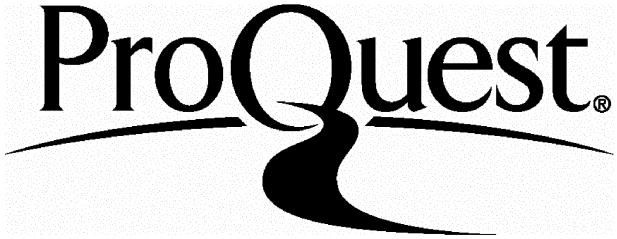
Use of mass spectrometry to identify small molecules, proteins and intermolecular interactions in biological systems

Sibylle Heidelberger

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Department of Pharmaceutical and Biological Chemistry
School of Pharmacy
University of London

June 2011



Plaigerism Statement

This thesis describes research conducted in the School of Pharmacy, University of London between 2004 and 2011 under the supervision of Professor David Thurston, Dr. Mire Zloh, Professor William J. Griffiths at the University of Swansea and Dr. Paul Long at King's College, University of London. I certify that the research described is original and that any parts of the work that have been conducted by collaboration are clearly indicated. I also certify that I have written all the text herein and have clearly indicated by suitable citation any part of this dissertation that has already appeared in publication.

Sibylle Huidlberger — 26/07/2011

Signature — Date

Abstract

The use of mass spectrometry has been increasing in biological sciences over the last few decades as the technology has been getting more sensitive and specialised. From intact analysis of proteins to quantification of peptides or small molecules, the use of mass spectrometry continues to increase as it spreads through more disciplines in biological sciences.

Here, four different questions were attempted to be answered starting with the less complex of identifying hydroxysterols in plasma as a method of identifying changes in the blood stream. Within the plasma ratios were created of 24S- and 27-hydroxycholesterol with cholesterol acids to create a relative quantification technique for analysis. The complexity increases with the identification of proteins involved in the synthesis of cholesterol and its metabolites in rat brain followed by identifying potential proteins involved in the synthesis of a UV absorbing molecule shinorine in the cyanobacteria *Anabaena variabilis*. Both proteomics analyses were able to identify some of the proteins involved in their respective pathways, while the cyanobacteria also quantified the levels of shinorine that were present by HPLC. While a few proteins were identified, it becomes obvious that more separation and sensitivity is needed to identify more of the pathways as they are not in high abundance in their respective proteomes.

The final analysis performed was to understand the interaction of the protein STAT3β with a known and published inhibitor, STATTIC. Using in-gel digest, STATTIC was found to bind to a cysteine residue away from the SH2 domain of STAT3β.

Within each question, mass spectrometry has been shown to be a valuable tool, however it is also shown to have drawbacks. It is shown most valuable when used to complement other techniques such as NMR, SDS-PAGE and Western blots, fluorescence assays and HPLC.

Acknowledgments

I would like to thank all the people who have helped, supported and inspired me during my doctoral studies.

I would like to thank Professor William J. Griffiths for giving me the opportunity to come over here to work and study and guided me through my studies. I would also like to thank Professor David Thurston for his supervision and support and for Dr. Paul Long for giving me the opportunities to work on a rather fantastic project.

I would like to give a special thank you to Dr. Mire Zloh for his support, guidance and friendship and his belief in me, which have been invaluable.

I am very grateful to the School of Pharmacy for hiring me and giving me this opportunity.

A special thank you to Professor David Thurston's PPI group whose training and help with protein purification was greatly appreciated.

A great big thank you for all the friends and colleagues that I have made at the school over the years, in particular Dr. Teresa Barata, Dr. Cathrin Seibert, Dr. Dyeison Antonow, Dr. Cathy Lane, Anique Raether, Dr. Sarah Cousins and Emmanuel Samuel.

I am very grateful to my friends Anita Kilgour, Sam Portlock, Lisa Hagen, Dr. Emma Sharp, Dr. Simon Dawson, Adam Phillips, Dr. Piku Basu, Dr. Zoe Lang, and Alan Arrowsmith for their friendship, support, strength and their help over the years. It has been invaluable.

A very deep and sincere gratitude goes out to my wonderful family. I would not be here without the unconditional support, love and belief of my mother, father, grandmother, brother, sister-in-law and my niece and nephew all of whom I've missed very much. I could not have done this without all of you.

Table of Contents

List of Figures

List of Tables

Abbreviations

1-DE	1-dimensional gel electrophoresis
2-DE	2-dimensional gel electrophoresis
α-CA	α-cyano-4-hydroxycinnamic acid
A_{595}	Absorbance at 595 nm
AA	ammonium acetate
A. microphthalm	*Arcopora microphthalm*
A.variabilis	*Anabaena variabilis ATCC 29413*
ABC	ammonium bicarbonate
ACN	acetonitrile
AQUA	absolute quantification
APS	ammonium persulfate
ATP	adenosine triphosphate
BPI	base peak ion
BSA	bovine serum albumin
°C	degrees Celsius
CE	collision energy
$CHCl_3$	chloroform
CI	chemical ionisation
CID	collision induced dissociation
CO	carbon monoxide
Da	Daltons
DC	direct current
DEAP	diethylaminopropyl
DHB	2,5-dihydroxybenzoic acid
DMSO	dimethylsulfoxide
DTT	dithiothreitol
e	electron
E. coli	*Escherichia coli*
ECD	electron capture dissociation
EDTA	ethylenediaminetetraacetic acid
ETC	electron transfer dissociation
EI	electron impact
ESI	electrospray ionisation
EtOH	ethanol
FA	formic acid
FT	fourier transform

g	grams
g	acceleration due to gravity
gp130	glycoprotein 130
GST	glutathione-S-transferase
HCl	hydrochloric acid
H_2O	water
hr	hour
HMDB	Human Metabolome Database
HPA	3-dihydropicolinic acid
HPLC	high performance liquid chromatography
HUPO	Human Proteome Organisation
IC_{50}	half maximal inhibitory concentration
ICAT	isotope coded affinity tag labelling
IE	ion exchange
IEF	isoelectric focusing
I.D.	inner diameter
IFN	interferon
Igepal	Non-idet P40 nonylphenoxypolyethoxyethanol
IgG	immunoglobulin G
IL6	interleukin 6
IMAC	immobilised metal affinity column
IPG	immobilised pH gradient
IPTG	isopropyl-β-D-1-thioglactopyranoside
IR	infra-red
ITRAQ	isotope tags for relative/absolute quantification
JAK	janus kinase
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	lysogeny broth
LC	liquid chromatography
LDLR	low density lipoprotein receptor
M	molar concentration (moles/litre)
mA	milliamps
MAA	mycosporine-like amino acid
MALDI	matrix-assisted laser desorption/ionisation
Met	methionine
mg	milligram
min	minute
mL	millilitre

mm	millimetre
mM	millimolar
MeOH	methanol
MOPS	3-(N-morpholino)propanesulfonic acid
ms	millisecond
MS	mass spectrometry
MS/MS	tandem MS
MS^n	multiple tandem MS fragementation
mRNA	messenger ribonucleic acid
MWCO	molecular weight cut-off
m/z	mass-to-charge
NaCl	sodium chloride
NAD	nicotinamide adenine dinucleotide
Nano-ESI-LC-MS/MS	nanoelectrospray liquid chromatography tandem mass spectrometry
ng	nanogram
nl	nanolitre
nm	nanometers
nM	nanomolar
NMR	nuclear magnetic resonance
O.D	outer diameter
OD	optical density
ORF	open reading frame
P450	cytochrome P450
PAGE	polyacrylamide gel electrophoresis
PAR	photosynthetically active radiation
PLGS	ProteinLynx Global Server
PMF	peptide mass fingerprinting
PMSF	phenylmethanl sulphonyl fluoride
PPI	protein-protein interactions
ppm	parts per millions
PTM	post-translational modification
QTOF	quadrupole time-of-flight
RIC	relative ion chromatogram
RF	radio frequency
ROS	reactive oxygen species
RP	reverse-phase
rpm	rotations per minute

SA	sinapinic acid
SDS	sodium dodecyl sulphate
SH2	Src Homology 2
SILAC	stable isotope labelling by amino acid in cell culture
S. pistillata	*Stylophora pistillata*
SPE	solid phase extraction
SRM	selective reaction monitoring
STAT	signal transducer and activator of transcription factor
STD-NMR	saturation transfer difference nuclear magnetic resonance
TEMED	N,N,N',N'-tetramethylethylenediamine
THAP	2,4,6-trihydroxyacetone phenone monohydrate
TIC	total ion chromatogram
TOF	time-of-flight
Tris	trishydroxymethylaminomethane / 2-amino-2-hydroxymethyl-1,3-propanediol
TS	tris saline
µl	microlitre
µM	micromolar
µsec	microsecond
UV	ultraviolet
v-src	viral sarcoma
V	volts

Amino Acid Abbreviations

Amino Acid	One-letter code	Three-letter code
Alanine	A	Ala
Arginine	R	Arg
Asparagine	N	Asn
Aspartic Acid	D	Asp
Cysteine	C	Cys
Glutamic Acid	E	Glu
Glutamine	Q	Gln
Glycine	G	Gly
Histidine	H	His
Isoleucine	I	Ile
Leucine	L	Leu
Lysine	K	Lys
Methionine	M	Met
Phenylalanine	F	Phe
Proline	P	Pro
Serine	S	Ser
Threonine	T	Thr
Tryptophan	W	Trp
Tyrosine	Y	Tyr
Valine	V	Val

To my family

1 Introduction

1.1 Mass Spectrometry in Proteomics and Metabolomics

Total systems biology refers to a complete overview of all components, including DNA, mRNA, proteins and metabolites with the aim of understanding their functions and relationships within a biological system i.e., how they change in a quantitative manner. In any biological system, there are many phenomena involved in the profile of a single biological molecule e.g. from post-translational modifications of a protein to its cleavage and translocation.[1] Proteins and metabolites both are pivotal in the function of any cell, tissue or organism and are involved in diverse processes such as respiration, transport of molecules and signalling events throughout the biological system. It has been observed that a protein can interact with many other proteins or metabolites, the result of which may be an event or reaction that can lead to the signalling of further events within the biological system.[2]

While an overall understanding of the system is the main aim, tackling such a problem as a whole can be very difficult, so the research tends to be divided into subsections that can be further scrutinised and this has led to the developments of disciplines such as genomics, proteomics and metabolomics.[1]

In the post-genomic era where 'omics' terminologies are constantly being created, two terms: proteomics and metabolomics have also become quite commonplace. Proteomics refers to a study of any group of proteins which are being investigated, whether it is on the cellular level, a tissue, an organism or a subset of proteins. Metabolites, while dependent on the genome, have been studied for many years before the term 'metabolomics' was coined. From the analysis of glucose in blood for diabetics to the current analysis of small molecule biomarkers for disease in plasma, the study of metabolomics has become important for the understanding of biological systems.[3,4]

The expansion of these fields has only been possible with the developments of new instrumentation and techniques, which over the last 30 years has led to an explosion of data and better understanding of biological systems, processes and their interactions whether it be for a cell, tissue, treated systems, or even disease states. One of the major contributors to these fields has been mass spectrometry and its analytical and quantitative abilities.[2,5,6]

1.2 Mass Spectrometry

Mass spectrometry has boomed in its use from everything such as accurate mass analysis of synthesised compounds and intact mass analysis of large biopolymers to identification of metabolites and sequencing of peptides for protein identification. In almost all cases, it is a sensitive and rapid means for obtaining information from molecules. The ability of MS to obtain fragmentation and accurate mass data can be used to confirm structure and molecular formula of small molecules.[7-9] Mass spectrometry involves the separation, measurement and detection of ions based on their mass-to-charge (*m/z*) ratio. A mass spectrometer is made up of five components: an inlet, an ionisation source, a mass analyser, an ion detector and a computer.[10]

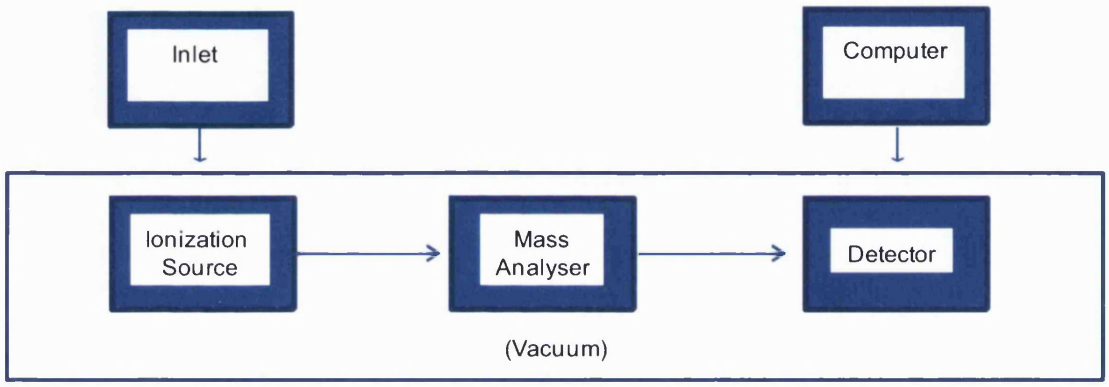


Figure 1.1: The components of a mass spectrometer

The inlet and ionisation source introduce the sample into the mass spectrometer from atmospheric pressure into a vacuum where it passes through the mass analyser and is detected at the other end. The acquired data is translated into a spectrum of intensity versus *m/z* ratio.[5,10, 11]

1.2.1 Ion Sources

The ionisation source transfers molecules to the gas phase as ions. There are a wide variety of techniques for ionisation including the earlier ionisation sources such as electron ionisation (EI),[10,12] chemical ionisation (CI)[10,12] atmospheric pressure chemical ionisation (APCI),[10] and fast atom bombardment (FAB).[10] The more common ionisation techniques such as ESI and MALDI are used for analysis of biological systems.[13] Each of these techniques has advantages and disadvantages as well as their own niches in chemistry and biology where they are most used. ESI and MALDI are both routinely used in metabolomics and proteomics as they are considered 'soft ionisation' techniques, as they allow for product ions as well as the intact mass to be analysed. ESI can easily be coupled to liquid and gas chromatographic techniques for separation of analytes. MALDI is also used for large biopolymer analysis as well as protein digest for identification.[13]

1.2.1.1 ESI

ESI, or electrospray ionisation is a technique that allows for ionisation at atmospheric pressure.[14] An electrospray is produced when liquid is passed through a small diameter needle or capillary in the presence of a strong electric field.[15] A potential difference is created between the capillary and a counter-electrode to produce the electric field releasing the analyte mixture as a plume of fine droplets.(Figure 1.2).[14,16]

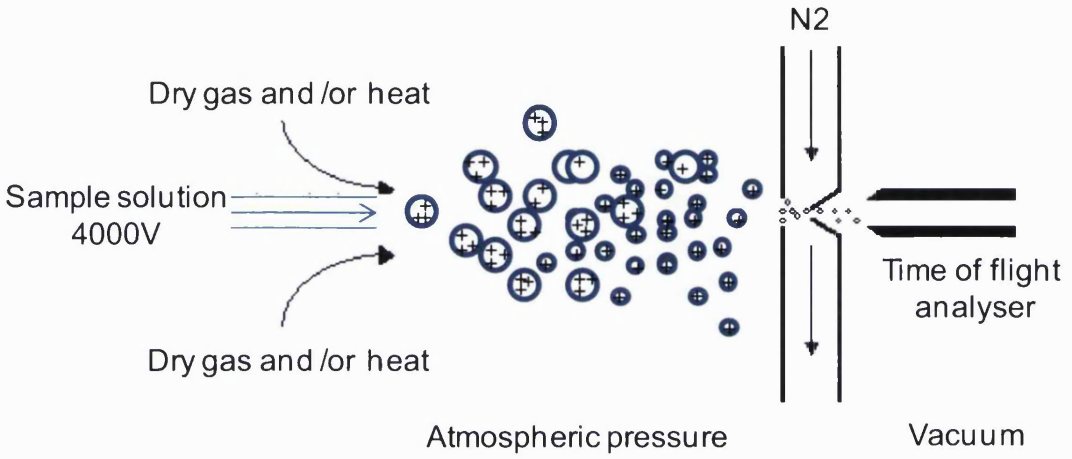


Figure 1.2: Example of the ionisation process on the tip of an electrospray ion source. A positive charge is applied to the capillary which results in positively charged ions to be dispersed from the tip. These migrate away from the capillary as small droplets which are reduced to size and eventually disappear leaving only the ionised analyte. Image modified from Lane, 2005.[13]

In the case of a positive charge applied to the capillary, positive ions are produced while a negative charge results in negative ions.[17,18] A Taylor cone is produced at the end of the capillary tip and the droplets that are formed disperse and move towards the counter electrode. The spray is desolvated by heat and counter-current flow gas as it moves through the atmospheric pressure chamber.[14,17-19] As the solvent is removed from the droplets, the surface density charge of the droplets increases, resulting in an increase in Coulombic forces which finally exceeds the Raleigh stability limit and the surface tension is overcome. The droplets explode, resulting in smaller and smaller droplets until the release of gas phase ions.[14,17,20] As a result, multiply charged ions are formed by both charge accumulation and by electrochemical processes at the probe tip. Ions are produced depending on the ability of the compound to hold a charge, either positive or negative. Larger molecular weight analytes, such as proteins can hold more charges and become multiply charged.[14] This allows for high molecular weight compounds to be analysed on a low *m/z* range instrument. At a high resolution, the charge state and the molecular weight of proteins can be determined. Lower molecular weight compounds such as peptides, do not require as high resolution for charge state determination. Multiply charged ions allow for improved mass accuracy. Calculation of the original mass of the compound can easily be done from a knowledge

of *m/z* and charge state or by using two peaks of different charge states.[16] Software is available that allows for deconvolution or calculation of the original mass of the multiply charged ion.[10,17,21]

ESI and nano-ESI (ESI at very low flow rates) have, in routine use, good sensitivity for femtomole analysis of peptides using nano-ESI. In fact, low flow ESI has been shown to provide an increase sensitivity that is believed to be due to the formation of smaller droplets.[16,21,20] This allows for even less sample to be used as the sensitivity is increased. ESI is easily adapted for LC and nano-LC. Although ESI can analyse molecules of wide range of molecular weights, including low molecular weight samples, this technique has disadvantages as it is very intolerant to salt and spectra of mixtures of high molecular weight compounds can be difficult to comprehend.[13,14,16,19]

1.2.1.2 MALDI

MALDI refers to matrix-assisted laser desorption/ionisation mass spectrometry. Unlike ESI, it involves a sample in a solid form rather than a liquid form. MALDI and large molecular weight analyses using this technique was first introduced in 1988 by Tanaka, Karas and Hillenkamp and has been used for many different biopolymers.[22,23] Currently a large variety of samples can be analysed by MALDI such as peptides and proteins[24-32], oligonucleotides[33-37], post translational modifications[3,38], carbohydrates[39], lipids and polymers as well as imaging of tissues, viruses and bacteria.[40-44] The use of MALDI for analysis keeps increasing with new and novel techniques and instruments constantly being developed. A search on MALDI gives a phenomenal number of publications per year, a consequence of the increased use of this technique.

The principle of MALDI involves transfer of ionisation energy from a matrix to an analyte. The acquisition of the MALDI spectrum involves two steps, sample preparation and data acquisition. The first involves mixing the compound to be analysed with a solvent containing a matrix – a small molecule with a strong absorption to the laser wavelength used either UV or IR. The matrix is usually a large weak organic acid which is in excess to the sample.[45,46] The mixture is spotted onto a metal plate and the droplet is allowed to dry, creating a visible solid deposition on the target plate. Three main types of spotting methods are used. Droplets are created by either mixing prior to spotting onto the target plate (dried droplet), spotting matrix onto the plate followed by a drop of sample (layer) or sandwich method which involves spotting matrix, sample and matrix again, allowing each droplet to dry before spotting the next layer. The dried mixture forms crystals which are normally coloured in nature.[35,47]

Once the sample is dry, the plate is loaded into the MALDI where it is under vacuum. Irradiation of the mixture by the laser results in energy being absorbed by the laser absorbing matrix crystals, causing excitation of the molecules. The intensity of the laser also causes sublimation of matrix and sample into the gas phase in a plume (Figure 1.3). The co-crystallisation in the presence of excess matrix allows the sample to vaporise but not requiring direct absorption of the laser energy by the analyte, thus reducing the amount of laser induced fragmentation. The exact mechanism of desorption and ionisation of the sample and the method in which the matrix transfers its energy is not well understood, although gas-phase photoionisation, excited state proton transfer, ion-molecule reactions, desorption of pre-formed ions are all suggested theories. One of the more widely accepted methods of ion formation involves gas-phase proton transfer in the expanding plume of matrix molecules.[35,48-50]

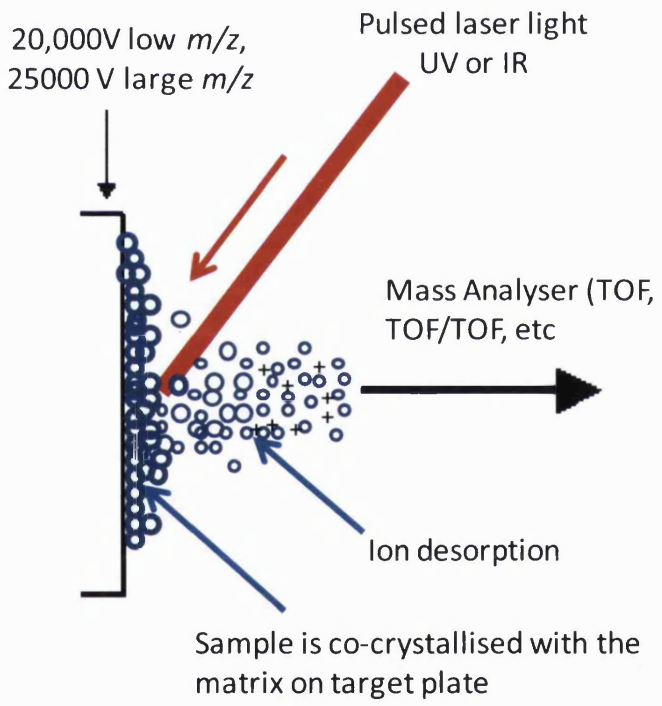


Figure 1.3: Desorption and ionisation of sample from a MALDI target plate. The mixture of matrix and sample are spotted onto a target plate and allowed to dry before being ablated by a laser of UV or IR wavelength which the matrix is able to absorb. The matrix and sample are desorbed and ionised in a plume as they leave the plate and enter the mass analyzer.

Through the desorption/ionisation process, the sample tends to maintain only one charge leading to $[M+H]^+$ ion forming. Large proteins have been known to carry 2 or more charges.

MALDI is more tolerant to salt than ESI, allowing for some analyses in complex buffer mixtures. Matrix type is dependent on the nature of the sample to be analysed. Matrices must be able to absorb either UV or IR energy,[51] must have a low mass for sublimation, must be stable and must lack chemical reactivity. Common matrices used for UV MALDI are listed in Table 1.1.

Table 1.1: Common UV MALDI matrices.[47]

Matrix	Applications
α-cyano-4-hydroxycinnamic acid (α-CA)	Peptides, proteins, organic compounds
3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid) (SA)	Proteins, high mass biopolymers
2,5-dihydroxybenzoic acid (DHB)	Peptides, proteins, carbohydrates, polyethylene glycol and other polymers
3-hydroxypicolinic acid (HPA)	Oligonucleotides
Trihydroxyacetonephenone (THAP)	Oligonucleotides, peptides

Once desorbed and ionised, the charged molecules are directed by electrostatics from the MALDI source into the mass analyser where they are analysed. Mass range for a MALDI is dependent on the mass analyser used, typically a TOF type. At the low *m/z* range, MALDI analysis is inhibited by the ionisation of the matrix molecules at < 600 *m/z* although small molecule analysis is still possible with specific types of matrices. The upper limit of a MALDI, if using a TOF mass analyser, is theoretically infinite. TOF allows for accurate mass determination when run in reflector mode, which can be used for protein identification using peptide mass fingerprinting or Edman degradation. In the linear mode, MALDI allows for the theoretical analysis of any high molecular weight polymer as long as it is able to hold a charge, and ionise into the gas phase for analysis.[22,46,52]

Resolution in reflector mode (Section 1.2.2.3) on a TOF is such that isotopic resolution is visible and is useful for protein identification using peptide mass fingerprinting. At high *m/z*, the mass accuracy with an internal calibrant is +/- 0.1% for most MALDI-TOF instruments allowing for analysis of large molecular weight biopolymers such as IgG, as well as being able to identify post-translational modifications on smaller proteins.[23, 50,53]

1.2.2 Mass Analysers

Mass analysers are the key component to a mass spectrometer, separating *m/z* of small or large molecules using their *m/z* either by a magnetic or electric field. The most common types of mass analysers are quadrupoles, ion traps and time-of-flight (TOF). The mass analysers are required to be sensitive, and ideally provide high resolution and mass accuracy.[5] New instrument types have also been developed which have

increased sensitivity and resolution allowing for higher mass accuracy: FT-ICR[54,55] and Orbitrap[56-58] are two examples of the new technology that has been developed.

1.2.2.1 Quadrupole

A quadrupole is made up of 4 rods either circular or hyperbolic. Pairs of rods are held adjacent to each other and maintain opposite polarities (Figure 1.4).

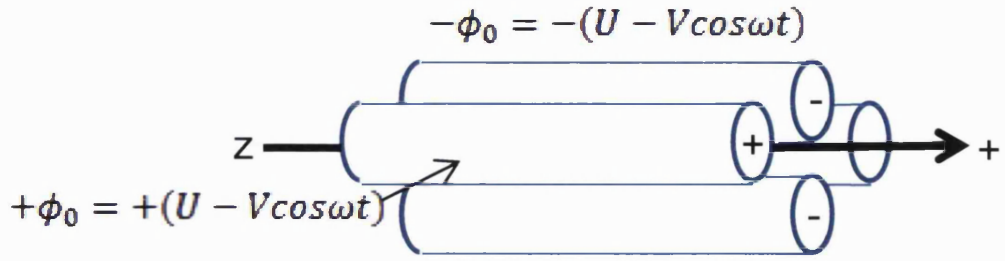


Figure 1.4: Schematic of a quadrupole. RF and DC voltages applied along x and y axis changing to scan across the field and allowing only ions with stable trajectories in the scanned region to pass through. Φ_0 represents the potential applied to the rods, ω is the angular frequency (rad/s = $2\pi\nu$ where ν is the frequency of the radio frequency, U is the direct current and V is the RF amplitude. Modified from Lane, C 2005.[13]

As the ions travel along the z-axis, they enter the space between the quadrupoles at a constant velocity in the z-direction. Along the x- and y-axis, they are subjected to acceleration which is a result of the applied electric field forces – radio frequency (RF) and direct current (DC). Parallel quadrupoles are connected and have the same DC and superimposed RF voltages. The opposite poles have the opposite DC voltage and the RF phase is shifted by 180°. The RF and DC voltage amplitudes are changed from a few volts to 1000 V when scanning the quadrupole, however the frequency does not change and the RF/DC voltage ratio remains constant while scanning. As the quadrupole is scanned, ions of a particular *m/z* start going through the quadrupole and fluctuate until a stable trajectory is established by the applied electric fields for specific *m/z*, allowing it to pass through the quadrupole. Hence the quadrupole is considered a scanning mass analyser, scanning across RF and DC allowing specific *m/z* through one at a time causing those that are not selected to have unstable trajectories resulting in the ions hitting the quadrupoles instead of going through the mass analyser to the detector.[10]

1.2.2.2 Ion Traps

An ion trap is essentially a quadrupole analyser bent on itself with the inner quadrupole rod as a point in the centre of the trap, the outer rod as the ring electrode and the top and bottom rods as the end caps (Figure 1.5).[59]

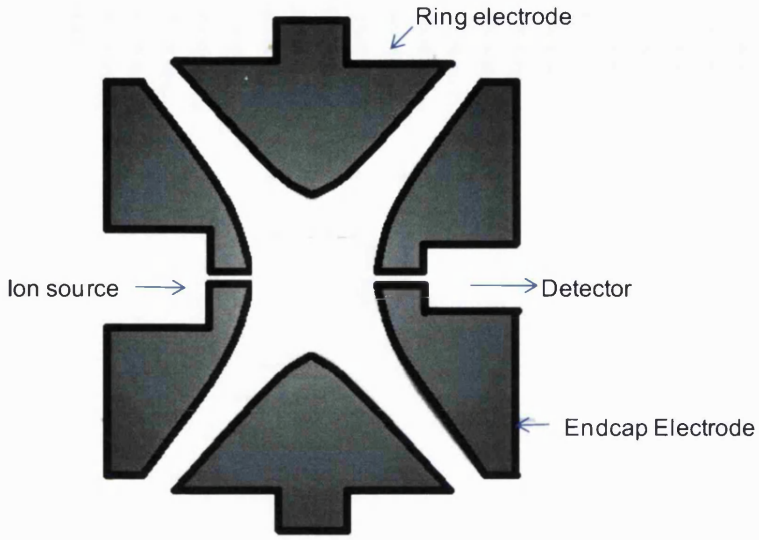


Figure 1.5: Schematic representation of an iontrap. Ring electrodes and endcap electrodes maintain ions of certain *m/z* in the trap for analysis and allow all others to be ejected to the detector. Selected ions can then be fragmented to acquire MS/MS data or MS^n data.

Ions which enter the ion trap are held by oscillating electric fields generated by the RF voltage applied to the ring electrode. Unlike a quadrupole, no DC voltage is applied. Ions are guided to the trap where they are sampled at the same time and are confined by the RF by taking on an oscillating frequency related to their *m/z*. Ions repel each other in the ion trap, expanding their figure-8-like trajectory as a function of time. Helium gas is used to prevent expansion of the trajectories, keeping the ions in the centre by dampening the excess energy.[12,59]

Scanning the RF frequency increases the amplitude and this increases the frequency and oscillations of the ions at the same time as an RF is applied to the end caps. Once selected by applying RF at the end caps, the oscillations become so large that the trajectory destabilises and the ions are ejected along the axis of the end cap.[12,59,60]

Tandem MS (MS/MS) or multistage MS (MS^n) is performed by selecting an *m/z* and ejecting the rest of the ions by changing the RF. Once selected, partial destabilisation of the oscillating frequency of the ion is carried out by allowing the resonance of the end caps to increase the energy of the ion. The addition of helium gas causes further destabilisation and induces fragmentation. From this mixture of fragment ions a second *m/z* can be selected and fragmented and this 'tandem MS' can continue until the signal is too low to go any further.

Newer types of ion traps such as linear ion traps[15,61] have better sensitivity and mass accuracy than the simple quadrupole ion traps.[12] Orbitraps are also a type of ion trap that have far greater sensitivity and mass accuracy making them a very useful tool.[56,58]

1.2.2.3 Time of Flight (TOF)

The time-of-flight (TOF) analyser was first described by Stephens *et al.*[62] Ions are expelled from the source in packets, such as from a laser pulse or a source focusing lens. The ions are accelerated by an electric potential and they travel a distance *(d)* prior to reaching the detector. The *m/z* is determined by measuring the time that ions take to get through a field-free region between the source and the detector. The relation between the kinetic energy of an ion leaving the accelerating cell, its mass *(m)* and its charge *(q = ze)* shows a dependency of the kinetic energy *(E_k)* on the potential energy *(U)*[63]:

$$\frac{mv^2}{2} = qU = zeU = E_k$$

Where v is the velocity of the particle.

The time required to fly distance d: $t = \frac{d}{v}$

depends on the *m/z* : $t^2 = \frac{m}{z}\left(\frac{d^2}{2Ue}\right)$

This can be used to calculate *m/z* using the measured time as the distance and applied potential remain constant during the acquisition.

Mass resolution is affected by factors that cause a distribution of flight times among ions of the same *m/z*. The distribution is the result of variations in time and space of the ionisation event, and variation in kinetic energy of the ions. Mass resolution is dependent on flight time, and increasing the flight tube is one way of increasing the resolution.[64,65]

A means of reducing kinetic energy spread among ions of the same *m/z* is to leave a delay between ion formation and extraction. Ions are allowed to expand into a field-free region and after a short period of time, a voltage is applied to extract ions from the source. This method is referred to as delayed pulse extraction. As ions with more energy within the extraction field-free region move closer to the detector, a pulse is applied after a short period of time and the delay allows more energy to be transmitted to the slower ions and less to the faster ions, allowing them to focus and move at the same velocity towards the detector.[64,65]

While this technique decreases peak broadening, it alone does not provide high-resolution. To improve the resolution further, an electrostatic mirror referred to as a reflectron is used to improve the resolution up to *m/z* 5000.

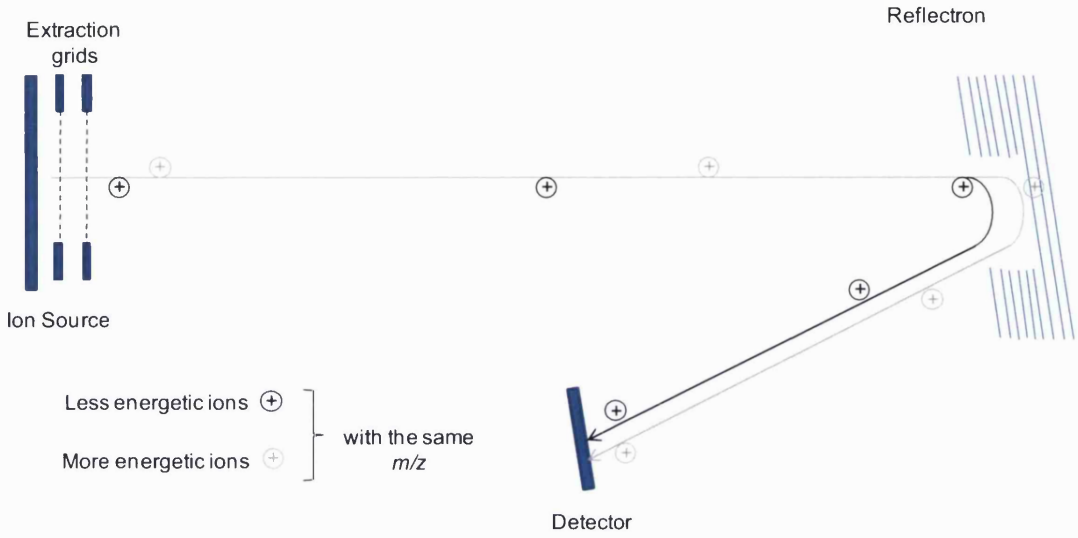


Figure 1.6: Schematic of a time-of-flight reflectron mass analyser. Samples travel through the flight tube, spreading out to be focused in the reflectron before reaching the detector. Modified from Lane, C, 2005.[13]

Despite the use of the delayed extraction, the ions moving towards the detector still have a small amount of spreading velocity (kinetic energy) that does not allow for good resolution. The reflectron corrects the dispersion by allowing ions of greater kinetic energy to penetrate further into a series of grids and electrodes that work as an ion mirror, retarding the ions. As the faster moving ions penetrate further into the reflectron, the slower do not penetrate as far, before all the ions are reflected back to the detector with the same kinetic energy and speed (Figure 1.6).[10]

1.2.3 Coupling Ion sources to mass analysers

MALDI analysis has normally been coupled to TOF analysers with both reflectron and linear methods of analysis. This allows for peptides to be analysed for peptide mass fingerprinting and intact protein analysis in linear mode. The use of TOF-TOF instruments allows for the peptides identified to be fragmented.[10,11]

ESI traditionally has been coupled to liquid chromatography for separation on the inlet front. Ion traps, triple quadrupoles (and single quadrupoles) as well as QTOFs, have been coupled to ESI to allow for precursor ion scanning and fragmentation of *m/z*. FT-MS are also normally set up with ESI allowing for online liquid chromatography separation.

While traditional instruments are still keenly used, new instrumentation and developments such as MALDI-ion traps or triple quadrupoles and other combinations have been developed. Instruments like FT-ICR have been coupled to new instruments such as Orbitrap creating a new, top of the line instrument which has great signal and sensitivity but with the increased cost.

1.2.4 Tandem Instruments

One of the best outcomes for the analysis and identification of proteins was the development of tandem instruments that take advantage of the abilities of individual instrumentation as well as adding an extra dimension to the abilities of each instrument such as allowing for tandem MS. Hybrid instruments refers to instruments with more than one type of mass analyser.

1.2.4.1 Triple Quadrupole

Triple quadrupoles are still very common instruments. They involve three tandem quadrupoles or hexapoles. The first is set as a mass filter, allowing all ions to pass through, while the second is referred to as a collision cell and is used to fragment *m/z* ions which pass through. The third quadrupole is again used as a mass filter, allowing ions to pass through to the detector. In this setup, tandem MS, selective reaction monitoring (SRM) and other experiments can be performed. Fragmentation of ions is performed using argon gas as the inducer – hence it is referred to as collision induced dissociation (CID) or collision associated dissociation.[60]

There are four possible tandem MS experiments that can be achieved using a triple quadrupole:

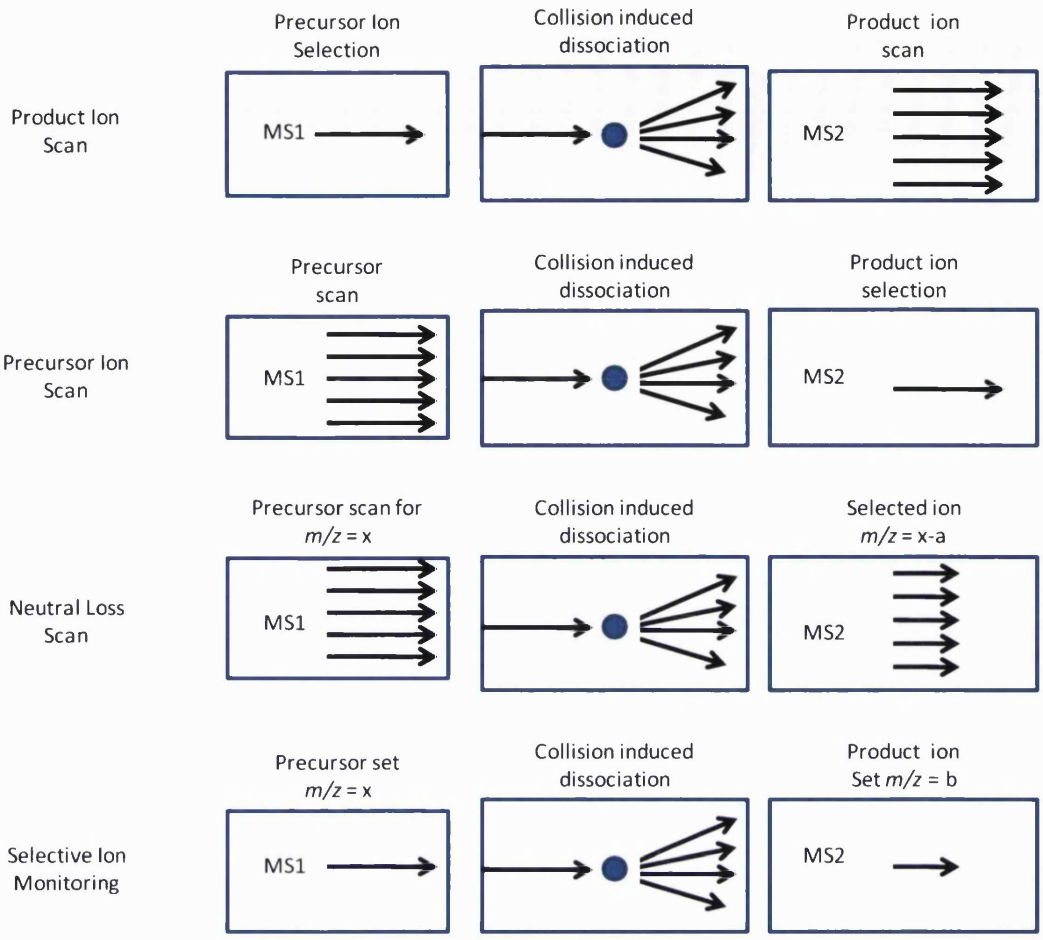


Figure 1.7: Four types of tandem analysis experiments that can be performed using a triple quadrupole.

- Product ion scan – a precursor ion is scanned and undergoes CID and all products ions are then scanned using MS2 and detected. This is the most common type of tandem experiments and can be done on all tandem MS instruments.
- Precursor ion scan – a product ion is set in MS2 while all precursor ions are scanned in MS1 to identify the precursor ion of interest.
- Neutral loss scan – both MS1 and MS2 are scanned together, MS2 looking for an offset mass from MS1.
- Selected reaction monitoring – MS1 and MS2 are set to selected masses, one of which is a product ion of the other. This selectivity increases the sensitivity of the experiment.

Triple quadrupoles have been used for protein identification, metabolomics, protein interactions and many other types of experiments, making this a versatile instrument especially for quantitative analysis.[60]

1.2.4.2 QTOF

A QTOF refers to a hybrid instrument with at least two quadrupoles in tandem followed by a TOF – usually a reflectron. The first quadrupole is normally set as an ion guide in MS mode or to select specific ions in MS/MS mode. The second quadrupole or collision cell can either be set as an ion guide in MS mode or set for CID in MS/MS mode. The TOF is the final mass analyser and is used for resolving both MS and MS/MS ions as they pass through. Replacing the final quadrupole with a TOF increases the resolution, accuracy and sensitivity of the instrument.[65]

1.2.5 Chromatographic Separation

While all mass spectrometers can work in standalone mode, a more convenient way of handling samples and separation is to couple a mass spectrometer to liquid or gas chromatography. In most cases, for proteomics, reversed-phased LC is used which is based on the hydrophobic interactions between the amino acid side chains and the hydrophobic surface. As the polar mobile phase flows through the column the peptides bind to the hydrophobic phase to minimise the exposure of their hydrophobic portions. As the organic content of mobile phase is increased, the peptides are eluted off in the order of increasing hydrophobicity.[20,66]

Metabolites can be analysed using a reverse-phase or normal phase approach for separation which is dependent on their nature.[67,68] Hydrophilic metabolites are analysed in a fashion similar to that of peptides, using mobile phases with higher aqueous content and increasing organic solvent for separation. More hydrophobic metabolites are separated using normal phase, allowing the mobile phase to start with higher organic concentration than that used for proteomics.

1.2.5.1 Capillary and nano HPLC

Nano and capillary HPLC have proven to be very useful tools for proteomic and metabolomics studies as the low volume of nanospray increases the sensitivity in MS experiments. This enables better analysis using less sample; an important point when working with unstable and low abundance molecules.

Nano-LC as it states, uses nanolitre flow rates (50 - 400 nL/min) while capillary flow (1-50 µL/min) is somewhat higher. In most cases, capillary HPLC is split to get to low nanolitre flow for nanospray, as pumps which go down to nanoflow are not very common. The same separation methods apply here as they do with HPLC, however the size and shape are on a smaller scale. Mobile phases for proteomics and metabolomics can vary accordingly to the analytes of interest as long as the buffers and solvents are mass spectrometry compatible. In other words, buffers must be a low

concentration of volatile salts such as ammonium acetate (AA) while acids, bases and organic solvents must also be volatile such as formic acid, acetic acid and acetonitrile. In most instances, unless exceedingly hydrophobic analytes are used, some sort of aqueous buffer or acidic aqueous environment is used. While the issue of volatility is somewhat limiting, it still allows for sufficient diversity to enable the analysis and separation of a wide variety of analytes.

1.2.6 Fragmentation

Structure elucidation of small molecules or peptides can be done by fragmenting the *m/z* of interest and interpreting the fragment ions. Ionisation methods such as EI and CI cause source fragmentation which is useful for small molecule identification as it is very reproducible. For larger molecules or thermally labile molecules, CID is considered a better approach.

1.2.6.1 CID

One of the most common types of fragmentation used is CID, or collision induced dissociation, a type of fragmentation common for ESI – QTOF, triple quadrupole and iontraps.[69] Fragmentation occurs when an *m/z* of interest is selected and a neutral or inert gas is applied into either the collision cell in the case of tandem instruments or directly into the ion trap.[70]

Sequencing of peptides for proteomics involves the use of tandem instruments which allows for the selection, fragmentation and analysis of product ions. In the case of a hybrid instrument such as a QTOF the fragmentation involves collision induced dissociation using a neutral and inert gas such as argon,[63,71] while ion traps use helium.

As the ions enter into the first quadrupole, a selected ion is allowed to pass through and all the others are lost. As the selected ion enters the hexapole collision cell, an offset DC voltage is applied to ions, which increases their kinetic energy. The ions collide with a neutral gas such as argon, and fragment. The fragments then pass through to the last mass analyser where they are detected.[10,69]

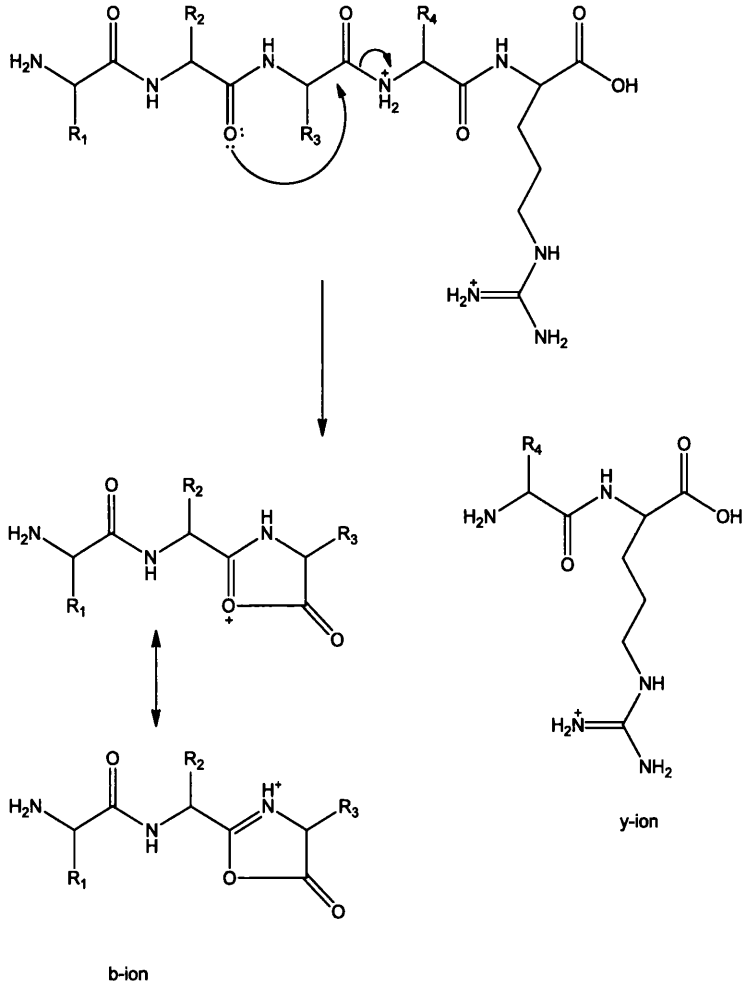


Figure 1.8: Fragmentation of the peptide backbone using CID. Protonation of the peptide backbone induces fragmentation and cleavage of the amide bond.

Fragmentation is most easily explained through peptide fragmentation theory (Figure 1.8). As ions are transferred to the gas phase, desolvation cools the ions, lowering their energy. Dissociation of the peptide is promoted by a mobile proton, which is needed at the cleavage site. As energy is added, it alters the initial protein population and the protons mobilise, increasing the population of protonated peptides with higher energy. These protons are normally located on the backbone heteroatoms and initiate cleavage to b- or y- type ions (Figure 1.9). Immonium ions are also formed, resulting from multiple cleavages of the peptide chain and can be identified at the lower end of the *m/z* spectrum. The data that is created is usually processed and stored into a format that can be used for searching using available databases.[69,70]

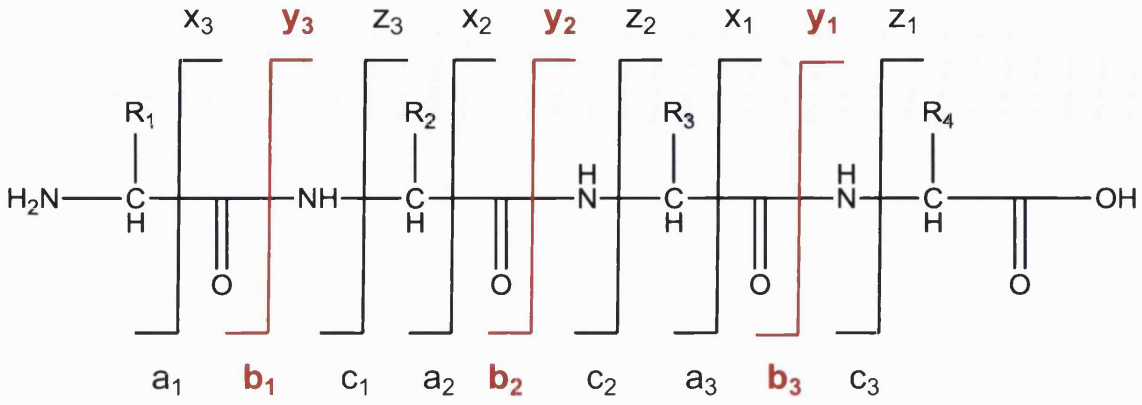


Figure 1.9: CID fragmentation pattern of a peptide explained. Red denotes the b- and y-ions typical of CID.

