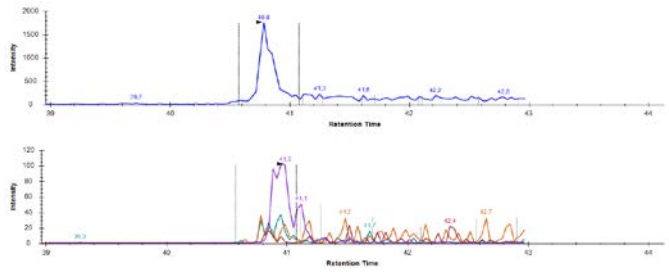


Figure 3.21: An example of a peak for precursor 284.3 (top panel in blue) and the measured product ions (bottom panel in multicolours).

The other biomarker candidate in this group is 403.17 (+1) which was targeted with the following MRM transition set:

Precursor	>	Product ions
403.17 (+1)	>	69.03
403.17 (+1)	>	83.05
403.17 (+1)	>	120.06
403.17 (+1)	>	156.14
403.17 (+1)	>	160.14

The results for this MRM showed a clear peak at 28.9min in the first batch and the same peak eluted at 28.0min in the repeat batch (Figure 3.22). In this MRM there appeared to be some interference on the MS1 scan, however at the MS/MS level a clear peak with co-eluting product ions was observed. Where the integrated peak was present in one sample it was present in the repeat injection indicating that this MRM is reproducible, however in sample NT_L157 the peak and its product ions were observed in the repeat, but not in the first injection. A similar inconsistency appeared in the same sample for one of the LTBI candidates (189.0) indicating that this may be more of a problem with the sample than with the MRM itself. This biomarker candidate was observed in 2/8 ATB samples, and in 5/8 in the NTB samples. Table 3.10 summarises these results.

Table 3.10: Results summary for 403.17 (+1) MRM assay.

No. of samples observed in 7/33	No. of repeat observations 6/7	RT batch 1 ~28.9 mins	RT batch 2 ~28.0 mins	RT CV 28%

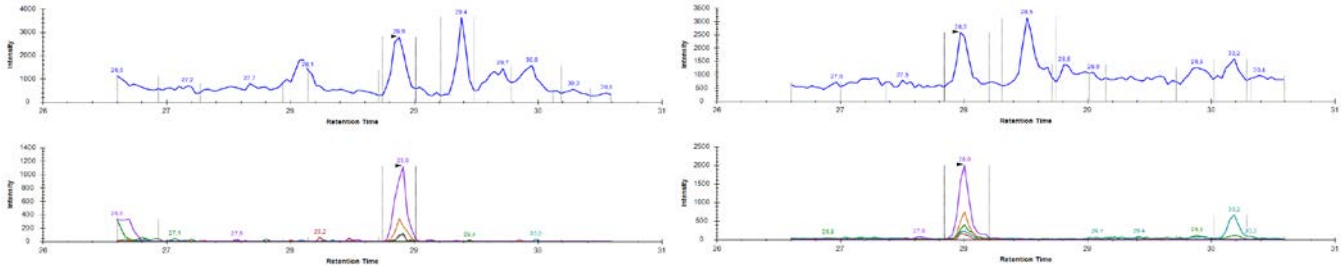


Figure 3.22: An example of a peak for precursor 403.17 (top panel in blue) and its coeluting product ions (bottom panel in multicolours) appearing at 28.9 minutes (left panel) and at 28.0 minutes in its repeat injection two weeks later (right panel).

3.4.4.3 Active TB biomarker candidates

The first biomarker candidate in this group is 284.1 (-1) which was measured using the following MRM transition set:

Precursor	>	Product ions
284.1 (-1)	>	101.02
284.1 (-1)	>	109.04
284.1 (-1)	>	127.05
284.1 (-1)	>	128.03
284.1 (-1)	>	145.06

The results for this MRM showed a clear peak at ~12.7min in the first batch and the same peak eluted at ~12.1min in the repeat batch (Figure 3.23). In this MRM there appeared to be some interference on the MS1 scan, however at the MS/MS level a clear peak with co-eluting product ions was observed. The product ion peaks are not identical in shape which may be caused by interference from other molecules; i.e. the product ions may be from two separate but similar precursors eluting at the same time. The product ion chromatograms are consistent and with no library available to make a comparison, it is impossible to decide which transitions to discard. Given the interference, it is recommended that this MRM just be used for presence/absence studies rather than quantification until the structure of this candidate biomarker can be elucidated, and a more accurate MRM generated. Where the integrated peak was present in one sample it was present in the repeat injection indicating that this MRM is reproducible. This biomarker candidate was observed in 8/8 ATB samples, in 9/13 LTBI samples, and in 9/12 NTB group samples. Table 3.11 summarises the results from this MRM assay.

Table 3.11: Results summary for 284.1 (-1) MRM assay.

No. of samples observed in 26/33	No. of repeat observations 26/26	RT batch 1 ~12.7 mins	RT batch 2 ~12.1 mins	RT CV 12%

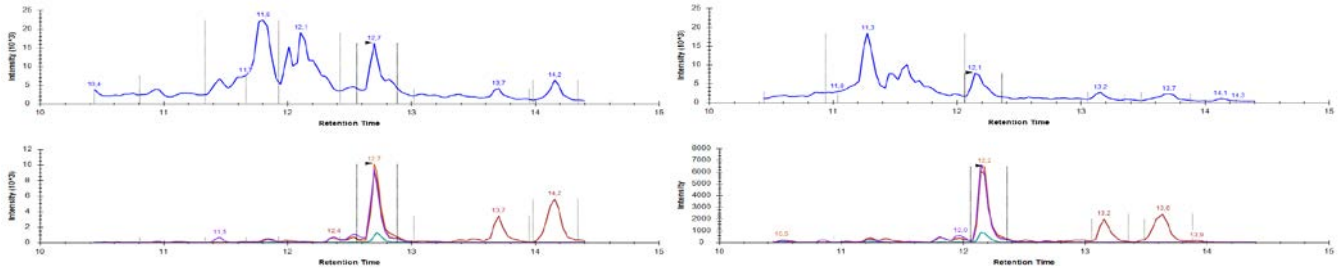


Figure 3.23: An example of a peak for precursor 284.1 (top panel in blue) and its coeluting product ions (bottom panel in multicolours) appearing at 12.7 minutes (left panel) and at 12.1 minutes in its repeat injection two weeks later (right panel).

The next biomarker candidate in this group is 432.2 (-1) which was analysed using the following MRM transition set:

Precursor	>	Product ions
432.2 (-1)	>	130.06
432.2 (-1)	>	133.07
432.2 (-1)	>	160.08
432.2 (-1)	>	204.07

The results for this MRM showed a clear peak at ~10.9min in the first batch and the same peak eluted at ~10.3min in the repeat batch (Figure 3.24). In this MRM there appeared to be some interference on the MS1 scan, however at the MS/MS level a clear peak with co-eluting product ions was observed. In some cases, as can be seen in Figure 3.24 on the right panel, the product ions did not form the same peak shape. This only seemed to happen when the product ion intensity was very low (>500) so this may have been interference from noise rather than co-eluting precursors. This MRM did not show very high reproducibility; with a total of five samples having it present in the first batch and not the second. Where it was present in the repeat injection, the intensities were substantially lower than in the first injection. This biomarker candidate was observed in duplicate in 4/8 ATB samples, in 4/13 LTBI samples, and in 8/13 NTB samples. Table 3.12 summarises the results for this MRM assay.