Single protonated peptides need a very strong amino group to relocate the proton and require more energy if an extra proton is not present. Doubly charged peptides tend to initiate fragmentation at the C-terminal residue and the N-terminal amine.

Fragmentation of small molecules and metabolites, while more complex, is also feasible and allows for more structure elucidation when identifying small molecules of interest. CID is useful when compared to standards or similar compounds of interest.[69]

1.3 Proteomics

The study of proteomics has evolved out of the sequencing of genomes of various organisms.[5,10] Currently, proteomics can be defined as the determination of gene products i.e. proteins, and their cellular functions. Sequencing of entire genomes resulted in the discovery of new and previously unidentified genes with unknown function, expression profile and identity. The study of gene products is complicated by the fact that the sequence of a genome does not truly reflect the number of proteins that can be transcribed, as the transcriptional and translational processes from DNA to protein are complicated by splicing events. Additionally, the translation from mRNA to protein may show little correlation between mRNA and protein expression levels, limiting the usefulness of mRNA microarrays. Furthermore, the prediction methods and software used to determine the function of genes are not very accurate and so cannot yet be considered reliable.[1]

The complexity of the proteome is confounded by the presence of both high and low abundance proteins making it very difficult to identify the low abundant ones, thus requiring sensitive analytical tools and separation techniques. Mass spectrometry has become a clear method of choice for protein detection, although this was only made

possible by the availability of gene/genome sequence databases for searching the data acquired.[5,72]

Expression proteomics is the most common type of studies performed. It results in a global profile of the expression of proteins from a genome at a given moment. In the past, this particular type of proteomics experiment was performed using 2D-gel electrophoresis, however the current trends for analysis involve 1D-gel electrophoresis with subsequent analysis by MS or a combination of tandem LC or solid phase separation techniques followed by MS. 2D or 3D-LC separations have become a popular choice, whether it be peptides, proteins, or metabolites.[1] New analytical techniques and software have been developed which allow for the investigation into differential expression proteomics involving the comparison of a proteome amid two or more states (disease or other).[73]

Experimentally, considerable attention must be paid to sample preparation in proteomics as well as the resulting quality of data. Most proteomic and genomic studies performed, especially on human samples, are considered observational studies as the proteome or genome is being observed at a particular time during a particular point of a lifetime. While the genome may be regarded as static, this is certainly not true of the proteome. Additionally, sample composition can be altered through minor changes in the sample collection as is observed by the differences between blood samples that are taken from patients while they are sitting, standing or having just eaten. These effects make it generally difficult to perform quantitative studies without taking extra care in the collection, preparation and the handling of the samples.[74]

While much new information is being obtained from proteomics, the examination of eukaryotic tissues and cells is complicated by post-translational processing of proteins and gene-splicing that are not directly apparent from the DNA sequences, making correlation of gene to protein somewhat difficult. In the case of post-translational modifications these can include phosphorylations, acetylations and glycosylations amongst others, which are the foundation of signalling within the cell. To complicate things further, the state of proteins is in constant flux, making it hard to determine what point is being viewed.[5]

Although there are many points to consider, it should be noted that separation of any kind can help reduce the complexity and allow for the identification of more proteins from the system.[75]

1.3.1 Protein Separation

The use of separation for proteomics has been questioned as more groups are working on techniques that do not require separation such as blood droplet, urine or tissue analysis.[40,43,76] However the traditional methods of analysis require some sort of separation to remove unwanted molecules and to extract the proteins of interest for further analysis.[77] This tends to be the rate-limiting step. Too little separation and only the most abundant proteins will be seen. Too much separation becomes time consuming and low abundance proteins can be lost due to excessive sample handling.[74] Proteins are expressed over a very large dynamic range i.e. 10 → 10^6 copies/cell. Many of the most interesting proteins, such as growth factors and receptors, are found in very low number of copies in the cell.[1,6, 28,78]

1.3.1.1 Electrophoresis

SDS-PAGE or 1-DE gel electrophoresis (1-DE) is a very reliable, robust method for separating proteins in a mixture. It allows for the approximation of a relative molecular weight, relative concentration of pure proteins, and can be used as one step in a 2-dimensional separation for protein identification. On its own 1-DE can be used to identify a protein of interest by in-gel digestion of a "purified" protein and then analysing by peptide mapping, or by being the first step in immunoblotting when an antibody is available for binding.[79] Separation of organelles or separation of another type prior to 1-DE can also reduce complexity.

1-DE on its own is insufficient for whole proteome analysis. A second method of separation is required such as digestion and LC-MS/MS.[80] The classical method of using 2-dimensional gel electrophoresis (2-DE) for qualitative and quantitative expression proteomics from cellular lysates is still routinely used for comparative proteomics,[81,82,83,84] although it has many limitations. 2-DE uses two physical properties of proteins for separation. The first separation is done by isoelectric focusing (IEF) usually on an immobilised pH gradient (IPG) gel strip which separates by isoelectric point. The strip is then laid onto the top of an SDS-PAGE and the proteins migrate through the gel, separating by size.[85,86] While the technique is sound, it cannot separate more than ~1500 proteins, most of which are the high abundance proteins. It is also limited by the amount of starting material that can be applied onto the IEF strip, limiting the number of proteins that can be seen at any time on the gel. 2-DE also has difficulties with basic and hydrophobic proteins, which cause precipitation on the IEF strip, preventing separation.[10,87,88]

While 2-DE is useful, other 2-dimensional methods or combinations of methods have been used in current proteomic analyses.[87,89-94] Using multidimensional protein

identification techniques or MudPIT – on peptides or proteins allows for more separation. Most commonly SCX is used followed by reverse-phase LC – either online or offline and in tandem with a mass spectrometer.[82,95] The increase in identified proteins compared to 2-DE is higher however it is at the cost of time as these experiments take longer to run. MudPIT[82], 2-D LC-MS or 1-DE LC-MS[80] are types of general separation methods used to identify all the proteins within an entire proteome, nevertheless with different approaches, subsets of proteins can be analysed such as cysteine containing proteins, or phosphoproteins. In the case, where a simple subset of proteins of interest are desired for analysis, other techniques such as co-immunoprecipitation, affinity chromatography and immuno-affinity chromatography are used. These separation techniques allow for a reduction in the complexity of the sample by using characteristic properties of the subset of interest.

An illustration of the use of multiple dimensions of separation is provided by the work of Adachi *et al* who used urine as a sample for the identification of proteins in the human proteome. Prior to their work, the best identification was performed using 2-DE followed by MALDI[26] or LC-MS/MS. Only 150 proteins could be identified with reasonable confidence leaving over 1000 observed protein spots which could not be identified. The studies of Adachi *et al* showed that by using ultrafiltration, 1-DE, reverse phase LC, followed by MS analysis resulted in over 1500 proteins being identified- an incredible increase in identification.[96] Having sufficient amounts of sample to start with to do proper analysis, as is the case with urine samples, is very beneficial, although sometimes impossible to obtain.

1.3.2 Protein Identification

Two common methods to identify proteins are peptide mass fingerprinting (PMF) by MALDI-TOF[38,83] or by MS/MS which can be performed by ESI-LC-MS/MS, MALDI-TOF/TOF, ion trap or any other instrument capable of tandem MS.[96] In most cases, proteins are digested, in solution or in-gel by proteases or by chemical means into peptides (Table 1.2).[97]

Table 1.2: List of proteases, cleavage reagents and their specificities. X- refers to the N-terminal side of the amino acid to be cleaved, while -Y refers to the C-terminal side of the amino acid. The '/' is the side of the amino acid which undergoes cleavage.[9]

Protease	Amino Acid Specificity		Exceptions
Trypsin	X-**Lys**/-Y X-**Arg**/-Y		Does not cleave if Y = **Pro**
Endoproteinase Lys-C	X-**Lys**/-Y		Does not cleave if Y = **Pro**
Clostripain	X-**Arg**/-Y		
Endoproteinase Asp-N	X-/**Asp**-Y		Does not cleave if Y = **Ser**
CNBr	X-**Met**/-Y		Does not cleave if Y = **Ser, Thr, or Cys**
Glu-C (V8 Protease (E))	X-**Glu**/-Y X-**Asp**/-Y		Does not cleave if Y = **Pro**
Pepsin	X-**Phe**/-Y X-**Leu**/-Y X-**Glu**/-Y		Does not cleave if Y = **Ala, Val, Gly**
Endoproteinase Arg C	X-**Arg**/-Y		Does not cleave if Y = **Pro**
Thermolysin	X-/**Phe**-Y X-/**Met**-Y X-/**Ile**-Y	X-/**Ala**-Y X-/**Leu**-Y X-/**Val**-Y	Does not cleave if X = **Pro**
Chymotrypsin	X-**Phe**/-Y X-**Trp**/-Y	X-**Tyr**/-Y X-**Leu**/-Y	Does not cleave if Y = **Met, Ile, Ser, Thr, Val, His, Glu, Asp**
Formic Acid	**X-**/Asp-**Y**		

Trypsin, a serine protease, is one of the most widely used proteases for digesting proteins as it cleaves specifically at the C-terminal of lysine or arginine residues giving a positive charge to the peptide.[98] These digests can then be analysed by MALDI-TOF to create fingerprints of peptide masses or can be further separated and fragmented using tandem MS. Peptide mass fingerprinting has its disadvantages since it relies on the mass accuracy of the instrument and it requires a pure protein as mixtures are difficult to analyse and identify individual components.

Using MS/MS on peptides increases the probability of identifying the protein as both mass of the peptide and the MS/MS data are used to compare to the peptide mass and MS/MS data of the theoretical proteins in the database, thus increasing the information used to identify the proteins with higher probability.

1.3.3 Protein Quantification

Quantification by mass spectrometry was not considered to be straight forward as many factors can affect the quality of the signal making it difficult unless related to an internal standard. Quantification can be done as a relative (comparison) or absolute (internal calibrant) quantification using gel or gel-free systems (Figure 1.10).[99] 2-DE uses comparisons between control and treated samples to identify changes in the proteome, which are then cut out, digested and analysed.[83,91,92,100]

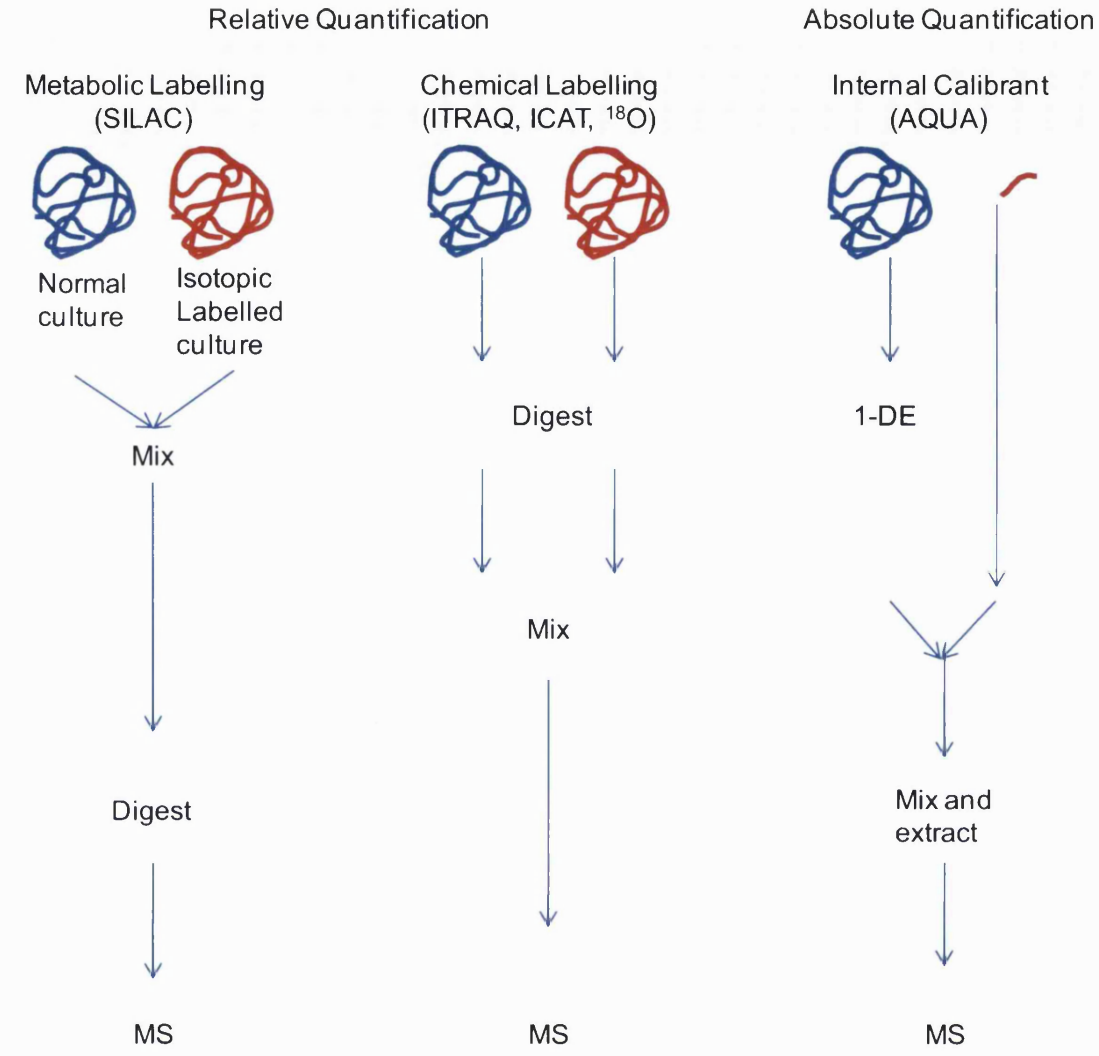


Figure 1.10: Quantification techniques in mass spectrometry: SILAC is added during cell growth while ITRAQ, ICAT, ^{18}O and AQUA are added to the sample and are not incorporated through metabolic labelling.

SILAC (stable isotope labelling by amino acid in cell culture) uses ^{13}C labelled culture media which can then be used to compare to normal ^{12}C culture media. The labelling allows for mixing of the protein samples and identifying directly by LC-MS/MS as heavy labelled peptides elute with the same retention time as the light version of the peptides.[99] The ICAT tag (isotope coded affinity tag) binds to cysteine residues, followed by a linker which is labelled heavy (D) or light (H) and a biotin molecule. The mass difference between the labels is 8 but their mass does not alter the retention times, allowing for both to elute at the same time. Both MS and MS/MS can be used for relative quantification.[99,101] ITRAQ (isotope tags for relative and absolute quantification) is another non-gel based technique for analysis. Using a tag with a reporter moiety of increasing mass of 1 Dalton, attached to a balance moiety of less one Dalton to give a tag of 145 Da, allows for multiple samples to be mixed together. ITRAQ binds to the N-terminal of peptides and upon fragmentation the reporter moiety

of the mixtures can be used to determine relative quantification.[99,101] With respect to ^{18}O labelling, samples are digested with tryptic beads in the presence of H_2O or $H_2{}^{18}O$ giving a mass difference in the resulting sample of 4 Da which can then be used for quantification.[102] AQUA (absolute quantification) also uses stable isotope labelled amino acids. A specific peptide sequence, which is unique to the protein of interest is usually labelled with ^{15}N at lysine or arginine residues. The peptide is added as an internal calibrant to the extraction from gel or in-solution, allowing for absolute quantification of one protein at a time.[99,103,104]

Label free quantification is also starting to emerge more and more which involve more statistical analysis. Looking at peptide intensities over many samples has proven to be one way to carry out label free quantification, while a second method uses the average total intensity count (TIC) for all the MS/MS spectra to identify a protein as a quantitative measure.[99]

1.3.4 Protein Interactions and imaging

Protein-protein, protein-small molecule interactions,[105] intact organism and tissue imaging are other fields where mass spectrometry has been surging. Imaging predominantly involves SELDI or MALDI techniques with novel software for imaging the tissue.[40,42-44,63,76,106] Coupling these two to TOF/TOF gives an opportunity to image and to partially identify the analytes on the surface – whether it is protein or lipid.

Understanding the interactions or identifying proteins that interact with each other; identifying inhibitors of these interactions and quantifying those interactions are also key for biological sciences, in particular medicinal sciences.[33] Normal biochemistry approaches would require a UV absorbing/colorimetric molecule to be used which, would change colour upon reaction or inhibition.[32,37,107,108] If the protein can be bought or its antibody generated, affinity chromatography can be used to purify the complexes out of a mixture. Biotinylation of proteins or creating GST-fusion proteins allows for larger production of proteins which can be used for immunoprecipitation-like purifications. Whole cell or tissue lysate, can be eluted through and complexes can be pulled down or partially pulled down in the column for identification. Non-covalent interactions although difficult to maintain, can be performed both by electrospray or MALDI.[32] Covalent modifications by small molecules can readily be done on either type of ionisation source.

1.3.5 Protein Modifications

As post-translational modifications (PTM)s are important for signalling, their identification is extremely important to understand the entire signalling and activation

pathways of biological activites such as glycolysis or apoptosis. Separating PTMs from non-PTMs is also very useful to remove the high abundance of non-modified proteins from the low abundance of the modified signalling proteins. PTMs such as phosphorylation, glycosylation[109,110] and acetylation occur in low abundance which makes them difficult to identify. In the case of phosphorylation, CID analysis is difficult as the phosphate group is easily cleaved off the peptide, leaving identification of the location nearly impossible. Digestion using 2 different proteases allows for an approximate, if not definite identification of PTMs by defining amino acids to which it is not bound. New fragmentation techniques such as ECD[111-114] and ETD[115,116] have removed this difficulty, but the low abundance of PTMs still requires some separation prior to analysis either in peptide or protein form. Top down proteomics prevents excess loss of PTMs and allows for identification without having to digest. In some cases, specific separation techniques such as IMAC (immobilised metal affinity columns) have improved the identification of phosphorylated peptides.[117,118] Other methods such as immuno-affinity chromatography have made identifying post-translational modifications more feasible.

1.3.6 Bioinformatics

Databanks of genomic data have been key to enabling the development of proteomics. Search engines such as MASCOT[119], X!Tandem[120], and PLGS[121] have allowed for tandem MS and PMF data to be analysed against the information obtained by sequencing of genomes. The complexity of identifying a protein solely on the digested masses of the protein is far less reliable than using tandem MS. The more proteins in a databank that is being searched, the more likely the identification may be random. Purity of the sample and accurate masses of the peptides are both key for a positive identification using PMF. Mixtures are far more easily analysed using LC-MS/MS tandem instruments. The use of peptide mass, charge and its corresponding product ions allow for more confidence in the determination of the sequence of the peptide and the identification of the protein. The addition of PTMs to the search increases the complexity of the analysis, especially if the modifications are not complete on the protein but rather variable. This can increase the length of time for the search as well as increase the likelihood of false positives, thereby increasing the amount of manual intervention in order to confirm the presence of the PTM. As data has become more complex with new fragmentation techniques, as well as PTM identification[122] and tags for quantification, the search engines have had to develop advanced searching algorithms, adding new tools and creating new statistics to increase confidence in the results.

1.3.7 Future Challenges and expectations

Technologies have been expanding quickly, and data size and the ability to analyse the data is becoming a bottle neck as the speed of searches is limited to the speed of the servers. Vast amounts of data require longer times to search and while the automated analysis is getting better, is has not been perfected, requiring some manual work to screen the results.

Better and more open communication across groups such as the Human Proteome Organisation (HUPO) would allow for a means of creating standardised ways to test methods and demonstrate the functionality of developed methods. HUPO has started with a major project to sequence the genome which resulted in multiple labs participating and has from this point, diverged into other groups for tissue specific analyses, all of which can be compared from their websites: http://www.hupo.org/research/ links to all the various initiatives that HUPO has been working on.

New technologies such as FT-ICR, Orbitrap, ETC, ECD, UPLC[114] and techniques such as imaging, have given mass spectrometers better sensitivity, faster analysis time and have created a new technique for analysis which has benefitted users. The increase in data has increased the data analysis bottleneck as data file sizes and the time required to search and confirm results can take longer than the actual running of the sample. More confidence in the algorithms would allow for less manual scrutiny and faster computers would certainly allow for faster searching. However it must be said that the ability of database search engines to develop novel technologies and methods of identification has been vital in increasing and expanding the knowledge of new techniques and has aided in their implementation.

1.4 Metabolomics

Metabolomics or metabonomics, while relatively new terms, have in fact been around and studied for a long time. Metabolomics merely considers the study of the metabolome, or the metabolites within a system[123] while metabonomics is described as 'the quantitative measurement over time of the metabolic responses of an individual or population to drug treatment or other intervention.[124] When considering the number of proteins within a proteome, the number of metabolites is significantly larger and the types and diversity of the metabolites is such that the whole metabolome cannot be studied at once, but rather, split into areas or groups of metabolites of interest. These can be lipids,[125] groups of lipids such as phospholipids, glycolipids, or sterol based or even metabolic-process based.[8,123,126] Qualitative and quantitative data are both of

interest first in identifying the metabolite and secondly quantifying it for a perspective on its activity whether it be in a normal state, a disease state or an induced state.[3,127]

There are two types of approaches to metabolomics: targeted and untargeted. Untargeted[108] metabolomics maximises the coverage of all metabolites in a system, but in doing so, makes a compromise with respect to sensitivity and specificity to maximise the diversity. Less method development is required; however more data analysis is needed for the massive amounts of data that is created. Data analysis can be a challenge as experimental artefacts occur as well as being dependent on extensive computational tools to interpret the complex data.[9,128]

Targeted metabolomic approaches detect and quantify specific metabolites of interest which allows for maximum specificity and sensitivity.[129,7,130] It also allows for the use of standards to create suitable methods of extraction, separation and analysis as well as creating calibration curves and defining fragmentation. Absolute quantification is possible using dilute isotopes or calibration curves.[7,9] In clinical studies, there are many examples of metabolic identification and quantification from amino acids, and drug metabolism to enzymatic activity and insulin studies.[9,108,123,131,132]

1.4.1 Identification and Separation

Separation and purification are required to an extent with untargeted metabolomics and are far more important for targeted metabolomics. Untargeted metabolomics may require removal of proteins and salts from mixtures such as urine to allow for identification of small molecules present – as observed by the work of Griffiths *et al,*[7, 130] who use a quick purification such as a solid phase C18 extraction followed by LC-MS of human bile. In comparison, targeted analysis of oxysterols in brain and plasma not only require a far more complex separation, but their neutrality requires further derivatisation and purification in order to be able to identify them on the mass spectrometer.[4,133]

Lipid based molecules in comparison to molecules found in urine or other aqueous fluids are somewhat more difficult for analysis. Metabolites found in urine can be resolved on HPLC using water and acetonitrile. Lipids however, are not as water soluble, and cannot use such a method. Lipids are easily separated from other molecules using a method such as an ethanol precipitation, to precipitate out most of the hydrophilic molecules and proteins and leaving those that have hydrophobic characteristics.[7] These can then be separated based on their lipophilicity using a C_{18} SPE cartridge or HPLC. More hydrophilic lipids will elute first such as oxysterols, bile acids and steroids, while more hydrophobic such as cholesterol, triglycerides and

phospholipids are retained which can then be eluted using ethanol or a mixture (1:1 v/v) of chloroform and methanol.[7,130]

There is no one single separation technique that can be applied to all metabolites, which is far more complex for analysis than protein identification. The use of other separation types such as lipidex-DEAP can separate compounds based on acidity or a second solid-phase C_{18} can be employed to further separate out groups of interest.[7,67,134] HPLC can also be employed to separate out unwanted groups, which can be of interest when the compounds are visible using UV / fluorescence/ IR / visible light.[9]

1.4.2 Analysis and Quantification

Metabolites and other small molecules have traditionally been analysed using EI and CI[135,136] and GC mass spectrometry[8,9,66,137] which is limited to molecules that are volatile or that can be derivatised to be volatile, targeting key functional groups for derivatisation which does not allow for a global metabolomic analysis. Thermally labile molecules are also unsuitable for GC-MS as they would degrade and so would be better suited to HPLC-MS or MALDI analysis.[138] However a GC column offers better resolution and reproducibility than an HPLC column and analysis can easily be obtained by EI which would allow for identification based on libraries of molecules as well as from the structural information obtained by the fragmentation of the molecule. One drawback is that with EI, the molecular ion is not always present, although this can be circumvented by using CI which would allow for $[M+H]^+$ or $[M+NH_4]^+$ molecular ions to be present.

Quantification can be performed using standards when available or using techniques such as isotope dilution mass spectrometry.[7,138,139]

1.4.3 Bioinformatics

Data analysis of metabolites can be defined by various factors. Exact mass and reproducible retention times tend to be initial identifiers that can be quantitatively compared while MS/MS can aid in identification as well as quantification. Results are analysed using a variety of different statistical techniques such as univariate and multivariate analysis. Databases such as KEGG (http://www.genome.jp/kegg/), Metlin (http://metlin.scripps.edu/) and HMDB (http://www.hmdb.ca/) allow for same analysis and identification as protein databases, however they may cause multiple identifications and have insufficient mass accuracy for proper identification. Modification for neutral molecules or molecules that give low signal are required for proper analysis and standardisation of these modifiers has not yet been accomplished.

The addition of the modifier changes the molecule enough to also change the way it fragments making searching difficult unless someone has used that specific modifier previously and has input it into the database.

1.4.4 Future Challenges and Expectations

Like proteomics, it is hoped that search engines and databases will become better and more adept at identifying metabolites from samples, although in this case, it's a far more complex problem as different groups of metabolites behave differently. Neutral metabolites require a chemical modification to increase sensitivity for mass spectrometry, however these methods are not standardised, nor are the use of the chemical modifiers. This complexity of samples, modification methods and fragmentation make it very difficult to share and compare data to other groups as well as making it difficult to compare data to database resources unless the exact modification and instrument type were used.

The sharing of both data and experimental processes as well as incorporating them into databases for searching is key to identifying and quantifying metabolites. The bottleneck of data, is searching which may or may not be able to identify metabolites from untargeted metabolomics data – requiring manual assistance for proper identification. Targeted metabolomics allows for standards and calibration curves to be used, making identification, even if done manually, a somewhat less complicated issue. More groups such as the human metabolome project (http://www.hmdb.ca/) are required to help with the standardisation, identification and data-sharing.

1.5 Aims of this Project

This thesis demonstrates the use different mass spectrometry techniques in projects carried out in the School Pharmacy over a period of six years. Each of these projects has been an integral part of the research aims of different groups, and dealing with different aspects of metabolomics and proteomics problems. As the experience of the researcher grew, the complexity of the projects and biological systems under study have been increasing. The projects were aiming to:

1. Identify and quantify oxysterols in plasma as a means of monitoring 24S-hydroxycholesterol produced from the brain.
2. Identify proteins involved in the synthesis and metabolism of cholesterol in rat brain.
3. Identify a UV absorbing molecule shinorine and identify the proteins involved in its synthesis using the bacteria *Anabaena variabilis.*
4. Identify the nature of intermolecular interactions between STAT3β, an oncogenic protein, with a known inhibitor STATTIC.

2 Oxysterol Identification in Biood

2.1 Introduction Blood Metabolomics

Blood is composed of a mixture of white and red blood cells and platelets suspended in plasma which contains, water, amino acids, proteins, carbohydrates, lipids, hormones, vitamins, electrolytes, and dissolved gases.

Although there are a lot of components to blood, there are certain ones which are in far higher concentration than others, making it difficult to identify those in lower abundance. Of the plasma proteins found at high concentrations, the majority are albumins, globulins and fibrinogens. With respect to metabolites and small molecules, the ones found at high concentrations are predominantly cholesterol which is found at 5 mM,[123] glucose 8.77 mM, and lactate 2.4 mM.[140]

The use of blood for identification of biomarkers of disease or for metabolite analysis of drugs is becoming increasingly popular, as is the analysis of urine. Both of these fluids are readily available. Although blood sampling is somewhat more invasive, it is still not as invasive as the need to biopsy tissue or organs to look for disease. In the case of the brain, it's impossible to biopsy prior to death. In most cases, biological fluids still require pre-treatment which is necessary to remove matrix and other interfering molecules prior to analysis by e.g. LC-MS. The analysis by LC-MS must be robust to identify the molecules of interest and deal with the complexity of the mixture.

One particular molecule, 24S-hydroxycholesterol is of great interest. 24S-hydroxycholesterol is a cholesterol metabolite which is formed from cholesterol by cytochrome P450 46A1, found exclusively in the brain. Its formation is considered to be the main method for removal of cholesterol from the brain. This oxysterol could potentially be used as a biomarker if it was possible to extract it out of blood and quantify it. The levels of 24S-hydroxycholesterol in plasma have been determined, along with other oxysterols (Table 2.1). With particular reference to neurodegenerative diseases such as Alzheimers, there is a demyelination process which occurs in the brain, which could potentially be monitored in blood as the level of 24S-hydroxycholesterol should increase with increasing cholesterol levels in the brain.[141,142]

Table 2.1: Oxysterols and their concentrations found in plasma.

Name	Chemical Formula	Mass	Concentration[1]	GP-hydrazone *m/z*
Cholesterol	$C_{27}H_{46}O$	386.35	2 mg/mL[123]	518
Cholesterol sulphate	$C_{27}H_{46}O_4S$	466.31	50-300 ng/mL[123]	598
7α-hydroxycholesterol	$C_{27}H_{46}O_2$	402.35	40 ng/mL[123]	534
7β-hydroxycholesterol	$C_{27}H_{46}O_2$	402.35	5 ng/mL[123]	534
24S-hydroxycholesterol	$C_{27}H_{46}O_2$	402.35	80 ng/mL[143]	534
			70 ng/mL[144]	
			60 ng/mL[129,139]	
25-hydroxycholesterol	$C_{27}H_{46}O_2$	402.35	3 ng/mL[123]	534
27-hydroxycholesterol	$C_{27}H_{46}O_2$	402.35	120 ng/mL[143]	534
			150 ng/mL[129,139]	
3β-hydroxycholest-5-en-27-oic acid	$C_{27}H_{44}O_3$	416.33	100 ng/mL	548

[1]Levels are for "total" oxysterols and sterols corresponding to the sum of free molecules and fatty acylesters (measured following hydrolysis).

In plasma, there are at least 6 hydroxycholesterols that can potentially be identified and quantified with sensitive MS methods. Some, such as 25-hydroxycholesterol and 7β-hydroxycholesterol are present in such low quantities that a very sensitive method is required for these (Table 2.1). The others listed are at much higher concentrations, which makes quantification far simpler. Separation of oxysterols from cholesterol is vital to avoid autoxidation of cholesterol into hydroxycholesterols, which could then create artefacts in the analysis.[68]

2.2 Aims

Using blood, a separation method can be developed to allow for the identification of oxysterols; in particular 24S-hydroxycholesterol, 27-hydroxycholesterol as well as 3β-hydroxycholest-5-en-27-oic acid and 3β,7α-dihydroxycholest-5-en-27-oic acid, which can then be compared to each other to create a method of relative quantification using a derivatisation method and LC-MS/MS.

2.3 Materials and Methods

2.3.1 Reagents

Cholesterol, 19-hydroxycholesterol, 24S-hydroxycholesterol, 25-hydroxycholesterol, 27-hydroxycholesterol were purchased from Steraloids Inc. and Sigma-Aldrich Ltd, Cholesterol oxidase (2 mg/ml in H_2O, 44 units/mg protein) and triethylamine-sulfuric acid 2M:2M solution (TEAS) were purchased from Sigma Aldrich Ltd. Girard P Reagent was purchased from TCI Europe (25 g, Tokyo, Japan). Unisil (activated silicic acid), 200-300 mesh, acid washed was purchased from Clarkson Chemicals (Williamsport, PA, USA) and SepPak C_{18} and tC_{18} (1 g, 6 cc) were purchased from Waters (Milford, MA, USA). Acetic acid (glacial), hexane, ethyl acetate, isopropanol (IPA), chloroform ($CHCl_3$), dichloromethane (CH_2Cl_2), ethanol (EtOH), and potassium dihydrogen phosphate were purchased from Fisher (Waltham MA, USA). A dry-ice roto-evaporator was purchased from Büchi (Rungis, France) and the vacuum concentrator was purchased from Jouan/Haerus (Thermo Fischer Scientific, Waltham MA, USA).

2.3.2 Plasma Sample Preparation

2.3.2.1 Sample Preparation

Plasma samples were provided by GSK with institutional review board ethical approval. Four plasma samples were provided for the initial study.

Plasma, 0.8 mL was added dropwise to 10.5 mL of 99.9% EtOH in an ultrasonic bath over 10 minutes (in duplicate runs, 600 ng of 19-hydroxycholesterol (60 µL, 10 ng/µL) was added to the sample). The mixture was diluted to 70% EtOH by adding 3.7 mL water and ultrasonicated for another 2 minutes before centrifuging at 50,000 x *g* for 10 minutes at 4°C. The supernatant was decanted into a fresh 15 ml centrifuge tube.

2.3.2.2 Fractionation of Plasma Sample

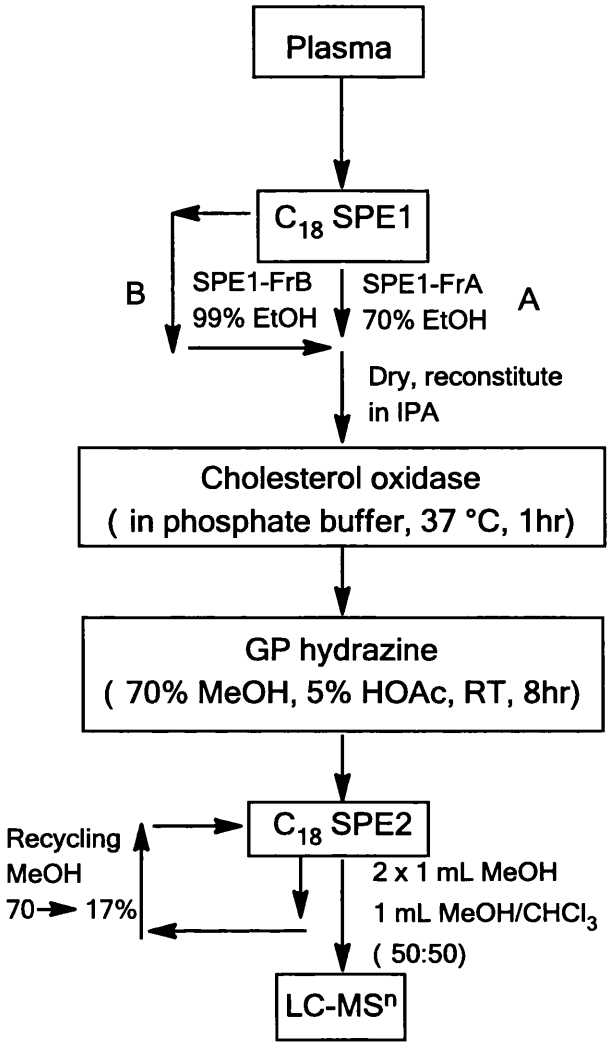


Figure 2.1: Separation of Plasma Samples for hydroxysterol analysis. Samples were separated into two fractions from SPE1 (FrA and FrB) and each was oxidised and derivatised before being run on a second SPE2 column and recycled from 70% methanol to 17% methanol.

A bed of 1 g Sep-Pak tC_{18} was prepared in a gravity flow glass column (0.8 cm x 7 cm) using 70% EtOH. The bed was rinsed with five column volumes of 70% EtOH (18 mL). Plasma mixture in 70% EtOH (15 mL) was added onto the column and allowed to flow through at 0.25 mL/min. The flow-through was collected and combined with four column washes of 5 mL 70% EtOH (35 mL). (FrA)

The column was eluted with 5 mL of 99.9% EtOH and collected. (FrB)

The column was eluted with 20 mL of 99.9% EtOH which was collected. (FrC)

The fractions were dried using a dry ice roto-evaporator.

2.3.2.3 Oxidation of Oxysterols using Cholesterol oxidase from *Streptomyces sp.*

The dried samples were reconstituted into isopropanol (100 µL) and 50 mM phosphate buffer, pH 7 (1000 µL) was added. To this mixture, 3.0 µL cholesterol oxidase was added (2 mg/mL in H_2O, 44 units/mg of protein) and the reaction mixture was left at 37°C for 1 hour. The reaction was stopped with 2000 µL MeOH.[145]

2.3.2.4 Derivatisation of Oxysterols using Girard P reagent

To the mixture above, 150 µL glacial acetic acid was added and mixed well. Approximately 150 mg of Girard P reagent was added before mixing and leaving the reaction overnight in the dark.[145]

2.3.2.5 Purification

A new column bed was prepared using Waters Sep-Pak C_{18} (1 x 0.8 cm). The bed was washed as follows:

10 mL 50:50 $CHCl_3$/MeOH
10 mL 100% MeOH
10 mL 10% MeOH
5 mL 70% MeOH

The derivatised sample was applied onto the column and the effluent was collected. The sample flask was washed with 4 mL 70% MeOH which was then applied to the column and collected with the sample for a total volume of 7 mL. An equal amount of H_2O (7 mL) was added to the effluent from the column to give 14 mL of 35% MeOH.

The column was washed with 1 mL 35% MeOH and was collect with the sample effluent for a total volume of 15 mL of 35% MeOH. The sample (15 mL of 35% MeOH) was recycled onto the column and the effluent was collected. An equal amount of H_2O (15 mL) was added to the effluent to give 30 mL 17.5% MeOH. The column was washed with 1 mL 17% MeOH which was collected with the effluent to give 31 mL of 17.5% MeOH. The sample was recycled onto the column (31 mL of 17.5% MeOH) and the effluent was collected into a single container. The column was washed with 10 mL 10% MeOH and the effluent was collected.

The oxysterols were finally eluted from the column as follows:

1 mL MeOH (FR1)

1 mL MeOH (FR2)

1 mL 50:50 $CHCl_3$/MeOH (FR3)

Fraction 3 was dried down using no heat on the vacuum concentrator.[145]

2.3.2.6 Nano-ESI of standards

Oxysterol analysis was performed on an LCQ^{duo} ion trap mass spectrometer (Thermo Fisher, Hemel Hempstead) using a nano-ESI source. Initial screening of eluates was performed on all fractions from the second C_{18} SPE extraction to identify oxysterols using metal-coated glass capillary tips which have had the tip cracked and were cut down to 5 cm in length. Data was acquired using a tune method developed for oxysterols with the following settings: spray voltage 1.2 kV, capillary temperature 200 °C. No sheath gas or auxiliary gas was used. A full scan of *m/z* 125-600, MS^2 and MS^3 for cholesterol and for selected oxysterols 24- , 25- and 27-hydroxycholesterol was performed using $[M]^+$ 518/534, MS^2 $[M]^+ \rightarrow$, MS^3 $[M]^+ \rightarrow [M-79]^+ \rightarrow$.

2.3.2.7 Nano-ESI-MS^n of samples

Nano-ESI-MS^n was performed using the LCQ^{duo} ion-trap mass spectrometer (Thermo Fisher, Hemel Hempstead) coupled to an Ultimate 3000 Capillary HPLC system (Dionex, Camberley, UK). The system is composed of a SRD-3600 on-line solvent degasser, LPG-3600 low pressure dual micro-gradient pump, and WPS-3000 well plate autosampler. For separation, a Hypersil Gold C_{18} column (180 µm x 100 mm, 3 µm particle size, Thermo Fisher) was used. Four µl of the MeOH eluates was diluted with 16 µl of MeOH and 13 µl of 0.1% formic acid (FA) was added to give a solution of 60% MeOH, 0.04% FA. From this, 2 µL was injected onto the column. The MeOH/$CHCl_3$ (50:50 v/v) fraction was dried and reconstituted into 100% MeOH before being diluted in the same fashion and injected onto the capillary column.

Mobile phases used for the elution of the hydroxycholesterols from the column were A: 0.1 % FA in 65% MeOH, B: 0.1 % FA in 95% MeOH, and transport solvent 0.1 % FA in 65% MeOH. The gradient and flow rates were as follows:

Table 2.2: LC gradient used for separation of oyxsterols on LC-MS.

Time (min)	% A	% B	Flow rate µl/min
0	80	20	0.8
10	30	70	0.8
15	20	80	0.8
30	20	80	0.8
30.5	80	20	0.8
50	80	20	0.8

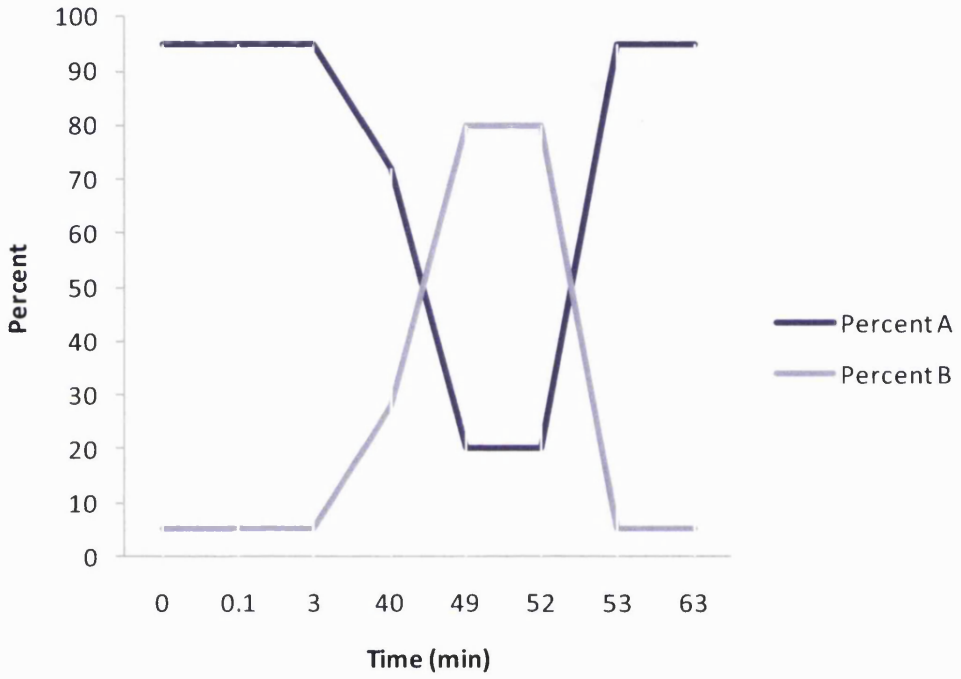


Figure 2.2: Schematic representation of the nano-LC Gradient used for LC-MS separation.

A linear gradient was run for a total of 50 minutes (Figure 2.2, Table 2.2) The sample tray was held at 4 °C and the sample was injected using a 5 µL sample loop. Once injected, the injector and needle were flushed and washed with IPA: MeOH (1:1 v/v). For analysis, the LCQ ion-trap was setup with an ESI Pico-Tip emitter (360 µm O.D, 75 µm I.D, 15 µm I.D tip, New Objective). Capillary potential was set at 2 kV while the heated capillary temperature was set to 200 °C. No sheath gas, or auxiliary gas was used. Data was acquired for MS, MS^2 and MS^3 and recorded in a data dependent mode using both exclude and include lists with an acquisition range of *m/z* 50 – 600. Exclude list included the following *m/z*: 416, 430, 446, 504, 520. For MS and MS^n, 3 microscans were averaged with a maximum injection time of 200 ms and an isolation width for the parent ion was set to 2.0. A normalised collision energy of 45% was used for all tandem MS data. An include list for MS consisted of three *m/z*: 534, 548 and

564, which were then subjected to MS^2. The *m/z* product corresponding to $[M-79]^+$ was then selected for MS^3 (Table 2.3).

Table 2.3: Include list of masses for MS, MS^2, MS^3 for selection and quantification

MS (ion selection for MS^2)	MS^2 (ion selection for MS^3)	MS^3 (for quantification)
534	455	437
548	469	451
564	485	467

A minimum of 4 blank injections were run between samples to ensure no carry over.

2.3.2.8 Relative Quantification of Ions

Data was processed using Xcalibur Qual Browser version 1.2© (Thermo Fisher, Milford MA, USA). Individual ions (nine in total) from Table 2.3 were extracted from the BPI chromatograms to be integrated for peak area and peak height. The corresponding values were entered into a table and subsequent ratios were calculated.

2.4 Results

2.4.1 Optimisation of nano-HPLC method using standards

Optimisation of nano-HPLC-MS was based on previously published methods[68,146] using derivatised 25-hydroxycholesterol. A standard concentration of 15 ng/µL was used with varying injection volumes to determine approximate retention times of the hydroxycholesterols (Table 2.4) corresponding to a retention time of 19 minutes.

Table 2.4: Peak area and height of 25-hydroxycholesterol 15 (ng/µL) from nano-ESI capillary LC-MS^3.

Conc	RT	*m/z* 534		*m/z* 455		*m/z* 437	
	(min)	Area	Height	Area	Height	Area	Height
5 ul	19.91	3.84E+07	4.42E+05	1.46E+07	1.38E+05	7.87E+06	7.14E+04
5 ul	18.89	1.33E+09	2.63E+07	6.45E+08	1.22E+07	3.13E+08	6.04E+06
2 ul	19.12	7.26E+08	1.06E+07	3.38E+08	4.98E+06	1.65E+08	2.29E+06
2 ul	19.01	6.26E+08	9.16E+06	3.00E+08	4.13E+06	1.46E+08	1.91E+06
2 ul	19.30	3.52E+08	4.58E+06	1.76E+08	2.26E+06	8.77E+07	1.10E+06

As can be seen from Table 2.4, the system requires a "conditioning" injection to cover all the column active sites before reproducible data is forthcoming. A nano-ESI-LC-MS^3 method was used to separate and identify the three different *m/z* ions. Reconstructed ion chromatograms were created for derivatised 25-hydroxycholesterol and their corresponding spectra were compared to a previously created library for confirmation and identified (Figure 2.3).[147]

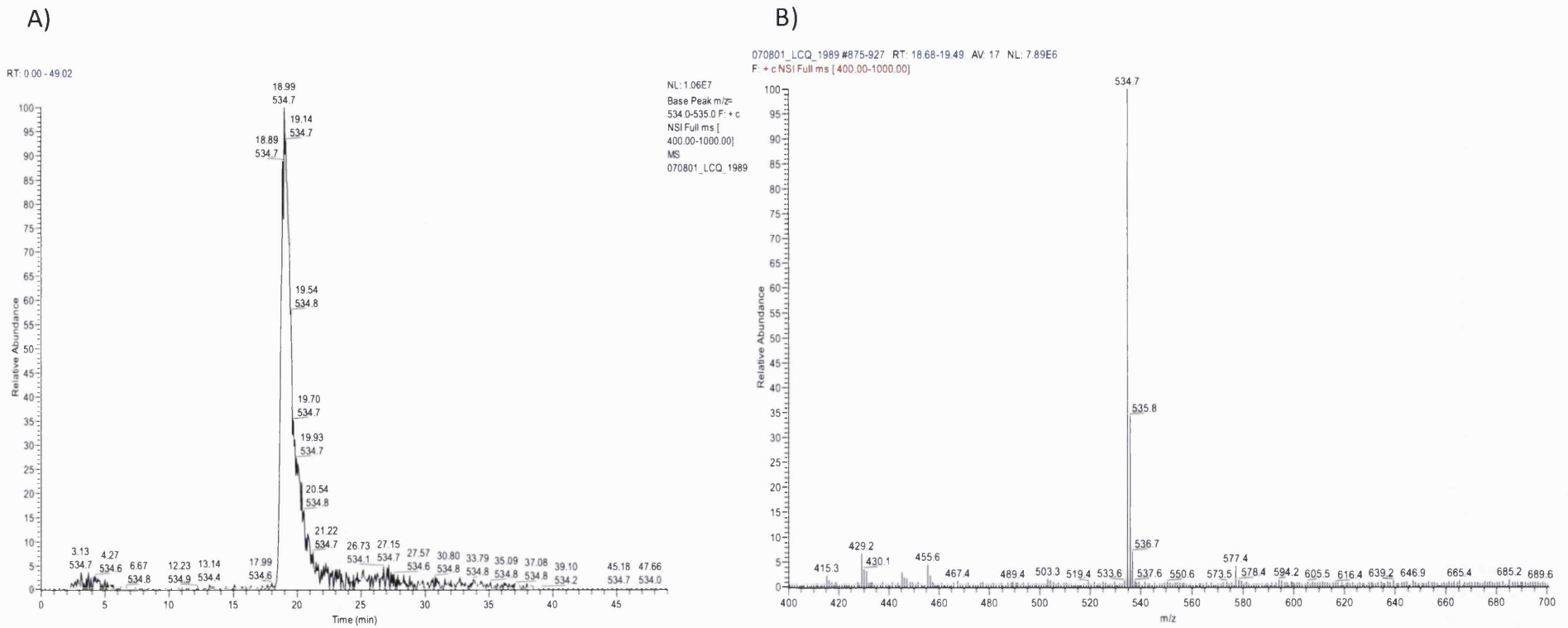


Figure 2.3: Chromatograms (left) and spectra (right) of standard 25-hydroxycholesterol. (A) RIC of *m/z* 534 and (B) its corresponding MS spectrum.
(Continued on the next page)

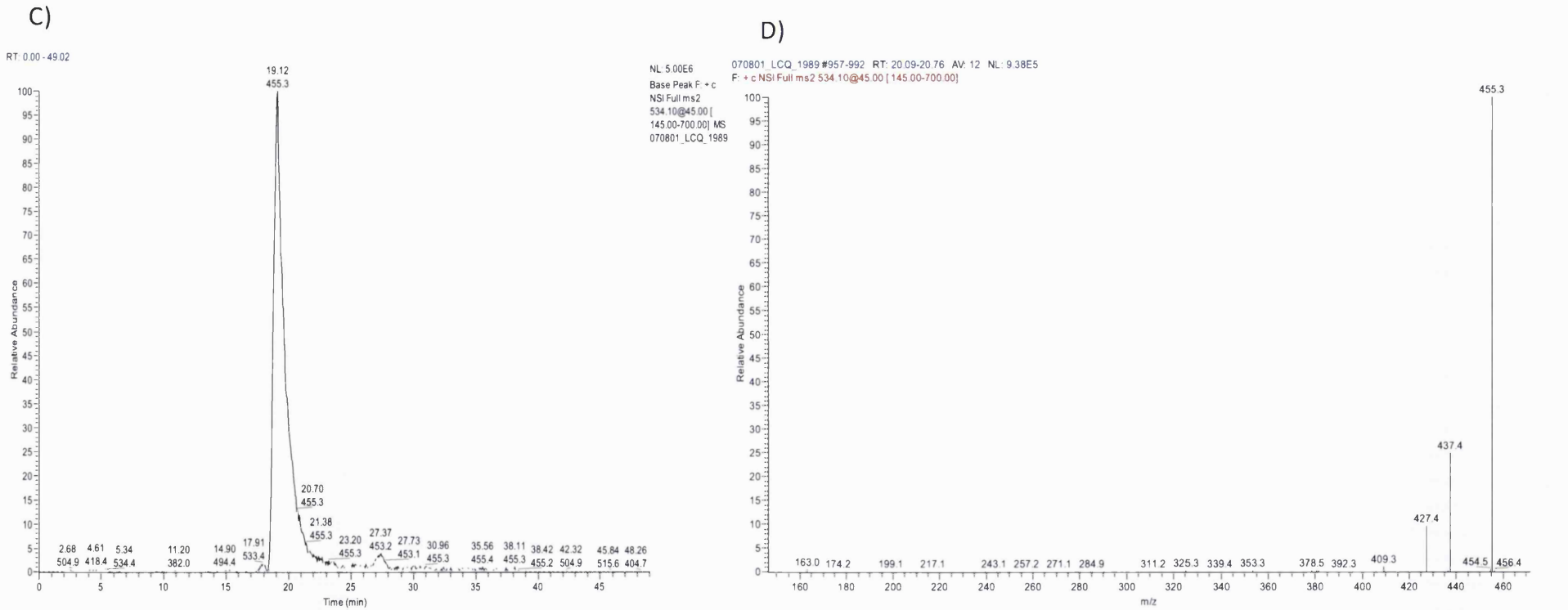


Figure 2.3: Chromatograms (left) and spectra (right) of standard 25-hydroxycholesterol. (C) RIC of MS^2 of 534→, and (D) its corresponding spectrum.
(Continued on the next page)

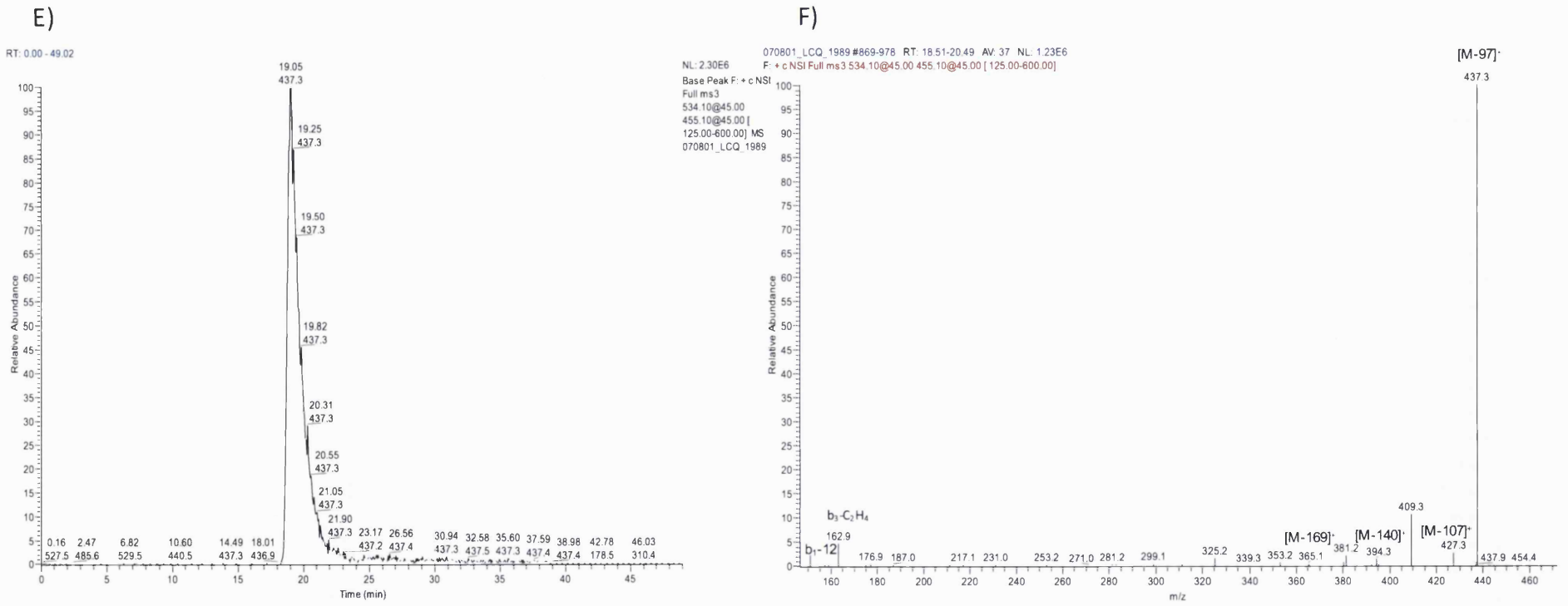


Figure 2.3: Chromatograms (left) and spectra (right) of standard 25-hydroxycholesterol. (E) RIC of MS^3 534->455-> and (F) the corresponding spectrum.

Initially the two fractions, FrAFr1 and FrA1Fr2 from the plasma sample were analysed using MRM in nano-ESI-LC-MS.

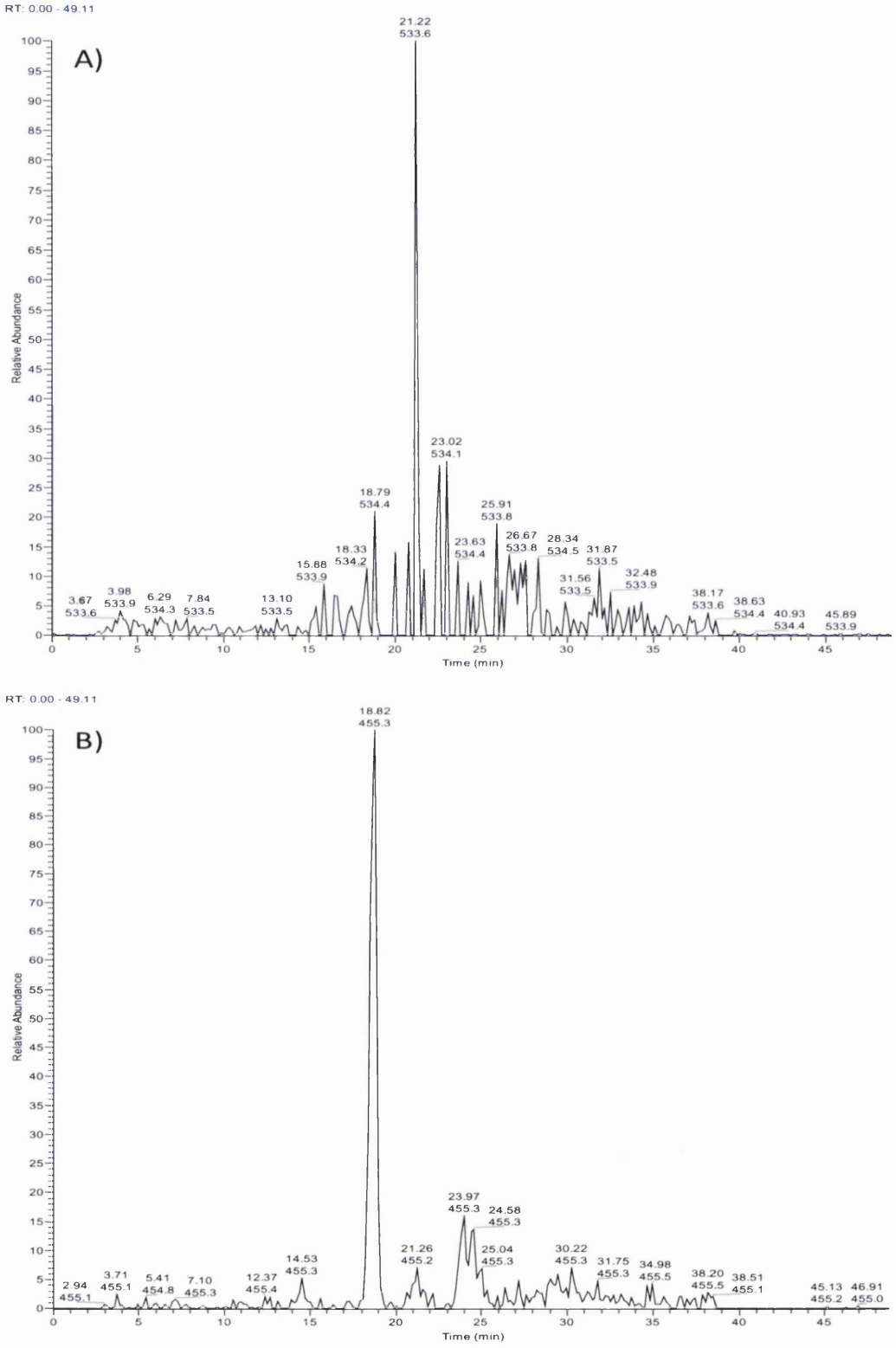


Figure 2.4: Chromatograms of plasma sample 1 FrAFR2. (A) RIC 534, (B) RIC MS^2 534 ->
(Continued on the next page)

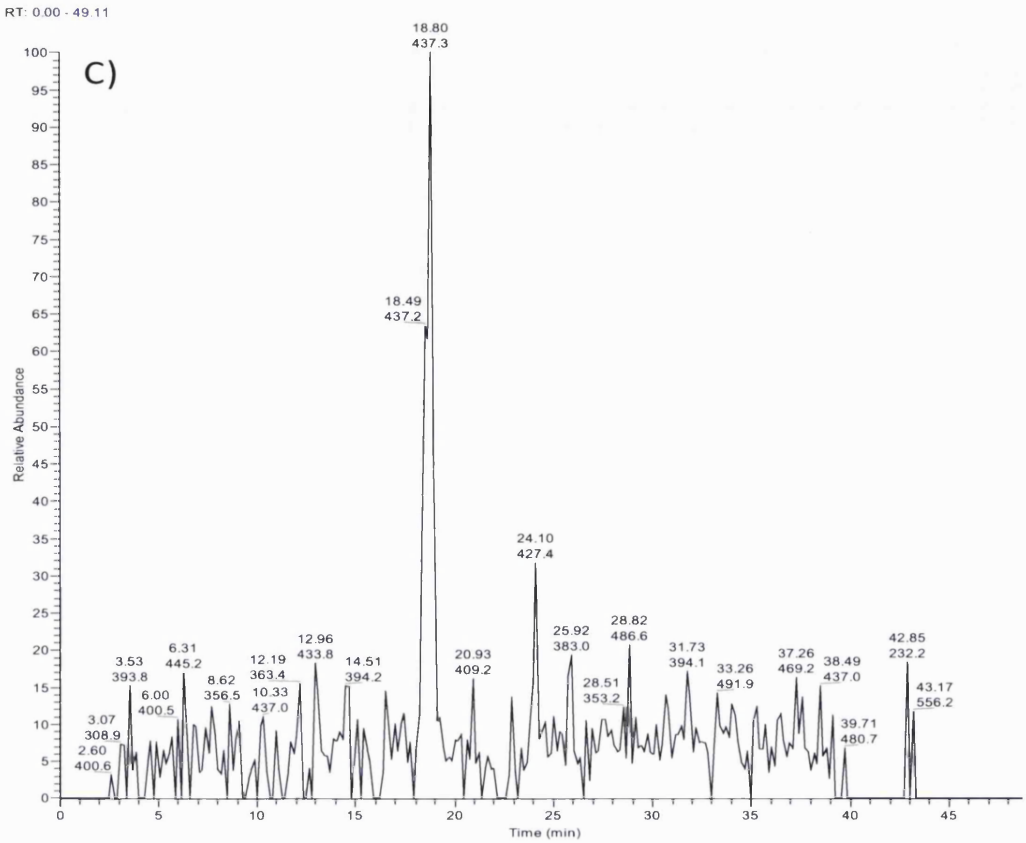


Figure 2.4: Chromatograms of plasma sample 1 FrAFR2. (C) RIC MS^3 534-> 455 ->

Three different components – monohydroxycholesterols *m/z* 534, monohydroxycholest-5-en-27-oic acid *m/z* 548 and dihyrocholest-5-en-27-oic acid *m/z* 564 were screened (Table 2.5).

Table 2.5: Initial run of sample 1 FrA1Fr1 and FrAFr2; 2-step separation prior to nano-ESI-LC-MS^3.

Fraction	*m/z*	RT/min	MS^2,MS^3/ RIC	Area	Height
FrAFr1	534	18.63	455	3.42E+05	2.69E+04
			437	3.90E+05	5.73E+03
	564	16.69	485	1.41E+07	3.48E+05
			467	1.22E+06	2.53E+04
	548	18.58	469	1.25E+06	3.39E+04
			423	2.38E+05	5.67E+03
FrAFr2	534	18.82	455	1.06E+07	3.55E+05
			437	1.40E+06	4.23E+04
		23.97	455	3.43E+06	5.71E+04
			427	4.21E+05	1.34E+04
		30.22	455	1.22E+06	2.54E+04
			437	2.72E+05	5.90E+03
	564	14.27	485	3.24E+06	8.60E+04
			467	7.83E+05	1.29E+04
		16.72	485	3.57E+07	6.98E+05
			467	3.48E+06	6.20E+04
	548	18.92	469	1.72E+07	5.91E+05
			423	1.14E+06	3.68E+04

Sample 1 gave a positive analysis for hydroxycholesterols identified at 19 minutes. 25-hydroxycholesterol was used as a marker for the approximate retention times that were used to identify the derivatised hydroxycholesterols. The peak at 19 minutes was determined to be a combination of 24S-hydroxycholesterol and 27-hydroxycholesterol (Figure 2.5). Two other retentions times for *m/z* of 534 were found at both 24 min and 30 min. From Table 2.5 it is evident that the majority of the monohydroxycholesterol eluted in FrAFr2, while the acids eluted predominantly in FrAFr1.

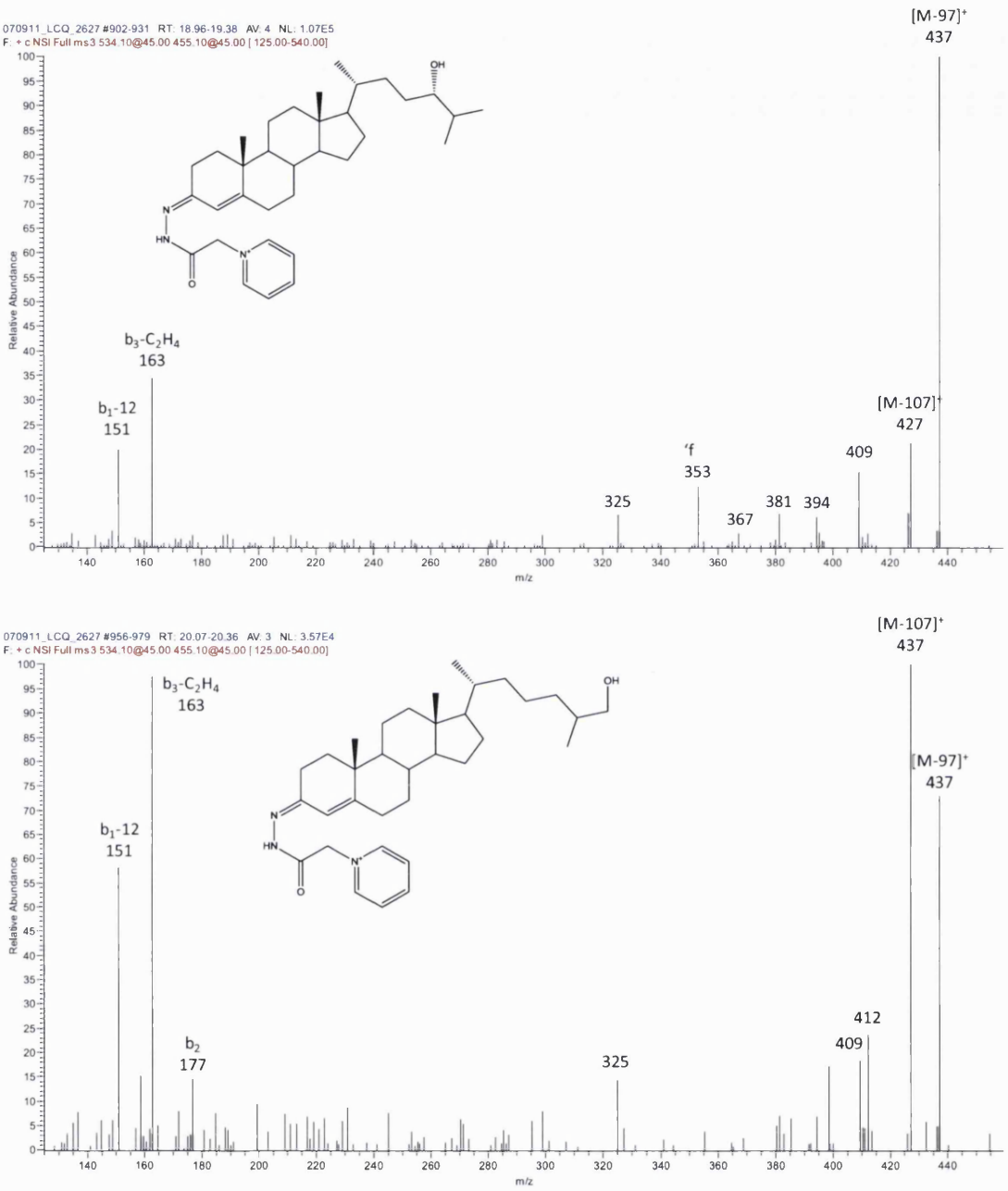


Figure 2.5: Comparison of MS^3 data for 24S-hydroxycholesterol (upper) and 27-hydroxycholesterol (lower) from sample 1 FrAFr2. The upper spectrum is from the leading edge of the peak at 18 min and the lower spectrum is from the tail of this peak.

To confirm the above result a second plasma sample was worked up and analysed using the same nano-LC-MS analysis as seen above (Figure 2.6,Table 2.6).

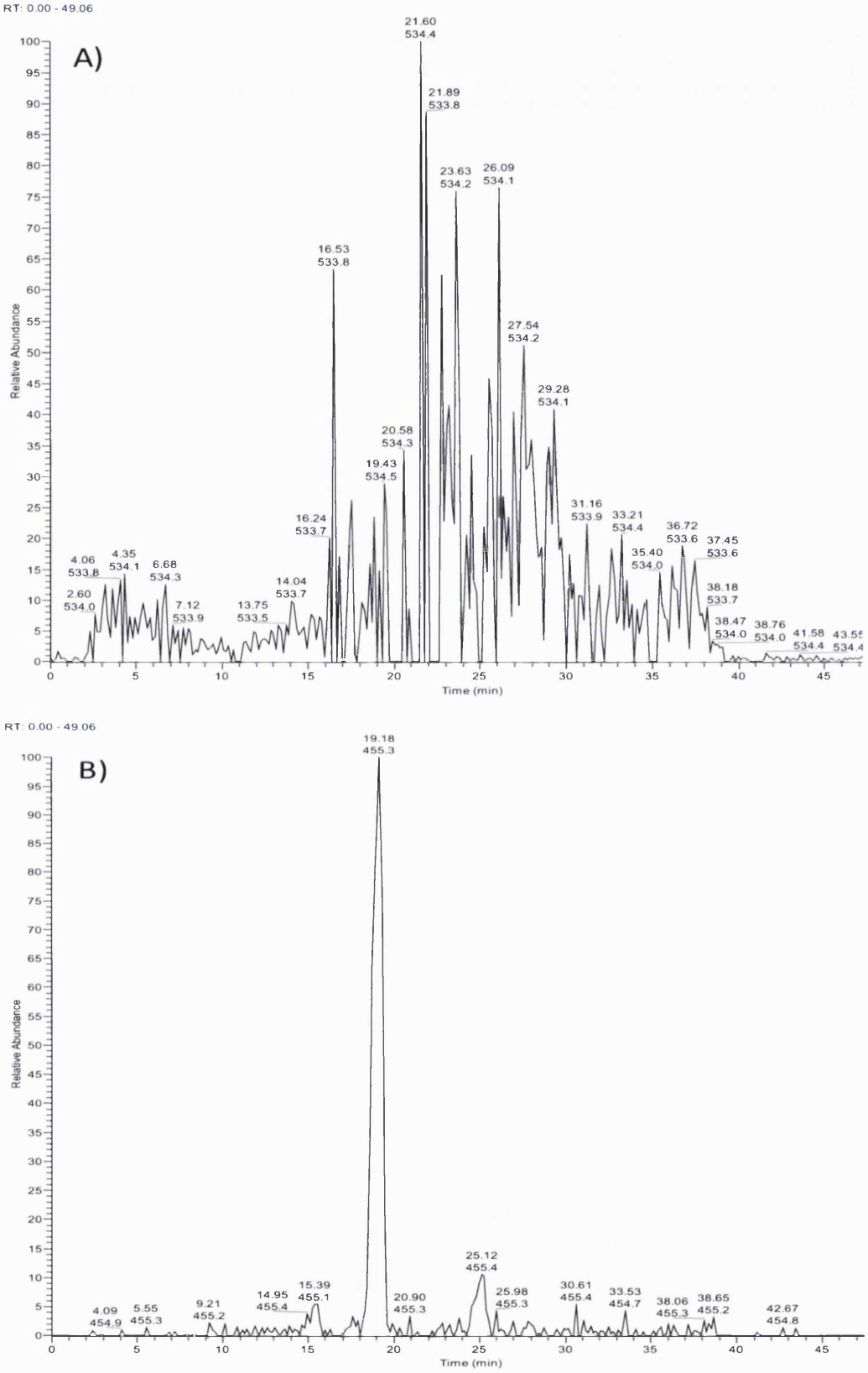


Figure 2.6: Chromatograms of plasma sample 2 FrAFR2. (A) RIC 534, (B) RIC MS^2 534 ->
(Continued on the next page)

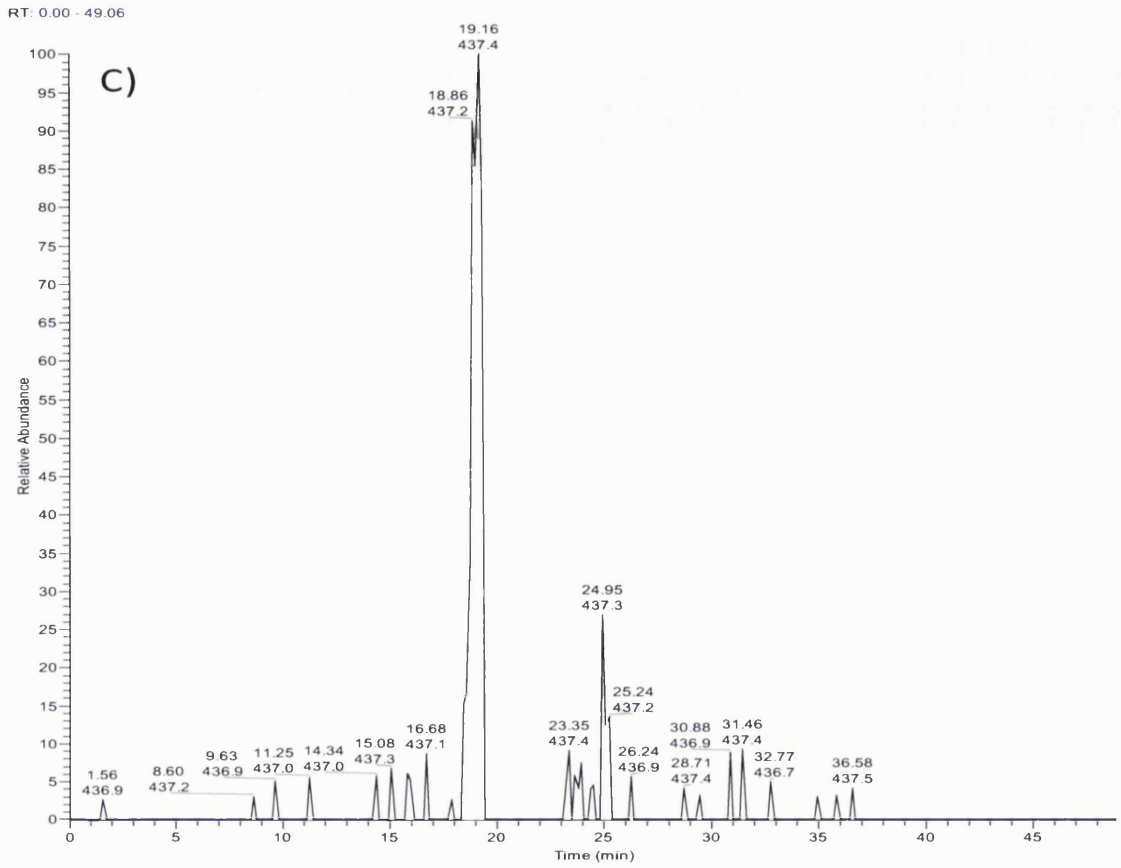


Figure 2.6: Chromatograms of plasma sample 2 FrAFR2. (C) RIC MS^3 534-> 455 ->

Table 2.6: Analysis of sample 2 FrAFr1 and FrAFr2 by nano-ESI-LC-MS.

Fraction	*m/z*	RT/min	MS^2,MS^3/ RIC	Area	Height
FrAFr1	534	18.65	455	1.67E+05	8.64E+04
			437	-	
	564	16.9	485	7.80E+07	2.87E+06
			467	6.10E+06	2.31E+05
		17.91	485	3.16E+07	9.70E+05
			467	2.57E+06	8.27E+04
	548	18.69	469	3.48E+06	1.38E+05
			423	1.97E+05	9.49E+03
FrAFr2	534	19.18	455	1.23E+07	2.93E+05
			437	1.41E+06	3.40E+04
		25.12	455	1.40E+06	3.08E+04
			427	1.87E+05	9.16E+03
	564	16.82	485	1.98E+07	9.20E+05
			467	1.75E+06	6.21E+04
		17.98	467	2.19E+07	6.15E+05
			467	1.75E+06	4.81E+04
	548	19.29	469	4.45E+07	1.88E+06
			423	2.68E+06	1.02E+05

As with sample 1, more than one peak was seen for *m/z* 534, as there is a split in the peak at 19 minutes when the MS^3 data was observed. This again indicates that 24S-hydroxycholesterol and another hydroxycholesterol were present. The peak at 30 minutes, seen in sample 1 is not present in sample 2. As well the peak at 25 minutes seen in both samples did not have enough MS^3 data for identification.

2.4.2 Reproducibility: System Suitability and Method Suitability

A test of method suitability for quantification was performed using a new sample of plasma (sample 3) and aliquoting it into 3 x 0.8 mL. Each aliquot in turn was separated and purified according to section 2.3.2.5. Fractions were each injected onto the system and their retention times, areas and heights were calculated for each ion of interest as well as their selected MS^2 and MS^3 ions. Fractions FrAFr1/Fr2/Fr3 were combined and their averages were taken as were fractions FrBFr1/Fr2/Fr3. As can be seen from Table 2.7 the metabolites corresponding to ions of interest at 534, 548 and 564 eluted predominantly in FrA.

Table 2.7: Integrated Peak Areas and Heights of Replicate Sample Preparation of Sample 3.

		MS			MS^2			MS^3		
		534	564	548	455	485	469	437	467	451
FrA	Area	3.97E+08	6.66E+08	1.43E+08	8.85E+07	1.50E+08	4.72E+07	5.30E+06	1.02E+07	4.52E+06
Fr1/Fr2/Fr3		7.82E+08	1.19E+09	1.19E+09	1.93E+08	3.77E+08	3.14E+08	1.41E+07	2.89E+07	9.54E+06
		3.49E+08	6.80E+08	8.24E+08	1.51E+08	1.70E+08	2.17E+08	1.14E+07	1.21E+07	6.94E+06
	Mean	5.09E+08	8.45E+08	7.20E+08	1.44E+08	2.32E+08	1.93E+08	1.03E+07	1.71E+07	7.00E+06
	CV	0.47	0.35	0.74	0.36	0.54	0.70	0.44	0.60	0.36
	Height	2.84E+06	6.29E+06	8.09E+06	8.72E+05	1.94E+06	2.61E+06	1.12E+05	1.59E+05	1.16E+05
		7.28E+06	1.90E+07	2.11E+07	2.09E+06	6.51E+06	6.24E+06	2.82E+05	6.26E+05	2.46E+05
		4.10E+06	8.01E+06	1.59E+07	1.68E+06	3.07E+06	4.40E+06	1.96E+05	2.26E+05	1.63E+05
	Mean	4.74E+06	1.11E+07	1.50E+07	1.55E+06	3.84E+06	4.42E+06	1.97E+05	3.37E+05	1.75E+05
	CV	0.48	0.62	0.44	0.40	0.62	0.41	0.43	0.75	0.38
		MS			MS^2			MS^3		
		534	564	548	455	485	469	437	467	451
FrB	Area	-	-	-	2.28E+06	2.16E+06	4.68E+05	8.07E+04	-	6.05E+04
Fr1/Fr2/Fr3		-	2.53E+07	2.25E+05	1.33E+06	5.05E+05	3.57E+05	1.14E+05	3.89E+04	1.08E+04
		-	-	3.77E+07	5.92E+06	4.78E+06	3.20E+06	4.24E+05	2.94E+05	-
	Mean	-	8.42E+06	1.26E+07	3.18E+06	2.48E+06	1.34E+06	2.06E+05	1.11E+05	2.38E+04
	CV	-	1.73	1.72	0.76	0.87	1.20	0.92	1.44	1.36
	Height	-	-	-	5.51E+04	2.58E+04	1.66E+04	5.85E+03	-	6.86E+03
		-	5.06E+05	8.16E+03	2.15E+04	1.88E+04	1.96E+04	6.87E+03	2.98E+03	1.22E+03
		-	-	5.62E+05	7.35E+04	1.00E+05	5.52E+04	1.15E+04	1.46E+04	-
	Mean	-	1.69E+05	1.90E+05	5.00E+04	4.83E+04	3.04E+04	8.07E+03	5.86E+03	2.69E+03
	CV	N/A	1.73	1.70	0.53	0.93	0.70	0.37	1.32	1.36

The [M]+ ions 534, 564 and 548 were difficult to integrate from the baseline in the RIC from the MS spectra, however the ions for MS^2 and MS^3 gave better chromatographic peak shape allowing for both peak area and height to be determined. Areas were calculated based on the determined retention times of 19 minutes, 17 minutes and 19 minutes consecutively for the three parent ions and fragment ions. If two chromatographic peaks were seen around the same retention time, both were combined because they were inseparable.

Mean and coefficient of variations were calculated for the replicate sample preparations below for all FrA fractions. FrB analysis was expected to be very variable as most of the analytes were found in the FrA set of fractions.

A comparison was made between the peak areas and heights from FrA Fr1/Fr2/Fr3 and FrAFr1/Fr2 to see how much FrAFr3 contributes to the analysis.

Table 2.8: Comparisons of Peak Area and Peak Height from Combined Fractions FrA Fr1/Fr2/Fr3 and FrA Fr1/Fr2 of Sample 3.

		MS			MS^2			MS^3		
		534	564	548	455	485	469	437	467	451
Fr1Fr2Fr3	Area	3.97E+08	6.66E+08	1.43E+08	8.85E+07	1.50E+08	4.72E+07	5.30E+06	1.02E+07	4.52E+06
		7.82E+08	1.19E+09	1.19E+09	1.93E+08	3.77E+08	3.14E+08	1.41E+07	2.89E+07	9.54E+06
		3.49E+08	6.80E+08	8.24E+08	1.51E+08	1.70E+08	2.17E+08	1.14E+07	1.21E+07	6.94E+06
	Mean	5.09E+08	8.45E+08	7.20E+08	1.44E+08	2.32E+08	1.93E+08	1.03E+07	1.71E+07	7.00E+06
	CV	0.47	0.35	0.74	0.36	0.54	0.70	0.44	0.60	0.36
Fr1/Fr2		3.97E+08	6.66E+08	1.43E+08	7.81E+07	1.49E+08	3.36E+07	4.90E+06	1.00E+07	4.13E+06
		7.82E+08	1.19E+09	1.13E+09	1.87E+08	3.76E+08	3.08E+08	1.37E+07	2.87E+07	9.33E+06
		3.49E+08	6.80E+08	7.58E+08	1.46E+08	1.69E+08	2.10E+08	1.11E+07	1.20E+07	6.68E+06
	Mean	5.09E+08	8.45E+08	6.77E+08	1.37E+08	2.31E+08	1.84E+08	9.90E+06	1.69E+07	6.72E+06
	CV	0.47	0.35	0.74	0.40	0.54	0.76	0.46	0.61	0.39
Combined Mean		5.09E+08	8.45E+08	6.98E+08	1.41E+08	2.32E+08	1.88E+08	1.01E+07	1.70E+07	6.86E+06
Combined CV		0.42	0.32	0.66	0.34	0.49	0.65	0.40	0.54	0.33
		MS			MS^2			MS^3		
		534	564	548	455	485	469	437	467	451
Fr1/Fr2/Fr3	Height	2.84E+06	6.29E+06	8.09E+06	8.72E+05	1.94E+06	2.61E+06	1.12E+05	1.59E+05	1.16E+05
		7.28E+06	1.90E+07	2.11E+07	2.09E+06	6.51E+06	6.24E+06	2.82E+05	6.26E+05	2.46E+05
		4.10E+06	8.01E+06	1.59E+07	1.68E+06	3.07E+06	4.40E+06	1.96E+05	2.26E+05	1.63E+05
	Mean	4.74E+06	1.11E+07	1.50E+07	1.55E+06	3.84E+06	4.42E+06	1.97E+05	3.37E+05	1.75E+05
	CV	0.48	0.62	0.44	0.40	0.62	0.41	0.43	0.75	0.38
Fr1/Fr2		2.84E+06	6.29E+06	8.09E+06	7.71E+05	1.92E+06	2.45E+06	9.28E+04	1.54E+05	1.09E+05
		7.28E+06	1.90E+07	2.05E+07	2.02E+06	6.47E+06	6.15E+06	2.74E+05	6.18E+05	2.39E+05
		4.10E+06	8.01E+06	1.54E+07	1.63E+06	3.03E+06	4.31E+06	1.83E+05	2.18E+05	1.55E+05
	Mean	4.74E+06	1.11E+07	1.47E+07	1.47E+06	3.80E+06	4.30E+06	1.83E+05	3.30E+05	1.68E+05
	CV	0.48	0.62	0.43	0.43	0.62	0.43	0.49	0.76	0.39
Combined Mean		4.74E+06	1.11E+07	1.49E+07	1.51E+06	3.82E+06	4.36E+06	1.90E+05	3.33E+05	1.71E+05
Combined CV		0.43	0.56	0.39	0.37	0.56	0.38	0.41	0.68	0.35

The ion currents for Fr1/Fr2/F3 and Fr1/Fr2 were essentially similar indicating that Fr3 can be ignored (Table 2.8). It is noted that the CV values were rather high, indicating that direct quantification using the cylindrical ion-trap instrument was likely to be quite imprecise, and quantitative values will thus be of an approximate nature. Accuracy was likely to be considerably improved with the inclusion of an internal standard in which case sample preparation, injection and detector variations would be corrected for.

System Suitability was performed using sample 1 FrAFr1 and FrAFr2. Each sample was injected 6 times to ensure reproducibility. Retention times, areas and heights were calculated for each ion of interest (Table 2.9).

Table 2.9: Retention Times for Sample 1 System Suitability 6 Injections for FrAFr1.

Mass	RT (min)	RT (min)	RT (min)	RT (min)	RT (min)	RT (min)	Mean RT (min)	CV
534	20.45	-	-	18.25	-	-	19.35	0.08
455	20.48	19.16	19.3	19.31	19.15	19.31	19.45	0.03
437	19.88	18.85	19.28	19.13	19.27	18.99	19.23	0.02
564	17.69	16.65	16.79	16.94	16.49	16.65	16.87	0.03
485	17.76	17.61	17.74	17.75	17.74	17.6	17.70	0.00
467	18.02	17.73	17.72	17.58	17.71	17.58	17.72	0.01
548	20.3	19.28	19.42	-	-	19.28	19.57	0.03
469	20.14	19.11	19.25	19.25	19.24	19.26	19.375	0.02
451	19.97	19.23	19.23	19.37	19.31	19.23	19.39	0.01

In each case, the variability of the retention times was considered to be minor as they remained very steady. System suitability was performed and the peak areas and heights were determined. To correct for injection variations we measured the ratios of 534/548, 564/548, 455/469, 485/469, 437/451, and 467/451 standardising to the hydroxycholestenoic acid. The system suitability performed included combining peak areas and peak heights for Sample 1 FrA Fr1/Fr2 for all 9 ions of interest and then comparing their ratios to 548/469/451. Means and coefficient of variations were calculated for all the data (Table 2.10).

Table 2.10: System Suitability of Sample 1 FrAFr1. Calculated ratios based on 548/469/451 ions of the hydroxycholestenoic acid.

	MS			MS^2			MS^3		
	534.00	564.00	548.00	455.00	485.00	469.00	437.00	467.00	451.00
AREA	0.64	1.10	1.00	0.74	1.29	1.00	3.20	4.84	1.00
	-	-	-	0.64	2.22	1.00	4.42	8.05	1.00
	-	-	-	0.66	1.82	1.00	2.26	4.11	1.00
	0.49	2.95	1.00	0.71	1.91	1.00	2.19	5.17	1.00
	-	-	-	0.64	1.72	1.00	2.28	5.79	1.00
	-	-	-	0.47	0.96	1.00	2.70	2.89	1.00
MEAN	0.57	2.02	1.00	0.64	1.65	1.00	2.84	5.14	1.00
CV	0.18	0.65	0.00	0.15	0.28	0.00	0.30	0.34	0.00
HEIGHT	-	-	-	0.50	1.33	1.00	2.29	3.97	1.00
	-	-	-	0.44	1.95	1.00	2.43	4.17	1.00
	-	-	-	0.48	1.69	1.00	1.92	2.95	1.00
	-	-	-	0.46	1.51	1.00	1.12	3.56	1.00
	-	-	-	0.53	1.71	1.00	1.38	4.50	1.00
	-	-	-	0.35	0.59	1.00	2.09	1.61	1.00
MEAN	-	-	-	0.46	1.46	1.00	1.87	3.46	1.00
CV	-	-	-	0.13	0.33	0.00	0.28	0.30	0.00

Values for the CVs and means that were calculated showed some variance within the data for system suitability (Table 2.10). MS data for area and height was incomplete as peak areas and heights were not well defined. Ratios determined for MS^2 and MS^3 areas and heights were well defined giving better coefficient of variations.

2.4.3 Analysis of other samples

A final three samples as well as sample 2 were analysed according to the defined methods. Retention times for the three *m/z* of interest (*m/z* 534, 564 and 548) as well as relative ion chromatograms for each peak were created and the peak area and peak heights were combined and calculated for the fractions FrAFr1/Fr2/Fr3 as well as for the FrBFr1/Fr2/Fr3, fractions. Where possible, samples were run 3 times and coefficient of variations were calculated.

Table 2.11: Ratios of Hydroxycholesterols to Hydroxyacids for sample 2.

	MS			MS^2			MS^3		
	534.00	564.00	548.00	455.00	485.00	469.00	437.00	467.00	451.00
FrA Fr1/Fr2									
AREA	N/A	N/A	N/A	0.59	2.70	1.00	1.30	3.41	1.00
Height	N/A	N/A	N/A	0.58	1.67	1.00	1.13	2.06	1.00
Mean	N/A	N/A	N/A	0.59	2.19	1.00	1.21	2.73	1.00
CV	N/A	N/A	N/A	0.02	0.33	0.00	0.10	0.35	0.00
FrB Fr1/Fr2									
Area	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Height	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Mean	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
CV	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Both areas and heights for FrA showed a moderate correlation between the ratios calculated for areas and for heights when compared to each other. Unsurprisingly, FrB combined fractions do not show any ions.

Table 2.12: Ratios of Hydroxycholesterols to Hydroxyacids for sample 4.

FrA FR1/FR2/FR3		MS			MS^2			MS^3		
	Sample 1	534.00	564.00	548.00	455.00	485.00	469.00	437.00	467.00	451.00
AREA	1	0.00	1.57	1.00	0.29	1.73	1.00	0.63	5.09	1.00
	2	0.00	4.64	1.00	0.35	1.64	1.00	0.86	4.31	1.00
	MEAN	0.00	3.11	1.00	0.32	1.69	1.00	0.75	4.70	1.00
	CV	N/A	N/A	0.00	N/A	0.04	0.00	0.22	0.12	0.00
Height	1	0.00	0.30	1.00	0.16	1.84	1.00	0.48	5.66	1.00
	2	0.00	2.34	1.00	0.21	2.40	1.00	0.52	4.13	1.00
	MEAN	0.00	1.32	1.00	0.18	2.12	1.00	0.50	4.89	1.00
	CV	N/A	1.09	0.00	0.19	0.19	0.00	0.06	0.22	0.00
FrB FR1/FR2/FR3										
Area	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	2	N/A	N/A	N/A	0.86	2.04	1.00	N/A	N/A	N/A
	MEAN	N/A	N/A	1.00	0.86	2.04	1.00	N/A	N/A	N/A
	CV	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Height	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	2	N/A	N/A	N/A	1.00	2.02	1.00	N/A	N/A	N/A
	MEAN	N/A	N/A	N/A	1.00	2.02	1.00	N/A	N/A	N/A
	CV	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Sample 4 (Table 2.12) showed moderate reproducibility through CVs for peak area and heights. There is very little if any of the ions seen in FrB combined fractions. Comparison of the means for peak area and peak heights for a particular ion shows a good comparison.

Table 2.13: Data for Sample 5

	MS			MS^2			MS^3		
	534.00	564.00	548.00	455.00	485.00	469.00	437.00	467.00	451.00
FrA Fr1/Fr2/Fr3									
AREA	N/A	1.57	1.00	0.29	1.73	1.00	0.63	5.09	1.00
Height	N/A	0.30	1.00	0.16	1.84	1.00	0.48	5.66	1.00
Mean	N/A	0.94	1.00	0.22	1.79	1.00	0.56	5.37	1.00
CV	N/A	0.95	0.00	0.41	0.04	0.00	0.19	0.08	0.00
FrB Fr1/Fr2/Fr3									
Area	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Height	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Mean	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
CV	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

The MS data is very variable, however MS^2 and MS^3 data shows that both area and heights are comparable to each other. FrB combined fractions show nothing present in the fractions.

Table 2.14: Date for Sample 6

	MS			MS^2			MS^3		
	534.00	564.00	548.00	455.00	485.00	469.00	437.00	467.00	451.00
FrA Fr1/Fr2/Fr3									
AREA	N/A	4.64	1.00	0.35	1.64	1.00	0.86	4.31	1.00
Height	N/A	2.34	1.00	0.21	2.40	1.00	0.52	4.13	1.00
Mean	N/A	3.49	1.00	0.28	2.02	1.00	0.69	4.22	1.00
CV	N/A	0.47	0.00	0.35	0.27	0.00	0.35	0.03	0.00
FrB Fr1/Fr2/Fr3									
Area	N/A	N/A	N/A	0.86	2.04	1.00	N/A	N/A	N/A
Height	N/A	N/A	N/A	1.00	2.02	1.00	N/A	N/A	N/A
Mean	N/A	N/A	N/A	0.93	2.03	1.00	N/A	N/A	N/A
CV	N/A	N/A	N/A	0.11	0.01	0.00	N/A	N/A	N/A

As with sample 5, there is very little seen in FrB combined fractions and the ratios of area and heights from the MS data for FrA combined fractions show great variability. However, there is a good agreement between peak area and height ratios for MS^2 and MS^3 data for FrA combined data.

2.5 Discussion

The use of separation techniques is a common way to increase signal of low abundance compounds which would be masked by those at higher levels. In this case, the ability to analyse oxysterols in plasma is very much dependent on the ability to remove cholesterol from the sample as cholesterol poses problems when trying to analyse oxysterols by LC-MS. The first is the level of cholesterol in the blood/plasma which is approximately 2 mg/mL. This concentration far exceeds that of any oxysterol found in blood and can thus create difficulty with the dynamic range. Cholesterol, being a neutral molecule is not amenable to analysis in positive electrospray, even at the high concentrations it is found in blood. Seeing hydroxycholesterols, which are also neutral molecules and which are far lower in abundance is far more difficult. Thus, to see these low abundant molecules, more sample would need to be injected, which would cause cholesterol to saturate the system, causing carryover on the column.

Another major problem with analysing oxysterols in a cholesterol mixture is the autoxidation which cholesterol may undergo, which would change the levels of hydroxycholesterols found in the sample. In order to prevent this, removing cholesterol from the oxysterol fraction is vital for good replicate data.

The method that was used for separation of samples involved dilution of plasma in 70% ethanol, extracting cholesterol on the C_{18} column, with oxysterols in the flow-through, and final elution of cholesterol in 100% EtOH. Only the first two fractions were oxidised, derivatised and purified. In all cases, with all samples that were screened this way, cholesterol did not appear in the FrA, but the *m/z* 534 which corresponds to any hydroxycholesterol present, only appears in high concentration in fraction FrA.