Table 3.12: Results summary for 432.2 (-1) MRM assay.

No. of samples observed in 15/33	No. of repeat observations 10/15	RT batch 1 ~10.9 mins	RT batch 2 ~10.3 mins	RT CV 10%

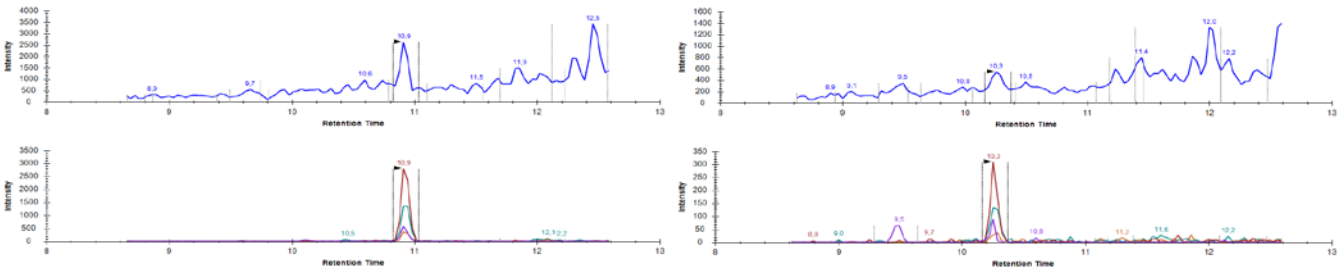


Figure 3.24: An example of a peak for precursor 432.2 (top panel in blue) and its coeluting product ions (bottom panel in multicolours) appearing at 10.9 minutes (left panel) and at 10.3 minutes in its repeat injection two weeks later (right panel).

The next biomarker candidate in this group is 685.436 (+1) which was targeted with the following MRM transition set:

Precursor	>	Product ions
685.436 (+1)	>	72.89
685.436 (+1)	>	75.06
685.436 (+1)	>	184.33
685.436 (+1)	>	673.35

The results for this MRM showed one peak in one sample at 40.8min (Figure 3.25). This same peak was not observed in the repeat injection, so it may have been there due to chance. This MRM should not be used to further measure this biomarker candidate as it is unlikely to be measuring anything of biological significance.

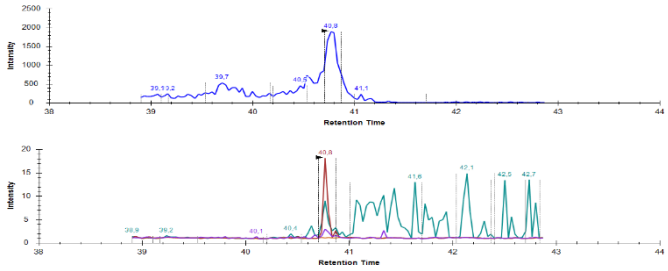


Figure 3.25: The peak for precursor 685.436 (top panel in blue) and its coeluting product ions (bottom panel in multicolours) appearing at 40.8 minutes.

The next biomarker candidate in this group is 785.35 (-1) which was targeted with the following MRM transition set:

Precursor	>	Product ions
785.35 (-1)	>	113.02
785.35 (-1)	>	175.02

The results for this MRM showed no peaks in any of the samples and very little consistency between experiments. This was somewhat expected as the results from the preliminary MRMs did not give strong evidence for this being the biomarker candidate.

The next biomarker candidate in this group is 817.3 (+1) which was measured using the following MRM transition set:

Precursor	>	Product ions
817.3 (+1)	>	95.05
817.3 (+1)	>	109.07
817.3 (+1)	>	123.08
817.3 (+1)	>	137.06
817.3 (+1)	>	151.08
817.3 (+1)	>	177.1
817.3 (+1)	>	179.11

The results for this MRM showed a clear peak at ~17.1min in the first batch and the same peak eluted at ~16.4min in the repeat batch (Figure 3.26). Another peak appeared at 16.3min in some of the NTB

samples, however the ratio of product ions for this peak was different to that of the other observed peaks and this is therefore unlikely to be the same molecule. This MRM was reproducible with every observation being made in both the sample and its repeat. This biomarker candidate was observed in just three of the 33 clinical extracts, and all these observations were made in the ATB group. Table 3.13 summarises the results from this MRM assay.

Table 3.13: Results summary for 817.3 (+1) MRM assay.

No. of samples observed in 3/33	No. of repeat observations 3/3	RT batch 1 ~17.1 mins	RT batch 2 ~16.4 mins	RT CV 17%

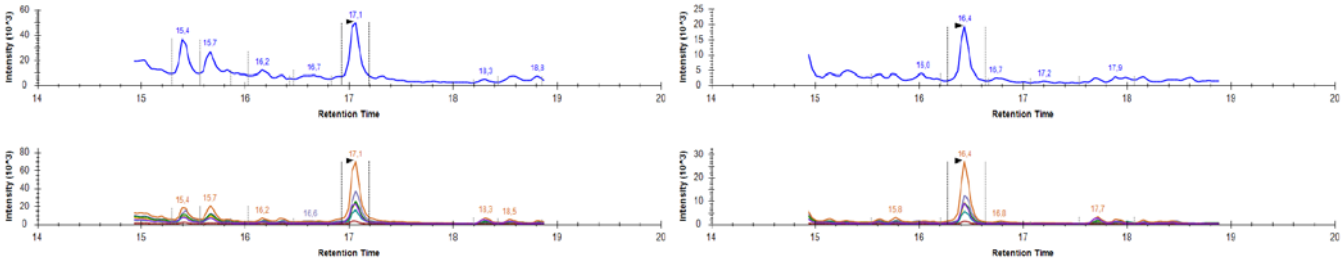


Figure 3.26: An example of a peak for precursor 817.3 (top panel in blue) and its coeluting product ions (bottom panel in multicolours) appearing at 17.1 minutes (left panel) and at 16.4 minutes in its repeat injection two weeks later (right panel).

The final biomarker candidate in this group is 1273.613 (+1) which was targeted according to the following MRM transition set:

Precursor	Product ions
1273.613 (+1)	525.19
	581.25
	637.33

The results for this MRM showed a clear peak at ~29.4min in the first batch and the same peak eluted at ~28.7min in the repeat batch (Figure 3.27). This MRM was somewhat reproducible as it was measured in 27 of the 33 samples, however in four of these samples it was not seen in duplicate. This biomarker candidate was observed in duplicate in 7/8 ATB samples, in 7/13 LTBI samples, and in 9/12 NTB samples. Table 3.14 summarises the results for this MRM assay.

Table 3.14: Results summary for 817.3 (+1) MRM assay.

No. of samples observed in 27/33	No. of repeat observations 23/27	RT batch 1 ~29.4 mins	RT batch 2 ~28.7 mins	RT CV 29%

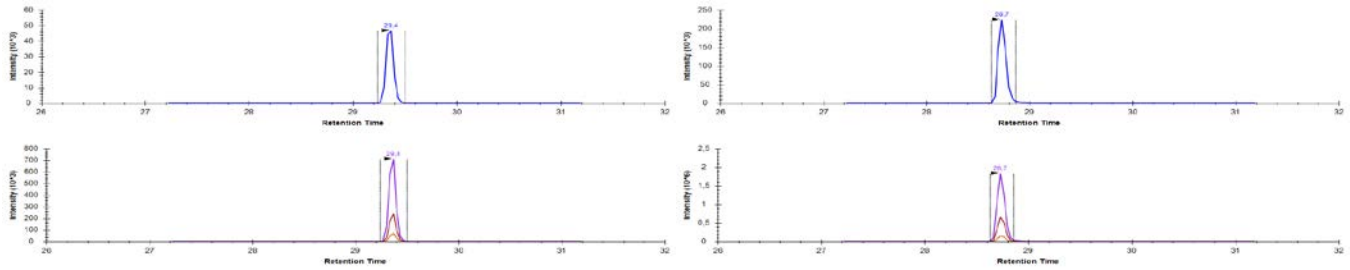


Figure 3.27: An example of a peak for precursor 1273.613 (top panel in blue) and its coeluting product ions (bottom panel in multicolours) appearing at 29.4 minutes (left panel) and at 28.7 minutes in its repeat injection two weeks later (right panel).