A standard of 25-hydroxycholesterol was run on the column to determine the retention time expected for the hydroxysterols which came out around 19 min (Table 2.4) while any residual cholesterol was seen at 25 min (data not shown). HPLC runs of the sample 1(Figure 2.4, Table 2.5), showed good separation of the peaks of the *m/z* of interest. Selective ion monitoring was performed for *m/z* 534 – derivatised hydroxycholesterols, *m/z* 548 – derivatised 3β-hydroxycholest-5-en-27-oic acid and *m/z* 564 – derivatised 3β,7α-dihydroxycholest-5-en-27-oic acid. Each of these ions was in turn selected for MS^2 and their $[M-79]^+$ ion was selected for MS^3 (Table 2.3). The subsequent chromatogram was used to identify the compound of interest.

Retention times were measured (Table 2.5, Table 2.6) and the reproducibility of the retention times was determined for replicate runs (Table 2.9), giving overall low

coefficient of variation for most ions. In some cases, the $[M]^+$ mass of the parent ion is not reported, nor is the retention time, as a relative ion chromatogram for the particular *m/z* does not yield a specific peak, in which case, peak area and height cannot be determined. Instead it was observed that using the MS^2 or MS^3 which are more specific gave much better reproducibility for analysis (Table 2.7). A comparison was made for the difference that using FrA-Fr1/Fr2 and FrA-Fr1/Fr2/Fr3 made to the results. In general, the addition of Fr3 to the data did not change the mean very much suggesting that most of the ions were found in FrA1- Fr1 and FrA1-Fr2 (Table 2.8).

System reproducibility and method reproducibility were both performed to determine the suitability of the methods. Method suitability (Table 2.7) involved preparing one sample three times and analysing using the method at hand. In all cases, it can be seen that the MS ions give a higher CV than the MS^2 and MS^3 ions. The data for MS^2 and MS^3 is much more reliable.

Plasma, while containing high levels of cholesterol, also contains many other small molecules which can be of interest. These molecules could potentially be used as biomarkers, or to identify changes which may help characterise illness or disease before it becomes detrimental or chronic to the patient. The molecules of interest here are those belonging to the steroid family – in particular those with a cholesterol backbone. Previous studies have shown that there are key oxysterols and cholestenoic acids that are present in plasma.[123,129,139,143,144]

In particular, of interest, are 24S-hydroxycholesterol and 27-hydroxycholesterol. 24S-hydroxycholesterol is a membrane-permeating form of cholesterol hydroxylated by CYP46A1 for release through the blood brain barrier for either recycling into a bile acid or to be excreted in urine. This particular hydroxycholesterol is only formed in the brain and can potentially be used as a marker for disease. Previous studies have found that the 24S-hydroxycholesterol is found at concentrations between 60-80 ng/mL[129,139,143,144] while the concentrations of 27-hydroxycholesterol, which is excreted from most mammalian cell types and is the most prevalent hydroxycholesterol in plasma is found at concentrations of 120 – 150 ng/ mL.[129,139,143] Of note, is the concentration of 3β-hydroxycholest-5-en-27-oic acid which is found at 100 ng/mL. Its hydroxylated form, 3β,7α-dihydroxycholest-5-en-27-oic acid has not been quantified in plasma. Relative ratios of the 24S-hydroxycholesterol and 27-hydroxycholesterol to 3β-hydroxycholest-5-en-27-oic acid can be determined by dividing by the acid giving rise to ratio of 0.6-0.8 : 1 and 1.2 – 1.5 :1 respectively for literature data. Combined ratios of 24S/27-hydroxycholesterol to the acid would then be 2.0 – 2.2 : 1. The relative peak areas

and heights of experimental data falls between 0.04 – 2.84 for 24S / 27-hydroxycholesterol and 0.3 – 5.66 for 3β, 7α-dihydroxycholest-5-en-27-oic acid (Table 2.10, Table 2.8, Table 2.11, Table 2.12, Table 2.13, Table 2.14). The two monohydroxylated cholesterols are not resolvable, thus their combined areas and heights were used for analysis against 3β-hydroxycholest-5-en-27-oic acid and their MS^2 and MS^3 data were compared to a library for confirmation of their identity.[147]

2.5.1 Interpretation of MS data

Key cholesterol products found in this study were 24S-hydroxycholesterol, 27-hydroxycholesterol, 3β-hydroxycholest-5-en-27-oic acid and its metabolite 3β,7α-dihydroxycholest-5-en-27-oic acid. As with cholesterol, these molecules do not ionise well for analysis by mass spectrometry including by positive ESI. In order to increase their ionisation capabilities, Griffiths *et al* developed a method using cholesterol oxidase to oxidise a hydroxyl group into a ketone which in turn is derivatised into a Girard P hydrazone through a Schiff Base reaction. The hydrazones contain a quaternary nitrogen ion which increases the signal intensity in positive mass spectrometry methods. Identification of the derivatised compounds has been with the help of reference standards generated within the lab[147] and comparing the MS^3 data obtained from the capillary LC-MS to the identification from the reference standards. The hydroxycholesterol hydrazones give a *m/z* of 534 and give specific MS^2 and MS^3 patterns and ratios of ions which in turn can be used to identify isomers from each other as seen using 24S-hydroxycholesterol as an example (Figure 2.7).

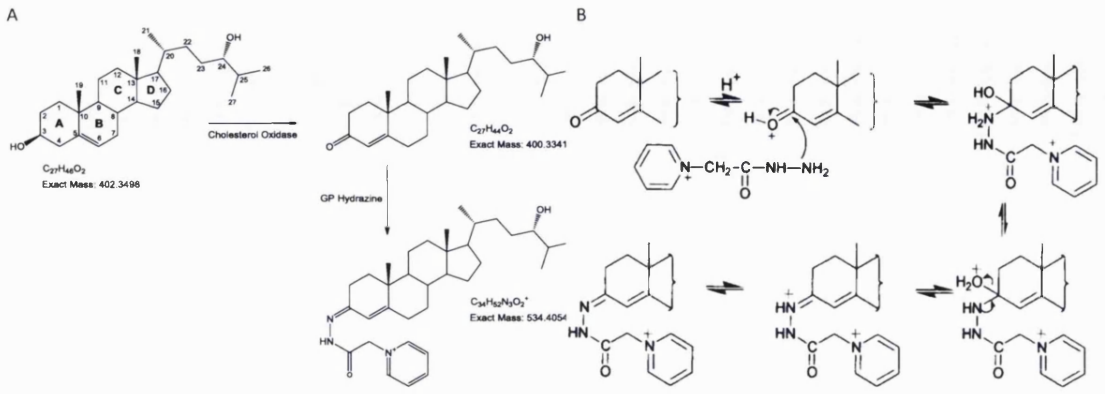


Figure 2.7: Oxidation and derivatisation using cholesterol oxidase followed by Girard P reagent on 24S-hydroxycholesterol

The MS^2 spectra are dominated by $[M-79]^+$ and $[M-107]^+$ ions which correspond to the loss of pyridine and pyridine plus carbon monoxide respectively. In many cases $[M-97]^+$ can also be seen which corresponds to the loss of the pyridine and a water molecule. Fragmentation nomenclature of cholesterol derivatives is based upon the cholesterol ring nomenclature (Figure 2.7). At the low *m/z* range, $[M-79]^+$ gives rise to characteristic fragmentation ions 151; 163; and 177 corresponding to the cleavage in the A and B rings; b_1-12, b-28 and b_2 respectively from the $[M-79]^+$ ion (Figure 2.8).

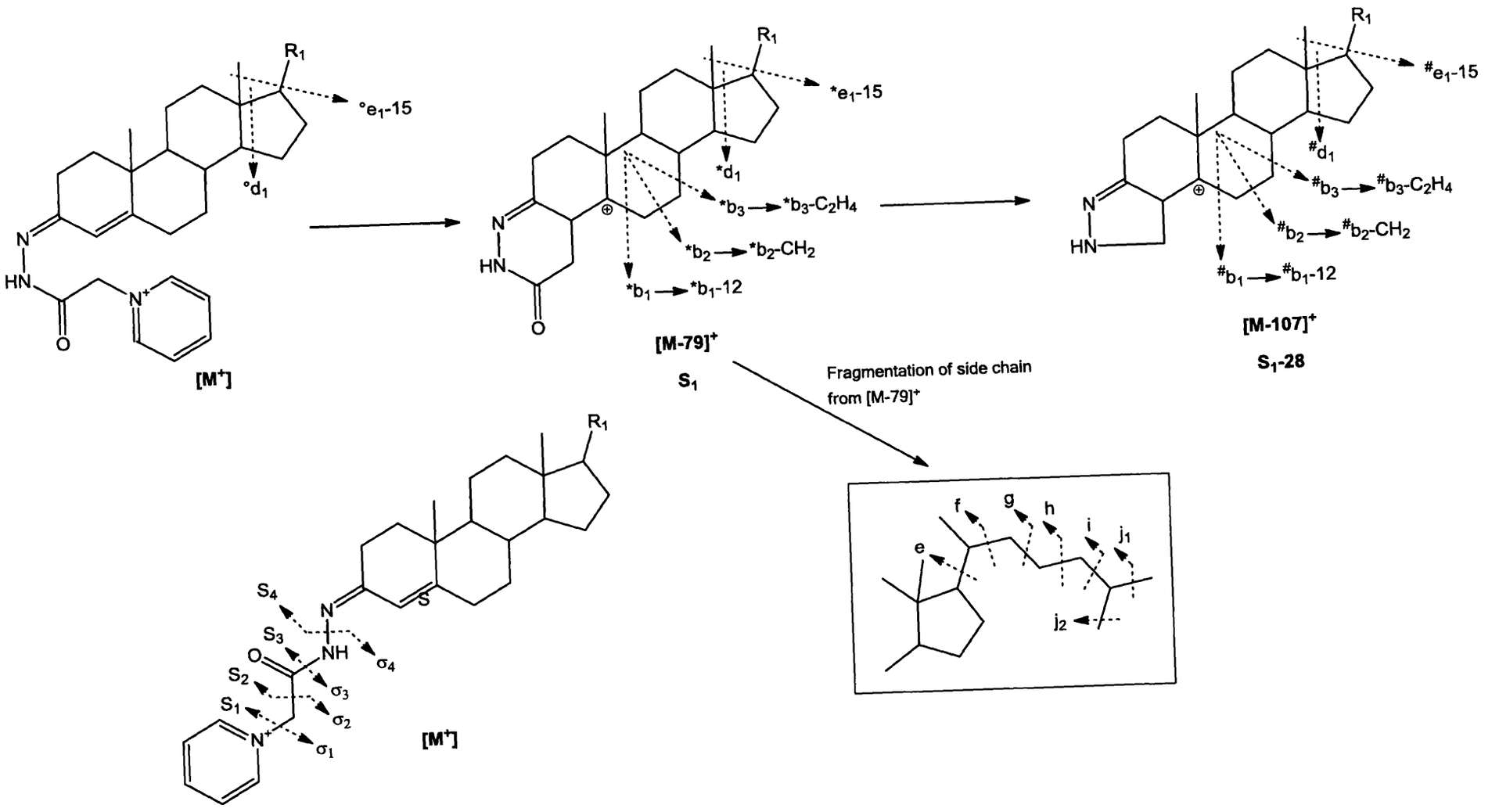


Figure 2.8: MS^2 and MS^3 fragmentation of sterol GP hydrazones using 24-hydroxycholesterol as an example.

Identification of key hydroxycholesterols and acids present in the sample was performed after the separation on the nano-ESI-LC-MS had been optimised to allow for better identification of the key compounds. Derivatised hydroxycholesterols were compared to a library created in the lab[147] using MS^2 and MS^3 data from the LCQ.

Table 2.15: Summary of fragmentation ions for hydroxysterols of interest.

		MS^2			MS^3								
Hydroxysterol	M^+	M-79	M-97	M-107	M-97	M-107	M-122	M-125	M-137	M-140	M-151	M-154	M-169
24-hydroxycholesterol	534	455	437	427	437	427	412	409	——	394	——	380	365
25-hydroxycholesterol	534	455	437.3526 (15)	427	437	427	——	409	——	394	——	380	365
27-hydroxycholesterol	534	455	437	427	437	427	412	409	397	394	383	380	365
3β-hydroxycholest-5-en-27-oic acid	548	469	451	441	451	441		423		408		395	
3β, 7α-dihydroxycholest-5-en-27-oid acid	564	485	467	457	467	457		439					395

	MS^3											
Hydroxysterol	'*h^2	'*g	*f'	'*f	*e'	"e'	'*e	*d_1-12	*c_2	*b_2	*b_3-28	*b_1-12
24-hydroxycholesterol	381	367	——	353	327	——	325	273	231	177	163	151
25-hydroxycholesterol	381	——	355	353	——	——	325	——	231	177	163	151
27-hydroxycholesterol	381	——	355	353	327	284	325	273	231	177	163	151
3β-hydroxycholest-5-en-27-oic acid										177	163	151
3β, 7α-dihydroxycholest-5-en-27-oid acid											179	151

M-79, Loss of C_5H_5N
M-97, Loss of $C_5H_5N + H_2O$
M-107, Loss of C_5H_5N + CO
M-122, Loss of C_5H_5N + CO +NH
M-125, Loss of C_5H_5N + CO + H_2O
M-137, Loss of C_5H_5N + CO + CH_2O
M-140, Loss of C_5H_5N + CO + H_2O + NH
M-151, Loss of C_5H_5N + CONH +CH_2NH
M-154, Loss of C_5H_5N + CONH +CH_3OH
M-169, Loss of C_5H_5N + CO +H_2O +NH+$NHCH_2$

Figure 2.8 shows the fragmentation of the molecule denoting the location of the fragmentation of b-h.

An asterix preceding letter in Table 2.15 refers to fragmentation by loss of pyridine moiety, while hash refers to that which has lost the pyridine and a CO molecule.

A prime preceding the letter refers to hydrogen transfer from the ion to the neutral molecule.

A prime post letter refers to the neutral molecule donating a hydrogen to the ion.

At an approximate retention time of 19-20 minutes, hydroxycholesterols are eluted from the column. From the MS^3 data acquired of *m/z* 534 -> 455, two hydroxycholesterols appear to be present: 24S-hydroxycholesterol and 27-hdyroxycholesterol (Figure 2.9).

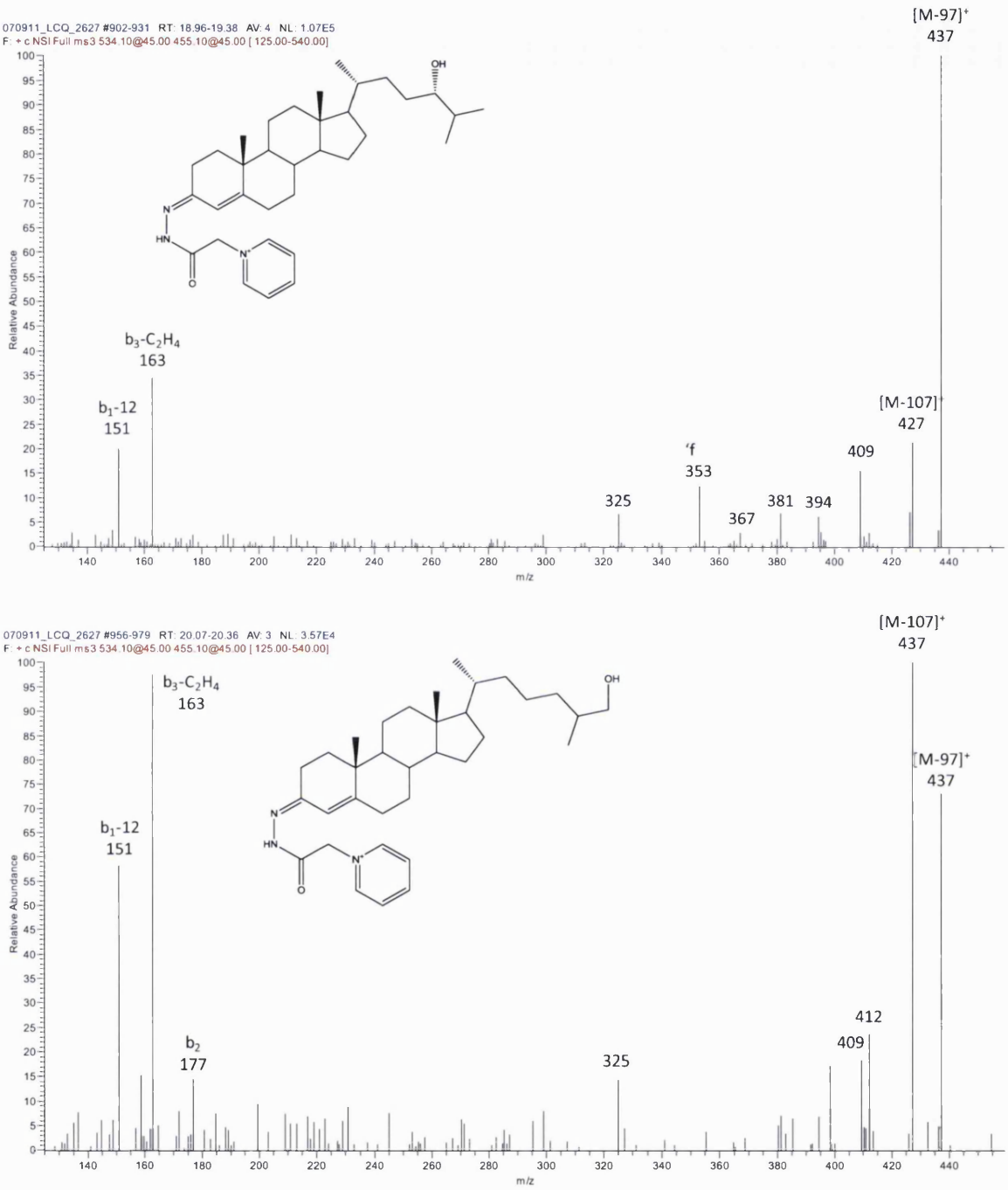


Figure 2.9: 24-hydroxycholesterol (upper) and 27-hydroxycholesterol (lower) MS^3 of *m/z* 534→455→ from sample 2 FrAFr2. The upper spectrum is from the leading edge of the peak at 1.8 min and the lower spectrum is from the tail of this peak.

A comparison of MS^2 and MS^3 data of *m/z* 534 shows similarity between 24-hydroxycholesterol and 27-hydroxycholesterol as both give the same ions of 151 and 163, but the 24-hydroxycholesterol has the base peak of 437 and a peak of 353, while the base peak for 27-hydroxycholesterol is 427 and has another peak of 412 present which is in agreement with the standards in the library.

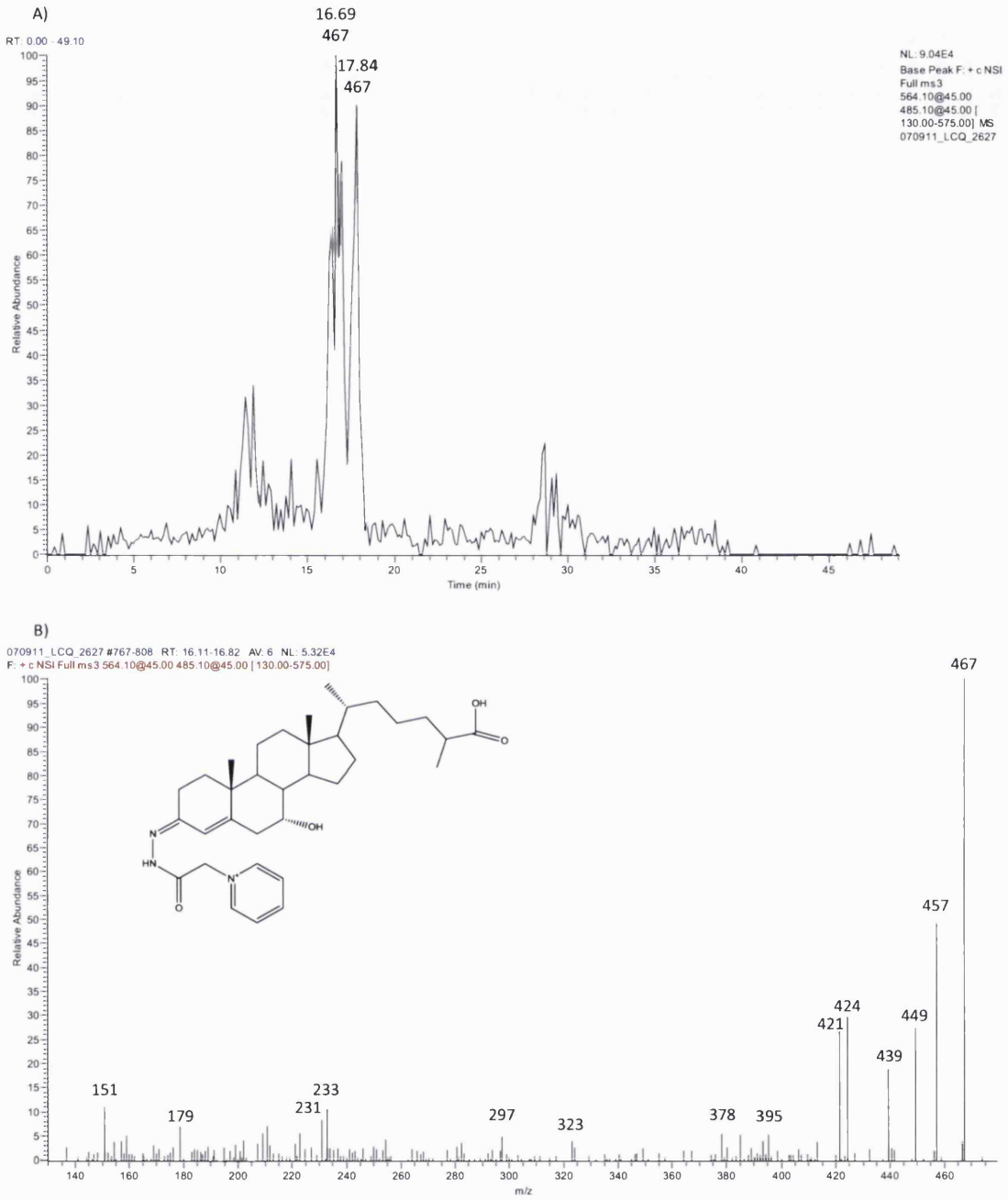


Figure 2.10: A) Chromatogram of MS/MS data of 564→485→ into two peaks: B) retention time 16.75. (Continued on the next page)

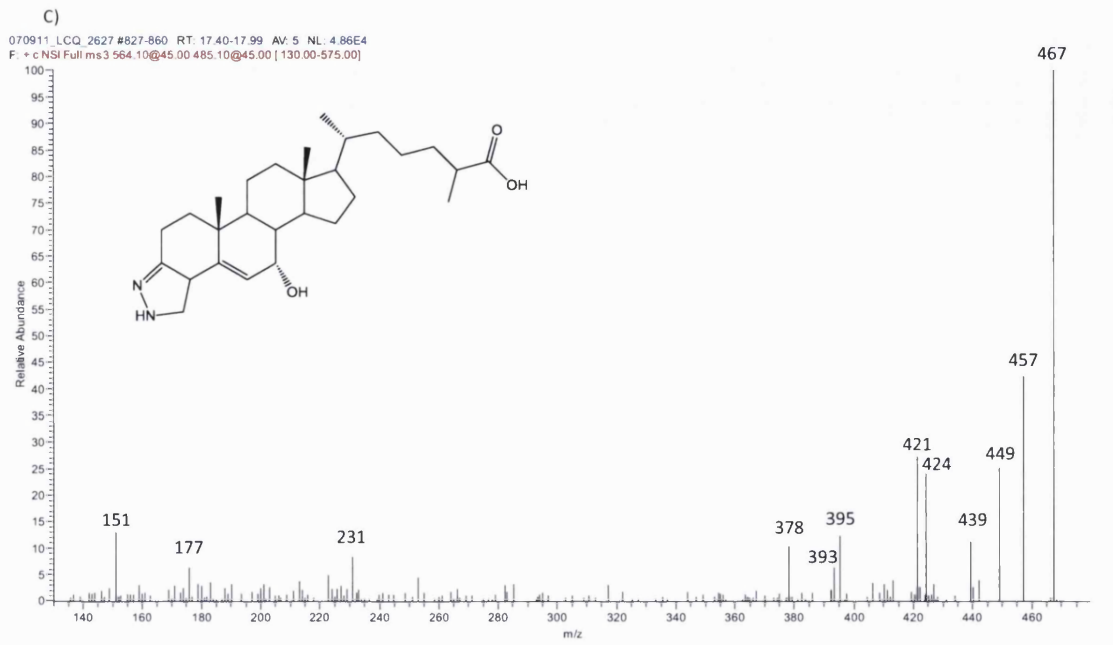


Figure 2.10: Chromatogram of MS/MS data of 564→485→ into two peaks: C) retention time 17.63.

The *m/z* 564 corresponds to a 3β,7-dihydroxycholestan-5-en-27-oic acid. Two peaks are visible in the chromatogram when looking at MS^3, however their spectra are near identical and are difficult to differentiate between possibly the 7α and 7β isomers (Figure 2.10).

The final *m/z* 548 was determined to be 3β-hydroxycholestan-5-en-27-oic acid. As seen here in Figure 2.11.

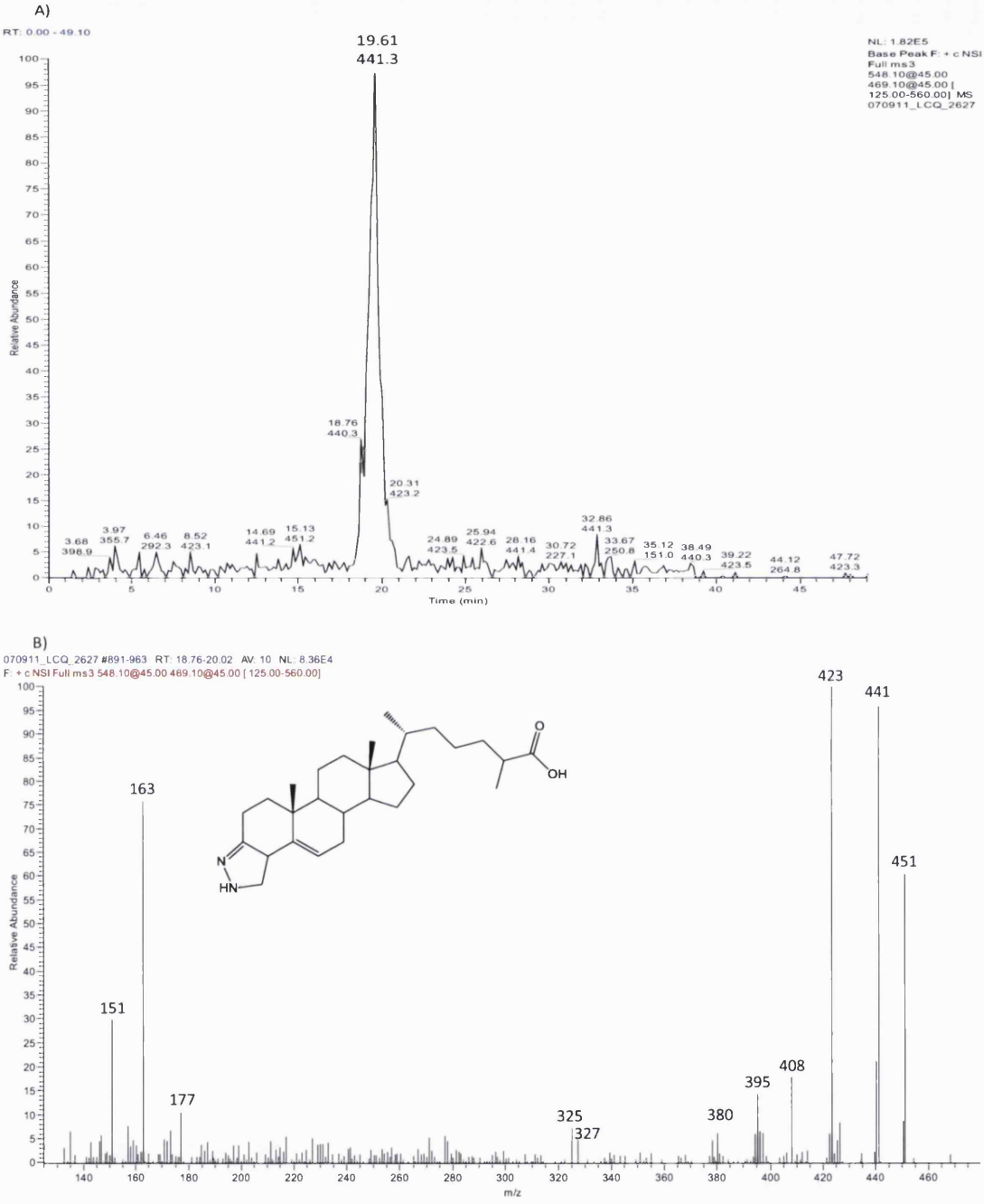


Figure 2.11: A) Chromatogram of MS/MS data of 548→469→ and B) MS/MS spectra

2.6 Conclusions and Future Work

The use of a reference library is key to identifying compounds of interest, in particular the compounds here were oxidised and derivatised to increase signal intensity for analysis by mass spectrometry. Without reference libraries to compare to, identification becomes far more difficult. The libraries however in each case would have to be done in a comparable fashion i.e. the same type of fragmentation would need to be done with the same type of derivatisation to be able to confirm the structures.

In this case, 24S-hydroxycholesterol, 27-hydroxycholesterol, 3β-hydroxycholest-5-en-27-oic acid and 3β,7-dihydroxycholest-5-en-27-oic acid were identified by elucidation of their structures based on their MS^2 and MS^3 data as well as by comparing to an in-house library.

The LC-MS analysis of plasma samples with selective ion monitoring for the ions and their MS^2 and MS^3 ions gave sufficient separation of the hydroxycholesterols from the acids. While the two hydroxycholesterols were not well separated from each other, their average peak areas and peak heights were used to create ratios against 3β-hydroxycholest-5-en-27-oic acid, a very common hydroxycholesterol acid found in plasma. This ratio, should be consistent for healthy individuals and can be used as a basis for identifying changes in a healthy individual which may then be characterised by further testing. The changes would reflect predominantly on the changes in 24S-hydroxycholesterol, which is predominantly expressed in the brain and could reflect neurological disease, such as Alzheimers or damage to the myelin sheaths. The expression of 27-hydroxycholesterol occurs in most mammalian tissues and could also reflect problems with either an increase in expression or a decrease in metabolism and may reflect a potential for artherosclerosis occurring. The ratios were determined for 6 individual samples which did not vary greatly when comparing the peak areas and heights of the MS^2 and MS^3 ions but the MS ions show more variablility as their peak shape was not well resolved (Figure 2.4, Figure 2.6). In the case of the samples, there were more ions of *m/z* 534, 564 and 548 than just the three ions of interest.

The three ions and four hydroxycholesterols of interest were identified and their ratios were determined using a robust method which allowed for the analysis of 6 different samples.

Future work could involve resolving the 27- and 24S-hydroxycholesterol peak which could then be used to compare ratios against each other as well as identify other

compounds which were seen in the chromatogram but were not identified, but may also be of interest. Identification of the lower abundance hydroxycholesterols would also benefit from the use of a linear ion trap or another more sensitive and accurate instrument. A combination of better resolution and sensitivity would improve both identification and quantification of each metabolite.

3 1-DE Analysis of Proteins Involved In Sterol Synthesis, Metabolism and Transport in Brain

3.1 Brain Proteomics

The brain with all of its complexity makes it one of the most interesting tissues for study. Its role is vital to the viability and function of evolved species such as vertebrates. It is the focal point for thought, movement, perceptions, memory and learning and, its importance cannot be overstated.[148,149]

The brain is a system composed of many different cell types and over the years, research into the brain development and aging has been furthered by newer and newer technologies, but is still at the edge of pushing the technologies currently in practice with its complexity. This complexity is partially derived from the varying cell types and partially from the fact that approximately 30-50% of the genome is expressed in the mammalian nervous system, leading to a phenomenally large number of gene products to be identified – a monumental task for the current separation techniques. While the overall processes of the brain are of great interest because of its important function, there is also a social importance directly associated with neurodegenerative diseases such as Alzheimers, Parkinson's, Multiple Sclerosis and prion diseases. The brain also has a very limited capacity for regeneration, making the prevention of trauma or disease from attacking the brain vital. Identifying and understanding these changes on a molecular level would allow for hope that better prevention and/or treatments could be created.[148,149] These include both proteomic and metabolomic studies to better understand the intricacies involved with normal and abnormal brain function. The interest in mapping the proteome of the brain in human and mouse has become substantial over the years, involving major collaborations to combine, process and validate the results that are obtained to create a 2-dimensional map of the brain proteome.[150] These huge collaborations have also included profiling of the mRNA and subsequent correlation of the mapped differentially expressed genes with their corresponding proteins.[148,150] The acquired data requires substantial amounts of time to process and interpret the results; removing the false positives and any poor identification to get a better perspective of the proteome, followed by comparisons with literature and gene ontology to confirm proteins and identify those that may be involved with disease states.[148,149]

As a result, a pilot study was developed to study mouse brain at various ages and parts of the human brain with and without disease. The pipeline establishes a strategy for bioinformatics. These samples were given to 15 test laboratories across the world

using a standard protocol which was annotated based on the changes each group made. From these groups, a variety of different analyses were performed creating qualitative and quantitative data from looking at the whole proteome to differential analysis and comparisons. All the data that was created by the different laboratories was subjected to central processing to allow for sharing of data across the groups.[151,152,153,154] Data was submitted to the Data Collection Center (DCC, Medizinisches Proteom-Center, Bochum Germany) for reprocessing after which all data was available publicly on the PRIDE database (http://www.ebi.ac.uk/pride) to be used as a reference for future analyses. Each group used their own instruments to analyse the data and in doing so, many of them performed smaller experiments in tandem which provided for more information about the samples, techniques and analyses used. These ranged from types of isoelectric focusing[149], pattern profiling on 2-DE,[84,90,94,155] to 2-D gel electrophoresis (2DE) to quantitative techniques such as Cy3-Cy5[91] and using 1DE with LC-MS/MS[156] or 2D LC-MS/MS[80] for qualitative analysis of the entire proteome.

When dealing with such large data sets, a sizeable group of people need to be involved to be able to process and analyse the data for the entire proteome. Alternatively, many groups did not focus on the whole proteome, but focused on differential expression[83,89,91,92,157] between two proteome states, or purification of a subset of proteome such as the phosphoproteome[117,118] or the glycosylproteome[109,110,134] or search the proteomic data for a subset of proteins of interest[158] such as the proteins involved in the synthesis of cholesterol and other sterols, their metabolism and their transport.

3.1.1 Cholesterol and Steroids in the Brain

Neuronal function, which is core to the brain, requires adequate cholesterol (Figure 3.1) for its function and for normal neuronal development, plasticity and synaptic transmission.[159]

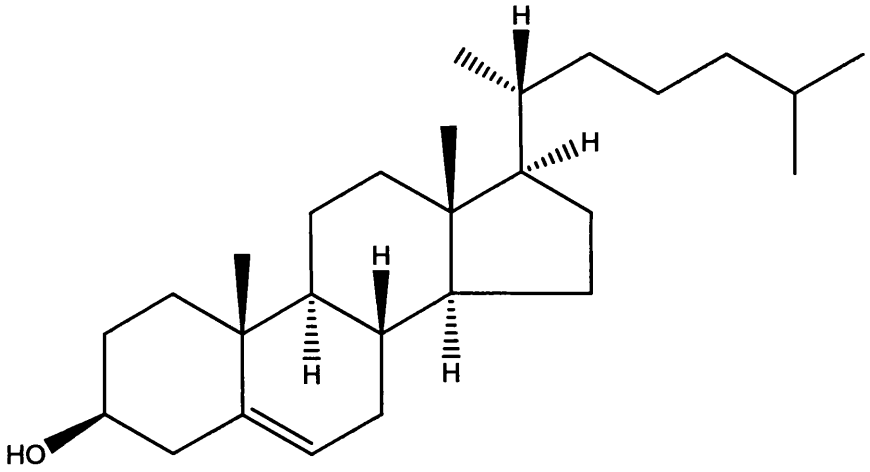


Figure 3.1: Structure of Cholesterol

The blood brain barrier prevents the uptake of lipoprotein cholesterol complex into the brain from circulation, requiring *de novo* synthesis to occur *in situ.*[160] In the CNS, more than 99.5 % of the cholesterol is unesterified and is divided mainly into 2 pools: the myelin sheaths formed by the oligodendrocytes and the plasma membranes of the astrocytes and neurons. Myelin is composed of approximately 70% lipid, cholesterol, phospholipids and glycosphingolipids in a ratio of 4:4:2. Up to 70% of the cholesterol in the brain is found in myelin and up to half of the white matter is composed of myelin, making the brain the most cholesterol-rich organ.[161,162]

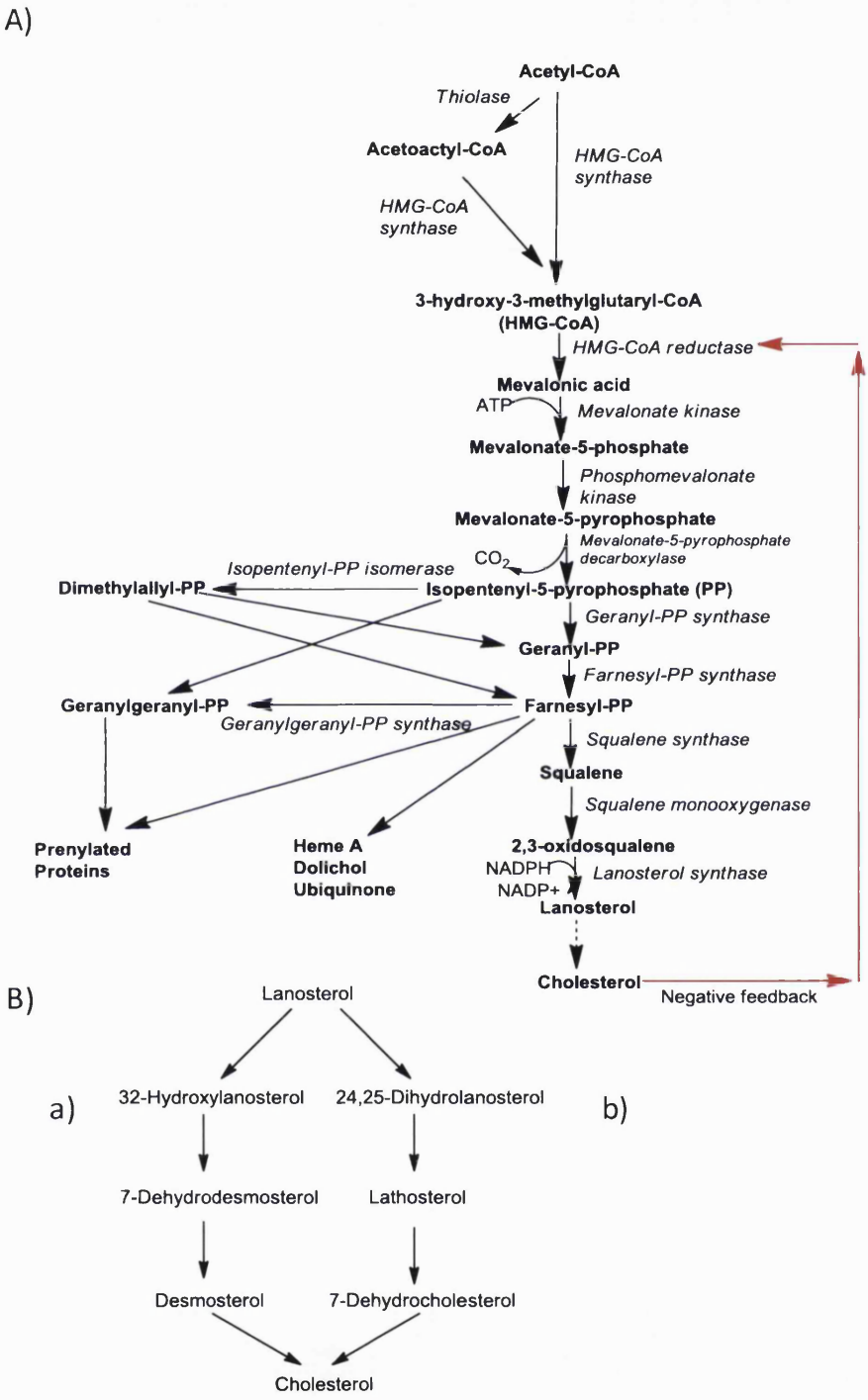


Figure 3.2: Schematic representation of cholesterol synthesis. A) Synthesis of cholesterol via the mevalonate pathway.[163] B) Synthesis of cholesterol via (a) 7-dehydrodesmosterol and (b) 7- dehydrocholesterol pathways towards cholesterol synthesis. Adapted from *Seth et al, 2006*

Synthesis of cholesterol in the CNS (Figure 3.2A) is high in the developing fetus and declines into adulthood. Most of the cholesterol is then recycled with a long half-life of 6 years. Astrocytes synthesise 2-3 times more cholesterol than neurons. Oligodendrocytes synthesise more cholesterol than astrocytes and use their cholesterol to create myelin sheaths. In the CNS, the oligodendrocytes primarily use the 7-dehydrodesmosterol pathway for synthesis of cholesterol while the glial cells predominantly use the 7-dehydrocholesterol pathway (Figure 3.2).[161]

Extracerebrally, cholesterol can then be transformed into a variety of different products such as glucocorticoids, mineralocorticoids, androgens and estrogens, oxysterols and bile acids depending on the location and tissue. Glucocorticoids are compounds such as cortisol which affects carbohydrate, protein and lipid metabolism. They influence a variety of functions such as inflammatory response and stress. They are predominantly synthesised in the cortex of the adrenal gland. Mineralocorticoids are synthesised in the same place but regulate salt and water in the biological system via the kidneys while androgens and estrogens are predominantly synthesised and released from the testes and ovaries and regulate sexual development and function.[164]

Oxysterols are found in low concentrations in the body and are predominantly formed by hydroxylations of cholesterol by members of the cytochrome P450 family. These oxysterols increase the permeation of cholesterol across membranes and allow for easier transport to the liver for the synthesis to bile acids through to cholic acid.[123]

One of the products found in the blood which is specific to the brain is 24S-hydroxycholesterol, produced by CYP46A1. (Figure 3.3) CYP46A1 hydroxylates the 24 position and, to a lesser extent, the 25 and 27 positions. Synthetic pathways for cholesterol modifications in the brain are shown in Figure 3.3, with products including neurosteroids and hydroxysteroids that have been indentified and quantified in the brain of both rat and human.[68,126,144]

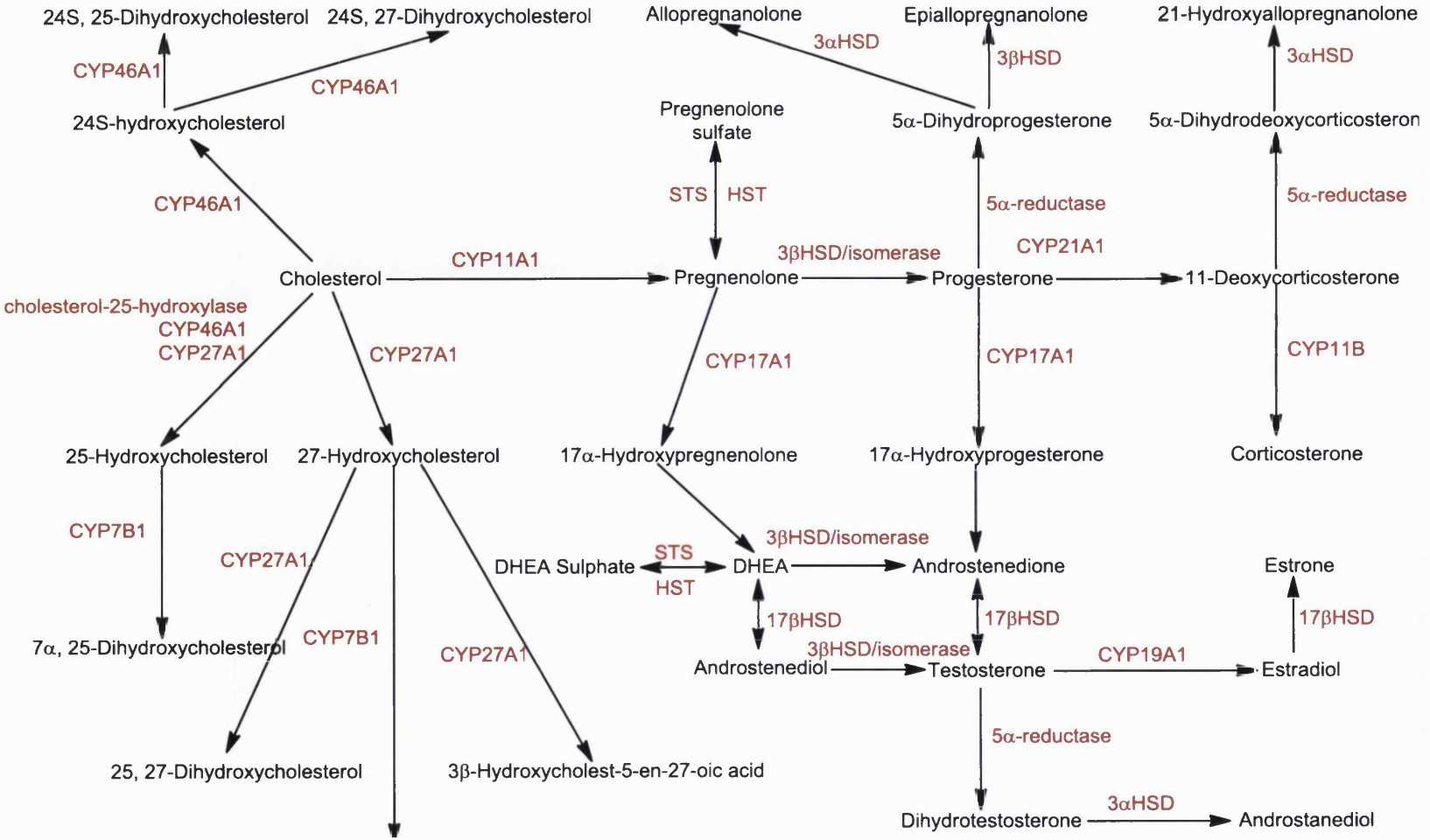


Figure 3.3: Pathway of synthesis of sterols and neurosteroids from cholesterol in the brain. Enzymes are shown in red. HSD: hydroxysterol dehydrogenase, CYP: cytochrome P450, STS: steroid sulfatase sulfohydrolase, HST; hydroxysteroid sulfotransferase. Adapt from *Griffiths et al*.[123]

Cholesterol synthesised in the brain when in excess must be transported out from the brain and excreted to maintain the important steady-state. Lipoproteins, which are globular and micelle like in structure, have a core of non-polar triacylglycerides and cholesteryl esters which are surrounded by amphiphilic proteins and phospholipids. Three types: VLDL (very low density lipoprotein), ILDL (intermediate density lipoprotein) and LDL (low density lipoprotein) transport endogenous triacylglycerides and cholesterol from the liver to tissues while HDL (high density lipoprotein) transports cholesterol from tissues to the liver. Apolipoproteins are proteins that are loosely associated with lipoproteins, usually enveloping the micelle structure. As VLDL loses triacylglycerides, cholesterol and apolipoprotein, it becomes ILDL and then LDL, thus recycling its use through the system and moving lipids into various tissues. ApoE is the major lipoprotein in plasma that is involved in transport of cholesterol. It has also been found in the CSF suggesting that there is circulation of cholesterol occurring in the CNS. Other transporters found and expressed in the CNS include many of the lipoprotein transporters, ATP binding cassettes (ABCA1), low density lipoprotein receptor related proteins (LDLR), and apolipoproteins.[165] Knockouts of the LDL receptor and apoE are not associated with any brain abnormalities, however a VLDLR and ApoE receptor 2 knockout do show a specific phenotype.

ApoE and ApoA-I are present in spherical discoidal particles in the CNS. The role of ApoA-I is not clear, although it has been suggested that it is expressed in the spinal cord. As ApoA-I is a potent mediator of cholesterol efflux in the rest of the body, it may have a role in cholesterol metabolism in the brain.