3.5 Conclusions

The final output of this Chapter are MRM assays developed for 10 out of 70 biomarker candidates, which were selected from an original list of ~2000 molecular features. This massive reduction in candidates is expected when progressing from biomarker discovery to verification (C. E. Parker & Borchers, 2014) and this list is expected to decrease further when this panel goes through final validation. Below, in Table 3.15, is a list of the final transitions for each biomarker candidate, with the expected retention time.

The LTBI biomarker candidate, *m/z* 189.0, shared a common fragment, *m/z* 144.05, with homoarginine or targinine, chemical formula $C_7H_{16}N_5O_2$. Further work is required to validate this finding although it is an interesting one as homoarginine is an inhibitor of alkaline phosphohydrolase (Lin & Fishman, 1972), and has been measured in urine, associated with liver disorders (Shoda et al., 1990; Shoda et al., 1988; Gatti & Gioia, 2008) and has also been associated with pulmonary disease (Bregy et al., 2018). Alternatively, targinine is not a well-researched molecule and has been observed in blood or in food sources such as pulses and grains (Human Metabolome Database, 2017).

Table 3.15: Final transition sets for MRMs for the refined list of biomarker candidates.

Biomarker class	Precursor *m/z*	Charge state	Product ions	Retention time
LTBI	146.119	+1	74.06 100.08 102.09 127.09	8.3 – 8.9 mins
	188.176	+1	90.95 116.95 141.96	9.7 – 10.2 mins
	189.0	+1	130.09 131.98 144.05 161.07 171.06	5.6 – 6.6 mins
	233.12	+1	128.95	1.7 – 2.0 mins
	332.281	+1	213.89 225.89 236.89 251.89 253.0	15.8 – 16.4 mins
LTBI / Active TB	403.17	+1	69.03 83.05 120.06 156.14 160.14	28.0 – 28.9 mins
Active TB	284.1	-1	101.2 109.4 127.05 128.03 145.06	12.1 – 12.7 mins
	432.2	-1	130.06 133.07 160.08 204.07	10.3 – 10.9 mins
	817.3	+1	95.05 109.07 123.08 137.06 151.08 177.1 179.11	16.4 – 17.1 mins
	1273.613	+1	525.19 581.25 637.33	28.7 – 29.4 mins

Chapter 4 – Conclusions and Future Work

4.1 MRM assays for known standards

The first part of this thesis (Chapter 2) developed MRM assays for heavy labelled standards. These standards can be used to absolutely quantify octanoic acid, lauric acid, myristic acid, palmitic acid and stearic acid in biologically derived samples. It was also found that in the case of all the standards, a longer C18 chain length was directly proportional to a later retention time. Therefore it is considered that this is an ideal set of standards for retention time prediction and cross-column calibration, as demonstrated in the experiment in Chapter 3.

4.2 MRM assays for biomarker candidates *de novo*

The second part of this thesis (Chapter 3) aimed to develop MRM assays for biomarker candidates with unknown structure or identity. Initially 2000 molecular features were identified in a previous experiment and this list was refined to 70 of the statistically most significant candidate biomarkers. The methods employed in this thesis refined this list to 10 biomarker candidates which produced clear and reproducible MRM assay results. The methodology applied here can in principle be used to generate an MRM assay on any panel of unknown compounds. Because there are no standards available for these compounds, these MRM assays cannot be used for absolute quantification of the biomarker candidates but are suitable for relative quantification.

This list of 10 candidates showed no molecule that performed well as a singular biomarker and the recommended end-point for this list would be a diagnostic panel able to measure unique biosignatures for different disease states. Given the age of the clinical samples analysed, firm quantitative conclusions could not be drawn on each biomarker.

4.2.1 Biomarker development pipeline and future work

In mass spectrometry-based proteomics, the biomarker pipeline progresses from discovery to verification to validation and finally to clinical evaluation (C. E. Parker & Borchers, 2014). These definitions will be applied to this lipidomic project as a similar logic is being followed. The discovery work for this project had been done previously where a priority list of 70 biomarker candidates was produced. The work described here achieved the verification of a subset these biomarker candidates where MRM assays were developed, and then tested for scheduling and reproducibility. The suggested next phase would be validation where a biomarker signature becomes associated with a known TB diagnostic outcome. This would require between 500 and 1000 fresh clinical samples (Surinova et al., 2011), stratified as they were in this project. Validation can take up to 12 months and the outcome is a defined biomarker panel for clinical evaluation. Clinical evaluation – the final phase – is where the test is evaluated in the intended patient population and the sensitivity and specificity of the test is defined.

4.2.2 Structural elucidation and identification

The ability to generate an MRM for unknown compounds is important in the field of lipidomics and metabolomics, where molecule identification is time consuming, expensive, and requires high levels of expertise. This experiment refined a list of some 2000 compounds to a much more manageable list of 10 molecules of interest. For these MRM assays to be used for absolute quantification, they would ultimately need to be identified so that commercially available isotopically labelled standards could be used (Simpson et al., 2009).

Lipidomic identification and quantification remains a challenge given the complexity and diversity of lipid species. Isobaric species are common amongst lipids which means that MS1 identity alone is unreliable and further separation methods and fragmentation are necessary (Bielow et al., 2017; Hancock et al., 2017). The current methods for identifying lipids from shotgun mass spectrometry data are class-specific, use only MS1 values, or require prior knowledge of the lipid class being investigated (Husen et al., 2013; Kochen et al., 2016; Taguchi et al., 2007). Given that novel MRM assays for 10 analytes result from this thesis, a targeted approach may be more appropriate in future studies.

Mass spectrometry-based techniques could be coupled with two-dimensional orthogonal chromatographic separations to isolate the lipid compounds of interest (Zheng et al., 2018) with a further ion mobility spectrometry (IMS) step to characterise the lipids and separate lipid isomers (Baker et al., 2014). A significant challenge in structural elucidation of lipids is the verification and quantification of carbon-carbon double bonds. These bonds cannot be characterised in standard mass spectrometry experiments, and specialised fragmentation methods are required. Methods which have shown promise in the past are electron transfer dissociation (Liang et al., 2007), metastable atom-activated dissociation (Li et al., 2016), electron-induced dissociation (Jones et al., 2015), radical-directed dissociation (Pham et al., 2012), and ultraviolet photodissociation (Klein & Brodbelt, 2016). These techniques should be employed with caution however as data analysis is not standardised and requires a specialised expertise. Structural elucidation and identification of the lipids would be an important supplement to the data generated in this project, however such an endeavour is beyond the scope of a Masters Thesis.

References

Aebersold, R., & Mann, M. (2003). Mass spectrometry-based proteomics. *Nature, 422*(6928), 198–207. doi.org/10.1038/nature01511

Agilent. (2007). *How many data points to be acquired across an LC peak.* Retrieved from https://www.agilent.com/cs/library/support/documents/FAQ_Approved_PDF_Template_enough_datapoints.pdf

Aguilar, M. (2004). *Reversed-phase high-performance liquid chromatography*. Totowa, NJ: Humana Press Inc.