ApoE is derived locally from glial cells with little contribution from plasma. The lipoproteins secreted by the glial cells provide lipids to the neurons for membrane synthesis. Although ApoE is not required to maintain brain homeostasis, it is critical with the development of neurodegenerative disorders and recovery from brain injuries. ApoE binds lipoprotein in an isoform-specific manner which can affect cholesterol efflux from neurons. Of the common human isoforms, and a major risk factor for Alzheimers disease is ApoE4, which is the cause of 40-60% of the genetic variation of the disease. One of the main reasons is believed to be due to its poor effectiveness at transporting cholesterol to the neurons, unlike ApoE3 which is far more efficient at transport.[166]

ATP-binding cassettes are also considered to be important, shuttling cholesterol from glial cells to neurons. ABCA1 and ABCA2 and ABCG1 are three prominent transporters in the brain from this superfamily.[159]

There are two main ways that cholesterol can be removed from the CNS. The first involves the formation of hydroxylated cholesterols. In humans, the prominent hydroxylated sterol excreted across the blood-brain-barrier is 24(S)-hydroxycholesterol.[161] Approximately 6-7 mg of 24(S)-hydroxycholesterol is moved from the CNS every 24 hours, and the uptake from the liver has been used to show that it is exclusively from the brain. The rate of its production is 2/3 that of the synthesis of cholesterol. Cytochrome P450 46A1, the enzyme responsible for the formation of the 24(S)-hydroxycholesterol is essentially expressed only in neurons of the cerebral cortex, hippocampus and the dentate gyrus of the brain.[165] While deletion of this enzyme was thought to cause an inherent disease state, it was found that the synthesis of cholesterol reduced itself by approximately 35% and that the suppression resulted more in the reduction of cholesterol in the grey matter rather than the heavily myelinated areas. It has been suggested that 24-hydroxycholesterol could be used as a surrogate marker for brain cholesterol homeostasis by monitoring its levels in plasma from the CNS to the liver.[167] Its transport to the liver is the same as that for cholesterol, using ApoE and LDLR and the ATP binding cassettes to move from the brain.

3.2 Aims

As the brain proteome is slowly uncovered, it is the membrane-associated proteins which are of particular interest, in particular those involved in sterols and steroids biosynthesis, metabolism and transport. As cholesterol is synthesised *de novo* in the brain, those enzymes involved with its biosynthesis and metabolism must also be present as must transporters and receptors. Using 1D-gel electrophoresis prior to tryptic digestion and subsequently followed by 1D-LC-MS/MS of the tryptic peptides, membrane-bound proteins, including those involving cholesterol and its various pathways can be identified.

3.3 Materials and Methods

3.3.1 Materials

HPLC grade H_2O, acetonitrile, 99.8% ethanol, ammonium hydrogen carbonate, glycine, phosphoric acid and 99% BSA were purchased from Fisher Scientific (Waltham MA, USA). Trypsin was purchased from Promega (Southampton, UK). NuPAGE® Novex Pre-Cast Gel System 4-12 % gels with 10 lanes, NuPage® Reducing Agent (500 mM DDT), NuPage® Sample Buffer (4X), NuPage® MOPS (20X) buffer and NuPage® SimplyBlue Safe Stain were bought from Invitrogen (Paisley, UK). Bio-Rad protein assay was purchased for Bio-Rad (Perth, UK). Tris was bought from Melford (Chelsworth, Ipswich, UK). Formic acid (99%), sucrose, glycerol (molecular biology grade) and glucose (molecular biology grade) were purchased from VWR (Arlington Heights, IL, USA). Siliconised tubes, iodoacetamide, Coomassie Brilliant Blue G-250 and Sigmafast protease inhibitor tablets were purchased from Sigma-Aldrich (St. Louis, MO, USA).

3.3.2 Cellular Fractionation:

Three rat brains (stored at -80°C) with a net wet weight of 5.59 g were thawed and washed 3 X with homogenisation buffer (0.01 M Tris-HCl buffer, pH 7.4, 15% glycerol, 0.25 M sucrose, protease inhibitor cocktail tablets) to remove any excess blood. The tissue was cut into small pieces and homogenised using a glass/Teflon homogeniser. A total of 40 mL homogenisation buffer was used to prepare the homogenate and rinse out the homogeniser into 2 x 40 mL falcon tubes. The tubes were balanced, centrifuged (Sigma Centrifuge, rotor 11162, 10 minutes, 4°C, 2400 x *g*) and the pellet (nuclei, cell debris) was discarded (Figure 3.4). The supernatant was split again into 2 centrifuge tubes and equated using homogenisation buffer and centrifuged (Beckmann J2-21 Centrifuge, JA20 Rotor, 20 minutes, 4°C, 12,250 x *g*). The pellet was kept (mitochondrial fraction) and was dissolved in 2 x 1 mL buffer (0.1 M Tris-HCl pH 7.4, 15% glycerol, 2 protease inhibitor cocktail tablets) and stored at -80°C. Using 4 ultracentrifuge tubes, the supernatant was aliquoted and the tubes were filled to ¾ their volume. The supernatant was then centrifuged for a final time using an ultracentrifuge (Beckmann L8-M Ultracentrifuge, SW Ti 45 rotor, 60 minutes, 4°C, 151,457 x *g*) to separate the microsomal fraction (pellet) and the cytosolic fraction (supernatant). The microsomal fraction was re-dissolved in a total volume of 2.0 mL microsomal buffer and

was divided into 200 µL aliquots that were stored at -80°C. A biological replicate was generated using a further three rat brains with a net total wet weight of 5.02g. The rat brains were combined and processed in the same fashion as the first set of three rat brains.

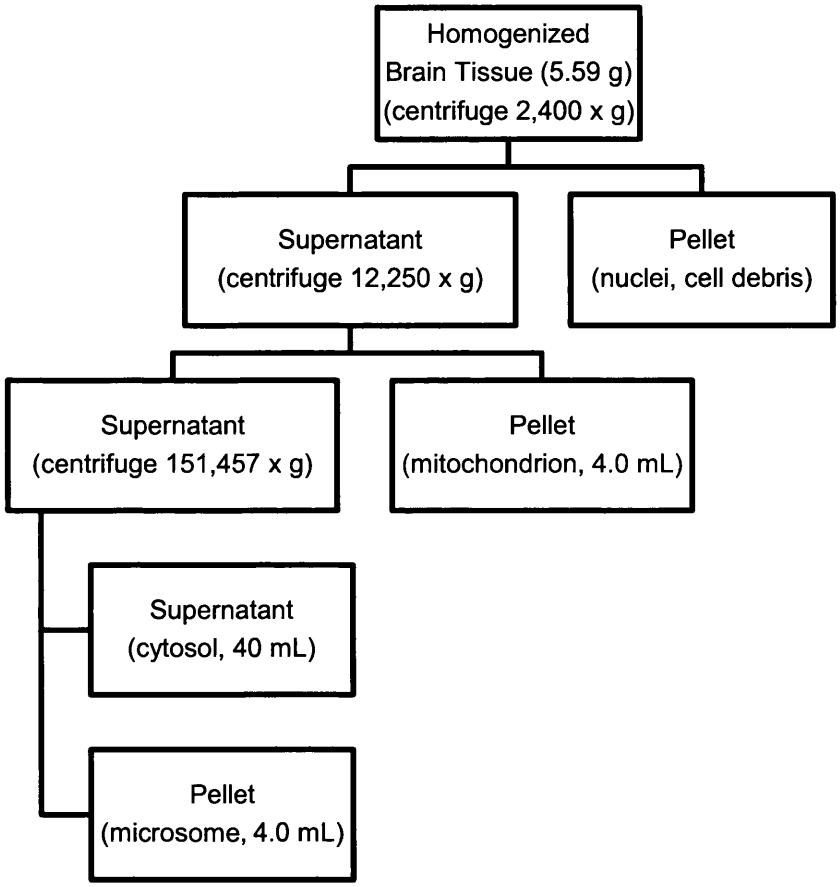


Figure 3.4: Centrifugation steps for the cellular fractionation of rat brain.

3.3.3 Protein Assay

The protein concentration in each fraction was determined by using a Bradford Assay. A stock solution of Coomassie Brilliant Blue G-250 was prepared by dissolving 30 mg in 100 mL of absolute ethanol and 50 mL of concentrated phosphoric acid (10 M). This was diluted to a final volume of 1000 mL with ethanol and then filtered (0.22 micron, Millipore, Billerica, MA, USA).

Using a stock solution of BSA (400 µg/mL) and H_2O as a solvent, the following dilutions of BSA were made up to a volume of 1 mL: 5, 10, 15, 20, 25, 30, 40, 50, 60 µg/mL.

Calibration of the UV spectrophotometer (Perkin Elmer, Waltham, MA, USA) at λ = 595 nm was performed by mixing 100 µL of the BSA standard with 1 mL Bradford reagent. From the calibration plot generated, a slope was determined for the linear part of the calibration curve. This, in turn, was used to determine the approximate concentration of the samples.

Samples were prepared by diluting with H_2O to a volume of 100 µL and mixed with 1 mL Bradford reagent before their absorbance's were read at λ = 595 nm.

Calculation for the concentration of the samples was as follows:

$$\frac{A_{595}}{slope\ of\ curve} \times dilution\ factor = protein\ concentration$$

3.3.4 SDS-PAGE

Using a general protocol, each subcellular fraction underwent another stage of separation using NuPAGE® Novex Pre-Cast Gel System (Invitrogen, Paisley, UK). A maximum volume of 25 µg protein could be loaded onto the gels. Sample preparation for 25 µg protein involved denaturation and reduction using 2.0 µL Invitrogen NuPage® Reducing Agent (500 mM DDT), 5.0 µL Invitrogen NuPage® Sample Buffer (4X) and HPLC grade water to make the sample up to 18 µL. Each sample was incubated at 70°C for 10 minutes. A solution of 200 mM iodoacetamide (Sigma-Aldrich, Dorset, UK) was prepared just prior to use. A total volume of 2 µL of 200 mM iodoacetamide was added to each sample to make a final volume of 20 µL. The samples were incubated in darkness for 60 minutes to encourage alkylation of cysteine residues to prevent the re-formation of cysteine bonds.

Each sample, with a maximum of 25 µg of protein, was loaded onto 10-well NuPAGE® 4-12% Bis Tris gel. Pre-stained protein molecular weight marker (5 µL) was loaded into a free lane (SeeBlue® Plus2 Pre-Stained Standard, Invitrogen, Paisley, UK) as a reference.

NuPAGE® Novex Pre-Cast Gel System (Invitrogen, Paisley, UK) can use two distinct running buffers, one which gives better resolution for mid-sized proteins (MOPS) and

one for smaller molecular weight proteins (MES). As each sub-cellular fraction contains a mixture of both large and small proteins, MOPS (20X) was prepared as the running buffer.

The gel was run using a constant voltage (200 V) with a starting current of 110-115 mA and a finishing current of 60-70 mA. Normal running time for gels was between 50-60 minutes.

Staining was performed using the general microwave staining protocol developed by Invitrogen™. The gel was placed in 100 mL of H_2O and was loosely covered and put into the microwave on high for 45 seconds. The gel was shaken for 1 minute on a gel rocker and the water was discarded. This was repeated twice. Once the rinsing steps were completed, 20 mL of Invitrogen™ SimplyBlue™ SafeStain was added and the gel was once more microwaved for 50 seconds. The gel was washed multiple times in 100 mL H_2O until the background was clear and the gel bands distinct. A scan of the gel was made and retained prior to in-gel digestion.

3.3.5 In-Gel Digest and Extraction

Bands were excised and cut into approximately 1 mm^3 pieces which were placed into 1.5 mL siliconised tubes. Destaining of the bands involved the addition of 500 µL of destaining solution (50:50 50 mM ammonium bicarbonate: ethanol) and placing the tubes on the vortex for 30 minutes. Samples were centrifuged and the subsequent process was repeated until the gel pieces were clear. The clear pieces were dried by adding 100 µL of 100% acetonitrile, vortexing 10 minutes, discarding the supernatant and repeating the process a minimum of a second time before the samples were placed onto the heated vacuum concentrator (Jouan, Buckinghamshire, UK) for 30 minutes at 40°C to complete the drying process.

Sequencing grade modified was prepared by removing a 20 µg solid aliquot from the freezer and dissolving it in 100 µL re-suspension buffer (Promega, Southampton, UK). While the re-suspension buffer was acidic to prevent trypsin from undergoing autolysis, it was also kept on ice to prevent any further degradation. The trypsin (0.2 µg/µL) was further aliquoted out into 5 µL aliquots which were stored in the -20°C freezer until they were required. The aliquot was then buffered to its active pH by adding 20 µL 50 mM

ammonium bicarbonate to it. A total volume of 5-10 µL (0.1 – 0.2 µg trypsin) was added to each 1.5 mL tube containing dry and brittle gel pieces to allow for maximum absorption of the tryptic solution. The samples were kept at 4°C for 30 minutes before a final 20 µL of 50 mM ammonium bicarbonate (pH 8.0) was added onto the gel pieces to prevent drying out. The samples were placed in an incubator at 37°C overnight.

Following digestion, samples were removed from the incubator and 25 µL extraction solvent (5% FA in 50:50 ACN/H_2O) was added to each individual tube. The tubes were placed on the vortex for 10 minutes followed by a quick centrifugation on a benchtop centrifuge (Heraes Instruments, Fresco Biofuge, 4°C, 13,000 x *g*, 1 minute) to pellet the gel pieces to the bottom. The supernatant was transferred to new 1.5 mL siliconised tubes and kept. To improve the extraction process, 30 µL of 100% ACN was added to gel-containing tubes to remove more peptides. The samples were placed on the vortex for 5 minutes before centrifuging and transferring the supernatant to their corresponding tubes. The process was repeated a further 2 times to ensure that most of the peptides had been extracted. Once completed, the samples (pooled supernatants) were dried (SpeedVac, Jouan, 40°C) to remove all the organic solvent.

Re-suspension of the digest was in 0.1% FA in H_2O. Prior to sample analysis, each digest was placed on the vortex and centrifuged on a benchtop centrifuge (Heraeus Instruments Biofuge Pico, 13,000 x *g*, 1 minute) to remove any particulate before transferring to polypropylene autosampler vials.

3.3.6 Capillary LC-MS/MS Data Dependent Acquisition

All samples for protein identification via MS/MS analysis were dissolved in 25 µL of 0.1% FA and 5 µL were injected onto the Waters CapLC system coupled to the front end of the QTOF Global Ultima™. The Waters CapLC comprises an autosampler and a LC-pump system that is connected directly to the mass spectrometer through a switching valve. The total run time for one injection was 63.0 minutes. As the sample was injected through the sample loop, it was pushed through to the pre-column (5 mm x 0.3 mm 5 µm, PepMap C18 Guard Column, Dionex, Camberley, UK) by pump C (0.1% FA) (Table 3.1: Auxiliary Pump C Curve) where salt and other small molecules were removed.

Table 3.1: CapLC Auxiliary C Flow Timetable and Stream Select Events.

Time	Flow µl/min	Stream Select Position
0.10	15.0	Position 1
3.00	15.0	Position 2
3.10	1.0	Position 2
53.0	1.0	Position 1

Once desalted (3.0 minutes), the stream select valve (Figure 3.5) switched to allow pumps A (95% H_2O, 5% ACN, 0.1% FA) and B (95% ACN, 5% H_2O, 0.1% FA) to flow through the pre-column.

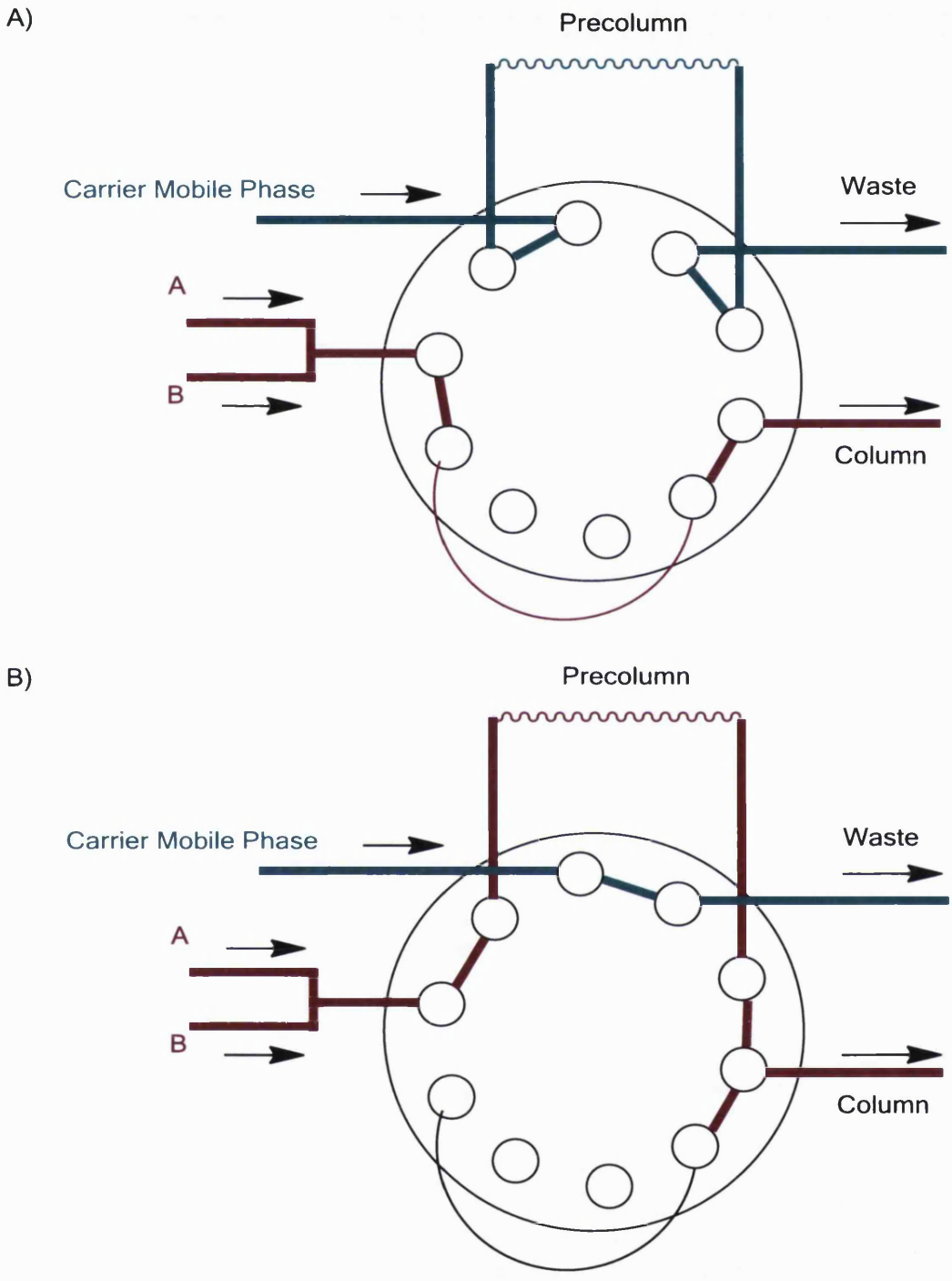


Figure 3.5: Stream Select Switching Valves. (A) shows the valves in the initial position where the flow from the carrier mobile phase goes through the pre-column and goes to waste to desalt. After three minutes, the stream select valve switches over (B) and the flow from both A and B are mixed together and go through the pre-column and into the column to be finally eluted into the mass spectrometer.

Upon switching the valve, the flow from the pre-column moved through to the nano-column (150 x 0.075 (i.d) mm, 3 µm C18, Dionex) and into the mass spectrometer. As the gradient was formed through the mixing T with mobile phases A and B, (Table 3.2, Figure 3.6) the peptides were eluted from the columns.

Table 3.2: Waters CapLC Gradient Timetable.

Time	Percent A	Percent B	Flow
0.10	95.0	5.0	6.0
3.00	95.0	5.0	6.0
40.0	72.0	28.0	6.0
49.0	20.0	80.0	6.0
52.0	20.0	80.0	6.0
53.0	95.0	5.0	6.0
63.0	95.0	5.0	6.0

The gradient formed by A and B contains less than 50% organic solvent for most of the time as this is optimal for the elution of most of the peptides from the column. Initial flow from the CapLC is at 6 µl/min, which is split prior to the column so the flow through the column to the mass spectrometer is only 200 nL/min.

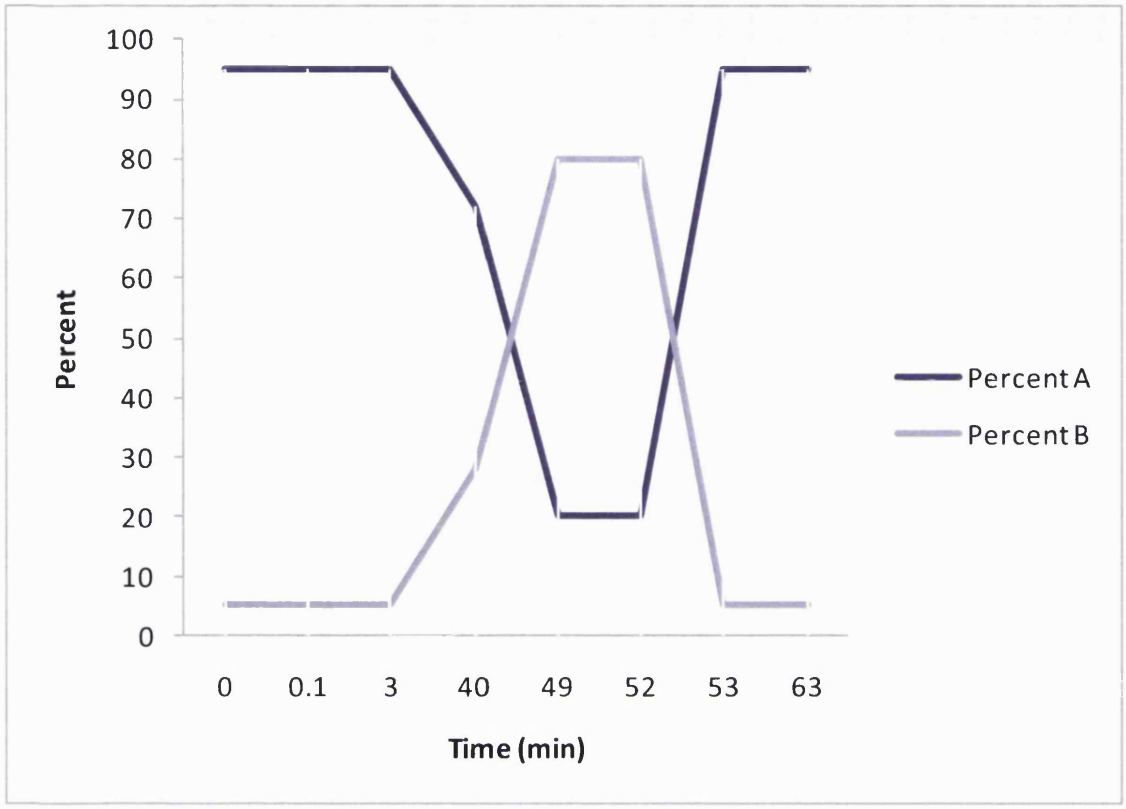


Figure 3.6: Representation of the gradient formed between mobile phase A and mobile phase B.

Peptides eluted off the column were directly sprayed into the mass spectrometer for analysis.

3.3.7 Mass Spectrometry (MS/MS)

All data were acquired using the Waters QTOF Global Ultima™ using a nanospray source for increased sensitivity which allows for less sample to be injected. The acquisition and processing software used was Waters MassLynx Version 4.0.

Samples were run using two different approaches for MS/MS analysis. The first used, (method 1) acquired data over a full *m/z* range with MS/MS analysis. The second approach (method 2) involved setting up three methods to analyse the full *m/z* range via three smaller *m/z* ranges. Four methods prepared to cover the smaller *m/z* ranges were *m/z* 400-605, 600-805 and 800-1000. The second approach, although it allowed for a lot more peptides to be analysed, took much longer to cover the same m/z range used in the first instance.

Table 3.3: Parameters for the analysis of samples by the QTOF (Method 1).

Waters QTOF Global	Parameters	
ACQUISITION	Survey Start Time	3.0
	Survey Ends Time	60.0
	Survey Ion Mode	ES Mode
	Survey Polarity	Positive
MS SURVEY	Survey Start Mass	400.0
	Survey End Mass	1600.0
	Intensity Threshold	10
	Survey Scan Time	1.0 seconds
	Survey Interscan Time	0.1 seconds
	Survey Data Format	Continuum
	Survey Use Tune Page CV	Yes
	Survey Cone Voltage	35.0 V
MS/MS	MSMS Start Mass	50.0
	MSMS End Mass	1700.0
	Number of Components	4
	MSMS to MS Switch Criteria	Intensity Falling Below Threshold
	Switchback Threshold	3.0 counts/second
	Use MSMS to MS Switch After Time	YES
	MSMS Switch After Time	3.3 seconds
	MSMS Scan Time	1.0 seconds
	MSMS Interscan Time	0.1 seconds
	MSMS Data Format	Continuum
	Use Tune Page Cone Voltage	YES
	MSMS Cone Voltage	35.0 V
	Use MS/MS ipr File	NO
PEAK DETECTION	Peak Detection Window	1.0
	Use Include by Charge State	YES
	Charge State(s)	2,3
	Number of Include Components	60
	Charge State Tolerance Window	3.0
	Charge State Extraction Window	2.0
	Discard Survey Data	NO
COLLISION ENERGY	Use Charge State Recognition	YES
	Maximum Charge State	4
	Charge State 1 Filename	Default_CS_1*
	Charge State 2 Filename	Default_CS_2*
	Charge State 3 Filename	Default_CS_3*
	Charge State 4 Filename	Default_CS_4*
EXCLUDE	Detected Precursor Inclusion	Using Real Time Exclusion
	Detected Precursor Inclusion	Include After Time
	Include after Time	60.0 seconds
	Use Exclude Mass List	NO
	Exclude Window	+/- 1500.0 mDa
	Exclude Retention Time Window	10 seconds

*Charge state files used for charge state recognition can be found in Appendix B.

Table 3.3 shows the general experiment setup for method 1. Differences between method 1 and method 2 were in the Survey Start Mass and Survey End Mass. In the general method (method 1), the survey start mass and survey end mass were as listed. For method (2), they were varied as follows: 400-605, 600-805, 800-1000. Spectra were acquired in MS mode and up to 4 individual peptides could be fragmented by MS/MS per MS scan. The mass spectrometer would fragment the multiply charged masses once their intensity reached 10 counts/scan or above and would fragment for a

total of 3.3 seconds or until the intensity fell below 3.0 counts/scan. That particular mass and a window of 1.5 Da around it were excluded for 60 seconds, allowing the mass spectrometer to fragment as many different components as possible during the run time.

3.3.8 Calibration

Prior to analysis, Glu-Fibrinopeptide (Sigma-Aldrich) was used for calibrating the mass spectrometer and to ensure that sensitivity was optimal. A concentration of 100 fmol/µL Glu-Fib (peptide sequence EGVNDNEEGFFSAR) was infused. A calibration file was prepared by fragmenting the $[M+2H]^{2H+}$ ion of Glu-Fib. The ion fragments were processed and compared to a theoretical fragment ion peak list for calibration. To ensure the LC-MS system was optimising properly, 100 fmol of bovine serum albumin, pre-digested, was injected into the LC-MS and the base peak chromatogram generated to ensure satisfactory performance (i.e. resolution, elution times and sensitivity).

3.3.9 Data Processing

The resulting raw data was processed using MassLynx ProteinLynx V4.0. The system was setup using Peptide Auto and the parameters used were from the file Bill_Dec10.mlp. Processing parameters included combining all sequential scans with the same precursor and processing all combined scans. Mass measurment of the combined scans involved spectral smoothing using Savitzky Golay method with a 3.00 channel window. The smoothing was performed twice. A centroid peak list was then created using the top 80.00% of the peak with a minimum peak width at half height of 4. Once the data had been processed, it was then combined into a single pkl file which could then be used for searching against databases. PKL is an extension for a text file created by Masslynx which lists all the MS data (*m/z* and charge) and tandem mass spectrometry data associated with that *m/z* and charge. The data was then searched against MASCOT software installed locally.

3.3.10 MASCOT

MASCOT (Matrix Science, London, UK), a probability based search engine, was used to search all the data. It can use any available database which is in FASTA format.

The fundamental basis of MASCOT is calculation of the probabilities that an observed match between an experimental spectrum and a theoretical spectrum from a sequence entry is a random event. In MASCOT the probability that an observed match is a chance event is calculated based on ($P<0.05$), although the probability is listed as a score which is calculated by -10log(P). In this case, the lower the probability (of a random match), the higher the score should be. The match with the lowest probability to be a random event is the best match however the significance of that match depends on the size of the used database of theoretical spectra. Mascot can take into account missed cleavage sites as well as post-translational modifications. One of the most important values that must be manually entered is the *m/z* error window as too large a window increases the randomness and too small a window misses valid matches.[1]

Using the international protein index (IPI), the data files were searched using the following parameters:

Database: IPI_RAT	Taxonomy: none
Enzyme: trypsin	Maximum missed cleavages: 1
Fixed modifications: none	Variable modifications: -Oxidation (Met) -Carbamidomethylation (Cys) -Acetyl (Protein N-term)
Peptide Tolerance: +/- 0.3 Da	MS/MS tolerance: +/- 0.3 Da
Peptide charge: Mr	Report top: Auto hits

A decoy database of the IPI rat was used to determine the false-positive rate of identification. The decoy database contains the forward-normal sequences but with the amino acid sequences reversed. The parameters were the same as the original search parameters. A requirement of a less than 5% false-positive rate was required for each analysis.

3.3.11 Proxeon

Once the searches were performed and the data was saved, text files of the IPI results were created for searching using Protein Center (Proxeon Bioinformatics, Odense, Denmark). The identified proteins are then converted into gene symbols and annotated using the universal gene ontology (GO). These genes were then divided into molecular categories to determine their functions and their locations.

3.4 Results

The Bradford Assay detects the binding of Coomassie Brilliant Blue G-250 to the amino groups of lysine and arginine residues and it can be used to determine an approximate concentration of the proteins present in the sample. The calibration for the Bradford Assay using BSA as a standard shows a linear relation which was then used to determine the protein concentration in the samples.

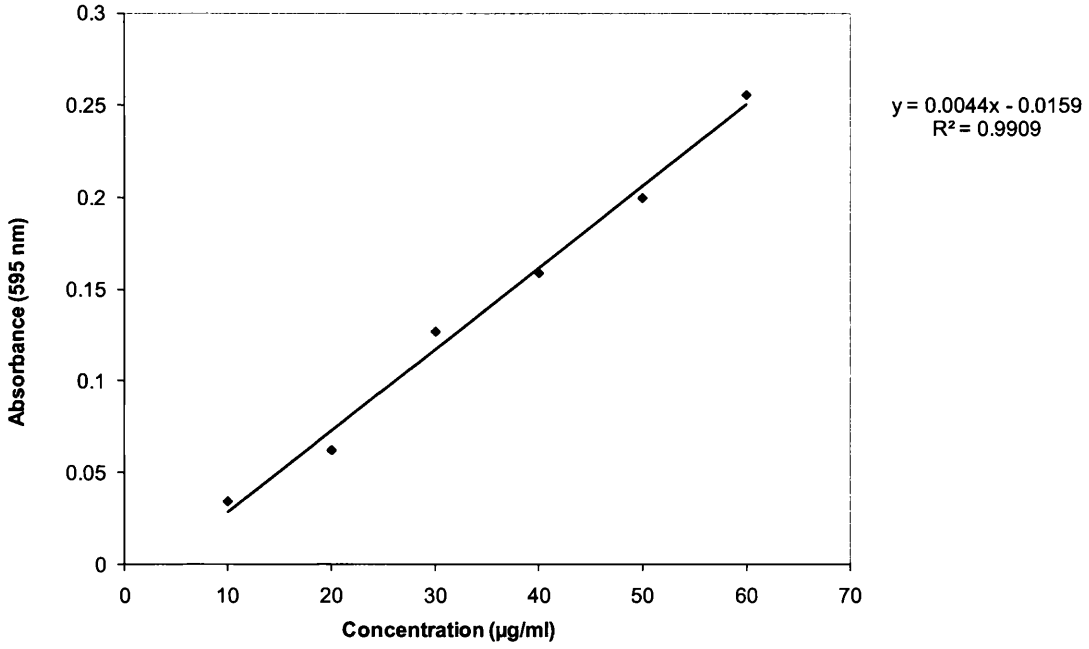


Figure 3.7: Protein Assay Curve for determination of protein content in cellular fractions.

Above a concentration of 70 µg/mL, the relation of the BSA absorbance to concentration is not linear. All the samples were analysed multiple times to ensure that their absorbance fell within the linear portion of this calibration curve.

The approximate total concentration of proteins in the samples and the amount of protein loaded onto each gel lane as determined by Bradford assays are given (Table 3.4, Table 3.6).

Table 3.4: Brain Sample 1: Total mass of 3 rat brains: 5.59 g, Sprague-Dawley female rats (7-8 weeks old, mature) were killed by CO. Weights were taken at time of autopsy. Total extracted protein from the combined brains: 238.6 µg.

Subcellular Fraction	Protein Content µg/µL	Volume of Sample / mL	Total Protein Content / µg	Total Protein Loaded onto Gel 1 / µg	Total Protein Loaded onto Gel 2 / µg
Microsome	3.57	4.0	14.28	62.83	26.53
Mitochondrion	6.98	4.0	27.92	122.84	27.86
Cytosol	4.91	40	196.4	86.42	24.48

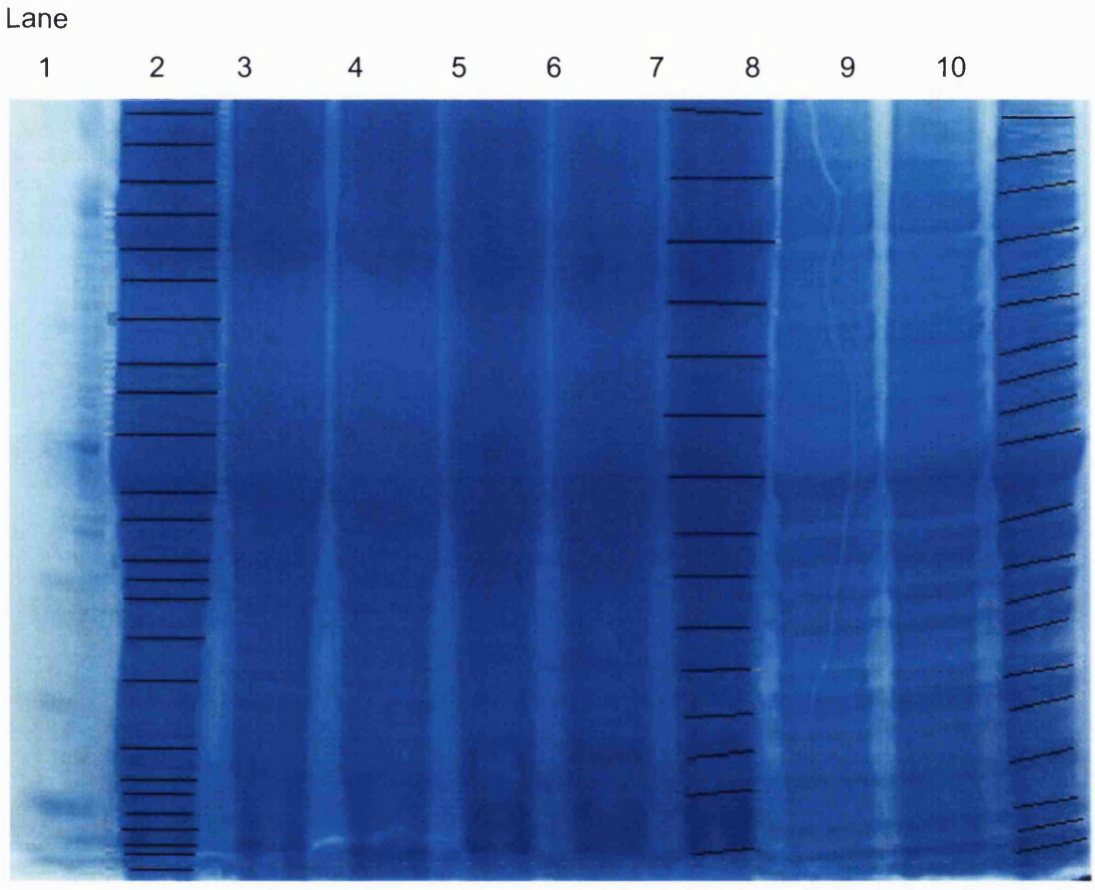
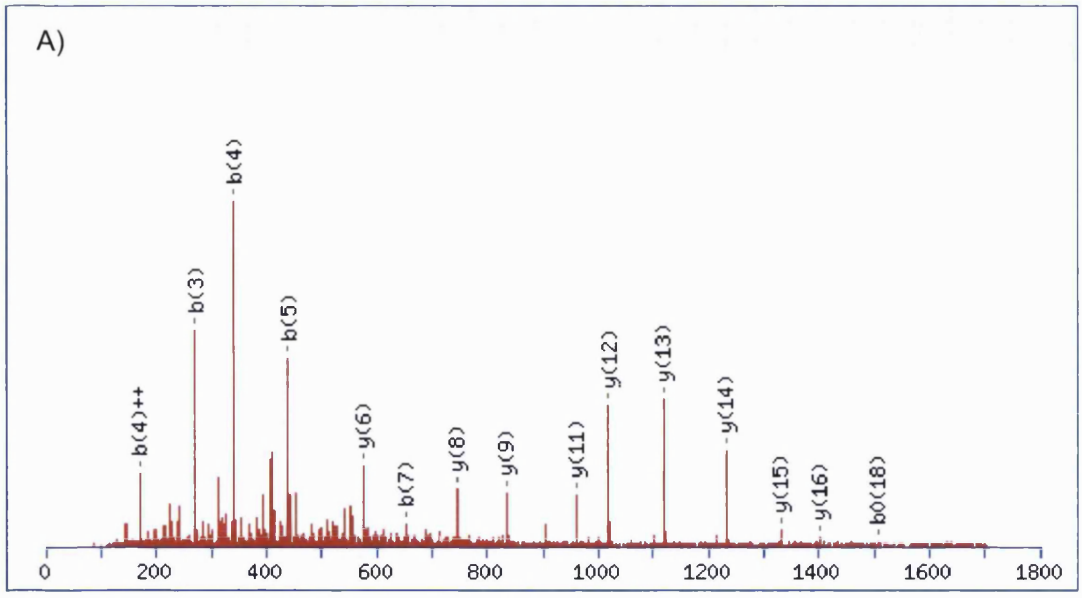


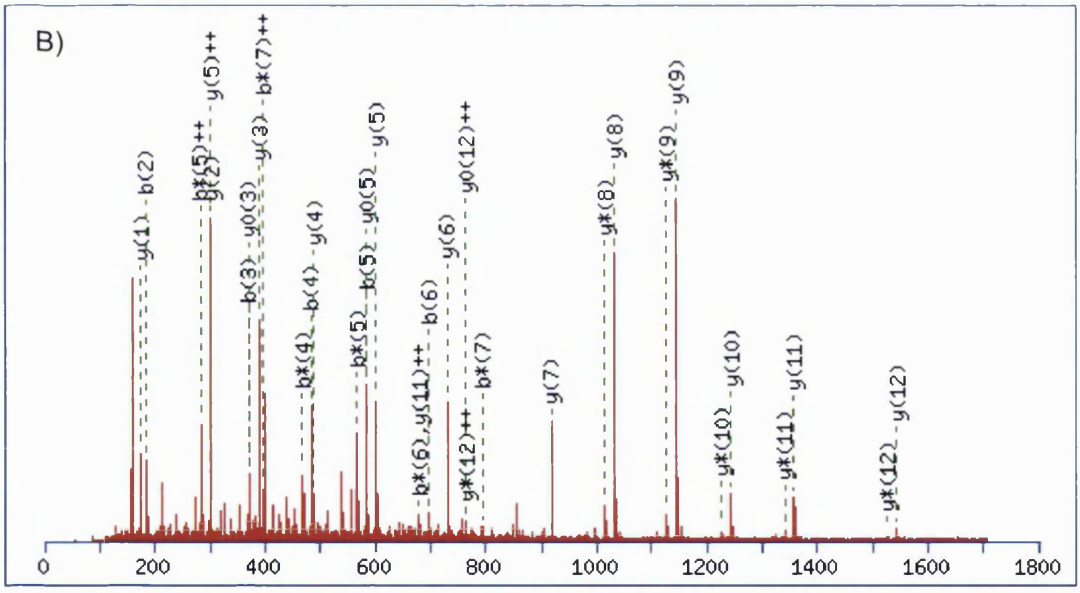
Figure 3.8: SDS-PAGE-1 of 3 subcellular fractions of the rat brain: Lane 1 – Molecular weight marker, Lane 2-4 microsomal fraction, Lane 5-7 mitochondrial fraction, Lane 8-10 cytosolic fraction.

Using the values in (Table 3.4), the maximum volume to be loaded onto each lane for each fraction was calculated and a gel was run (Figure 3.8). An entire lane per fraction was cut and digested. Lanes 2, 7 and 10 were cut, digested with trypsin and extracted for LC-MS/MS protein identification.

The list of identified proteins from each band and lane were scrutinised and considered as potential proteins related to the synthesis of cholesterol, the formation of steroids in the brain or any transporters that are directly involved in the movement of cholesterol or its derivatives. From the list of 1385 individual IPI numbers identified, 41% (574) were found in more than one fraction and 33% (453) of the total proteins identified were identified using only one peptide. Of this list, 13 proteins - less than 1% of the total number of proteins identified - were identified as steroid specific proteins (Table 3.5). All proteins were screened manually and those that only had one peptide associated were scrutinised to ensure that the MS/MS data was viable. Figure 3.9 gives examples of MS/MS data of two peptides assigned to two different proteins; A)3-hydroxyacyl-CoA dehydrogenase type II and B) Superoxide dismutase [Mn], mitochondrial precursor.



3-hydroxyacyl-CoA dehydrogenase type II



Superoxide dismutase [Mn], mitochondrial precursor

Figure 3.9: Fragmentation patterns of A) a peptide from 3-hydroxyacyl-CoA dehydrogenase type II and B) Superoxide dismutase [Mn], mitochondrial precursor.

Table 3.5: Steroid proteins identified from the digestion of the microsomal, mitochondrial and cytosolic gel lanes. Gel 1, method 1.

IPI	Protein Name	Subcellular Fraction Location	Score	MW	Peptides	Coverage %	Confirm	Same Peptide Found
IPI00325136	Nonspecific lipid-transfer protein, mitochondrial precursor	Microsome Mitochondrial	103	58775	2	3.3	Yes	Yes
IPI00212523	Protein DJ-1	Microsome Cytosol	50	19961	1	7.4	Yes	Yes
IPI00189766	Progesterone receptor membrane component 1	Microsome	43	24649	1	3.6	Yes	No
IPI00370458	ATP-binding cassette sub-family F	Microsome	59	79806	1	1.8	No	No
IPI00208598	Vigilin	Microsome	156	141496	3	2.8	Yes	No
IPI00192286	ATP-binding cassette sub-family A member 2	Microsome	61	270755	1	0.6	No	No
IPI00369995	PREDICTED: similar to lipoprotein receptor-related protein	Microsome	887	504558	20	3.8	Yes	Yes
IPI00231253	3-hydroxyacyl-CoA dehydrogenase type 2	Mitochondrion	166	27098	3	11.9	Yes	Yes
IPI00190557	Prohibitin-2	Mitochondrion	199	33292	3	10.4	Yes	Yes
IPI00203310	Glutathione S-transerase Yb4	Cytosol	92	25627	2	7.3	Yes	No
IPI00197703	Apolipoprotein A-I precursor	Cytosol	48	30069	1	2.7	No	No
IPI00190701	Apolipoprotein E	Cytosol Microsome	269	35731	6	17.3	No	No
IPI00188158	Hydroxymethylglutaryl-CoA synthase, cytoplasmic	Cytosol	42	57397	1	1.7	No	No
IPI00200359	Sulfotransferase 4A1	Mitochondrion	44	33032	1	3.9	No	No
IPI00203310	Glutathione S-transferase Yb4	Cytosol	92		3	7.3	No	No
IPI00373197	Membrane associated progesterone receptor component 2	Microsome	92	23389	3	10.6	No	No
IPI00230942	Glutathione S-transferase Yb-3	Mitochondrion	153	25533	3	11.5	Yes	Yes

Once the list of interesting proteins was generated, the same band in a technical replicate lane (Figure 3.8, lanes 3, 6 and 9) was digested as well as the bands above and below it to try and identify the same protein again. Table 3.5 lists the outcome and demonstrated the repeatability.

A second sample resulting from a pool of three new rat brains (Brain sample 2) was prepared and further gels (Figure 3.10: gel 2: Brain sample 1 and gel 3: Brain sample 2) were run with the maximum allowed protein amount loaded onto the gel (approximately 25 µg). By loading the maximum resolving capacity on the gel, the resolution between most of the gel bands was maintained. Table 3.4 lists the amount loaded onto gel 2 from brain sample 1 while Table 3.6 lists the amount loaded onto gel 3 from brain sample 2.

Table 3.6: Brain Sample 2. 3 rat brains combined for a total mass of 5.02 g, 7-8 week old female Sprague-Dawley rats killed by CO. Weights taken at time of autopsy.

Fraction	Protein Content µg/µL	Volume of Sample mL	Total Protein Content µg	Total Protein Loaded onto Gel µg
Microsome	7.34	4.0	29.36	24.87
Mitochondrion	15.24	4.0	60.96	23.95
Cytosol	2.90	40	116	24.99

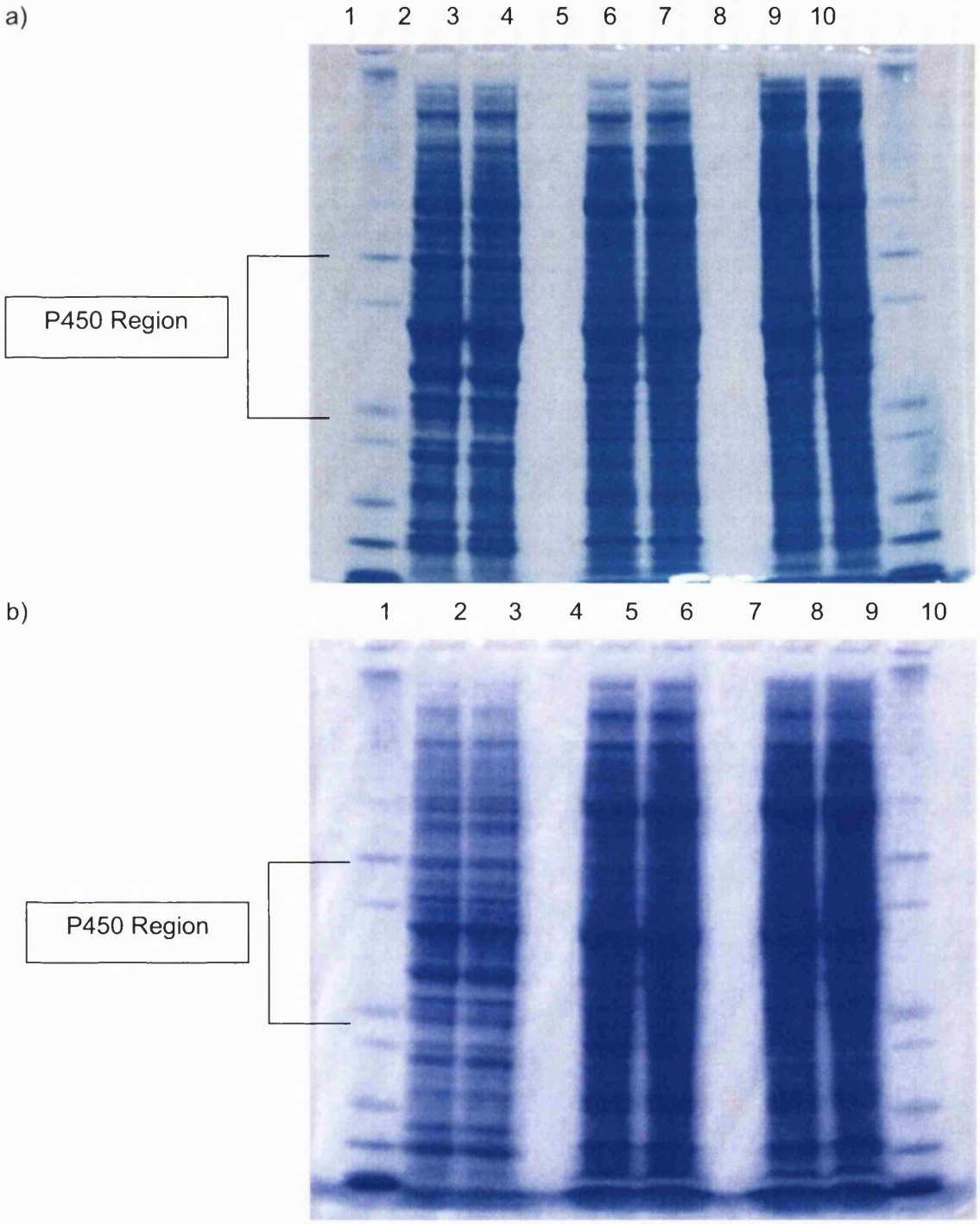


Figure 3.10: SDS-PAGE-2 and -3 of brain subcellular fractions. a) Brain sample 1 b) Brain sample 2. Lane 1 and 10: molecular weight marker, lane 2,3 – cytosolic fraction, lane 5,6 – mitochondrial fraction, lane 8,9 – microsomal fraction.

Using the molecular weight markers to approximate location, areas of the gel 2 and gel 3 were cut to correspond with the position of the identified proteins from gel 1. These bands were digested and run on the QTOF to investigate if the proteins could be identified again.

Table 3.7: Confirmation of protein identification from gel 2 and gel 3.

IPI	Protein Name	Subcellular Fraction Location	Confirm
IPI00325136	Nonspecific lipid-transfer protein, mitochondrial precursor	Microsome Mitochondrion	Yes
IPI00212523	Protein DJ-1	Microsome Cytosol	Yes
IPI00189766	Progesterone receptor membrane component 1	Microsome	Yes
IPI00370458	ATP-binding cassette sub-family F	Microsome	No
IPI00208598	Vigilin	Microsome	Yes
IPI00192286	ATP-binding cassette sub-family A member 2	Microsome	Yes*
IPI00369995	PREDICTED: similar to lipoprotein receptor-related protein	Microsome	Yes
IPI00231253	3-hydroxyacyl-CoA dehydrogenase type 2	Mitochondrion	Yes
IPI00190557	Prohibitin-2	Mitochondrion	Yes
IPI00230942	Glutathione S-transferase Yb-3	Mitochondrion	Yes
IPI00197703	Apolipoprotein A-I precursor	Cytosol	No
IPI00190701	Apolipoprotein E	Cytosol Microsome	Yes
IPI00188158	Hydroxymethylglutaryl-CoA synthase, cytoplasmic	Cytosol	No
IPI00200359	Sulfotransferase 4A1	Mitochondrion	No
IPI00203310	Glutathione S-transferase Yb4	Cytosol	No
IPI00373197	Membrane associated progesterone receptor component 2	Microsome	No

* Same peptide was found as that seen from results in Table 3.5.

Of all the proteins identified in gel 1, six were not identified a second time from gel 2 or gel 3. More interestingly were the few extra cholesterol-related proteins that were identified (Table 3.8).

Table 3.8: Proteins identified from the gel 2 and gel 3 that were not previously identified.

IPI Number	Protein Name	Subcellular Fraction Location	Score	MW	Number of Peptides	% Coverage	Confirmed
IPI00231860	ATP-binding cassette sub-family D member 3	Microsome	52	75136	1	2.1	YES
IPI00326948	Hsd17b4 protein	Microsome	40	81038	1	1.7	NO
IPI00196987	Beta-adrenergic receptor kinase 1	Microsome Cytosol	46	79734	2	1.3	NO
IPI00560892	Hadha protein	Mitochondrion	202	82613	3	5.3	NO

Important enzymes involved in steroid synthesis are the cytochromes P450. In order to increase our ability to identify some of these proteins, the P450 region was cut out of the gel and digested from gel 1 and gel 3. This was defined as the region between BSA (66.3 kDa) and Lactate dehydrogenase (36.5 kDa). These regions were analysed by mass spectrometry using 3 different *m/z* ranges per gel band. These ranges were *m/z* 400-605, 600-805 and 800-1000. The data from each gel lane was combined and processed using MASCOT. Table 3.9 lists the proteins involved with cholesterol and it's pathways that were identified from the P450 region of gels 1 and gel 3; microsomal and mitochondrial lanes.

Table 3.9: Proteins identified from the P450 region of gels 1 and 3 mitochondrial and microsomal fractions with functions related to sterols or steroids.

IPI Number	Protein Name	Subcellular Fraction Location	Score	MW	Number of Peptides	% Coverage	Confirmed
IPI00195129	PREDICTED: similar to Cytochrome p450 46A1	Microsome	65	56663	1	2.4	NO
IPI00231860	ATP-binding cassette sub-family D member 3	Microsome	52	75136	1	2.1	YES
IPI00193816	PREDICTED: similar to ATP-binding cassette sub-family E member 1	Microsome	47	67257	1	2.7	YES*
IPI00196988	Beta-adrenergic receptor kinase 2	Mitochondrion	78	79835	3	4.2	YES*
IPI00231200	NADPH--cytochrome P450 reductase	Microsome	97	76783	3	4.3	YES*
IPI00326948	Hsd17b4 protein	Microsome	88	81038	2	3.5	YES

* proteins that were previously identified and so are seen for the second time.

From all the data that was obtained from all the gels run, a final list of identified proteins was prepared. In total, 22 proteins were identified as interesting proteins involved in the synthesis of steroids. Their names, best scores and coverage are listed in Table 3.10.

Table 3.10: Final list of steroid proteins derived from all in-gel digestion and analysis performed.

IPI Number	Protein	Subcellular Fraction Location	Protein Score	Protein Mass	Protein Coverage	Peptides Matching	Confirmed Protein
IPI00325136	Nonspecific lipid-transfer protein, mitochondrial precursor	Microsome Mitochondrion	117	58775	3.5	2	Yes
IPI00370458	ATP-binding cassette, sub-family F	Microsome	59	79806	1.8	1	No
IPI00208598	Vigilin	Microsome	156	141496	2.8	3	Yes
IPI00192286	ATP-binding cassette sub-family A member 2	Microsome	61	270755	0.6	1	No
IPI00200359	Sulfotransferase 4A1	Mitochondrion	44	33032	3.9	1	No
IPI00203310	Glutathione S-transferase Yb4	Cytosol	92	25627	7.3	3	No
IPI00197703	Apolipoprotein A-I precursor	Cytosol	48	30069	2.7	1	No
IPI00190701	Apolipoprotein E precursor	Cytosol Microsome	269	35731	17.3	6	No
IPI00373197	Membrane associated progesterone receptor component 2	Microsome	92	23389	10.6	3	No
IPI00189766	progesterone receptor membrane component 1	Microsome	108	24649	10.3	4	Yes
IPI00231200	NADPH--cytochrome P450 reductase	Microsome	155	76783	5.9	3	Yes
IPI00230942	Glutathione S-transferase Yb-3	Cytosol	300	25533	29.4	5	Yes
IPI00212523	Protein DJ-1	Microsome Cytosol	241	19961	38.1	7	Yes
IPI00231253	3-hydroxyacyl-CoA dehydrogenase type-2	Mitochondrion	178	27098	18.8	3	Yes
IPI00190557	Prohibitin-2	Mitochondrion	425	33292	26.4	8	Yes
IPI00369995	PREDICTED: similar to lipoprotein receptor-related protein	Microsome	1099	504558	15.1	26	Yes
IPI00193816	PREDICTED: similar to ATP-binding cassette sub-family E member 1	Microsome	47	67257	2.7	1	No
IPI00196988	Beta-adrenergic receptor kinase 2	Mitochondrion	78	79835	4.2	3	No
IPI00231860	ATP-binding cassette sub-family D member 3	Microsome	52	75136	2.1	1	No
IPI00326948	Hsd17b4 protein	Microsome	88	81038	3.5	2	Yes
IPI00188158	Hydroxymethylglutaryl-CoA synthase, cytoplasmic	Cytosol	42	57397	1	1.7	No
IPI00195129	PREDICTED: similar to Cytochrome p450 46A1	Microsome	65	56663	2.4	1	No

Using all the data gathered, a final list of unique IPI numbers was created (see Master List) which included all LC-MS/MS run to date. Proxeon was used to sort the data and give the final resulting list of unique IPI numbers. Proxeon allowed for the GO molecular functions of the master list to be determined. Overall, most of the proteins identified were membrane proteins or involved in protein complexes (Figure 3.11). There were more mitochondrial proteins identified than cytosolic proteins and there was a large proportion with no annotation or classification.

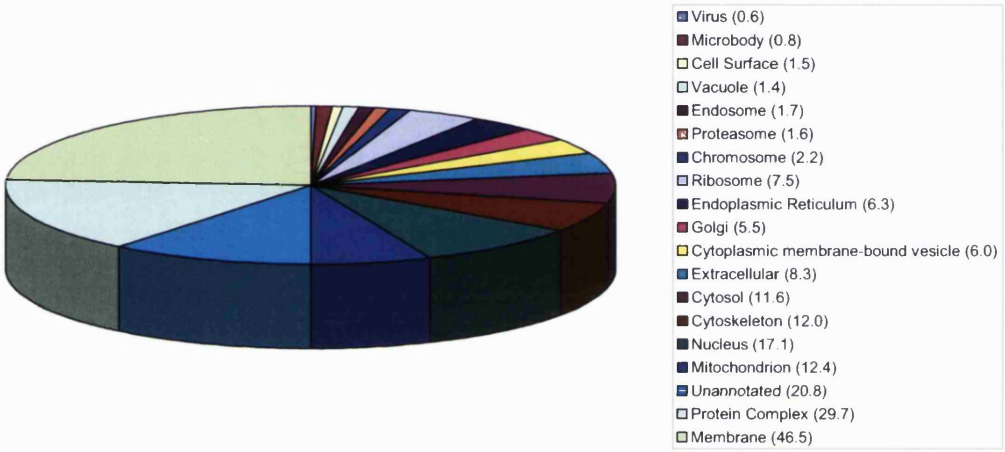


Figure 3.11: Diagram representation in percent of the GO molecular Functions of identified proteins in the combined data set.

In total, there were 2698 records that were searched for function via Proxeon. Of these, 1838 were unique IPI numbers; 855 of these were unique proteins and the rest were either repeats within the database or isoforms of each other. The comprehensive list of proteins can be found in Appendix B.

Table 3.11: Total proteins identified in replicate analyses.

Fraction	IPI Numbers Submitted	Unique IPI Numbers Identified
Mitochondrion	492	491
Microsome	1446	1408
Cytosol	760	756
All	2698	1838

3.5 Discussion

The aim of this study was to look at three different cellular fractions of the whole brain tissue homogenate and identify all the proteins possible (see Appendix B for entire list of proteins identified) and then identify those of interest which relate to cholesterol and other steroids which are found in the brain.

The method chosen included subcellular fractionation into cytosolic, microsomal and mitochondrial fractions, while discarding the nuclear and cellular debris. Samples were run in triplicate on an SDS-PAGE from which one lane from each subcellular fraction was cut and digested to produce a total of 77 bands (including one lane of each subcellular fraction) which were analysed using LC-MS/MS methods. All runs were carried out using C_{18} nano column on the capillary HPLC in-line with a mass spectrometer.

Three female rat brains were homogenised together and pooled for analysis. A maximum volume possible (25 µL) was used to prepare the initial gel (Figure 3.8), using the highest volume of the lysates possible. Low abundance proteins tend to be difficult to find in gel bands so high concentrations of protein were loaded on the gel. From the initial subcellular fractionation, 1-DE and then LC-MS/MS analysis, it became very obvious that there were many proteins that were present in high abundance that are not part of the cholesterol metabolism, transport etc. In particular, myelin, tubulin and many ribosomal proteins made it very difficult to identify other proteins (Appendix B). A total of 1385 proteins were identified from the analysis of the first run of three lanes (Figure 3.8). Peptides from the digest were analysed according to general lab protocols. The data was processed by the instrument into text files which could then be searched using MASCOT and IPI-Rat. Using the NCBI database would have resulted in a large number of redundant proteins to be identified which would have required more work post-processing to remove from the protein identification lists. Other databases do not necessarily allow for the searching of just one species, which was possible using IPI-Rat, reducing the length of processing time required per sample. From this first analysis, 17 steroid related proteins were identified (Table 3.5).

Of interest were proteins identified as transporters or receptors for steroids or enzymes involved in their metabolism. Most of the proteins in Table 3.5 are involved in cholesterol synthesis or steroid synthesis from cholesterol. Lipoprotein receptor-related protein 1 (IPI00369995) was found in the highest abundance, according to peptide count, and was

found frequently between the microsomal and mitochondrial fractions. The apolipoproteins were also of great interest for their interaction with low-density lipoprotein and their ability to transport cholesterol through the body. Apolipoprotein E (IPI00190701) is very important for this function in plasma as is apolipoprotein A-I (IPI00197703) in brain, both of which were identified in the sample.[159,161,165,166]

A second pool of three additional female rat brains were homogenised and prepared in a similar fashion to the first pool, to provide a biological replicate. Two more gels were prepared and run, one for brain sample 1 (Figure 3.10) and a second for brain sample 2 (Figure 3.10). Each cellular fraction was run on the SDS-PAGE in duplicate. The second and third gels were prepared and most of the proteins identified from the initial identification (Figure 3.8) were confirmed, with the exception of HMG-CoA synthase, Apolipoprotein A-I and ATP binding cassette subfamily F. These were all found with only one peptide, which may suggest that their abundance in the tissue is quite low and so difficult to identify (Table 3.5 and Table 3.7). Some additional proteins were also identified in this second run, which were then confirmed by the third gel (Table 3.8). These proteins were also found in low abundance and not seen in the first run gel.

At this stage of the research, none of the P450 family had as of yet been found (other than CYP46A1 in a single experiment and by a single peptide, see below). The P450 family is involved in cholesterol synthesis and steroidal synthesis. As they tend to be difficult to identify in most tissue with the exception of the liver, it was noted that this group of proteins needed more attention.

From what remained of the gels, the P450 region was cut up into approximately 5 bands, digested and extracted. These bands underwent LC-MS/MS three times, using narrow bands of *m/z* ranges: 400-605, 600-805, 800-1000 (method 2). Narrowing the ranges allowed the mass spectrometer to focus more on these ranges and would hopefully allow them to find some of the less abundant proteins/peptides that were not already been identified. The data was processed and the files were merged into a single file after being searched by MASCOT. The results (Table 3.9) showed that a few more interesting proteins had been found, one of which was cytochrome P450 reductase (IPI00231200, E.C.1.6.2.4), an important protein that is present with cytochrome P450s to reduce them back to their active state after they have completed an enzymatic reaction.

The protein of most interest that was found but unfortunately was not confirmed in a technical replicate was cytochrome P450 46A1 (IPI0000195129, E.C.1.14.13.98), a P450 specific to the brain which catalyses the 24S-hydroxylation of cholesterol, creating a more hydrophilic product which can then diffuse through the blood-brain-barrier into the blood, where it is transported to the liver to be cleared from the system.

From the first gel run, a total of 1385 proteins were identified, 13 of which (<1%) were from the steroidal pathways, transporters or receptors. The total number of proteins from the combined runs (gel 1, 2 and 3) gave a total of 1838 unique IPI numbers (a total of 2698 IPI submitted to Proxeon), and 21 (1.1%) steroidal proteins were identified. In each fraction membrane proteins made up the highest number of proteins identified, followed by proteins complexes and then un-annotated proteins. In the case of both cytosolic fractions and mitochondrial fractions, the proteins listed under these biological locations were not the most abundant. This is primarily due to the fact that many proteins were listed in more than one category, making it difficult to tabulate percentages. As many membrane proteins are found in the mitochondrion, a large number of them are listed in both membrane and mitochondrion. The same applies for most of the categories, giving a skewed perception of the amount in each group.
Mascot, the search engine used, incorporates probability based scoring which allows for three types of searching: Peptide mass fingerprinting, sequence query and MS/MS. Probability-based scoring is one of the best for searching as it 1) has a simple rule to judge whether the significance of a hit is real against false positives, 2) scores can be compared to other searches such as sequence homology and 3) the search parameters can be optimised with each iteration.[119]

Mascot is capable of using any FASTA formatted database, including SwissProt, Trembl and IPI. This allows for searching of nucleic acid databases that have been translated into all 6 reading frames.

When compared to other 2-D methods of separation which allowed for fractionation of the lysate into three cellular fractions it was observed that results obtained were very close to those observed in literature.[73,88,168] Recently Takagiri split rat brain into 6 regions (thalamus, hippocampus, frontal cortex, parietal cortex, occipital cortex, amygdala) and performed 1D SDS-PAGE nano-LC-QTOF MS/MS on each lane, cutting them up into 24 pieces. From these digests, 515 proteins were identified with greater than 95% confidence

with all single peptide assignments being manually scrutinised. While the number of proteins identified is significantly lower than the number identified here, it does show that some separation can help in the identification of lower abundance proteins.[79]

Sample preparation is very important in the analysis of cytochrome P450s. A comparison that was performed using three different techniques for sample preparation of brain microsomes, showed that the calcium aggregate model, not the Teflon pestle method used in this experiment, gave the highest level of cytochrome P450s and related proteins. This has been suggested to be due to loss of the thiol moiety during microsome preparation which can be alleviated by the addition of dithiothreitol into the medium. As well, the high lipid content of the brain renders them vulnerable to lipid peroxidation which may also interfere or damage the resulting cytochrome P450 extraction. The results suggests that with the calcium aggregate method, more than 2.5 times the amount of cytochrome P450 were found which, when considering the amount of these cytochrome P450 proteins in the liver is 3%, may make an major difference in being able to identify them in tissues with lower P450 abundance.[169] All results were searched using Mascot. The samples were run both with a database (forward) and a decoy database (forward, reverse) which contains the forward searching database but also the same peptide masses but sequences reversed. Results obtained were considered to be valid if the error tolerance was less than 5%.

3.5.1 Comparison of Current Data with Previous Studies

The HUPO Brain pilot study used three different ages of mouse brain and autopsied human brain tissue versus a biopsy of human brain tissue (front temporal lobe) which were distributed to 15 labs took who part across the world using standardised protocols for the preparation of the tissues to be analysed. Any changes in protocol needed to be noted and all data was centrally processed and analysed using the DCC in Germany (MPC Bochum, Germany). Techniques used varied from 2-DE, 2D-LC, 1-DE – LC, and included both differential analysis and mapping. While hard to compare differential analysis to mapping, it was noted that using 1-DE followed by LC-MS/MS gave the largest number (1422) non-redundant identified proteins. While this technique would not be useful for differential analysis – there is not enough resolution between bands to be able to successfully differentiate between changes – it has been shown to have a high level of mapping identification. The simple separation into soluble, membrane and DNA binding fractions used to run on the SDS-PAGE are similar to the soluble, microsomal and

mitochondrial used in the current study. The level of separation of the fractions is not the most important part, indeed the crude separation is used as another layer of separation to reduce the complexity of the sample. While the HUPO BPP group was able to identify 1422 non-redundant IPI proteins, the current study was able to increase that and identify 1838 non-redundant IPI proteins, an increase in over 400 proteins being identified.

3.6 Conclusion and Future work

The principle of this work was to look into the development of a method that would allow for identification of steroid-associated proteins from brain tissue which would allow for qualification and quantification. This in turn, would allow for disease states to be studied to look for differences in the brain, in particular with diseases such as Alzheimers where the cholesterol in the brain can be part of the cause of the disease occurring.

Preliminary work identified in total 1838 unique proteins from the combined analyses. However the percentage of proteins that were interesting for this particular project – steroidal proteins - was very low. Only 21 proteins; 1.1 % of the total proteins identified, belonged to this particular group of proteins. The majority of the proteins were high abundance proteins, in particular myelin and tubulin, which were found throughout the samples (See Appendix B). The proteins that were identified, with the exception of Cytochrome P450 46A1 were identified more than once. The cytochrome P450 46A1 was only identified when running the P450 region in small ranges.

It has become apparent that more separation is required in order to see most of the proteins of interest. From the data (Table 3.10) it is obvious that even those that were identified are in such low abundance that only a very few peptides from the proteins were observed. Techniques such as ion exchange or liquid isoelectric focusing could help separate proteins further to allow some low abundance proteins to be detected. As well, other techniques such as Western blotting can be used to confirm such hits as 46A1 that have only been identified once and with a very low coverage.

Future work would be to consider developing another dimension of separation prior to analysis. Key thoughts to keep in mind are the amount of time and money required to make such a separation viable for further work into quantification and analysis of disease-state tissues. Further work may be done on identification of metabolites in CNS or plasma for identification of potential metabolic biomarkers for identification of disease states instead of using protein biomarkers such as the use of 24S-hydroxycholesterol. This would allow for a robust analysis that may allow for rapid diagnosis. Work should continue with respect to the proteomics of the brain to fully understand how the disease state comes about.

4 1DE nano-LC/MS/MS Analysis of the proteome for *Anabaena variabilis ATCC 29413*

4.1 Ultraviolet light and it's potential for biological destruction

Over the years, there have been a number of studies done on the effects of UV on human health.[170-172] Under normal circumstances, an organism requires UV light to maintain health as it is vital to transform 7-deoxycholesterol into vitamin D which is important for bone health. It has also been shown that vitamin D inhibits certain internal cancers and indirectly, in some cases, autoimmune diseases such as multiple sclerosis, diabetes mellitus type I, rheumatoid arthritis and irritated bowel diseases.[170-172] However, high UV exposure has been connected to increase of skin cancers such as melanoma and non-melanoma skin carcinomas, cataracts and pteryguism. It is believed that UV exposure has also caused an increase in outbreaks of various infections such as herpes simplex virus or human papilloma virus.[170-172] Ultraviolet light can be classified into four different categories: UVC < 280 nm and is absorbed by the ozone, UVB 280 – 320 nm of which approximately 10% gets through the ozone, UVA 320 – 400 nm which is not filtered by the ozone and photosynthetically active radiation (PAR) 400 – 700 nm. While the PAR light is not considered to be dangerous, UVA – UVC are considered to be harmful for human health. With the slow decline of the ozone layer in varying degrees close to the poles, there is an increase in UVB getting through, thus increasing potential levels of cancer and biological damage in all living beings.[173]

Ultraviolet light is absorbed by conjugated bonds and is known to cause reactions leading to effects such as photobleaching, plastics decay and photochemical smog. In living organisms, it is absorbed by amino acids, nucleic acids, proteins, DNA/RNA and conjugated bonds in unsaturated fatty acids and lipids. Reactive oxygen species (ROS) such as singlet oxygens, hydrogen peroxide, superoxide and hydroxy radicals are all radicals that can be formed as a result of UV absorption. These have unpaired electrons, and are highly reactive. Radicals are capable of degrading proteins and DNA as well as destroying membrane lipids and any other receptive molecules.[170-172,174]

Many organisms are capable of protecting themselves against such damage through the synthesis of natural antioxidants which scavenge reactive oxygen species and any other radicals that are formed. Proteins such as superoxide dismutase and catalase, as well as small molecules such as carotenoids and ascorbic acid can all act as such antioxidants.[174]

Intertidal and marine organisms are exposed to the highest levels of ultraviolet radiation down to depths of 20 meters, so that phytoplankton and benthic organisms may also be affected.[175] Marine bacteria and other forms of marine life have the potential to create new ways of protecting themselves against such damage. These include expression, and regulation of antioxidant enzymes, accumulation and cycling of small molecule antioxidants, molecular repair and the formation of natural UV absorbing sunscreens.[173] Mycosporine-like amino acids (MAA)s are considered to be one of the best methods of UV protection for marine life.[173] These are small, cyclic molecules with aromatic properties, transparent to light and have varying λ_{max} within the UVB range. MAAs are capable of absorbing dangerous UV wavelengths and dissipating the absorbed energy without transferring to sensitive biomolecules or causing oxidative stress through the formation of oxygen radicals. MAAs are predominantly synthesised by cyanobacteria and algae, however they have been shown to accumulate in the epidermis, ocular and eggs, protecting organisms which consume the MAA synthesising bacteria and algae.[173,174,176,177]

Regulation of MAAs appears to be through UV exposure as MAA synthesis is stimulated upon UVA, UVB, white light or blue light exposure but red or green light does not increase MAA biosynthesis.[173] MAAs were suggested to be synthesised through the shikimate pathway and via 4-deoxygadusol which is structurally related to MAAs and are often found in high levels of fish and marine invertebrates as part of their defense against UV damage. The shikimate pathway is responsible for endogenous production of small aromatic molecules and various amino acids found in bacteria, fungi, plants and protozoans, while animals lack this pathway.

Glyphosphate is a small molecule capable of inhibiting 5-enoylpyruvyl shikimate-3-phosphate synthase at micromolar concentrations. Increasing the concentration of glyphosate to millimolar concentrations shows inhibition of 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase (DAHP) and 3-dehydroquinate (DHQ) synthase. Inhibition studies performed by Schick *et al* (1999) using *S. pistillata* and coral show at high levels (1 mM) glyphosate inhibited MAA synthesis. Specifically it inhibited cytosolic DAHP which requires Co^{2+} unlike the other found in *S. pistillata* which requires Mn^{2+} for activity. Portwich *et al* used radiolabelling to show that the shikimate pathway intermediate 3-dehydroquinate (DHQ) is the precursor for the six-membered carbon ring common to fungal mycosporines.[178] Starcevic *et al* recently used a method for suppression subtractive hybridisation PCR to look at genes upregulated by UV exposure in *A.*

microphthalma. Upon creation of the cDNA library, the cDNA sequences were aligned against an expressed sequence tags (EST) database resulting in 131 ESTs which have been upregulated. From these, one EST gave an alignment to an enzyme that catalyses O-methylation and dehydrogenase; steps which are expected to follow MAA synthesis after 3-DHQ, a branching point of the shikimate pathway. A second enzyme was identified for reduction of C_5 carboxylate groups, another step required for the synthesis of 4-deoxygadusol. A final EST sequence was found that defines a putative hydroxylase enzyme, another enzyme required for the synthesis of the proposed intermediate gadusol. Further studies into cyanobacteria showed that MAA production uses a fusion protein gene DHQ synthase and O-methyltransferase. Sequence alignment of the active site residues by Starcevic *et al* showed a strong correlation to DHQ synthase and not the EVS (2-epi-5-epi-valionate synthase) a protein with high level of homology to DHQ synthase.[179]

Very recently a paper came out discussing the idea that MAAs are produced by the shikimate pathway was discussed by Balskus and Walsh.[180,181] Using *Nematostella vectensis* a sea anemone whose shikimate pathway genes have been encoded, Balskus mined the genome cyanobacteria *ATCC 29413*, in an attempt to find a cluster of genes that encoded for an O-methyltransferase near a DHQ synthase homolog. BLAST searches of the O-methyltransferase and DHQ synthase using the genes from *N. vectensis* revealed clusters of homologs in genomes of dinoflagellates, cyanobacteria and fungi. These gene clusters revealed open reading frames (ORF)s for DHQ synthase homolog, an O-methyltransferase and two other conserved open reading frames in cyanobacteria. Upon further homology searching it identified a member of the ATP-grasp superfamily which are known to form peptide bonds and a nonribosomal peptide synthetase (NRPS) homolog. Using the 6.5 kb shinorine gene cluster of open reading frames, three constructs were created: the first used the full gene cluster, the second lacked the NRPS homolog and the third lacked the NRPS and the ATP-grasp homolog. Induction of expression of the intact gene cluster in *E. coli* resulted in the production of shinorine while the NRPS truncation formed mycosporine-glycine and the NRPS, ATP grasp homolog truncation formed 4-deoxygadusol. The synthesis of shinorine in *E.coli* suggested that no other genes or proteins are required for production. Biochemical analysis *in vitro* using 3-dehydroquinate and cofactors *S*-adenosylmethionine, nicotinamide adenine dinucleotide (NAD^+) and Co^{2+} failed to produce 4-deoxygadusol. Balskus and Walsh suggest that the homology between the DHQ synthase homolog and DHQ show variation in the active site residues, showing more homology with the homolog

2-*epi*-5-*epi*-valiolone synthase (EVS) rather than DHQ synthase. This suggested that the enzymes may share a common substrate from the pentose phosphate pathway intermediate sedoheptulose-7-phosphate. With this substrate and the conditions above, Balskus and Walsh were able to produce 4-deoxygadusol *in vitro*. With further testing of the two other enzymes *in vitro* they were able to produce mycosporine-glycine and shinorine giving strong evidence that the cluster is responsible for MAA biosynthesis however not through the shikimate pathway but through the pentose phosphate pathway.[180]

4.1.1 Anabaena variabilis *ATCC 29413*

Anabaena variabilis ATCC 29413, is a filamentous, heterocyst forming blue –green algae / cyanobacteria that grows in the shallows of salt water. It is the largest group of prokaryotes that can carry out oxygenic photosynthesis with water as an electron donor. They are capable of growing in freshwater, marine, terrestrial and hypersaline environments.[176] In particular, the *Anabaena* species is one of two species that are nitrogen fixing. This ability to split water and fix nitrogen makes them very attractive for engineering studies to increase NH_3 or H_2 production to be used as an alternative energy source.[182,183] It is capable of protecting itself from the harsh rays of the sun by producing a mycosporine-like amino acid (MAA), shinorine, which is able to absorb UV light.[184] MAAs are believed to be produced through a branching point of the shikimate pathway in some microorganisms and coral to provide protection from radical formation and oxidative damage of UV light.[173,185,186]

4.2 Aims

The pathway for MAA synthesis, has not been identified in any organisms. The exposure of *ATCC 29413* to UVB light is believed to switch relevant pathways on and induce the production of shinorine. The aims are to:

- demonstrate the successful induction of the pathways by quantifying shinorine using HPLC,
- identify the proteins from the shikimate pathway and shinorine synthesis by comparing exposed sample to a control sample subjected to proteomics techniques (SDS-PAGE, digesting, separating and fragmenting the peptides by capillary HPLC nano-ESI MS/MS).

4.3 Materials and Methods

4.3.1 Materials

Chemicals for media preparation were purchased from the following companies: $MnCl_2 \cdot 4H_2O$, $CuSO_4 \cdot 5H_2O$, $CaCl_2 \cdot 2H_2O$, $Fe_2(III)Cl_3 \cdot 6H_2O$ were purchased as AnalaR from BDH. H_3BO_3 was purchased from Lancaster, USA while $ZnSO_4 \cdot 7H_2O$ was obtained from Fluka and $Mg_2SO_4 \cdot 7H_2O$ was bought from Aldrich. The majority were purchased from Sigma-Aldrich (St. Louis, MO, USA): $Na_2MoO_4 \cdot 2H_2O$, Na_2CO_3, $CoCl_2 \cdot 6H_2O$, $NaNO_3$, citrate and Glufibrinopeptide B. Tris·HCl, NH_4HCO_3, KH_2PO_4, EDTA disodium salt, Bovine Serum Albumin (molecular biology grade) HPLC grade H_2O, acetonitrile and formic acid were purchased from Fisher Scientific (Waltham, MA, USA). Trypsin was from Promega (Southampton, UK). NuPAGE® Novex Pre-Cast Gel System, NuPage® Reducing Agent (500 mM DDT), NuPage® Sample Buffer (4X) were bought from Invitrogen (Paisley, UK). Bio-Rad protein assay was purchased for Bio-Rad (Perth, UK)

4.3.2 Media Preparation:

Culture media BG-11 for the cyanobacteria was prepared following the recipe listed in Table 4.1. Media was weighed and diluted to 900 ml H_2O prior to autoclaving. Upon autoclaving and cooling, 1 ml of trace metal solution (Table 4.2) was added to the media. The media was mixed well and the pH was adjusted to 7.4 using NaOH before adjusting to a final volume of 1 L.

Table 4.1: Culture Media (BG-11)

Ingredient	g/L	mM
$NaNO_3$	1.5	17.65
$K_2HPO_4 \cdot 3H_2O$	0.04	0.18
$MgSO_4 \cdot 7H_2O$	0.075	0.30
$CaCl_2 \cdot 2H_2O$	0.036	0.25
Ferric ammonium citrate	0.006	0.03
Citrate	*0.006*	0.03
EDTA (dipotassium magnesium)	0.001	0.003
Na_2CO_3	0.04	0.38
pH after autoclaving and cooling:	7.4	
Trace metal solution	1 ml	
Deionised water	To 1 L	

Table 4.2: Composition of Trace Metal Solution (g/L): masses weighed to prepare 100 ml.

Metal	Mass weighed (g)
H_3BO_3	0.286
$MnCl_2 \cdot 4H_2O$	0.181
$ZnSO_4 \cdot 7H_2O$	0.022
$Na_2MoO_4 \cdot 2H_2O$	0.039
$CuSO_4 \cdot 5H_2O$	0.008
$Co(NO_3)_2 \cdot 6H_2O$	0.005

The solution was filter sterilized.

4.3.3 Initial Culture

Initial growth of *ATCC 29413* was performed in a completely dark room over 6 weeks where the frozen glycerol culture was aseptically transferred to 6 150 mL Erlenmeyer flasks containing 50 mL culture media BG-11. The flasks were placed on a rotary shaker (New Brunswick Scientific, Innova 2000) at a constant 80 rpm. A green house light was used in 12 hour intervals to initiate growth.

4.3.4 Growth curve

A growth curve was established based on wet weight of the culture in triplicate. Using 5 ml of the established culture to inoculate 50 ml of new media, a curve was established by taking 1 ml aliquots on a daily basis over the course of 14 days. The wet weight was established after removal of the supernatant.

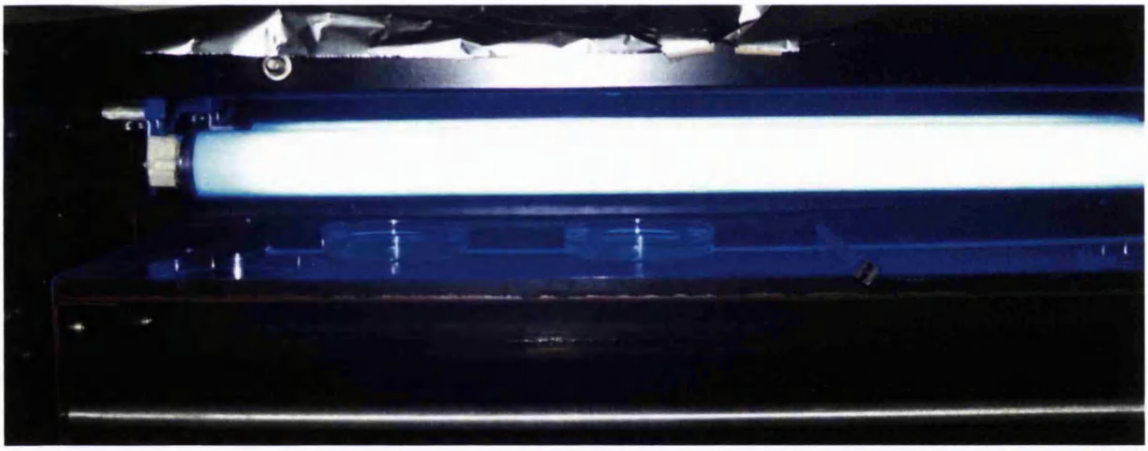
4.3.5 Establishing MAA induction

Using the established *ATCC 29413* culture, 5 - 250 mL flasks of 50 mL media were prepared and inoculated with 400 µL of culture. Flasks (Figure 4.1) were grown for 6 days before being pooled together and centrifuged at 4000 rpm 4 °C for 10 minutes (Heraeus Megafuge, DJB Labcare Ltd, Buckinhamshire, UK). The cells were then resuspended in 250 mL fresh BG-11 media before being split into 5 petri plates which were placed under a UV light (Phillips TL20W/05) providing UV B (280nm – 315nm) radiation at a distance of 4 cm (Figure 4.1B and C). The induction experiment started with alternating 12 hours UV light and dark for a total of 72 hours. Cells were centrifuged (Heraeus Biofuge Fresco, DJB Labcare Ltd, Buckinhamshire, UK) for 10 minutes at 4,000 x *g*, 4 °C before being transferred to a tared 2.0 mL centrifuge tube. Cells were centrifuged a second time for 5 minutes at 13,000 x *g*, at room temperature and all supernatant was removed. For the extraction of shinorine, 1.5 mL 100% MeOH was added to each tube. The tubes were

mixed (Scientific Instruments, Vortex Genie 2) long enough to break up the cell pellet and the tubes were left in the fridge overnight. The samples were centrifuged for 5 minutes at 13,000 x *g*, and the supernatant was transferred to a clean 2.0 mL centrifuge tube. The cell debris was left to dry. The supernatant was scanned using a UV/Vis spectrophotometer (Perkin Elmer Lambda 25) from 200- 700 nm using a quartz crystal cuvette. Samples that gave absorbance of 2 at 334 nm were diluted and analysed again.

A)

B)



C)

Figure 4.1: Growth and induction apparatus. A) Rotary shaker for culture growth. B and C) UV induction chamber.

4.3.6 MAA induction for shinorine analysis

A second set of flasks were prepared with 50 mL BG-11 and were inoculated with 400 µL of established *ATCC 29413*. Flasks were grown for 5 days before being centrifuged at 4000 x *g*, 4 °C, 10 minutes. The supernatant was discarded and the cells were transferred to a tared 2.0 mL centrifuge tube. The tubes were centrifuged at 13,000 x *g*, room temperature for 5 minutes and whatever supernatant was left was discarded. Cells were re-suspended in 100 % MeOH by mixing and the mixture was left in the fridge overnight. The cells were mixed and centrifuged at 13,000 x *g*, at room temperature for 5 minutes. The supernatant was collected (F1) in a separate tube and another 1.5 mL of 100% MeOH was added to the cell debris. The tube was vortexed and then centrifuged at 13,000 x *g* at room temperature for 5 minutes. The supernatant was kept (F2) and the cell debris was resuspended in 1.5 mL 100% MeOH and vortexed. The sample was centrifuged a third time at 13,000 x *g* at room temperature for 5 minutes and the supernatant was collected into a 2.0 mL centrifuge tube (F3). The cell debris was left to dry and the dry weight was tared. The three fractions were scanned using a UV/Vis spectrophotometer from 200 – 700 nm using a quartz crystal cuvette. Samples with an absorbance greater than 2 at 334 nm were diluted and checked again.

4.3.7 Shinorine samples preparation.

Shinorine samples were dried without heat on the vacuum concentrator (Jouan, RC-1022 heated concentrator, Heto cooling trap) for 1 hour. The three fractions from the same sample were dissolved in 1 mL H_2O which was sonicated (Camlab Transsonic TS701 H, Cambridge UK), vortexed (Fisher Scientific Vortex Genie 2, Waltham, MA, USA) and centrifuged at 13 000 x *g*, 4 °C for 5 minutes (Heraeus Biofuge Fresco, DJB Labcare Ltd, Buckinhamshire, UK). A Waters C_{18} SepPak column was equilibrated with 5 mL H_2O, 5 mL 100% MeOH and 5 mL H_2O again before the combined sample was loaded onto the column and collected in 1 mL fractions. A flow rate of 0.75 mL/min was used on the SepPak. Another 10 mL of H_2O was loaded onto the column and two fractions (1mL and 9 mL) were collected. The column was washed with another 5 mL 50% MeOH which was collected followed by 5 mL 100% MeOH which was also collected. The first two fractions of H_2O were checked for absorbance at 334 nm.

4.3.8 HPLC of shinorine standards and shinorine samples.

Shinorine standards were gifted by Malcolm Shick at the University of Maine, Department of Biological Sciences, Marine Sciences. A method for identification and quantification of shinorine was developed using a Waters HPLC (Waters Alliance quaternary pump and degasser, Waters 996 Photodiode Array Detector, Waters 600 Controller, Waters 717plus Autosampler) with a photodiode array detector (Waters 996) and autosampler. An isocratic method using H_2O + 0.02% AcOH at 0.5 mL/min, 20 minutes was created using a Phenomenex Luna 3 µ(2), 100 Å, 150 x 3 mm 3 µm column. Shinorine was monitored and quantified at 330 nm.[187]

Standard concentrations (0.12 – 12 pmol/µL) were injected in triplicate to create a calibration curve while system suitability used one concentration with 6 injections to check for reproducibility of peak area, peak height and peak retention time. Limit of detection and quantification were determined at the lower limit. Samples that had been cleaned using the Waters C_{18} SepPak were injected onto the system (10 µL).

4.3.9 MS of small molecules

Mass spectrometry data was acquired using borosilicate capillary tips (Thermo Scientific, Hertfordshire, UK) which were cracked prior to use. Samples were mixed 1:1 with 100% MeOH + 0.1% FA. Samples and standards were run using nanoESI on the Finnigan™ LCQ™ Duo (Thermo Scientific, Hertfordshire, UK) (1.8 kV capillary, shinorine tune method, no sheath gas, no auxiliary gas) controlled with Xcalibur™, and the Waters QTOF™ Global Ultima (Waters, Manchester, UK), (1.0 kV capillary, cone 50, collision 10, no api gas) controlled with MassLynx v. 4.1. Both instruments were in positive nano-electrospray modes. Fragmentation was attempted on both the LCQ (collision energy 40 eV) and the QTOF (CE 35).

4.3.10 Extraction of proteins from *ATCC 29413*

A set of 250 mL flasks were filled with 50 mL BG-11 and were inoculated with 400 µL of established *ATCC 29413.* The flasks were grown for 6 days alternating 12 hours light, 12 hours dark. Flasks were pooled together and centrifuged 4000 x *g*, 4 °C for 10 minutes and the supernatant was discarded. Fresh media was added to the cells and the cells were distributed equally into petri plates and placed under the UV light. The control samples were removed at time 0. The 72 hour samples were checked periodically to ensure that the media was not evaporating.

Once completed, the samples were removed and centrifuged at 4000 x *g*, 4°C for 10 minutes. The cells were transferred to a tared 2.0 mL centrifuge tube and centrifuged at 13,000 x *g* at room temperature for 5 minutes. The media was discarded. Lysis buffer (40 mM Tris pH 7.4, 10% glycerol, protease inhibitor pellet) was prepared and 1 mL was added to the cell mixture. The cells were mixed until suspended. The bacteria were transferred into a mortar and pestle cooled by liquid nitrogen (BOC, Guilford, UK) where they were ground up and then transferred to the tared centrifuged tube. The pellet was sonicated on ice at 15 mA (Soniprep 150, MSE, London, UK) 20s on 40s off for 10 minutes before being centrifuged at 13,600 x *g*, 4°C, 40 minutes (Beckman Max-E ultracentrifuge with MLA-80 rotor, London, UK). The supernatant was transferred to a new 2.0 centrifuge tube while the pellet was re-suspended in 3 mL lysis buffer and centrifuged a second time producing a pellet and more supernatant. Samples were stored at -80°C until use.

4.3.11 Bradford Protein Assay using microtiter plates

Using a stock solution of BSA (2.0 mg/mL) and H_2O as a solvent, the following dilutions of BSA were made: 5, 10, 15, 20, 25, 30, 40, 50, 60 µg/µL (Table 4.3). Using a microassay plate, 5 µL of standard or sample was transferred in triplicate onto the plate into separate wells. Reagent is prepared by mixing 1:4 Bio-rad Protein Assay with H_2O. Reagent was filtered using 0.22 µm (Millipore, Billerica, MA, USA) filters before use. The diluted protein assay reagent was added to each well to a volume of 200 µL. The plate was shaken gently and absorbance measured at 595 nm before 1 hour was over.

Table 4.3: BSA calibration curve prepared from 2 mg/mL BSA.

Concentration mg/ml	BSA (µl)	H_2O (µl)	Sample volume (µl)	Reagent volume (ml)
2.0	100	0	5	0.2
1.5	75	25	5	0.2
1.0	50	50	5	0.2
0.75	37.5	62.5	5	0.2
0.5	25	75	5	0.2
0.25	12.5	87.5	5	0.2
0.125	6.25	93.75	5	0.2

The results were put into Excel to determine protein concentrations.

Calculation for the concentration of the samples was as follows:

$$\frac{A_{595}}{slope\ of\ curve} \times dilution\ factor = protein\ concentration$$

4.3.12 SDS-PAGE

Using a general protocol from Invitrogen, each protein fraction was separated using the NuPAGE® Novex Pre-Cast Gel System. A maximum of 25 µg was loaded onto each lane, unless the concentration of the fractions was too low; in which case, the concentration loaded was normalised to the lowest concentration fraction. Samples were prepared by adding 2.0 µL Invitrogen NuPage® Reducing Agent (500 mM DDT), 5.0 µL Invitrogen NuPage® Sample Buffer (4X) and HPLC grade water to make the sample up to 18 µL. The samples were incubated at 70°C for 10 minutes. A solution of 200 mM iodoacetamide (Sigma-Aldrich, Dorset, UK) was prepared just prior to use. A total volume of 2 µL of 200 mM iodoacetamide was added to each sample to make a final volume of 20 µL. The samples were incubated in darkness for 60 minutes to encourage alkylation of cysteine residues to prevent the re-formation of cysteine bonds.

Samples were loaded onto a 10 or 12 well NuPAGE® 4-12% Bis Tris gel. Pre-stained protein molecular weight marker (5 µL) was loaded into a free lane (SeeBlue® Plus2 Pre-Stained Standard, Invitrogen, Paisley, UK) as a reference.

The gel was run using a constant voltage (200 V) using prepared MOPS (Invitrogen 20X) buffer with a starting current of 110-115 mA and a finishing current of 60-70 mA. Normal running time for gels was between 50-60 minutes.

The gel was stained using the general microwave protocol by Invitrogen. The gel was placed in a staining tray with HPLC grade H_2O. The gel was loosely covered and the microwave was set on high for 45 seconds. Carefully, the gel was removed from the microwave and placed on a gel rocker (Grant Boekel Gel Rocker, Grant Instruments, Cambridge, UK) for one minute. This was repeated twice more. After the third wash, 20 mL of Invitrogen™ SimplyBlue™ SafeStain (Invitrogen Paisley, UK) was added and the gel was once more placed in the microwave for 50 seconds and placed on the gel rocker for 5 minutes before washing the gel multiple times with H_2O until the background was clear and the gel bands distinct. A scan of the gel was made and retained prior to in-gel digestion.

4.3.13 In-Gel Digest and Extraction

From the first gel, lanes were cut into approximately 20 bands of each which were then further cut into 1 mm^3 pieces and placed into 1.5 mL siliconised tubes. Bands were destained using 500 µL of 50 mM ammonium bicarbonate: ethanol (1:1, v/v) and shaking the tubes for 1 hour. Samples were briefly centrifuged and the supernatant was discarded. The process was repeated until the gel pieces were clear. The clear pieces were dried by adding 100 µL of 100% acetonitrile, vortexing 10 minutes and discarding the supernatant. The process was repeated a second time before the centrifuge tubes containing the gel pieces were placed onto the vacuum concentrator for 30 minutes at 40°C to complete the drying process.

An aliquot of sequence grade modified trypsin (Promega, Southampton, UK) was diluted to 0.01 µg/ µl in 50 mM ammonium bicarbonate and 5 – 10 µl was added to each gel piece. The gels pieces were kept on ice for 30 minutes before a final 20 µL of 50 mM ammonium bicarbonate was added to the gel pieces. The samples were placed in an incubator (Innova 4300 incubator/shaker, New Brunswick Scientific, Hertfordshire, UK) at 37°C overnight.

After 12 – 16 hours, the gel pieces were removed from the incubator and 25 µL extraction solvent (5% FA in 50:50 ACN/H_2O) was added to each individual tube. The tubes were placed on the vortex for 10 minutes, centrifuged and the supernatant was transferred to new 0.5 mL siliconised tubes and kept. Acetonitrile (30 µL) was added to each gel piece before mixing for 5 minutes, centrifuging and transferring the supernatant to the 0.5 mL tube. The whole process was repeated 2 more times before the pooled supernatants were dried on a heated vacuum concentrator with 40 °C for 1 hour until dried.