Alshehry, Z. H., Barlow, C. K., Weir, J. M., Zhou, Y., McConville, M. J., & Meikle, P. J. (2015). An efficient single phase method for the extraction of plasma lipids. *Metabolites*, 5(2), 389–403.

Ankney, J. A., Muneer, A., Chen, X. (2018). Relative and absolute quantitation in mass spectrometry-based proteomics. *Annual Review of Analytical Chemistry, 11,* 49-77. doi.org/10.1146/annurev-anchem-061516-045357

Baker, P. R. S., Armando, A. M., Campbell, J. L., Quehenberger, O., & Dennis, E. A. (2014). Three-dimensional enhanced lipidomics analysis combining UPLC, differential ion mobility spectrometry, and mass spectrometric separation strategies*. Journal of Lipid Research*, *55*, 2432–2442. doi.org/10.1101/019281

Bang, D. Y., Byeon, S. K., & Moon, M. H. (2014). Rapid and simple extraction of lipids from blood plasma and urine for liquid chromatography-tandem mass spectrometry. *Journal of Chromatography A*, 1331, 19–26. doi.org/10.1016/j.chroma.2014.01.024

Bévilacqua, S., Rabaud, C., & May, T. (2002). HIV-tuberculosis coinfection. *Annales de Medecine Interne*, *153*(2), 113—118

Bidlingmeyer, B. A. (1980). Separation of ionic compounds by reversed-phase liquid chromatography an update of ion-pairing techniques. *Journal of Chromatographic Science, 18*(10), 525-539. doi.org/10.1093/chromsci/18.10.525

Bielow, C., Mastrobuoni, G., Orioli, M., & Kempa, S. (2017). On mass ambiguities in high-resolution shotgun lipidomics. *Analytical Chemistry, 89*(5), 2986–2994. doi.org/10.1021/acs.analchem.6b04456

Blesic, M., Marques, H., Plechkova, N. V, & Seddon, K. R., Rebelo, L. P. N, & Lopes, A. (2007). Self-aggregation of ionic liquids: Micelle formation in aqueous solution self-aggregation of ionic liquids. *The Royal Society of Chemistry*, *9*, 481-490. doi.org/10.1039/B615406A

Bregy, L., Nussbaumer-Ochsner, Y., Sinues, P. M., Garcia-Gomez, D., Suter, Y., Gaisl, T., Stebler, N., Gaugg, M. T., Kohler, M., & Zenobi, R. (2018). Real-time mass spectrometric identification of metabolites characteristic of chronic obstructive pulmonary disease in exhaled breath. *Clinical Mass Spectrometry, 7,* 29–35. doi.org/10.1016/j.clinms.2018.02.003

Cattamanchi, A., Dowdy, D. W., Davis, J. L., Worodria, W., Yoo, S., Joloba, M., Matovu, J., Hopewell, P. C., & Huang, L. (2009). Sensitivity of direct versus concentrated sputum smear microscopy tuberculosis. *Biomed Central Infectious Diseases*, *9*(53), 1–9. doi.org/10.1186/1471-2334-9-53

Cech, N. B., & Enke, C. G. (2001). Practical implication of some recent studies in electrospray ionization fundamentals. *Mass Spec Reviews, 20*, 362–387. doi.org/10.1002/mas.10008

Chegou, N. N., Sutherland, J. S., Malherbe, S., Crampin, A. C., Corstjens, P. L. A. M., Geluk, A., Mayanja-Kizza, H., Loxton, A. G., van der Spuy, G., Stanley, K., Kotze, L. A., van der Vyver, M., Rosenkrands, I., Kidd, M., van Helden, P. D., Dockrell, H. M., Ottenhoff, T. H. M., Kaufmann, S. H. E., & Walzl, G. (2016). Diagnostic performance of a seven-marker serum protein biosignature for the diagnosis of active TB disease in African primary healthcare clinic attendees with signs and symptoms suggestive of TB. *Thorax*, *71*(9), 785–794. doi.org/10.1136/thoraxjnl-2015-207999

Cominetti, O., Galindo, A. N., Corthesy, J., Moreno, S. O., Irincheeva, I., Valsesia, A., Astrup, A., Sair, W. H. M., Kussmann, M., & Dayon, L. (2016). Proteomic biomarker discovery in 1000 human plasma samples with mass spectrometry. *Journal of Proteome Research*, *15*(2), 389-399. doi.org/10.1021/acs.jproteome.5b00901

Corstjens, P. L. A. M., Tjon Kon Fat, E. M., de Dood, C. J., van der Ploeg-van Schip, J. J., Franken, K. L. M. C., Chegou, N. N., Sutherland, J. S., Howe, R., Mihret, A., Kassa, D., van der Vyver, M., Sheehama, J., Simukonda, F., Mayanja-Kizza, H., Ottenhoff, T. H. M., Walzl, G., & Geluk, A. (2016). Multi-center evaluation of a user-friendly lateral flow assay to determine IP-10 and CCL4 levels in blood of TB and non-TB cases in Africa. *Clinical Biochemistry, 49*(1), 22–31. doi.org/10.1016/j.clinbiochem.2015.08.013

Crutchfield, C. A., Thomas, S. N., Sokoll, L. J., & Chan, D. W. (2016). Advances in mass spectrometry-based clinical biomarker discovery. *Clinical Proteomics*, *13*(1), 1–12. doi.org/10.1186/s12014-015-9102-9

Damen, C. W., Isaac, G., Langridge, J., Hankemeier, T., & Vreeken, R. J. (2014). Enhanced lipid isomer separation in human plasma using reversed-phase UPLC with ion-mobility/high-resolution MS detection. *Journal of Lipid Research, 55*(8), 1772-1783. doi.org/10.1194/jlr.D047795

Davies, P. D. O., & Pai, M. (2008). The diagnosis and misdiagnosis of tuberculosis. *The International Journal of Tuberculosis and Lung Disease*, *12*(11), 1226–1234.

de Hoffman, E., & Stroobant, V. (2007). *Mass spectrometry principles and applications – Third edition.* Chichester: John Wiley & Sons Ltd.

Demarque, D. P., Crotti, A. E. M., Vessecchi, R., Lopes, J. L. C., & Lopes, N. P. (2016). Fragmentation reactions using electrospray ionization mass spectrometry: An important tool for the structural elucidation and characterization of synthetic and natural products. *Natural Product Reports, 33*(3), 432–455. doi.org/10.1039/c5np00073d

Drain, P. K., Losina, E., Coleman, S. M., Giddy, J., Ross, D., Katz, J. N., Freedberg, K. A., & Bassett, I. V. (2017). Clinic-based urinary lipoarabinomannan as a biomarker of clinical disease severity and mortality among antiretroviral therapy-naive human immunodeficiency virus-infected adults in South Africa. *Open Forum Infectious Diseases, 4*(3). doi.org/10.1093/ofid/ofx167

Fillatre, Y., Rondeau, D., Jadas-Hecart, A., & Communal, P. Y. (2010). Advantages of the scheduled selected reaction monitoring algorithm in liquid chromatography/electrospray ionization tandem mass spectrometry multi-residue analysis of 242 pesticides: A comparative approach with classical selected reaction monitoring. *Rapid Communications in Mass Spectrometry, 24*(16), 2453–2461. doi.org/10.1002/rcm.4649