The dried samples were dissolved in 20 µL 0.1% FA in H_2O, mixed and centrifuged prior to transferring the supernatant to a polypropylene autosampler vial (Waters, Manchester, UK).

4.3.14 Capillary LC-MSMS Data Dependent Acquisition

LC-MS/MS analysis was performed on a Waters CapLC system coupled to the front end of a Waters QTOF Global Ultima™. Per sample, 5 µL was injected using pump C (Table 4.4) which loads the sample onto the pre-column and de-salts the sample.

Table 4.4: CapLC Auxiliary C Flow Timetable and Stream Select Events

Time	Flow µl/min	Stream Select Position
0.10	15.0	Position 1
3.00	15.0	Position 2
3.10	1.0	Position 2
53.0	1.0	Position 1

After 3 minutes the stream select valve changes position putting the pre-column and column in tandem (Figure 4.2) allowing separation of the peptides using a gradient through the column (Table 4.5, Figure 4.3) into the ion source. Each sample has a run time of 63 minutes.

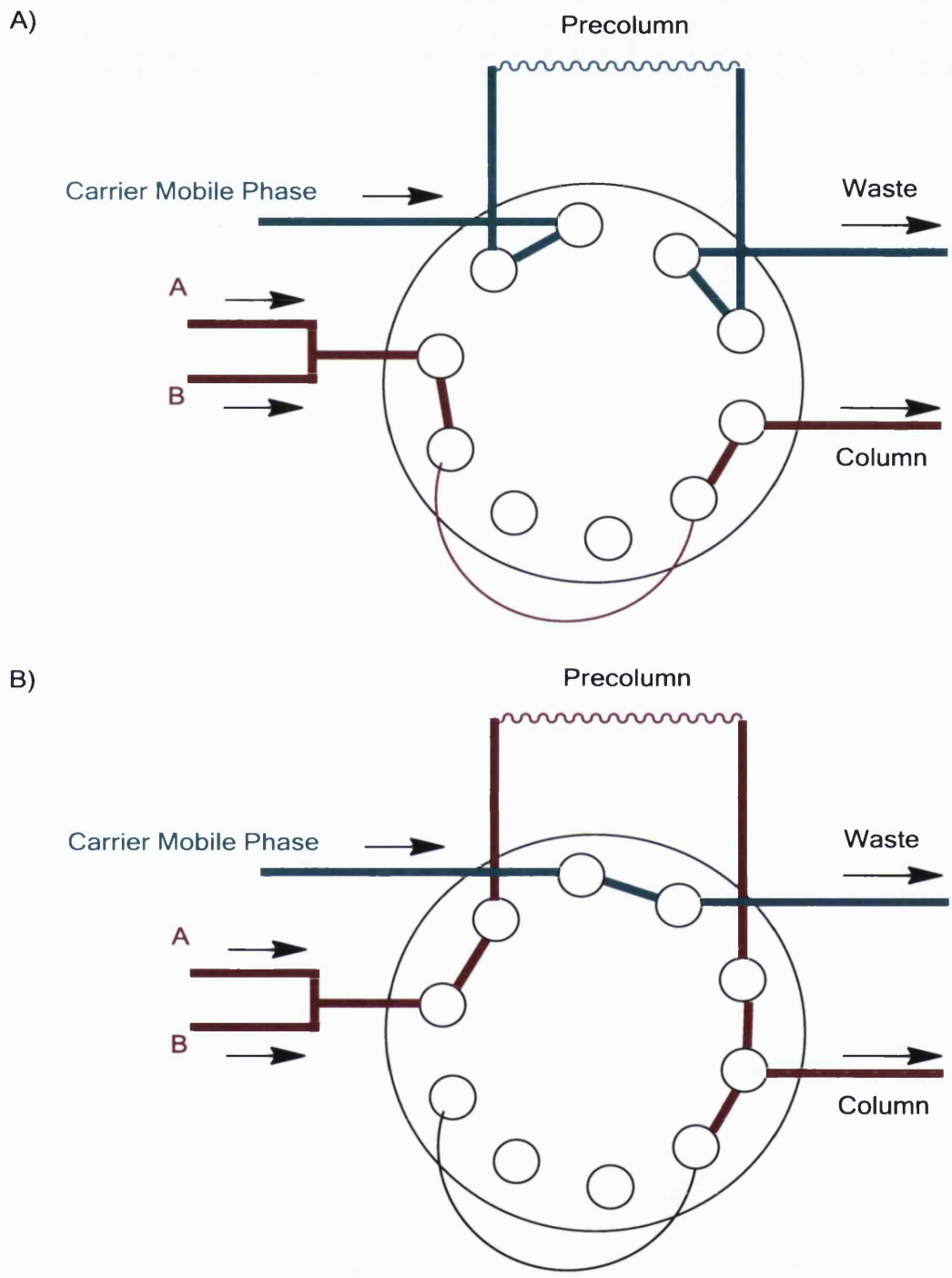


Figure 4.2: Stream Select Switching Valves. A) shows the valves in the initial position where the flow from the carrier mobile phase goes through the pre-column and goes to waste to desalt. After three minutes, the stream select valve switches over to B) and the flow from both A and B are mixed together and go through the pre-column and into the column to be finally eluted into the mass spectrometer.

Upon switching the valve, the flow from the pre-column moved through to the nano-column (150 x 0.075 (i.d) mm, 3 um C18, Dionex) and into the mass spectrometer.

Table 4.5: Waters CapLC Gradient Timetable

Time	Percent A	Percent B	Flow
0.10	95.0	5.0	6.0
3.00	95.0	5.0	6.0
40.0	72.0	28.0	6.0
49.0	20.0	80.0	6.0
52.0	20.0	80.0	6.0
53.0	95.0	5.0	6.0
63.0	95.0	5.0	6.0

Initial flow from the CapLC is at 6 µl/min, which is split prior to the column so the flow through the column to the mass spectrometer is only 200 nL/min.

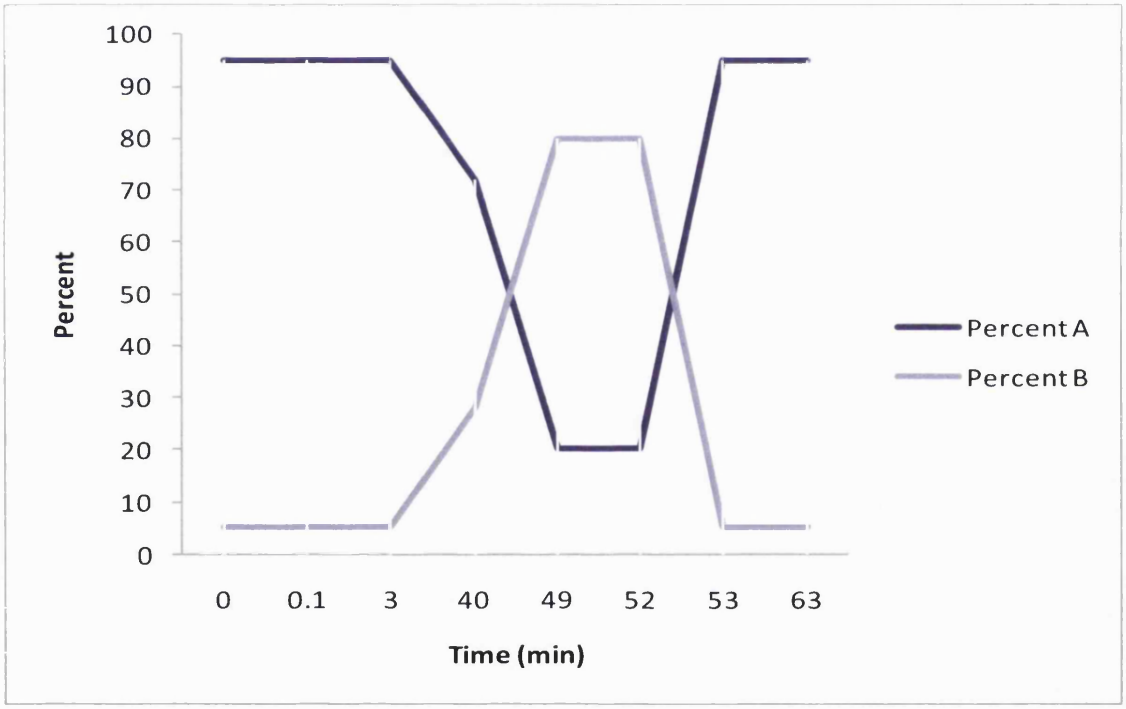


Figure 4.3: Representation of the gradient formed between mobile phase A and mobile phase B.

Peptides eluted off the column were directly sprayed into the mass spectrometer for analysis.

4.3.15 Mass Spectrometry (MS/MS)

All data was run using the Waters QTOF Global Ultima™ using a nanospray source for an increase in sensitivity. The acquisition and processing software used was Waters MassLynx Version 4.1. Samples were run acquiring data over a *m/z* range of 400-1600 with MS/MS analysis from 50-1700 (Table 3.3).

Table 4.6: Parameters for the analysis of samples by the QTOF

Waters QTOF Global	Parameters	
ACQUISITION	Survey Start Time	3.0
	Survey Ends Time	60.0
	Survey Ion Mode	ES Mode
	Survey Polarity	Positive
MS SURVEY	Survey Start Mass	400.0
	Survey End Mass	1600.0
	Intensity Threshold	10
	Survey Scan Time	1.0 seconds
	Survey Interscan Time	0.1 seconds
	Survey Data Format	Continuum
	Survey Use Tune Page CV	Yes
	Survey Cone Voltage	35.0 V
MS/MS	MSMS Start Mass	50.0
	MSMS End Mass	1700.0
	Number of Components	4
	MSMS to MS Switch Criteria	Intensity Falling Below Threshold
	Switchback Threshold	3.0 counts/second
	Use MSMS to MS Switch After Time	YES
	MSMS Switch After Time	3.3 seconds
	MSMS Scan Time	1.0 seconds
	MSMS Interscan Time	0.1 seconds
	MSMS Data Format	Continuum
	Use Tune Page Cone Voltage	YES
	MSMS Cone Voltage	35.0 V
	Use MS/MS ipr File	NO
PEAK DETECTION	Peak Detection Window	1.0
	Use Include by Charge State	YES
	Charge State(s)	2,3
	Number of Include Components	60
	Charge State Tolerance Window	3.0
	Charge State Extraction Window	2.0
	Discard Survey Data	NO
COLLISION ENERGY	Use Charge State Recognition	YES
	Maximum Charge State	4
	Charge State 1 Filename	Default_CS_1
	Charge State 2 Filename	Default_CS_2
	Charge State 3 Filename	Default_CS_3
	Charge State 4 Filename	Default_CS_4
EXCLUDE	Detected Precursor Inclusion	Using Real Time Exclusion
	Detected Precursor Inclusion	Include After Time
	Include after Time	60.0 seconds
	Use Exclude Mass List	NO
	Exclude Window	+/- 1500.0 mDa
	Exclude Retention Time Window	10 seconds

Table 3.3 shows the general experiment setup for the MS/MS method. Spectra were acquired in MS mode and up to 4 individual peptides could be fragmented by MS/MS per MS scan. The mass spectrometer was set to fragment the multiply charged masses once

their intensity reached 10 counts/scan and for a total of 3.3 seconds or until the intensity fell below 3.0 counts/scan. That particular mass and a window of 1.5 Da around it were excluded for 60 seconds, allowing the mass spectrometer to fragment as many different components as possible during the run time.

4.3.16 Calibration

Glu-Fibrinopeptide (peptide sequence EGVNDNEEGFFSAR) was used for calibration and optimisation of the mass spectrometer at a concentration of 100 fmol/µL infused at a flow rate of 0.3 µl/min. Using the $[M+2H]^{2H+}$ ion of Glu-Fibrinopeptide, the ion was fragmented and the resulting product ions were used for calibration.

BSA (100 fmol) was subsequently run on the LC-MS to ensure sensitivity and resolution are sufficient – peak width at half height must be less than 0.3 min and retention times compared to the last BSA run cannot be out by more than 0.5 min. Searching of BSA using ProteinLynx Global Server must give over 35% coverage of BSA to accept the system is running normally.

4.3.17 Data Processing

The resulting raw data was processed using MassLynx ProteinLynx V4.1. The system was setup using Peptide Auto and the parameters used were from the file Process.mlp. Processing parameters included combining all sequential scans with the same precursor and processing all combined scans. Mass measure of the combined scans involved spectra smoothing using Savitzky Golay with a 3.00 channel window. The smoothing was performed twice. A centroid peak list was then created using the top 80.00% of the peak with a minimum peak width at half height of 4. Once the data had been processed, it was then combined into a single pkl file which could then be used for searching against databases. PKL is an extension for a text file created by Masslynx which lists all the MS data (*m/z* and charge) and tandem mass spectrometry data associated with that *m/z* and charge.

4.3.18 Protein Lynx Global Server Search[121]

Using the raw data, a processing method, followed by search method were created. The processing method was the same as that used to create a pkl file:

Mass measure: Smooth by Savitzky Golay 3.00 channel, 2 smooths, centroid peak list using top 80% with minimum peak width at half height = 4.

The search was based on general search parameters used in the lab with the exception that the database was specifically for *ATCC 29413* which was extracted from the NCBI database in FASTA format. The formatting was modified to be readable using PLGS. A new database search was created with the following parameters:

Enzyme: trypsin	Maximum missed cleavages: 1
Fixed modifications: none	Variable modifications: -Oxidation (Met) -Carbamidomethylation (Cys) -Acetyl (Protein N-term)
Peptide Tolerance: +/- 0.3 Da	MS/MS tolerance: +/- 0.3 Da
Report top: 200 max	

4.3.19 X!Tandem

A second database was used to search the data. Using the pkl file generated by MassLynx 4.1, the data was searched online (http://h.thegpm.org/tandem/thegpm_tandem.html) using X!Tandem[120] which is run by the GPM. Specifically the database for Anabaena variabilis *ATCC 29413*, could be selected and the following parameters can be set:

Reverse sequence: include		
Fragment mass error: 0.3	peptide log(e) < -1	
	protein log (e) < -1	
Protein modifications (variable):	carbamidomethylation (Cys)	
	oxidation (Met)	
Protein cleavage:	trypsin	
Semi-style cleavage:	yes	
Spectrum conditioning:	remove redundant	yes
Spectrum synthesis:	yes	
Quad TOF error:	100 ppm	

Data from both databases were manually scrutinised and then compared to each other.

4.3.20 MASCOT and SCAFFOLD analysis

A third analysis was performed at the King's College London Proteomics Unit using the same data. Tandem mass spectra were extracted from the raw data by protein Lynx Global Server 2.2.5. Charge state deconvolution and deisotoping were not performed. All peak list files from MS/MS samples were analysed using Mascot Daemon (Matrix Science, London, UK version 2.2.03). Mascot was set up to search entries of uniprot_sprot_110118 database selected for bacteria (195743 entries) assuming digest with trypsin and up to 2 missed cleavages. Mascot was searched with fragment ion tolerance of 0.6 Da and a parent tolerance of 1.2 Da. Variable modifications used were methionine oxidation, iodoacetamide derivatisation of cysteine.

Scaffold (version Scaffold 3_0_04, Proteome Software Inc, Portland, Oregon, USA) was used to validate MS/MS based peptide and protein identification. Peptide identifications were accepted if they could be established at greater than 95% probablility as specified by the Peptide Prophet algorithm.[188] Unless stated, protein identifications were accepted if they could be established at greater than 99.0 % probability and contained at least 2 assigned peptides. Protein probabilities were assigned by the Protein Prophet algorithm.[189] Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped together to satisfy the principles of parsimony (minimal set of protein sequences which explain the maximum number of identified peptides).

4.4 Results

Once initial growth was established and a few generations of growth had passed, a growth curve was created to determine the growth phase of the bacteria (Table 4.7).

Table 4.7: Growth Curve for *ATCC 29413*

DAY	Replicate					
	1	2	3	4	Mean	log10
0	1.5	3.5	1.2	0.6	1.7	0.23
1	1.5	2.6	3.1	1.2	2.1	0.32
2	3.2	4	2.4	5	3.65	0.56
3	11.8	5.5	13.8	6.3	9.35	0.97
4	18.4	7.8	25.4	7.4	14.75	1.17
5	38.8	10.6	69.7	9.8	32.23	1.51
6	89.7	30.7	120.4	30.7	67.88	1.83
7	110.6	95.3	125.3	195.5	131.68	2.11
9	130.7	160.3	127.5	150.3	142.2	2.15
10	171	152.5	120.6	162.2	151.58	2.18
11	142.5	163.3	115.1	154.4	143.83	2.16
12	132.8	158.5	110.2	157.5	139.75	2.15
13	148.6	155.2	116.8	163.9	146.13	2.16
14	149.6	150.3	121.3	150.3	142.88	2.15

Up to day 8 the bacteria were still in log or exponential growth phase but they were able to maintain themselves in stationary phase for at least 6 days if not more. By day 14, they still have not reached the death phase. (Figure 4.4)

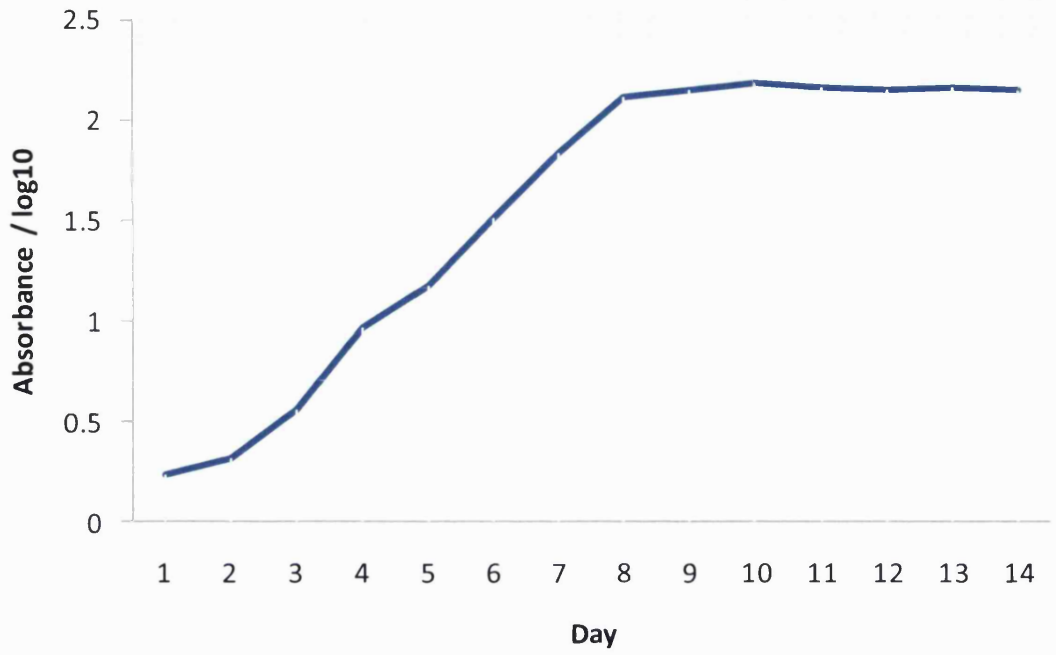


Figure 4.4: Growth Curve for *ATCC 29413*

Mid-log phase (6 days) was used for growth. Preliminary UV spectra indicated higher absorbance across the wavelengths associated with MAAs, chlorophyll and other UV absorbing materials (Table 4.8)

Table 4.8: Absorbances of various MAAs and other UV absorbing molecules

	λ	0hr*	24hr*	48hrs^	72 hour'
Mycosporine-gly	310	0.76	1.12	1.90	5.90
Palythine	320	0.77	1.53	2.38	9.30
Palythinol	332	0.85	1.88	2.98	12.72
shinorine	334	0.86	1.90	3.04	12.88
	338	0.87	1.88	3.10	12.87
palythene	360	0.86	0.56	2.06	3.30
	384	1.08	0.52	2.84	4.16
chlorophyll	434	1.64	0.95	4.69	7.13
carotenoids	473	0.95	1.12	2.23	3.97
biliproteins	618	0.34	0.12	0.81	1.12
chlorophyll	665	1.02	0.25	3.09	4.41

* recalculated for ½ dilution
^ recalculated for $^1/_3$ dilution
' recalculated for $^1/_8$ dilution

An example of a UV spectrum from 0 hr control (Figure 4.5) diluted by ½ and a sample obtained after 72 hour exposure (72 hour UV) (Figure 4.6) diluted by 1/8 show the incredible increase and changes in absorbance.

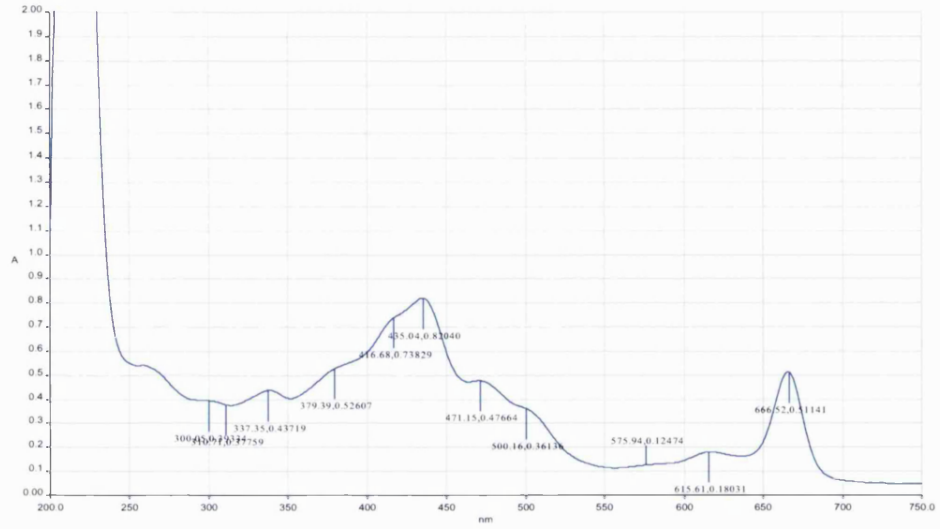


Figure 4.5: UV spectrum of a 0 hr control sample diluted by ½

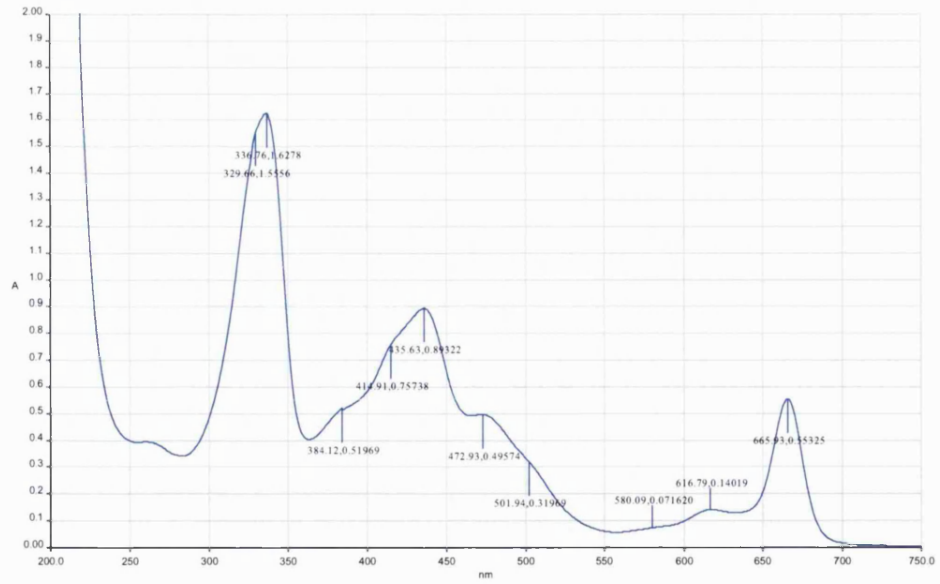


Figure 4.6: UV spectrum of 72 hour UV sample diluted 1/8.

The difference between 0 hr control and the 72 hour UV samples was a factor of 4 in dilution to be able to see the absorbance below 2 at 334 nm, the maximum of UV absorbance of shinorine.

Shinorine was extracted and purified by Waters C_{18} SepPak column which were then checked by scanning UV before being quantified by HPLC.

Once an appropriate method was created, the limit of detection and quantification were established in triplicate (Table 3.5). While the limit of detection seemed to be able to go further, 1.2 pmol was determined to be the lowest required for the scope of this analysis.

Table 4.9: Retention Times, Areas and Heights in triplicate for shinorine standards at 330 nm.

Shinorine	RT1		Height	RT 2		Height	RT 3		Height	RT / min		AREA		Height	
/ pmol	/min	Area 1	1	/ min	AREA 2	2	/min	AREA 3	3	MEAN	CV	MEAN	CV	MEAN	CV
1.2	10.44	3194	180	10.4	4678	252	10.40	4516	222	10.41	0.00	3194	0.53	180	0.17
2.4	10.28	9962	469	10.27	9795	464	10.30	9784	479	10.28	0.00	9847	0.50	471	0.02
3.6	10.22	1399	592	10.21	14825	619	10.17	13915	604	10.20	0.00	10046	0.79	605	0.02
4.8	10.14	19357	725	10.17	19791	733	10.15	20112	733	10.15	0.00	19753	0.50	730	0.01
12	10.32	38673	2244	10.31	41038	2423	10.37	41517	2406	10.33	0.00	40409	0.50	2358	0.04
24	10.30	93305	4739	10.26	92480	4709	10.29	94081	4718	10.28	0.00	93289	0.50	4722	0.00
36	10.17	147749	6275	10.20	144067	6201	10.19	141956	6149	10.18	0.00	144591	0.50	6208	0.01
48	10.10	196446	7227	10.10	201442	7301	10.14	179803	6698	10.11	0.00	192564	0.50	7075	0.05
120	10.35	977699	49104	10.30	993690	49791	10.25	994874	50303	10.30	0.00	988754	0.50	49733	0.01

At the upper end, 120 pmol did not fall within the linear range of the instrument for peak area or peak height. An upper limit of detection was also not determined, but for quantification, the upper limit was determined to be between 48 and 120 nmol.

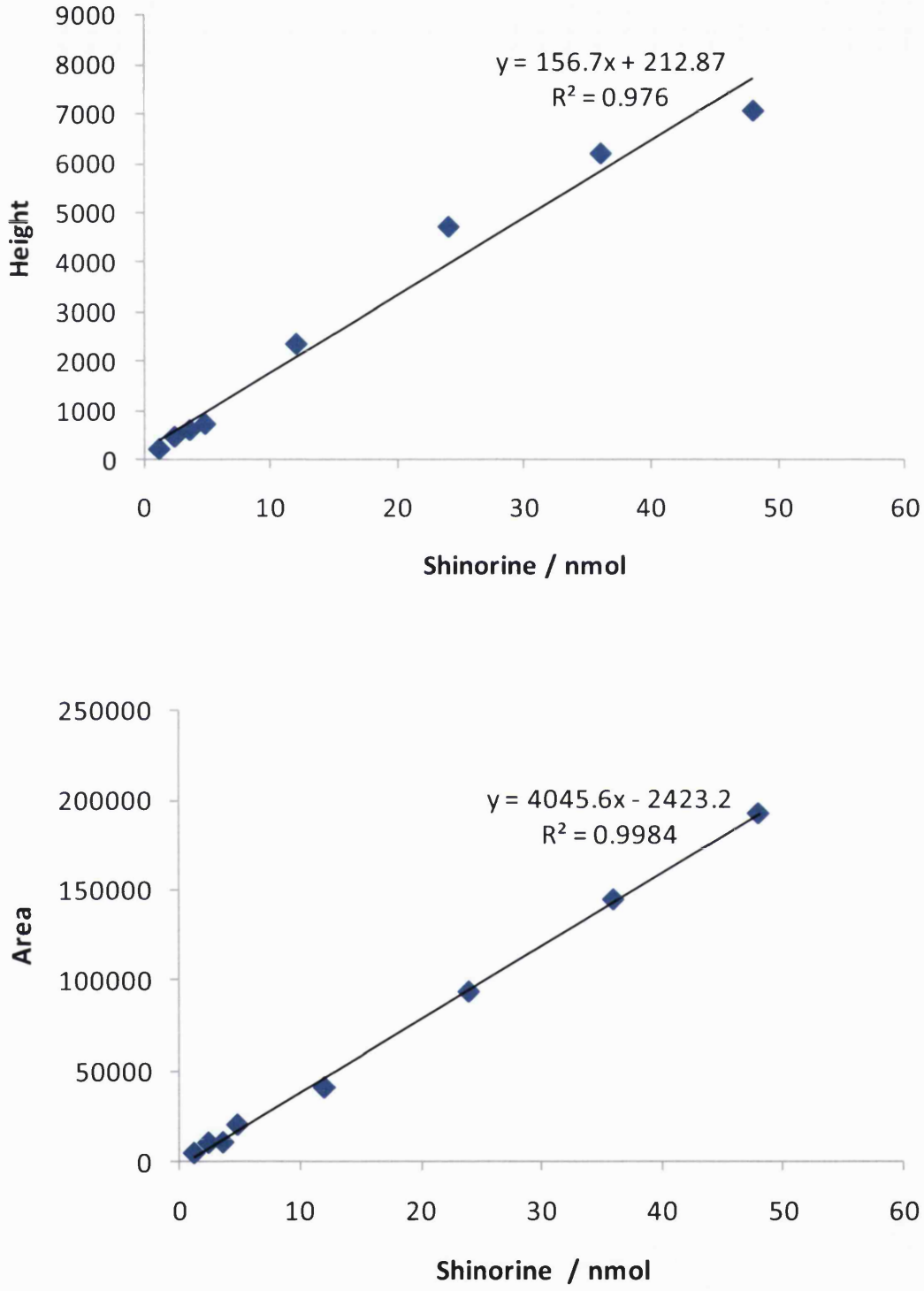


Figure 4.7: Linear curves for peak height (top) and peak area (bottom) from the calibration of shinorine at 330 nm.

System suitability of shinorine was based on 6 injections of shinorine standard and was performed on a separate day to determine the reproducibility of the method and system.

Table 4.10: System suitability of shinorine; Retention times and peak area and heights at 330 nm using 240 pmol shinorine (24 µM, 10 µl).

Shinorine	RT	AREA	HEIGHT
1	10.35	977699	49104
2	10.30	993690	49791
3	10.25	994874	50303
4	10.25	976052	49691
5	10.27	988377	49846
6	10.28	986138	49747
Mean	10.28	986138	49747
CV	0.00	0.01	0.01

The 6 replicate injections gave low CV values, showing the retention times, peak areas and peak heights are all reproducible.

Once the calibration curve was established, samples were injected onto the system (10 µL) in H_2O in triplicate to ensure good reproducibility (Table 4.11).

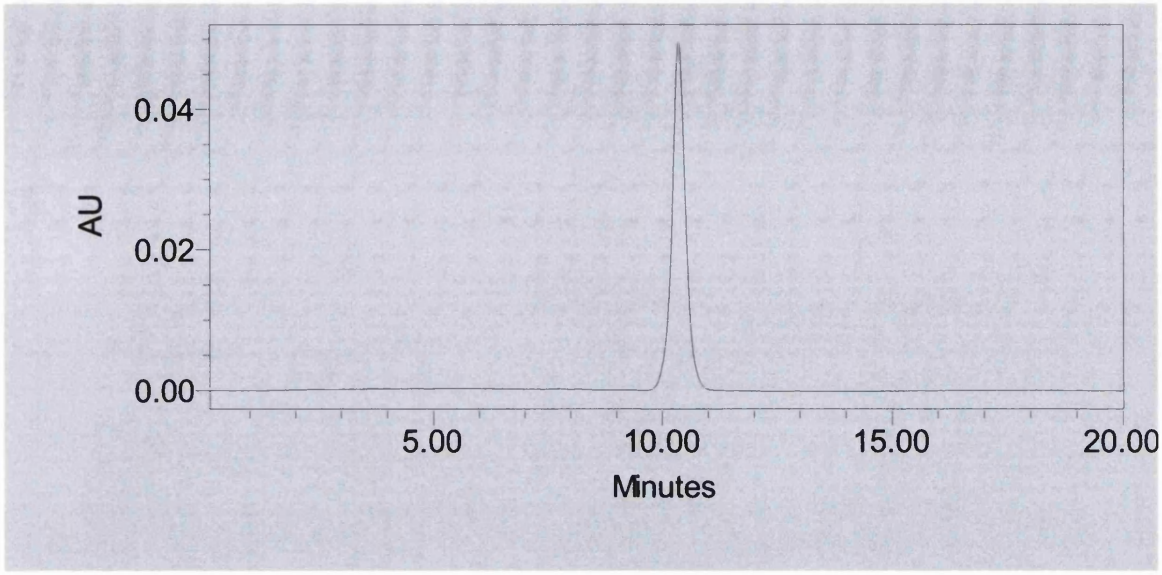
Table 4.11: Data from triplicate injections of 3 incubations of *ATCC 29413*

Incubation / hr	RT 1	Area 1	Height 1	RT 2	AREA 2	Height 2	RT 3	AREA 3	Height 3
Run 1									
72	9.90	3.54E+05	3013	9.98	3.77E+05	3881	9.52	3.88E+05	3872
48	10.16	8.19E+04	1959	10.16	8.30E+04	1945	10.16	8.22E+04	1935
24	10.13	4.85E+04	1693	10.17	3.94E+04	1633	10.21	4.99E+04	1691
0	10.15	5.64E+04	2082	10.17	5.08E+04	2811	10.16	5.00E+04	2851
0	10.20	4.59E+04	1705	10.20	4.57E+04	1732	10.21	4.59E+04	1695
Run 2									
72	9.91	1.81E+05	6987	9.96	1.80E+05	6945	9.98	1.80E+05	6712
48	10.21	9.49E+04	1977	10.32	9.48E+04	1947	10.20	9.49E+04	1912
24	10.22	5.12E+04	1730	10.15	5.26E+04	1722	10.12	5.19E+04	1734
0	10.11	2.56E+04	1444	10.13	2.56E+04	1381	10.12	2.55E+04	1399
Run 3									
72	9.98	2.09E+05	9402	9.98	2.03E+05	9287	9.97	2.09E+05	9332
48	10.15	1.22E+05	1956	10.18	1.21E+05	1988	10.17	1.21E+05	1957
24	10.26	2.14E+04	1553	10.19	2.13E+04	1539	10.21	2.14E+04	1528
0	10.12	1.92E+04	1639	10.15	1.92E+04	1676	10.13	1.93E+04	1663

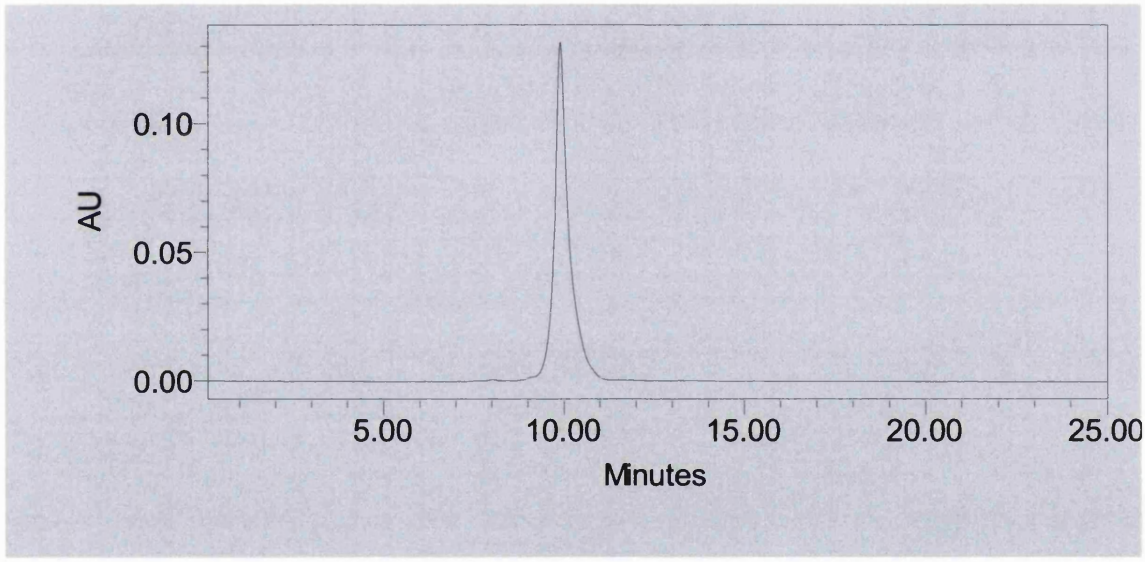
Table 4.12: Results from triplicate injections of 3 incubations of *ATCC 29413*

Incubation	RT		AREA		Height		Shinorine	Final	Cell dry weight	
/ hr	MEAN	CV	MEAN	CV	MEAN	CV	pmol	nmol	g	nmol/g cell
Run 1								calc.		
72	9.80	0.03	3.73E+05	0.50	3589	0.14	92.75	9.28	0.062	149.60
48	10.16	0.00	8.24E+04	0.50	1946	0.01	20.96	2.10	0.041	50.88
24	10.17	0.00	4.59E+04	0.51	1672	0.02	11.95	1.19	0.039	30.48
0	10.16	0.00	5.24E+04	0.50	2581	0.17	13.55	1.36	0.056	24.20
0	10.20	0.00	4.58E+04	0.50	1711	0.01	11.93	1.19	0.046	25.93
Run 2										
72	9.95	0.00	1.80E+05	0.50	6881	0.02	45.15	4.52	0.054	83.61
48	10.24	0.01	9.49E+04	0.50	1863	0.06	24.05	2.41	0.047	51.17
24	10.16	0.01	5.19E+04	0.50	1729	0.00	13.43	1.34	0.040	33.24
0	10.12	0.00	2.56E+04	0.50	1408	0.02	6.92	0.69	0.029	23.51
Run 3										
72	9.98	0.00	2.07E+05	0.50	9340	0.01	51.82	5.18	0.035	148.06
48	10.17	0.00	1.21E+05	0.50	1967	0.01	30.60	3.06	0.036	85.66
24	10.22	0.00	2.14E+04	0.50	1540	0.01	5.88	0.59	0.021	27.38
0	10.13	0.00	1.92E+04	0.50	1659	0.01	5.35	0.54	0.021	26.03

The mean and CV were calculated for the retention times, areas and heights. The number of moles were calculated per gram weight for comparison to literature values. Calculations were performed using the linear equation determined for peak area (Figure 4.7). The amount was then multiplied by 10 and then final moles were calculated based on the dry weight of the bacteria. An example of HPLC at 330 nm of shinorine standard and sample (Figure 4.8)



Standard



0 hr Control

Figure 4.8: Standard(upper) and a sample from *ATCC 29413* shinorine (lower) from the HPLC at 330 nm.

An attempt was made to analyse shinorine by nano-ESI. Initially the standard of shinorine was analysed for comparison and was even fragmented using tandem MS.

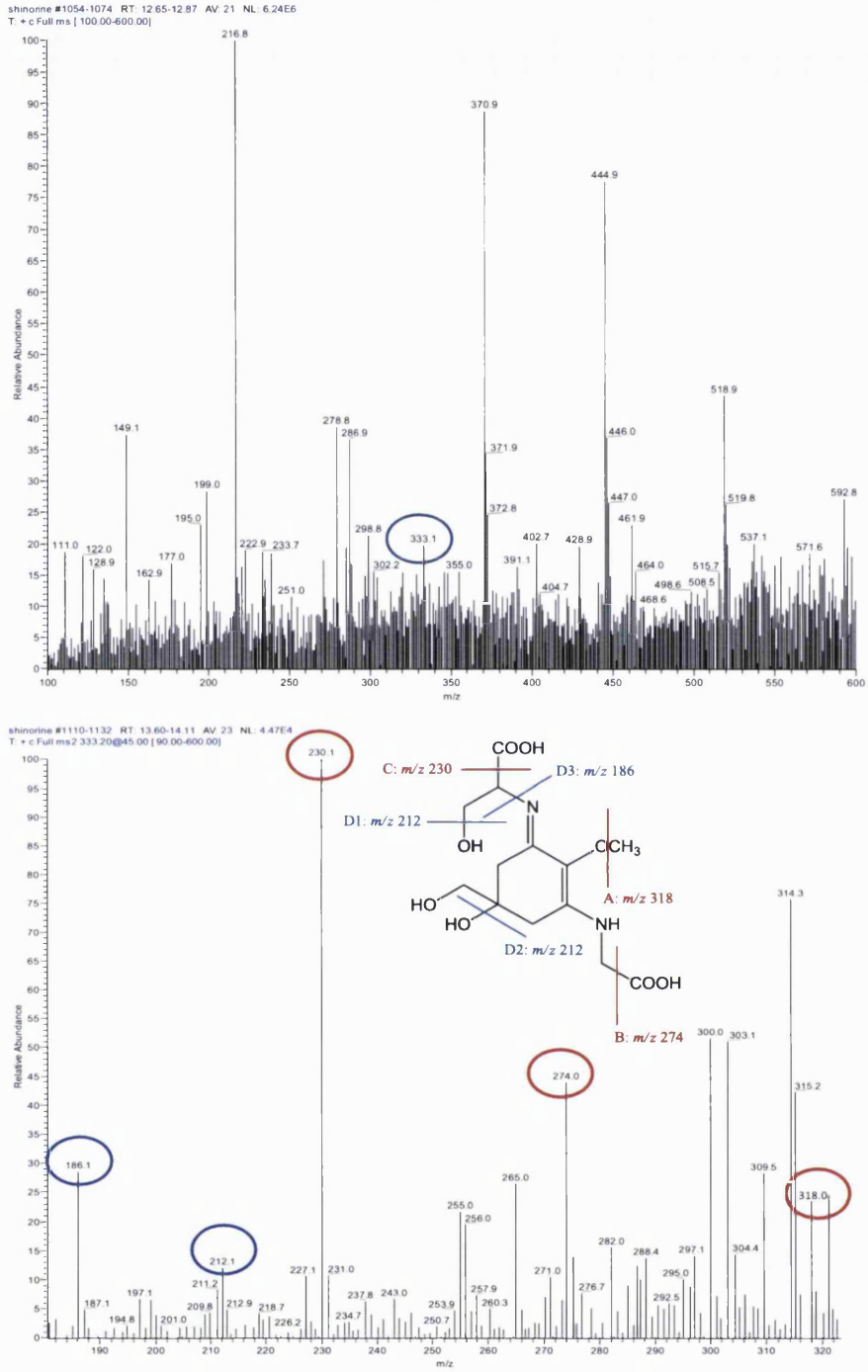


Figure 4.9: MS Spectra of standard shinorine acquired using the LCQ . Top: MS spectra showing 333, Bottom: MS/MS of 333 at 40 eV. Circled peaks are those that correspond to the peaks identified in previous studies. Initial loss of a radical methyl group A (red), is followed by a decarboxylation B (red) and another decarboxylation C (red). D1 and D2 (blue) are formed by loss of water from C while D3 (blue) is formed by the loss of a trimethylvinyl alcohol from C.[190, 191]

The spectrum was compared to a spectrum of the sample.

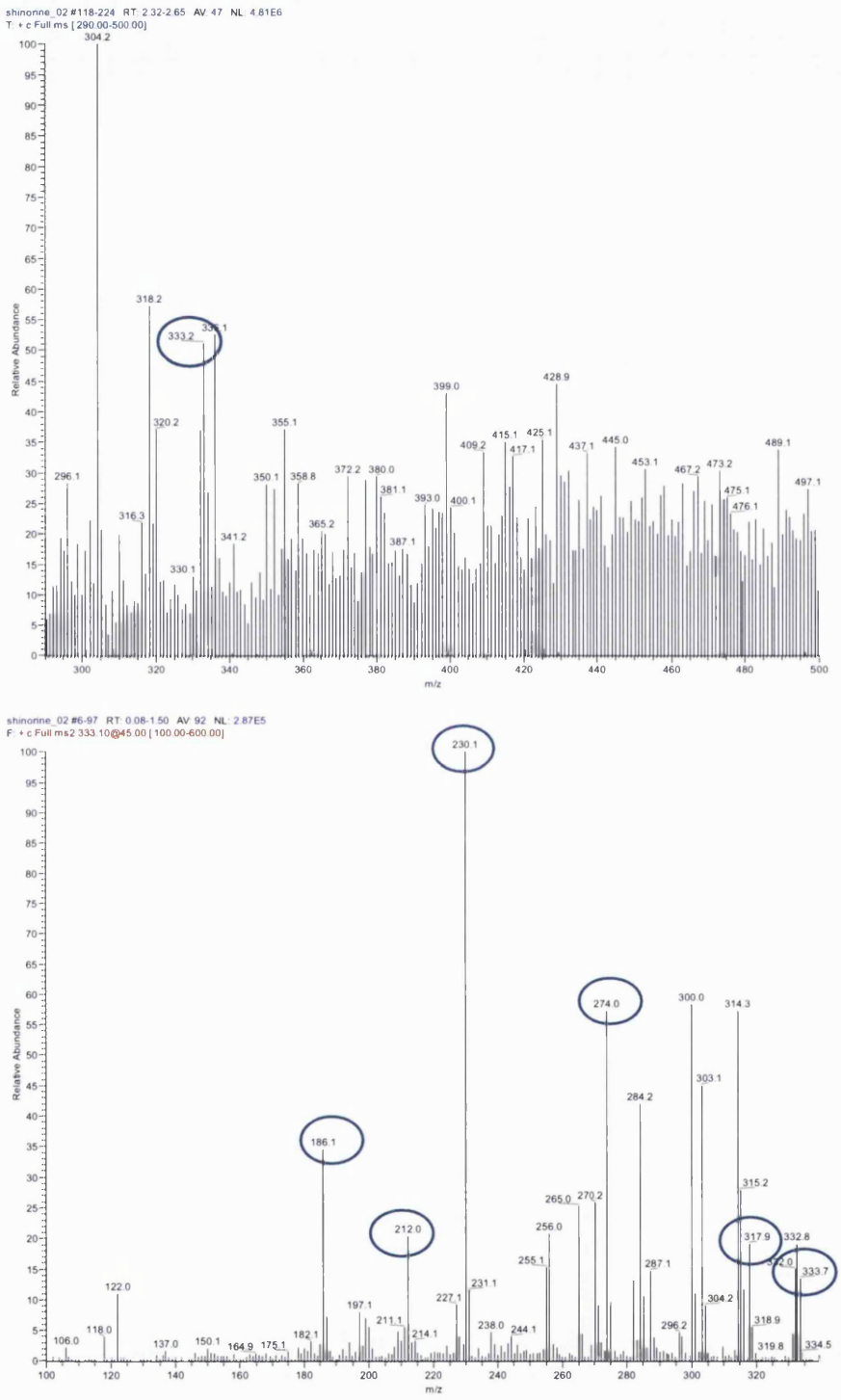


Figure 4.10: Sample 72 hour analysis of shinorine by LCQ. Top: MS spectra showing 333. Bottom: MS/MS of 333. Circled peaks are those that correspond to the peaks identified in Figure 4.9.

A comparison was made between literature[190, 191] MS/MS and the standards and samples run on the LCQ showing great similarity in the MS/MS pattern observed. A similar

comparison was made between literature[190, 191] MS/MS data and results obtained from the QTOF.