Gallien, S., Duriez, E., & Domon, B. (2011). Selected reaction monitoring applied to proteomics. *Journal of Mass Spectrometry,* 46(3), 298–312. doi.org/10.1002/jms.1895

Gatti, R., & Gioia, M. G. (2008). Liquid chromatographic analysis of guanidine compounds using furoin as a new fluorogenic reagent. *Journal of Pharmaceutical and Biomedical Analysis, 48*(3), 754-759. doi.org/10.1016/j.jpba.2008.07.017

Gillet, L. C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., Bonner, R., & Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular & Cellular Proteomics., 11*(6), O111.016717. doi.org/10.1074/mcp.O111.016717-1

Gillette, M. A., & Carr, S. A. (2013). Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nature Methods, 10*(1), 28–34. doi.org/10.1038/nmeth.2309.Quantitative

Goletti, D., Lee, M., Wang, J., Walter, N., & Ottenhoff, T. H. M. (2018). Update on tuberculosis biomarkers: From correlates of risk, to correlates of active disease and of cure from disease. *Official Journal of the Asian Pacific Society of Respirology*, *23*(5), 455–466. doi.org/10.1111/resp.13272

Hancock, S. E., Poad, B. L. J., Batarseh, A., Abbott, S. K., & Mitchell, T. W. (2017). Advances and unresolved challenges in the structural characterization of isomeric lipids. *Analytical Biochemistry, 524*, 45–55. doi.org/10.1016/j.ab.2016.09.014

Herzog, R., Schwudke, D., Schuhmann, K., Sampaio, J. L., Bornstein, S. R., Schroeder, M., & Shevchenko, A. (2011). A novel informatics concept for high-throughput shotgun lipidomics based on the molecular fragmentation query language. *Genome Biology*, *12*(1), 1–25. doi.org/10.1186/gb-2011-12-1-r8

Hoke, S. H., Morand, K. L., Greis, K. D., Baker, T. R., Harbol, K. L., & Dobson, R. L. M. (2001). Transformations in pharmaceutical research and development, driven by innovations in multidimensional mass spectrometry-based technologies. *International Journal of Mass Spectrometry, 212*(1-3), 135–196. doi.org/10.1016/S1387-3806(01)00499-7

Hsieh, E. J., Bereman, M. S., Durand, S., Valaskovic, G. A., & MacCoss, M. J. (2013). Effects of column and gradient lengths on peak capacity and peptide identification in nanoflow LC-MS/MS of complex proteomic samples. *Journal of The American Society for Mass Spectrometry, 24*(1), 148–153. doi.org/10.1007/s13361-012-0508-6

Human Metabolome Database. (2017). Showing metabocard for L-Targinine. Retrieved from http://www.hmdb.ca/metabolites/HMDB0029416

Husen, P., Tarasov, K., Katafiasz, M., Sokol, E., Vogt, J., Baumgart, J., Nitsch, R., Ekroos, K., & Ejsing, C. S. (2013). Analysis of Lipid Experiments (ALEX): A software framework for analysis of high-resolution shotgun lipidomics data. *PLoS ONE, 8*(11), 1–13. doi.org/10.1371/journal.pone.0079736

Jilge, G., Janzen, R., Giesche, H., Unger, K. K., Kinkel, J. N., & Hearn, M. T. W. (1987). Evaluation of advanced silica packings for the separation of biopolymers by high-performance liquid chromatography: III. Retention and selectivity of proteins and peptides in gradient elution on non-porous monodisperse 1.5-µm reversed-phase silicas. *Journal of Chromatography A, 397*, 71–80. doi.org/10.1016/S0021-9673(01)84990-6

Jones, A., Pitts, M., Al Dulayymi, J. R., Gibbons, J., Ramsay, A., Goletti, D., Gweinin, C. D., & Baird, M. S. (2017). New synthetic lipid antigens for rapid serological diagnosis of tuberculosis. *PLoS ONE, 12*(8), 1–28. doi.org/10.1371/journal.pone.0181414

Jones, J. W., Thompson, C. J., Carter, C. L., & Maureen, A. K. (2015). Electron-induced dissociation (EID) for structure characterization of glycerophosphatidylcholine and localization of acyl chains. *Journal of Mass Spectrometry, 143*(5), 951–959. doi.org/10.1017/S0950268814002131.Tuberculosis

Khaliq, A., Ravindran, R., Hussainy, S. F., Krishnan, V. V, Ambreen, A., Yusuf, N. W., Irum, S., Rashid, A., Jamil, M., Zaffar, F., Chaudhry, M. N., Gupta, P. K., Akhtar, M. W., & Khan, I. H. (2017). Field evaluation of a blood based test for active tuberculosis in endemic settings. *PLoS ONE, 12*(4), 1–12. doi.org/10.1002/jms.3698

Klein, D. R., & Brodbelt, J. S. (2016). Structural characterization of phosphatidylcholines using 193 nm ultraviolet photodissociation mass spectrometry. *Analytical Chem, 131*(20), 1796–1803. doi.org/10.1021/acs.analchem.6b03353

Kochen, Chambers, M. C., Holman, J. D., Nesvizhskii, A. I., Weintraub, S. T., Belisle, J. T., Islam, M. N., Griss, J., Tabb, D. L. (2016). Greazy: Open-source software for automated phospholipid MS/MS identification. *Analytical Chemistry, 88*(11), 5733-5741. doi.org/10.1021.acs.analchem.6b00021

Kozlowski, R. L., Campbell, J. L., Mitchell, T. W., & Blanksby, S. J. (2015). Combining liquid chromatography with ozone-induced dissociation for the separation and identification of phosphatidylcholine double bond isomers. *Analytical and Bioanalytical Chemistry, 407*(17), 5053–5064. doi.org/10.1007/s00216-014-8430-3

Kyle, J. E., Zhang, X., Weitz, K. K., Monroe, M. E., Ibrahim, Y. M., Moore, R. J., Cha, J., Sun, X., Lovelace, E. S., Wagoner, J., Polyak, S. J., Metz, T. O., Dey, S. K., Smith, R. D., Burnum-Johnson, K. E., & Baker, E. S. (2016). Uncovering biologically significant lipid isomers with liquid chromatography, ion mobility spectrometry and mass spectrometry. *Analyst, 141*(5), 1649–1659. doi.org/10.1039/C5AN02062J

Lange, V., Picotti, P., Domon, B., & Aebersold, R. (2008). Selected reaction monitoring for quantitative proteomics: A tutorial. *Molecular Systems Biology, 4*(222), 1744-4292 doi.org/10.1038/msb.2008.61

Lawn, S. D., & Gupta-Wright, A. (2015). Detection of lipoarabinomannan (LAM) in urine is indicative of disseminated TB with renal involvement in patients living with HIV and advanced immunodeficiency: Evidence and implications. *Transactions of the Royal Society of Tropical Medicine and Hygiene, 110*(3), 180–185. doi.org/10.1093/trstmh/trw008

Li, P., Hoffmann, W. D., & Jackson, G. P. (2016). Multistage mass spectrometry of phospholipids using collision-induced dissociation (CID) and metastable atom-activated dissociation (MAD). *International Journal of Mass Spectrometry, 403*, 1–7. doi.org/10.1016/j.ijms.2016.02.010

Liang, X., Liu, J., LeBlanc, Y., Covey, T., Ptak, A. C., Brenna, J. T., & McLuckey, S. A. (2007). Electron transfer dissociation of doubly sodiated glycerophosphocholine lipids. *Journal of the American Society for Mass Spectrometry, 18*(10), 1783–1788. doi.org/10.1016/j.jasms.2007.07.013

Lin, C. W., & Fishman, W. H. (1972). L-Homoarginine. An organ-specific, uncompetitive inhibitor of human liver and bone alkaline phosphohydrolases. *Journal of Biological Chemistry, 247*(10), 3082-3087.

Liu, C., Zhao, Z., Fan, J., Lyon, C. J., Wu, H., Nedelkov, D., & Zelazny, A. M. (2017). Quantification of circulating Mycobacterium tuberculosis antigen peptides allows rapid diagnosis of active disease and treatment monitoring. *Proceedings of the National Academy of Sciences of the United States of America, 114*(15), 3969-3974. doi.org/10.1073/pnas.1621360114

Löfgren, L., Ståhlman, M., Forsberg, G.-B. G.-B., Saarinen, S., Nilsson, R., & Hansson, G. I. (2012). The BUME method: a novel automated chloroform-free 96-well total lipid extraction method for blood plasma. *The Journal of Lipid Research*, 53(8), 1690–1700. doi.org/10.1194/jlr.D023036

Mass Bank of North America. (2007). Spectra Search. Retrieved from http://mona.fiehnlab.ucdavis.edu/spectra/search

Mlamla, Z. C. (2018). *Differential lipidomic profiling of Mycobacterium tuberculosis genotypic strains and urine of TB suspects for "bio-signatures" of TB disease through Liquid Chromatography-Mass Spectrometry* (Unpublished doctoral dissertation). University of Cape Town, Cape Town, South Africa.

National South African AIDS Council. (2011). *National strategic plan on HIV, STId and TB 2012-2016*. Retrieved from https://www.gov.za/sites/default/files/national%20strategic%20plan%20on%20hiv%20stis%20and%20tb_0.pdf

Nohara, D., Ohkoshi, T., & Sakai, T. (1998). The possibility of the direct measurement of micelle weight by electrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry*, *12*(23), 1933–1935. doi.org/10.1002/(SICI)1097-0231(19981215)12:23<1933::AID-RCM410>3.0.CO;2-N

Orgsoltab, M. (2017). Common organic solvents: Table of properties. Retrieved from www.organicdivision.org

Pabst, M., Bondili, J. S., Stadlmann, J., Mach, L., & Altmann, F. (2007). Mass + retention time = structure: A strategy for the analysis of *N*-glycans by carbon LC-ESI-MS and its application to fibrin *N*-Glycans. *Analytical Chemistry, 79*(13), 5051–5057. doi.org/10.1021/ac070363i

Parker, C. E., & Borchers, C. H. (2014). Mass spectrometry based biomarker discovery, verification, and validation – Quality assurance and control of protein biomarker assays. *Molecular Oncology, 8*(4), 840–858. doi.org/10.1016/j.molonc.2014.03.006

Parker, S. J., Rost, H., Rosenburger, G., Collins, B. C., Malmstrom, L., Amodei, D., Venkatraman, V., Raedschelders, K., Van Eyk, J. E., & Aebersold, R. (2015). Identification of a set of conserved eukaryotic internal retention time standards for data-independent acquisition mass spectrometry. *Molecular and Cellular Proteomics*, *14*(10), 2800-2813. doi.org/10.1074/mcp.O114.042267

Pham, H. T., Ly, T., Trevitt, A. J., Mitchell, T. W., & Blanksby, S. J. (2012). Differentiation of complex lipid isomers by radical-directed dissociation mass spectrometry. *Analytical Chemistry, 84*(17), 7525–7532. doi.org/10.1021/ac301652a

Rifai, N., Gillette, M. A., & Carr, S. A. (2006). Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nature Biotechnology*, *24*, 971–983. doi.org/10.1038/nbt1235

Rockwell, H. E., Gao, F., Chen, E. Y., McDaniel, J., Sarangarajan, R., Narain, N. R., & Kiebish, M. A. (2016). Dynamic assessment of functional lipidomic analysis in human urine. *Lipids, 51*(7), 875-886. doi-org.ezproxy.uct.ac.za/10.1007/s11745-016-4142-0

Schubert, O. T., Gillet, L. C., Collins, B. C., Navarro, P., Rosenberger, G., Wolski, W. E., Lam, H., Amodei, D., Mallick, P., MacLean, B., & Aebersold, R. (2015). Building high-quality assay libraries for targeted analysis of SWATH MS data. *Nature Protocols, 10*(3), 426-441. doi.org/10.1038/nprot.2015.015

Sethi, S., & Brietzke, E. (2017). Recent advances in lipidomics: Analytical and clinical perspectives. *Prostaglandins & Other Lipid Mediators, 128–129*, 8–16. doi.org/10.1016/j.prostaglandins.2016.12.002

Shoda, J., Mahara, R., Osuga, T., Tohma, M., Ohnishi, S., Miyazaki, H., Tanaka, N., & Matsuzaki, Y. (1988). Similarity of unusual bile acids in human umbilical cord blood and amniotic fluid from newborns and in sera and urine from adult patients with cholestatic liver diseases. *Journal of Lipid Research, 29*(7), 847-858.

Shoda, J., Tanaka, N., Osuga, T., Matsuura, K., & Miyazaki, H. (1990). Altered bile acid metabolism in liver disease: concurrent occurrence of C-1 and C-6 hydroxylated bile acid metabolites and their preferential excretion into urine. *Journal of Lipid Research, 31*(2), 249-259.

Simerville, J. A., Maxted, W. C., & Pahira, J. J. (2005). Urinalysis: A comprehensive review. *American Family Physician, 71*(6), 1153-1162.

Simpson, K. L., Whetton, A. D., & Dive, C. (2009). Quantitative mass spectrometry-based techniques for clinical use: Biomarker identification and quantification. *Journal of Chromatography B*, *877*(13), 1240–1249. doi.org/10.1016/j.jchromb.2008.11.023

Smith, R. D., Loo, J. A., Edmonds, C. G., Baringa, C. J., & Udseth, H. R. (1990). New developments in biochemical mass spectrometry: electrospray ionization. *Analytical Chemistry, 62*(9), 882-899. doi.org/10.1021/ac00208a002

Soyer, A., Özalp, B., Dalmıs, Ü., & Bilgin, V. (2010). Effects of freezing temperature and duration of frozen storage on lipid and protein oxidation in chicken meat. *Food Chemistry, 120*(4), 1025–1030. doi.org/10.1016/j.foodchem.2009.11.042

Statistics South Africa. (2015). Mortality and causes of death in South Africa, 2015. Retrieved from www.statssa.gov.za/?p=9604

Steingart, K. R., Schiller, I., Horne, D. J., Pai, M., Boehme, C. C., & Dendukuri, N. (2014). Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrance Database of Systematic Reviews, 1.* doi.org/10.1002/14657858.CD009593.pub3

Strimbu, K., & Tavel, J. A. (2010). What are biomarkers? *Current opinion in HIV and AIDS, 5*(6), 463-466. doi.org/10.1097/COH.0b013e32833ed177

Strupat, K., Scheibner, O., & Bromirski, M. (2013). *High-resolution, accurate-mass orbitrap mass spectrometry – definitions, opportunities, and advantages.* Retrieved from http://www.tecnofrom.com/moduloNotas/noti_41/archivosAdjuntos/TN-64287-HRAM-Orbitrap-MS-Terminology-Advantages-TN64287-EN_1_319.pdf

Surinova, S., Schiess, R., Huttenhain, R., Cerciello, F., Wollscheid, B., & Aebersold, R. (2011). On the development of plasma protein biomarkers. *Journal of Proteome Research, 10*(1), 5–16. doi.org/10.1021/pr1008515

Sweeney, T. E., Braviak, L., Tato, C. M., & Khatri, P. (2016). Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis*. Lancet Respiratory Medicine, 4*(3), 213–224. doi.org/10.1016/S2213-2600(16)00048-5

Taguchi, R., Nishijima, M., & Shimizu, T. (2007). Basic analytical systems for lipidomics by mass spectrometry in Japan. *Methods in Enzymology, 432,*185-211. doi.org/10.1016/S0076-6879(07)32008-9

Tanaka, N., Sakagami, K., & Araki, M. (1980). Effect of alkyl chain length of the stationary phase on retention and selectivity in reversed-phase liquid chromatography: Participation of solvent molecules in the stationary phase. *Journal of Chromatography A, 199*, 327–337. doi.org/10.1016/S0021-9673(01)91384-6

Teixeira, R. C., Rodriguez, M., de Romero, N. J., Bruis, M., Gomez, R., Yntema, J. B, Abente, G. C., Gerritsen, J. W., Wiegerinck, W., Bejerano, D. P., & Magis-Escurra, C. (2017). The potential of a portable, point-of-care electronic nose to diagnose tuberculosis. *Journal of Infection, 75*(5), 441-447. doi.org/10.1016/j.jinf.2017.08.003

Tipthara, P., & Thongboonkerd, V. (2016). Differential human urinary lipid profiles using various lipid-extraction protocols: MALDI-TOF and LIFT-TOF/TOF analyses. *Scientific Reports*, *6*, 33756. https://doi.org/10.1038/srep33756

Van, Q. N., Veenstra, T. D., & Issaq, H. J. (2011). Metabolic profiling for the detection of bladder cancer. *Current Urology Reports*, *12*(1), 34–40. doi.org/10.1007/s11934-010-0151-3

Watson, J. T., & Sparkman, O. D. (2007). *Introduction to mass spectrometry: Instrumentation, applications and strategies for data interpretation*. John Wiley & Sons.

World Health Organisation. (2015). WHO | Tuberculosis (TB). Retrieved from www.who.int/countries/zaf/en/

Wu, Z., Gao, W., Phelps, M. A., Wu, D., Miller, D. D., & James, T. (2004). Favorable effects of weak acids on negative-ion electrospray ionization mass spectrometry. *Analytical Chemistry, 76*(3), 839–847. doi.org/10.1021/ac0351670

Young, B. L., Mlamla, Z., Gqamana, P. P., Smit, S., Roberts, T., Peter, J., Theron, G., Govender, U., Dheda, K., & Blackburn, J. (2014). The identification of tuberculosis biomarkers in human urine samples. *European Respiratory Journal, 43,* 1719-1729. doi.org/10.1183/09031936.00175113

Zak, D. E., Penn-Nicholson, A., Scriba, T. J., Thompson, E., Suliman, S., Amon, L. M., Mahomed, H., Erasmus, M., Whatney, W., Hussey, G. D., Abrahams, D., Kafaar, G., Hawkridge, T., Verver, S., Hughes, E. J., Ota, M., Sutherland, J., Howe, R., Dockrell, H. M., Boom, W. H., Thiel, B., Ottenhoff, T. H. M., Mayanja-Kizza, H., Crampin, A. C., Downing, K., Hatherill, M., Valvo, J., Shankar, S., Parida, S. K., Kaufmann, S. H. E., Walzl, G., Aderem, A., & Hanekom, W. A. (2016). A prospective blood RNA signature for tuberculosis disease risk. *Lancet, 387*(10035), 2312–2322. doi.org/10.1016/S0140-6736(15)01316-1.

Zetola, N. M., Modongo, C., Matsiri, O., Tamuhla, T., Mbongwe, B., Matlhagela, K., Sepako, A., Catini, A., Sirugo, G., Martinelli, E., Paolesse, R., & Di Natale, C. (2017). Diagnosis of pulmonary tuberculosis and assessment of treatment response through analyses of volatile compound patterns in exhaled breath samples. *Journal of Infection, 74*(4), 367–376. doi.org/10.1016/j.jinf.2016.12.006

Zhang, Z., & Marshall, A. G. (1998). A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. *Journal of the American Society for Mass Spectrometry, 9*(3), 225–233. doi.org/10.1016/S1044-0305(97)00284-5

Zheng, X., Smith, R. D., & Baker, E. S. (2018). Recent advances in lipid separations and structural elucidation using mass spectrometry combined with ion mobility spectrometry, ion-molecule reactions and fragmentation approaches. *Current Opinion in Chemical Biology, 42*, 111–118. doi.org/10.1016/j.cbpa.2017.11.009

Zhou, N. E., Mant, C. T., Kirkland, J. J., & Hodges, R. S. (1991). Comparison of silica-based cyanopropyl and octyl reversed-phase packings for the separation of peptides and proteins. *Journal of Chromatography A, 548*, 179–193. doi.org/10.1016/S0021-9673(01)88600-3

Zivkovic, A. M., Wiest, M. M., Nguyen, U. T., Davis, R., Watkins, S. M., & German, B. (2009). Effects of sample handling and storage on quantitative lipid analysis in human serum. *Metabolomics, 5*, 507–516. doi.org/10.1007/s11306-009-0174-2

Zubarev, R. A., & Makarov, A. (2013). Orbitrap Mass Spectrometry. *Analytical Chemistry*, *85*, 5288–5296. doi.org/10.1021/ac4001223

Chapter 5 – Appendices

Table 5.1: MRM transitions for each standard with collision energies, from compound optimisatio

Compound	Precursor m/z	Voltage (V)	Cap. Temp. (°C)	Product Ions m/z	Collision Energy (V)
Octanoic Acid	158,229	3500	350	137,9 138,1	16 17
Lauric Acid	222,3	3500	350	149,5 164,9 151,3	6 24 45
Myristic Acid	254,4	3500	350	253,9	12
Palmitic Acid	286,5	3500	350	266,2	26
Stearic Acid	301,3	3500	350	283,4 283,0 284,7 212,8	23 24 21 61
Tocopherol Acetate	473,4	3200	320	165,1 207,1 147,1 149,1 137,1	24 19 30 28 48

Table 5.2: Inclusion lists targeted on the Q-Exactive.

1. Active TB (pos)			2. Active TB (neg)			3. LTBI (pos)			4. LTBI (neg)		
Mass [m/z]	**CS [z]**	**Polarity**	**Mass [m/z]**	**CS [z]**	**Polarity**	**Mass [m/z]**	**CS [z]**	**Polarity**	**Mass [m/z]**	**CS [z]**	**Polarity**
284,33110	1	Positive	1059,04000	-1	Negative	146,09270	1	Positive	745,13820	-1	Negative
304,29900	1	Positive	1088,12000	-1	Negative	188,19999	1	Positive	808,11750	-1	Negative
320,29220	1	Positive	1159,16000	-1	Negative	189,13480	1	Positive	898,15220	-1	Negative
403,23480	1	Positive	1206,06000	-1	Negative	233,07930	1	Positive	939,11020	-1	Negative
445,36370	1	Positive	1218,14000	-1	Negative	284,33110	1	Positive	879,12630	-1	Negative
707,28660	1	Positive	1220,13000	-1	Negative	304,29900	1	Positive	1191,77290	-1	Negative
721,50060	1	Positive	1280,09900	-1	Negative	304,29970	1	Positive	1160,15960	-1	Negative
974,57520	1	Positive	1282,09400	-1	Negative	320,29220	1	Positive	1148,09160	-1	Negative
685,43860	1	Positive	1310,17000	-1	Negative	332,33040	1	Positive	1170,14310	-1	Negative
792,33630	2	Positive	1399,22970	-1	Negative	403,23480	1	Positive	1177,17580	-1	Negative
817,32900	3	Positive	1402,21400	-1	Negative	445,36370	1	Positive	1218,13370	-1	Negative
899,60460	1	Positive	1291,09600	-1	Negative	707,28660	1	Positive	1170,14000	-1	Negative
903,41370	1	Positive	1057,04400	-1	Negative	721,50060	1	Positive	1177,18000	-1	Negative
935,50230	1	Positive	1071,13300	-1	Negative	959,96570	1	Positive	1218,13366	-1	Negative
942,63360	1	Positive	1293,15200	-1	Negative	974,57520	1	Positive	1218,13688	-1	Negative
972,47790	1	Positive	284,19400	-1	Negative				1340,07000	-1	Negative
1023,55440	1	Positive	432,22400	-1	Negative				1374,85000	-1	Negative
1184,50690	1	Positive	785,32200	-1	Negative				1059,04000	-1	Negative
1273,60950	1	Positive	852,37400	-1	Negative				1088,12000	-1	Negative
1293,60870	1	Positive	867,38400	-1	Negative				1159,16000	-1	Negative
1296,58680	1	Positive	1075,42200	-1	Negative				1206,06000	-1	Negative
1347,90420	1	Positive	1135,68400	-1	Negative				1218,14000	-1	Negative
			1385,86720	-1	Negative				1220,13000	-1	Negative
			1387,86000	-1	Negative				1280,09900	-1	Negative
			1412,88330	-1	Negative				1282,09400	-1	Negative
									1310,17000	-1	Negative
									1399,22970	-1	Negative

									1402,21400	-1	Negative
									1291,09600	-1	Negative
									1057,04400	-1	Negative
									1293,15200	-1	Negative

Table 5.3: Transition sets observed in each pooled sample in the DDA experiment. An MRM method was created for each pool based on these transition sets.

LAMp1		LAMn2		LAMn3		LTBI1		LTBI2		LTBI3		NonTB1		NonTB2		NonTB3	
Precursor m/z	Product m/z	Precursor m/z	Product m/z	Precursor m/z	Product m/z	Precursor m/z	Product m/z	Precursor m/z	Product m/z	Precursor m/z	Product m/z	Precursor m/z	Product m/z	Precursor m/z	Product m/z	Precursor m/z	Product m/z
284.3318	73,23	189,0654	100,08	146,1185	72,08	188,1758	72,94	146,0865	69,07	188,1758	72,94	189,0655	61,04	188,1756	72,94	188,176	72,94
(+1)	86,15	(+1)	111,05	(+1)	74,06	(+1)	90,95	(+1)	79,02	(+1)	90,95	(+1)	72,94	(+1)	90,95	(+1)	90,95
	109,10		129,06		86,06		113,97		94,54		113,97		90,95		113,97		113,97
	220,22		131,05		88,08		116,95		97,05		116,95		113,97		128,95		116,95
	230,12		144,05		93,04		131,98		108,56		131,98		129,06		131,98		131,98
	246,13		148,04		100,08		134,97		110,06		134,97		131,05		141,96		134,97
			153,13		102,09		141,96		111,07		141,96		131,98		146,96		141,96
			161,07		118,07		159,97		121,04		148,98		141,96		159,97		146,06
			171,06		127,09		163,94		121,97		159,97		148,04		163,94		148,98
			172,06		128,95		181,96		123,97		181,96		159,97		181,95		159,97
445,3675	51,04	403,1794	69,04	332,2811	98,99	284,17	73,03	188,1759	72,94	189,1598	72,94	284,2947	69,04	189,0657	72,94	189,0658	61,04
(+1)	84,73	(+1)	83,05	(+1)	213,89	(-1)	74,02	(+1)	90,95	(+1)	90,95	(+1)	85,03	(+1)	90,95	(+1)	72,08
	95,7		105,07		225,89		91		113,97		113,97		95,09		113,97		80,95
	151,8		119,05		236,88		101,02		116,95		121,97		98,99		121,97		100,08
	161,57		120,06		251,89		125,03		131,98		131,98		121,07		130,09		129,1
	189,05		129,06		253,89		127,05		134,97		141,96		166,06		131,98		131,05
	202,77		137,06		264,88		145,06		141,96		148,98		174,1		141,96		146,06
	216,65		251,05		276,88		192,11		159,97		159,97		209,12		148,98		149,06
	278,71		290,37		282,89		283,17		163,94		171,06		220,1		159,97		159,97
			329,17		284,89		285,17		181,96		172,98		223,14		172,98		171,15
817,3309	95,05	817,3361	95,05	284,15	74,02			233,1212	90,95	685,4362	92,22	403,1749	69,03			403,1748	69,03
(+1)	109,07	(+1)	109,07	(-1)	91			(+1)	128,95	(+1)	145,12	(+1)	93,07			(+1)	99,05
	137,06		137,06		113,02				146,96		166,06		105,07				105,07
	167,07		177,1		128,03				159,97		222,84		119,05				114,13
	177,1		179,11		130,09				163,94		320,57		121,07				119,05
	299,07		229,13		145,06				181,96		351,71		129,06				121,07
	364,09		247,14		185,12				191		362,43		147,07				129,06
	402,08		402,08		239,13				191,98		520,24		156,14				147,07
	404,09		419,11		241,14				207,04		594,21		160,14				156,14
	421,12		421,13		285,17				212,96		665,37		251,05				361,2
817,3313	95,05	284,1501	74,02					284,2956	84,05			284,1504	61,99			1273,601	147,12
(+1)	109,07	(-1)	91					(+1)	86,1			(-1)	74,02			(+1)	239,05
	123,08		130,09						95,09				91				337,03
	151,08		165,13						129,07				128,03				393,09
	179,11		179,14						132,1				137,02				434,12
	207,11		197,15						166,06				150,06				469,12
	299,07		223,13						174,1				194,05				525,19
	402,08		239,13						197,13				239,13				581,25
	419,1		241,14						220,1				283,19				637,31
	421,12		284,15						234,14				285,17				646,79
1273,613	147,12	432,2025	59,01					685,4362	52,07			432,2029	75,01			284,151	61,99
(+1)	239,05	(-1)	71,01					(+1)	72,89			(-1)	85,03			(-1)	74,02
	295,11		75,01						75,06				91				91
	337,03		85,03						105,96				113,02				128,03
	393,09		91						184,33				160,08				165,13
	434,12		103						261,97				174,09				195,14
	469,12		113,02						319,58				204,07				239,13

	525,19	129,02				370,74			213,11		241,15
	581,25	160,08				524,87			218,08		283,16
	637,33	204,07				673,35			257,08		285,17
284,1692	59,01				284,17	73,03		785,3507	75,01	432,2039	59,01
(-1)	73,03				(-1)	74,03		(-1)	91	(-1)	79,96
	74,02					91			113,02		91
	91					101,02			145,58		113,02
	101,02					109,04			175,02		117,07
	127,05					125,03			388,92		146,06
	145,06					127,05			534,59		160,08
	147,07					145,06			541,27		187,01
	239,13					192,11			719,82		204,07
	285,17					283,17			744,38		226,02
					432,2033	59,01					
					(-1)	85,03					
						91					
						113,02					
						130,06					
						133,07					
						160,08					
						193,03					
						204,07					
						216,13					