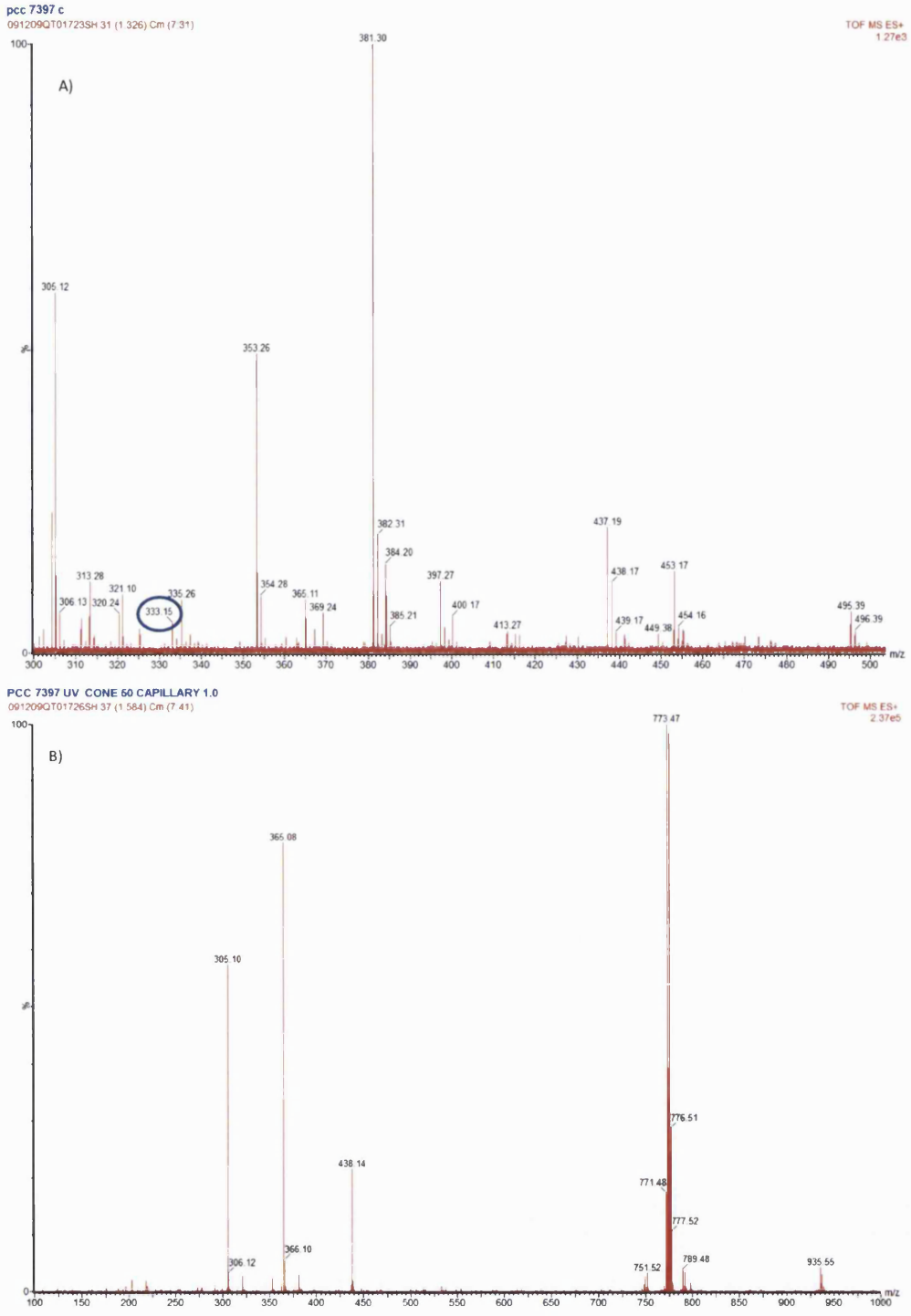


Figure 4.11: QTOF analysis of 0 hr control and 72 hr UV samples. A) MS spectra of 0 hr control sample. B) MS spectra of 72 hr UV sample. (Continued on the next page)

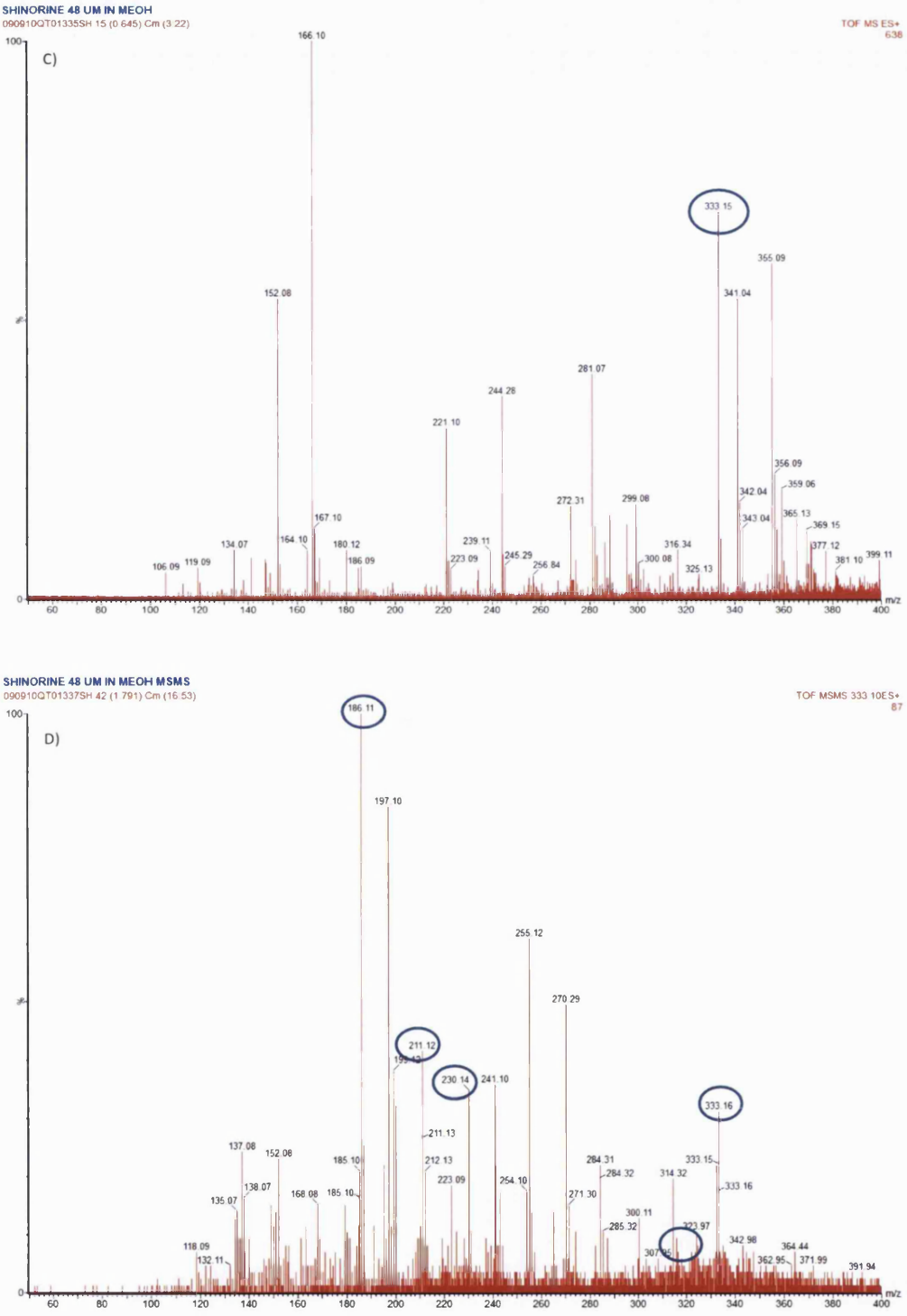


Figure 4.11: QTOF analysis of 0 hr control and 72 hr UV samples. C) MS spectra of Standard and D) MS/MS spectra of *m/z* 333 of standard.

The fragmentation data acquired on the QTOF was not sufficient to identify the fragments reliably. It was noted that concentrations required to do analysis by mass spectrometry

were much higher than those required for HPLC-UV experiments at greater than the 12 pmol/µL at which our standard was given.

At the same time as analysing the shinorine samples, protein samples from the bacteria were prepared and extracted for gel electrophoresis and proteomics analysis by QTOF. Concentration of protein was established using BSA as a standard in the microplate assay (Figure 4.12).

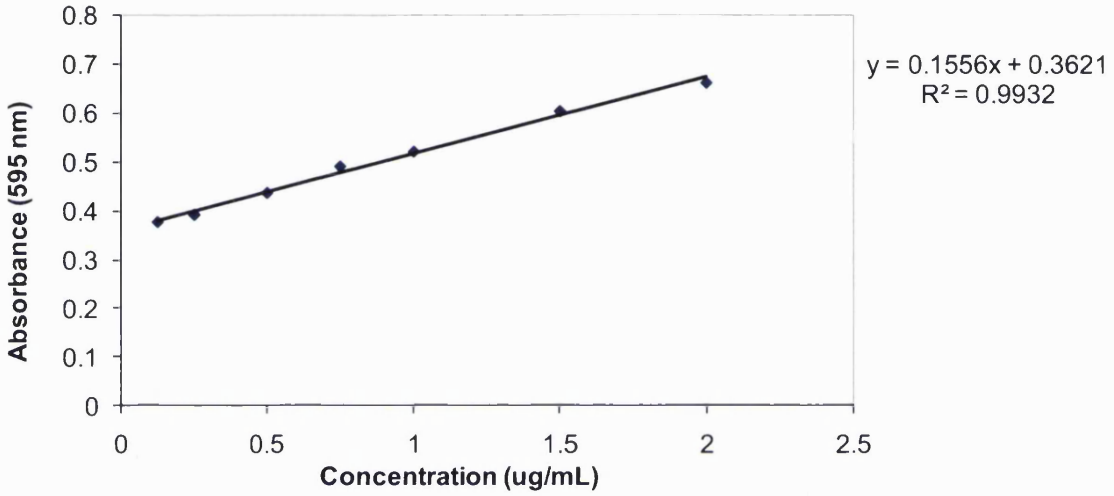


Figure 4.12: Protein Assay Curve for determination of protein content in fractions.

This calibration curve was used to calculate the concentration of protein in each fraction and sample (Table 4.13, Table 4.14)

Table 4.13: Protein concentration determined for fractions of *ATCC 29413*-1.

Sample	Protein Content µg/µL	Volume of Sample / mL	Total Protein Content / µg	Total Protein Loaded onto Gel 1 / µg
Control soluble	10.819	1.5	16228.5	25.1
Control insoluble	2.226	1.5	3339	25.0
UV soluble	8.909	1.5	13363.5	24.99
UV insoluble	2.139	1.5	3208.5	24.99

Table 4.14: Protein concentration determined for fractions of *ATCC 29413*-2

Sample	Protein Content µg/µL	Volume of Sample / mL	Total Protein Content / µg	Total Protein Loaded onto Gel 1 / µg
Control soluble	10.819	1.5	16228.5	25.1
Control insoluble	2.226	1.5	3339	25.0
UV soluble	8.909	1.5	13363.5	24.99
UV insoluble	2.139	1.5	3208.5	24.99

Each fraction was loaded onto the gel at the same concentration (Figure 4.13). Protein concentrations loaded onto the gel were normalised to the lowest concentration.

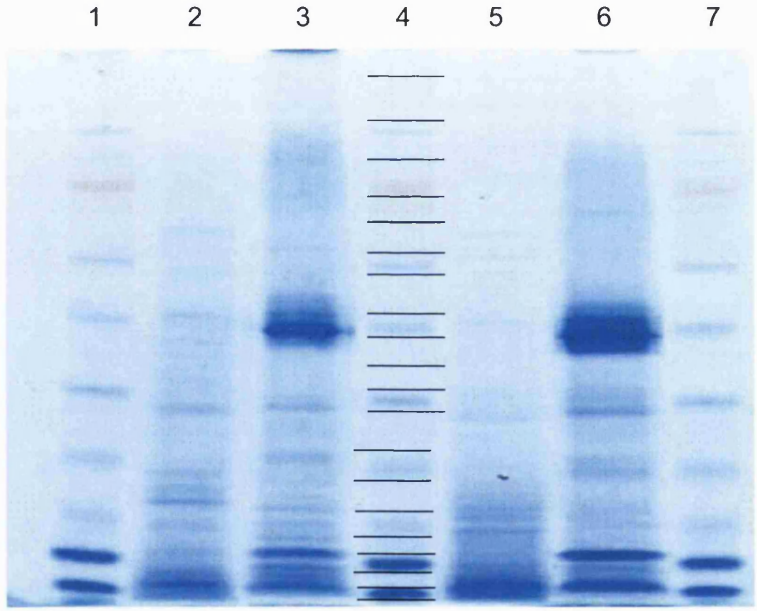


Figure 4.13: Sample separation of *ATCC 29413* by SDS-PAGE. Lanes 1,4, and 7 are standards. Lanes 5 and 6 are 0 hr control soluble and insoluble while lanes 2 and 3 are 72hr UV soluble and insoluble respectively. Black bands on lane 4 describe the bands that were cut.

Each sample (0 hr control or 72 hr UV) was extracted into a soluble and insoluble fraction, both of which were loaded and run on the SDS-PAGE. The second gel was also run and the identified lanes with proteins of interest were cut destained and digested (Figure 4.14).

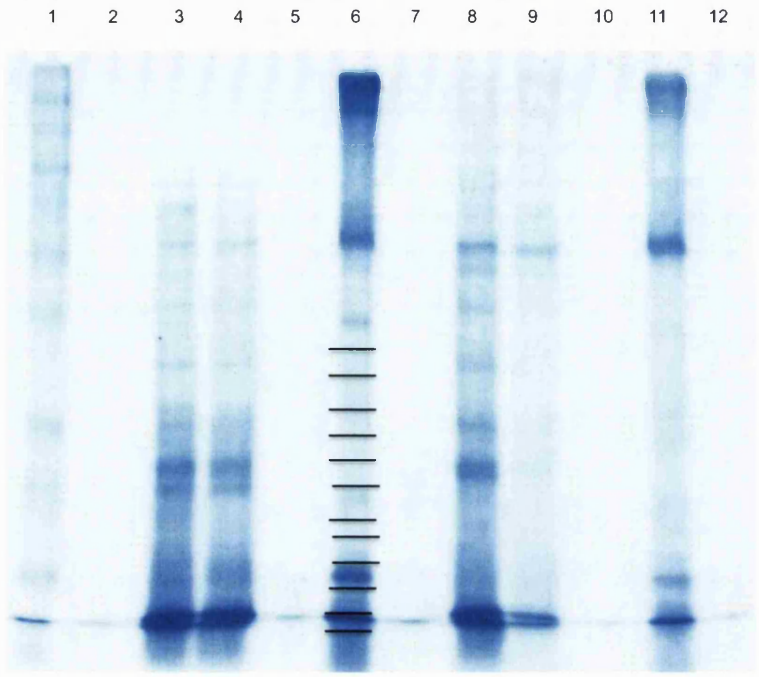


Figure 4.14: Second run of sample. Lanes 1, 6, 11 are standards, while Lanes 3 and 4 are Control 0hr soluble and insoluble fractions and lanes 8 and 9 are 72 hr UV soluble and insoluble.

Soluble fractions were blue while insoluble fractions were green. Samples were loaded and run on 4-12 % gels. The first gel (Figure 4.13) had the lanes cut into 20 bands digested, extracted and analysed by LC-MS/MS. Data was then scrutinised by 2 different database search platforms: X!Tandem and PLGS. An initial list of proteins that potentially may be part of the shinorine synthesis pathway was created using both databases (Table 4.15).

Table 4.15: Initial list of interesting proteins that could be part the pathway to shinorine synthesis.

Search engine	Protein description	Search engine
PLGS		X!Tandem
YP_320824.1	NADH dehydrogenase	gi|75906528|
YP_323679.1	branched-chain alpha-keto acid dehydrogenase subunit E2	gi|75909383|
YP_323161.1	inositol-5-monophosphate dehydrogenase	gi|75908865|
ABA22597.1	3-isopropylmalate dehydrogenase	gi|75909196|
YP_322444.1	dihydrolipoamide dehydrogenase	gi|75908148|
YP_323037.1	6-phosphogluconate dehydrogenase .	gi|75908741|
YP_322999.1	Iron-containing alcohol dehydrogenase	gi|75908703|
YP_320824.1	NADH dehydrogenase I subunit M	gi|75906528|
YP_322593.1	serine hydroxymethyltransferase	gi|75908297|
YP_322827.1	biotin carboxyl carrier protein	gi|75908531|
ABA23984.1	3-dehydroquinate synthase	gi|75910583|
YP_325141.1	3-deoxy-7-phosphoheptulonate synthase	gi|75910845|

These proteins were considered probable candidates to be part of the synthesis of shinorine. From this list, 2 proteins were found that directly related to what is known about shinorine synthesis (Table 4.16).

Table 4.16: Proteins of the shikimate pathway found only in UV treated bacteria identified directly from the proteomic analysis of control vs UV samples of *Anabaena variabilis*

Protein Identifier	Biochemical Function	Protein Name	Lane/Band
ABA23984.1	Synthase, phosphatase	3-dehydroquinate synthase	L2 B5
YP_325141.1	Synthase	3-deoxy-7-phosphoheptulonate synthase	L2 B9

An example of the MS/MS data from the shinorine proteins identified is one of the peptides from 3-deoxy-7-phosphoheptulonate synthase (Figure 4.15).

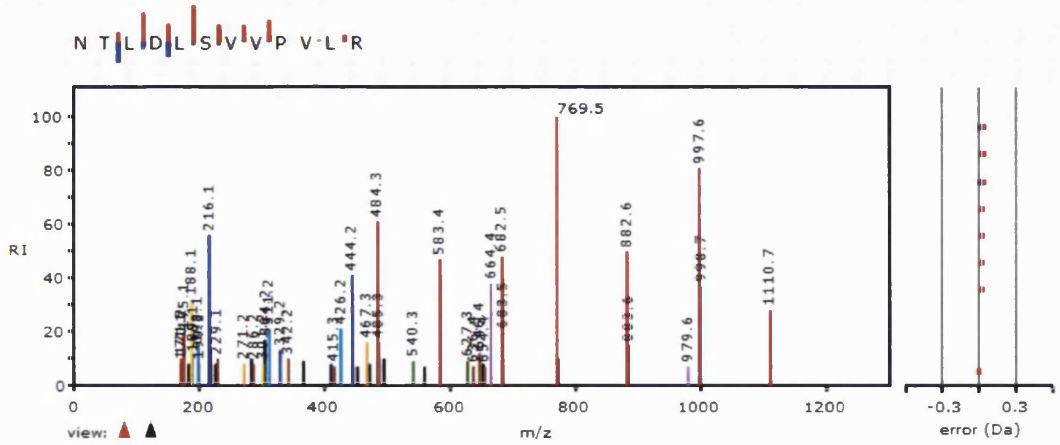


Figure 4.15: Fragmentation pattern of 3-deoxy-7-phosphoheptulonate synthase from X!Tandem search. Red ions denote y-ion series and dark blue denote b-ion series.

The bands with proteins of interest were cut from the second gel which was searched using both search engines in the same fashion as the bands from the first gel. In total, 1226 individual protein accession numbers were identified from the two proteins gels.

Table 4.17: Total number of individual protein accession numbers identified from *ATCC 29413*.

	PLGS	X!Tandem	Both
Control	98	86	98
UV	111	144	56
Both	253	216	180
Total	462	446	334

The proteins identified in the search against *ATCC 29413* genome were compared to those present in database for the organism *A. microphthalm* to identify proteins that may be involved in the production of shinorine. Homology among proteins is very common and can be used to find proteins with similar activity among different organisms. While the genome database of *ATCC 29413* does not show any proteins identified as specific to the synthesis of MAAs, the database from *A. microphthalm* does, so it can be used to look for homology between proteins and similar function. A homology search of the all the data against *A. microphthalm* database resulted in 7 proteins which may be involved in the synthesis of mycosporine like amino acids.

Table 4.18 Summary of identified shikimate pathway proteins from *Anabaena variabilis* using *A. microphthalm.*

Protein Identifier	Biochemical Function	Protein Name
YP_322169.1	Aminomethyl transferase	Glycine cleavage T protein
YP_323854.1	Dehydrogenase	3-beta hydroxysteroid dehydrogenase/isomerase
YP_322999.1	Dehydrogenase	Iron-contaning alcohol dehydrogenase
Q9XBL7.1	Oxidoreductase	NAD(P)H-quinone oxidoreductase subunit K
Q3MBD4.1	Oxidoreductase	NAD(P)H-quinone oxidoreductase subunit L
ABA24357.1	Oxidoreductase	Flavin oxidoreductase/ NADH oxidase
YP_321467.1	Hydrolase	NUDIX hydrolase

The number of proteins identified was spread evenly between the two search engines as well as between both of them. A second run on a gel was performed and the bands corresponding to the proteins of interest were cut out. (Figure 4.14). The proteins were analysed as described above and submitted for LC-MS analysis and database searching (Table 4.19).

Table 4.19: Proteins identified in SDS-PAGE-2 of *ATCC 29413*.

Protein	1st run	2nd
3β-hydroxysteroid dehydrogenase isomerise	many	many
Fe-containing alcohol dehydrogenase	L2 B9/10	L2 B3
NADPH quinone oxidoreductase	many	L8 B9
3-dehydroquinate synthase	L2 B9	x
3-deoxy-7-phosphoheptulonate synthase	L2 B9	X
Flavin oxidoreductase/ NADH oxidase	L5 B9	L2 B5
NUDIX hydrolase	L2 B6	L2 B9
UvrB/UvrC	L2 B2-6	L2 B3
Glycine cleavage T protein	L2 B3	x
NADH dehydrogenase I	many	many
branched-chain alpha-keto acid dehydrogenase subunit E2	L3 B6	L8 B9
inositol-5-monophosphate dehydrogenase	L2 B8-10	L8 B9
3-isopropylmalate dehydrogenase	L2 B10	x
dihydrolipoamide dehydrogenase	L2 B12	x
6-phosphogluconate dehydrogenase .	L2 B2	x
serine hydroxymethyltransferase	L2 B12	x
biotin carboxyl carrier protein	L2 B2	L8 B1/L9 B1
UvrB/UvrC	many	many

Table 4.20: Proteins of interest found from analysis using both databases.

	PLGS	X!Tandem	Both
Control	0	0	0
UV	9	9	9
Both	9	9	9
Total	18	18	18

Of those 1226 proteins, 23, or 1.8 % of the total proteins were considered potentially relevant to the shinorine pathway (Table 4.20). A final list of accession numbers identified by X!Tandem, PLGS and both databases were created and can be found in Appendix B.

A third search was performed using MASCOT and the data was validated using SCAFFOLD. The resulting list of proteins found are in Table 4.21 and examples of the data can be found in Appendix B.

Table 4.21: Results from searching shikimate pathway enzymes from MASCOT and SCAFFOLD.

#	Visible?	Identified Proteins (1680)	Accession Number	Molecular Weight	control	control	control	UV	UV	UV
Shikimate pathway enzymes										
1407	TRUE	AROE_RUBXD Shikimate dehydrogenase OS=Rubrobacter xylanophilus (strain DSM 9941 / NBRC 16129) GN=aroE PE=3 SV=1	Q1AWA3	29 kDa	0	0	39%	0	0	0%
1623	TRUE	AROE_SHESM Shikimate dehydrogenase OS=Shewanella sp. (strain MR-4) GN=aroE PE=3 SV=1	Q0HP98 (+1)	31 kDa	0	0	30%	0	0	0
151	TRUE	AROB_FUSNN Bifunctional 3-dehydroquinate synthase/phosphatase OS=Fusobacterium nucleatum subsp. nucleatum GN=aroB PE=3 SV=1	Q8RF47	77 kDa	0%	69%	71%	60%	33%	60%
457	TRUE	AROB_ANAVT 3-dehydroquinate synthase OS=Anabaena variabilis (strain *ATCC 29413* / PCC 7937) GN=aroB PE=3 SV=1	Q3M4V2	39 kDa	0	31%	0	52%	0%	0
1483	TRUE	AROA_SHESR 3-phosphoshikimate 1-carboxyvinyltransferase OS=Shewanella sp. (strain MR-7) GN=aroA PE=3 SV=1	Q0HV11	46 kDa	0	0	0	0	0%	36%
1013	TRUE	AROD_STAEQ 3-dehydroquinate dehydratase OS=Staphylococcus epidermidis (strain *ATCC* 35984 / RP62A) GN=aroD PE=3 SV=1	Q5HQR7 (+1)	27 kDa	0	0	0	52%	0%	0
1338	TRUE	PCHA_PSEAE Salicylate biosynthesis isochorismate synthase OS=Pseudomonas aeruginosa GN=pchA PE=3 SV=1	Q51508	52 kDa	0	0	42%	0	0	0
1643	TRUE	AROC_PASMU Chorismate synthase OS=Pasteurella multocida GN=aroC PE=3 SV=1	P57840 (+1)	38 kDa	29%	0	0%	0%	0	0%

From the search using MASCOT and SCAFFOLD, 8 proteins were identified from the shikimate pathway in a mixture of 0 hr control and 72 hr UV samples out of a total of 1680 proteins identified.

4.5 Discussion

Cyanobacteria, *Anabaena variabilis ATCC 29413*, are filamentous bacteria that grow in sea water and are capable of photosynthesis. Once their initial growth rate was established, a growth curve was determined (Table 4.7, Figure 4.4) with a log phase of 8 days and a stationary phase from 8 to 14 days, giving the bacteria a very long life which is very good for growth and extraction. Using a source of UVB light (12 hours light, 12 hours dark), the cyanobacteria were allowed to grow for 72 hours at which point shinorine was extracted. The high levels of shinorine after 72 hours were suggested in a paper by Singh *et al* (2008). The production of shinorine was confirmed by UV and was readily quantified by HPLC-UV as shinorine and other members of the group of mycosporine-like amino acids are UV absorbing (Table 4.8).

Initial analysis of the control sample and the UV induced sample showed an increase in absorbance across the UV spectrum from 200 – 750 nm (Figure 4.5, Figure 4.6) some of which may be attributed to shinorine and other molecules listed in Table 4.8. The estimated difference between the 0 hr control and 72 hr UV sample is at least a 4X increase in absorbance, which was better quantified using HPLC.

A calibration curve was created using a purified shinorine generously donated by Malcolm Shick from the department of biological sciences, marine sciences, University of Maine. Initial attempts to use a quantification method created by Shick *et al* (1997) proved unsuccessful as the shinorine was not retained on the available columns. Using an isocratic conditions using H_2O + 0.02% acetic acid with available columns was developed based on a recently reported method.[184,187]

The linear part of the calibration curve fell between 1.2 and 48 pmol. The calibration curves were generated for retention time, peak area and peak height in triplicate (Table 3.5, Figure 4.7) and system suitability was performed on a separate day showing good reproducibility of retention times, peak areas and peak heights (Figure 4.10). Samples were extracted to remove most other small molecules and fractions containing shinorine were identified by obtaining UV data prior to HPLC analysis. Samples were quantified by HPLC with sample 72 hour UV, requiring adequate dilution for reliable quantification. There is variability between the amount of shinorine quantified (Table 4.11, Table 4.12) however this could be due to the location of samples under the lamp as well as other factors relating to bacterial growth or extraction. Results are lower than those found by Shick *et al* (2008), as well as lower for those from Singh *et al* (2008) for 72 hours. In each case the levels of shinorine increased with increasing UV

absorbance. The samples showed no other molecules absorbing at 330 nm (Figure 4.8).

Nano-ESI on the LCQ and QTOF were performed using single use borosilicate tips. Shinorine standard initially provided in a solution was dried down and the concentration required to detect the shinorine by MS analysis was at least 24 pmol/µL. Shinorine was dissolved in MeOH + 0.02 % FA for both QTOF and LCQ. Shinorine standards were run on the LCQ (Figure 4.10) and QTOF (Figure 4.11 C & D) which showed that while the sample was clean by HPLC, it had much higher baseline noise by MS. Tandem MS data of the standards by LCQ (Figure 4.9) and QTOF (Figure 4.11 C& D) were not comparable to each other however; the tandem MS data from the LCQ standard (Figure 4.9) was comparable to the 72 hour UV sample (Figure 4.10). This in turn was compared to MS spectrum from Whitehead[191] and Cardozo[190] showing the common fragment patterns starting with a radical loss of $\cdot CH_3$ followed by continual losses of CO_2 and H_2O. In both cases, the fragmentation of even the pure shinorine is more complicated than those reported by either group, making it difficult to determine the fragmentation pattern. Unfortunately as seen by the MS spectra for the 72 hour sample (Figure 4.9, Figure 4.11 A & B), there are a lot more ions present in the sample. In the case of the QTOF (Figure 4.11 A & B) analysis, shinorine was not well detected to allow for MS/MS despite the fact it is the same sample as was successfully analysed by the LCQ (Figure 4.10).

With induction of shinorine confirmed, the protein extractions which were carried out in tandem with the shinorine experiment were run on a SDS-PAGE gel. The equivalent amounts of 0 hr control and 72 hr UV proteins (Table 4.13, Table 4.14) were loaded onto the gel as determined by Bio-rad protein assay (Table 4.3, Figure 4.12). Gradient gels of 4-12 % were run to increase band resolution. The initial gel had each lane cut into 20 bands, all of which were cut, destained, digested with trypsin and extracted. All samples were submitted for analysis on the QTOF mass spectrometer using a peptide C_{18} column and a run time of 63 minutes per sample. The data was searched using two databases searching platforms X!Tandem and PLGS, and later by MASCOT by another group at King's College, University of London.

First two search engines were used because neither had previously been used before on this system and it was unclear how they would perform. PLGS 4.1 was bought with the Waters QTOF Global Ultima™. The search engine was setup on a local server and databases can be uploaded as long as they have been put into a standard FASTA

format. In this case, the FASTA format from the NCBI database needed to be modified according to Waters suggestions for adding databases to the program.

X!Tandem was used on a web server which allowed for searching of a specific databank, in this case, cyanobacteria *Anabaena variabilis ATCC 29413*. Although this search engine can be setup on a local server as well, the web version had a software setup so that specific databases could be searched, and only searching parameters had to be modified. It was evaluated if X!Tandem would be appropriate to analyse the data obtained in this study.

Searching using PLGS could be done by one of two ways: using raw data or using a prepared pkl file which is simply a merged text file separating the individual peptides using a line, but including which each peptide, it's *m/z*, intensity and charge with a list of MS/MS ions produced by that peptide. This text file can be searched using most protein search engines. The raw data was treated using the same parameters to create the pkl file in order to be able to compare the two database searches. The raw data was smoothed and centred on the ions from the MS/MS data.

The setup for PLGS is far more simple than that for X!Tandem which gives far more variables, but the parameters were setup to ensure that searches are comparable. Both datasets were then manually checked and lists were prepared for each search. The searches were compared to each other to create a list of proteins that were identified from each search engine and a list of proteins that were identified from both search engines. Table 4.17 lists the total number of proteins found.

Unfortunately, NCBI protein database is redundant so while there appears to be different proteins between X!Tandem and PLGS, it can only be stated that different accession numbers were identified as their correlation may not be across and one search engine may chose one accession number while the second may chose another accession number for the same protein. The first pass analysis managed to identify a few proteins that could potentially be part of the shinorine pathway, such as isomerases, methylases, and dehydrogenases. Bacteria will have proteins that have more than one function. A separate group did homology searching of the genome from *Anabaena variabilis* against the proteins identified form the shinorine pathways in *A microphthalm*. Table 4.16 shows the confirmed proteins identified that were found in *A. microphthalm* and cross referenced to *Anabaena variabilis.* In total between the potential protein of interest, the identified proteins from the pathway and the homology searches, only 23 of 1226 accession numbers which were identified which may be part

of the synthesis of shinorine (Table 4.19, Table 4.20). When compared to the number of proteins identified using MASCOT and SCAFFOLD on the same data sets, a total of 1680 proteins were identified with 2 or more peptides and from these, 8 proteins were found that belong to the shikimate pathway with high confidence (Table 4.21). Barrios-Llerena (2006), performed shotgun proteomics on peptide digests and separated by SCX into 66 SCX fractions giving 45072 MS/MS of 646 proteins. In this case, 341 proteins had more than one peptide per identification, showing similar number of proteins identified. Ehling-Schulz *et al* [192] used terrestrial cyanobacteria and compared 2D gel electrophoresis spots of 0hr control and 72 hr UV treated bacteria. A comparison of the data obtained shows that 9 out of the 11 proteins identified by Ehling-Schulz were also identified by the 1-DE method here, confirming the changes in expression for those 9 proteins.

Data was searched for the EVS protein, however it was not seen in any of the searches that were performed while the DHQ in the shikimate pathway was identified in both 0 hr control and 72 hr UV samples. A comparison of the proteins found using X!Tandem and PLGS(Table 4.20) versus MASCOT/SCAFFOLD (Table 4.21) show an agreement in the identification of DHQ synthase, however the other confirmed protein found from PLGS, 3-deoxy-7-phosphoheptulonate synthase was not confirmed by the MASCOT search. Nor were the rest of the MASCOT proteins confirmed by comparing to the X!Tandem/PLGS search.

The discussion with respect to the protein cluster responsible for MAA synthesis has yet to be fully answered. While Balskus and Walsh believe it to be the cluster involved in EVS for the pentose phosphate pathway, cyanobacteria have a DHQ synthase, and a homolog of DHQ synthase. In *Styolophora pistillata* this homolog uses Mn^{2+} instead of Co^{2+} which may also hold true in cyanobacteria. The data presented here shows evidence for only one DHQ synthase present in sufficient concentration to be identified which follows the argument that the shikimate pathway and the work of Shick, Dunlap, Long and Bandarayanake is the method in which MAAs are biosynthesised.

4.6 Conclusion and Future work

The cyanobacteria *Anabaena variabilis ATCC 29413* in their shallow sea water environment have an ability to defend themselves from the powerful rays of the sun through the production of a mycosporine-like amino acid, shinorine which has absorbance at λ_{max} at 334 nm. This ability to produce a molecule that can prevent UV stress or oxidative stress is especially important to prevent DNA damage on a cellular level. Harvesting this ability has many potential uses. It was determined that the bacteria were able to produce between 200-1400 nmol shinorine per gram bacteria at 72 hours induction[173,174,184,187,193] which was confirmed by the results of the HPLC performed on shinorine in this study. As with the studies previously performed, there is variability between incubations. Extraction of shinorine proves to be quite easy as shinorine is very hydrophilic. Purification was performed using isocratic HPLC which does not show any other molecules absorbing at 330 nm. Analysis by mass spectrometry however, shows a lot more present in the mixture.

Identification of the proteins involved in shinorine synthesis proved to be more difficult than the quantification of shinorine itself. A total of 1226 individual accession numbers were identified but out of these 1226, only 23 were considered potential proteins in the biosynthetic pathway, and some were identified in the 0hr control samples while others were identified in the 72 hr UV samples. Searching of the same data by the bioinformatics groups at King's College, University of London resulted in 8 potential proteins or homologs that are related to the shikimate pathway being identified with only one in agreement with the results obtained from PLGS and X!Tandem. The EVS protein was not found by any searches.

The belief is that small organisms have proteins with multiple functions instead of separate proteins for each step. Proteins such as dehydrogenases, NADPH oxidase/reductases, flavin reductases as well as steroid dehydrogenases and isomerases have all been suggested as possibilities for identification. However in some cases, a repeat of the experiment did not confirm the presence of the proteins found in the original analysis.

The argument about the pathway MAAs are biosynthesised follows two courses: Balskus and Walsh believe that the pathway involves DHQ synthase homolog in a cluster with an O-methylation enzyme from the pentose phosphate pathway, while Shick, Dunlap, Long and Bandaranayake believe that the shikimate pathway and probably cluster of proteins are involved in MAA biosynthesis. At this point, while the

results show only enzymes from the shikimate pathway, the data is more supportive of Shick and Dunlap's theory.

In the future work, better separation is most certainly needed to improve coverage and increase the number of proteins identified, or use a 2-dimensional method such as 2-DE of 0 hr control and 72 hr UV treated to single out areas of difference instead of focusing on the whole proteome. A 2D-HPLC method would allow for more identification across the proteome which could also potentially find differences in the control vs. treated but would also increase the number of proteins and data created which increases the amount of time it would take to get through the data.

In order to prove which pathway shinorine is synthesised as both groups have valid arguments and data, mutants of the genes of interest need to be constructed. Only by eliminating the DHQ synthase and DHQ synthase homolog individually can it be determined which pathway is the major contributor towards shinorine.

Two other groups of proteins that would be of interest is the family of superoxide dismutases and NADPH: quinine oxidoreductases, some of which were found in the general data (Appendix B). These are known to be antioxidant proteins and radical scavengers. The protection of *Anabaena variabilis ATCC 29413* would not be dependent on just one method of protection, but would be a mixture of a variety of different methods including induction of these two families of proteins. Identifying their mechanism of upregulation enable a better understanding of their ability to protect themselves from the UV radiation and may help develop ways to protect people as well.[194]

A final direction would be to do more studies on the shinorine itself. As cyanobacteria are already used for nitrogen fixation by some engineers, making them into shinorine producers would not be so difficult. If it was possible, there is potential that testing of shinorine using *in vitro* and *in vivo models* would be possible either as a natural UV sunblock which can be applied on skin or potentially ingested to increase the overall UV protection and not just skin deep.[195]

5 Identification of binding sites of STATTIC in STAT3β

5.1 STAT3

STAT3 is a member of the STAT family, which are signal transducers and activators of transcription factors. They are latent cytoplasmic transcription factors that transmit signals from the membrane to the nucleus.[196] There are 7 isoforms which are encoded by distinct genes in mammalian cells.[197] STATs are activated by cytokine receptors or growth factor receptors and promote proliferation and other biological processes.[196,198-200]

STAT3 is similar in structure to other members of the family with a conserved amino terminus for cooperative binding to multiple DNA sites (1-130), a DNA-binding domain containing a specific palindromic IFN-γ-activated sequence element (400-500), a SH2 domain (600-700) involved in receptor recruitment and STAT3 dimerisation, a phosphotyrosine at Y_{705} and finally the carboxy terminal required for transcriptional activation.[201,202]

The protein has two phosphorylation sites, which are required for dimerisation, nuclear translocation and DNA-binding.[198,201] The activation of STAT3 is through phosphorylation of a critical Tyr_{705} which occurs through cytokine stimulation and is mediated by Janus kinase (usually JAK1), tyrosine kinases or Src family of kinases. The S_{727} phosphorylation is involved in transcriptional activation which is similar to that seen in STAT1.[201] The serine kinase involved in phosphorylating S_{727} has been difficult to identify as different signals appear to cause phosphorylation by a variety of different kinases.

In normal cells, STAT3 activation is rapid and transient, with the tyrosine phosphate dephosphorylated after one hour of activation.[200,203-206] STAT3 is involved in acute phase response of hepatoma cells, stimulation and proliferation of β lymphocytes, activation of terminal differentiation growth arrest in monocytes, and maintenance of pluripotency in embryonic stem cells. There is suggestion that STAT3 is a major signal transducer downstream of gp130 like receptors.[201]

An increasing number of cell lines and tumours have been shown to have STAT3 present. It is constitutively active and phosphorylated as seen with *v*-Src transformed cells[201] STAT3 contributes to carcinogenesis and tumour progression via upregulating gene expression and driving misregulation of key proteins.[197]

Interruption of the constitutively active STAT3 in oncogenesis is correlated with suppression of cell transformation and growth, and inducing apoptosis, showing an irreversible dependency to active STAT3 which is seen in breast cancers, prostate cancer, multiple myelomas, haematological and head and neck cancers.[199,201,204-207]

STAT3 activation occurs through a number of steps. Initially cells are stimulated by either growth factors or cytokines, causing receptor dimerisation and activation. Phosphorylation of the receptor's tail creates a docking site for the recruitment of monomeric, non-phosphorylated STAT3 via the SH2 domain. Activated tyrosine kinase, JAK phosphorylase is recruited to STAT3 at the Y_{705} near the C-terminal of the protein and the tyrosine is phosphorylated. The phosphorylated STAT3 is released from the receptor and a homodimer is formed through reciprocal binding of an SH2 domain of one monomer to the P_PYLKTK sequence containing the phosphotyrosine of the other monomer.[198] The homodimer translocates to the nucleus and forms a STAT3-DNA complex.[196]

Protein-protein interactions (PPIs) are a very attractive class of molecular targets. Most critical cellular functions such as growth, replication, signal transduction and many other cellular functions occur using multi-protein complexes. By targeting these interactions, through understanding how the protein complexes come together and why, enables various diseases to be identified and thus identifying more potential and specific targets for therapeutics.[105]

In the case of STAT3, each of these steps offers a place to inhibit the aberrant STAT3 activity and many different types of molecules have been made to try and inhibit various parts of the STAT3 activation developed through the knowledge gained through studying PPIs.[105,197] In the case of STAT3, peptidic or peptidomimetic have been the most common types of inhibitors studies, attempting to inhibit STAT3-STAT3 dimerisation by binding to the SH2 domain. Platinum based inhibitors have also been shown to inhibit STAT3 activity although their activity seems based upon cross-linking DNA. A third group of molecules is small molecule inhibitors of the STAT3 dimerisation.[105,196,197]

5.1.1 STAT3β and STATTIC

STATTIC (Figure 5.1) was a molecule that was identified through screening chemical libraries as a non-peptidic molecule showing selective inhibition of STAT3, by inhibiting

activation, dimerisation and nuclear transport, as well as increasing apoptotic rate of STAT3 dependent breast cancer cell lines.[105,196]

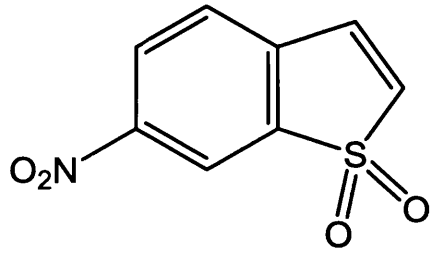


Figure 5.1: STATTIC molecule

Screening was performed by fluorescence polarization assays which resulted in 144 different small molecules inhibiting binding of a fluorescence labelled phosphotyrosine containing a peptide derived from gp130 receptor to STAT3 SH2 domain. STATTIC inhibited IL6 induced nuclear translocation of STAT3 assessed by immunolocalisation and inhibited DNA binding of the pre-phosphorylated STAT3 as determined by electrophoretic mobility shift assays with selectivity over STAT1.[105,196]

Activity was seen to be weak at 22°C, moderate at 30°C and high at 37°C with an IC_{50} 5.1±0.8 μM after 1 hour incubation and the activity was strongly reduced in the presence of DTT. Modification of the nitro group in STATTIC to an amine reduced the activity exponentially, suggesting that the double conjugated bond is required for activity and suggesting covalent modification (Figure 5.2) although evidence for covalent modification by mass spectrometry was not found at the time of publication.[105,196]

It was hypothesised that the most likely target for covalent binding of STATTIC to an amino acid would be cysteine residues as they are the most nucleophilic.

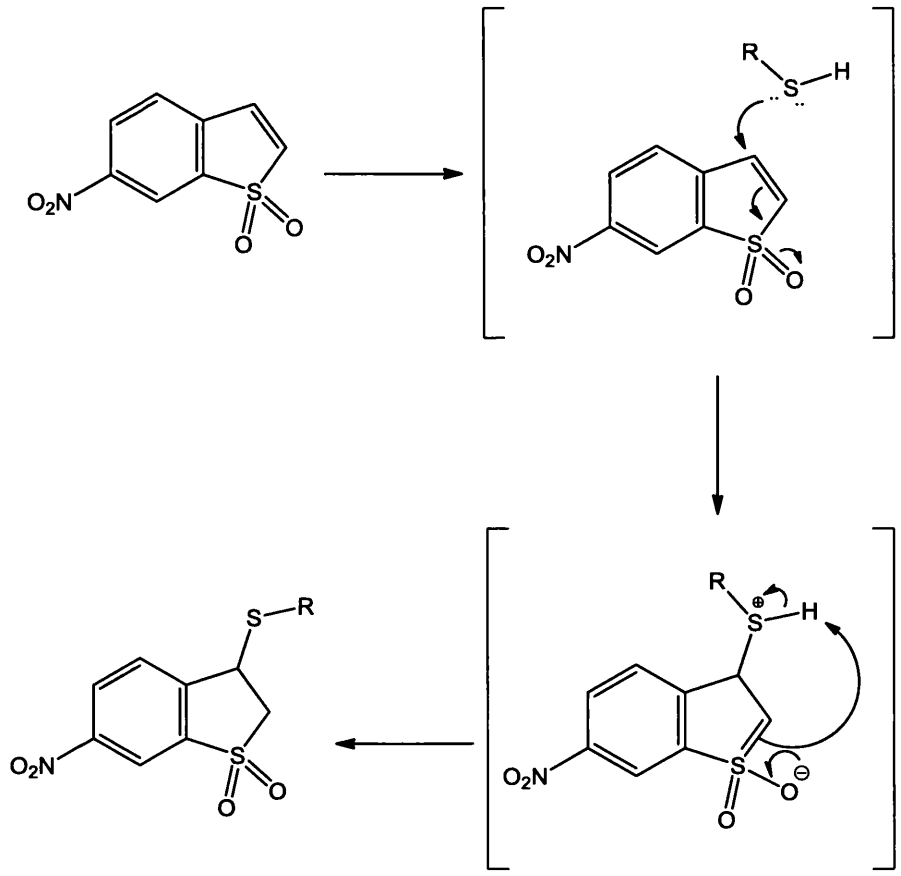


Figure 5.2: Possible mode of covalent binding of STATTIC to STAT3.

5.2 Aims

STATTIC with its highly conjugated system is likely a covalent modifier of STAT3, causing the inhibition of STAT3 dimerisation for activity. Using STAT3β, a truncated STAT3 protein, attempts will be made to prove the covalent binding of the STATTIC and identify the modified sites of STAT3 using mass spectrometry techniques.

5.3 Materials and Methods

5.3.1 Materials

Materials were purchased from the following companies:

HPLC grade H_2O, acetonitrile, methanol, 99.8% ethanol, ammonium hydrogen carbonate, biological grade NaCl, glycine and 99% BSA were purchased from Fisher Scientific (Waltham MA, USA). DMSO (ACS 99.9%), sinapinic acid (Fluka), ammonium acetate, α-cyano-4-hydroxycinnamic acid (Aldrich), TEMED, APS, PMSF, bromophenol blue, $MnCl_2$, EDTA, Igepal CA-630 and Sigmafast protease inhibitor tablets were purchased from Sigma-Aldrich (St. Louis, MO, USA). Formic acid (99%), KCl (BDH), SDS (BDH), TFA (Alfa Aesar), glycerol (molecular biology grade) and glucose (molecular biology grade) were purchased from VWR (Arlington Heights, IL, USA). Phosphate buffered saline tablets were purchased from Oxoid (Basingstoke Hampshire UK) and D_2O 99.8% was purchased from Cambridge Isotope Lab Inc. (Andover MA, USA). Trypsin was purchased from Promega (Southampton, UK). Bio-Rad protein assay was purchased for Bio-Rad (Perth, UK). Protogel was purchased from National Diagnostics (Atlanta Georgia, USA). K-15 STAT3 SC483 Rabbit polyclonal IgG antibody was purchased from Santa Cruz Biotech. ECl anti-rabbit IgG horseradish peroxidase and the ion exchange column, HiTrap™ QFF, 5 mL were purchased from GE Healthcare (Little Chalfont, Buckinghamshire, UK). DHB was purchased from Lancaster (Eastgate UK). Dialysis tubing, 6-8 kDa cutoff membrane was purchased from Spectrum laboratories (California, USA). Nitrocellulose membrane filter paper sandwiches were purchased from Invitrogen (Paisley, UK). Millex MP 0.22 µm were purchased from Millipore (Billerica MA, USA). Tris and DTT were purchased from Melford (Chelsworth, Ipswich, UK).

5.3.2 Growth of STAT3β

STAT3β was grown by the Biology PPI group of Professor David Thurston at the School of Pharmacy. Briefly, 2 x 10 mL LB seed were inoculated with 10 µL of uSTAT master cell bank (BL21 Rosetta cells) and grown in the presence of ampicillin and chloramphenicol overnight at 37°C. One of the 10 mL seeds was then used to inoculate 10 L electrolab fermenter in the presence of ampicillin and chloramphenicol (100 µg ampicillin, 20 µg chloramphenicol). The seed was gently mixed in the fermenter at 250 rpm at 37°C. The cells were induced with 1mM working concentration of IPTG when the OD reached 0.6 and the temperature was turned down to 21°C. The bacteria were left to grow overnight before harvested by centrifugation at 3600 x *g* for 25 minutes at 4°C.

5.3.3 Extraction of Unphosphorylated STAT3β

To each gram of pellet, 10 mL of extraction buffer (Table 5.1) was added to resuspend the pellet. The sample was placed into a chilled beaker of ice and sonicated for 5 minutes at 15 mA with repeating 15 seconds on and 15 seconds off using a medium sized sonicator probe (Camlab Transsonic TS701 H, Cambridge, UK).

Table 5.1 Extraction buffer (100 ml)

Required Concentration	Stock Concentration	Volume from Stock
20 mM Hepes-KOH pH 7.6	1 M	2 ml
0.1 M KCl	2 M	5 ml
10% Glycerol		10 ml
1 mM EDTA	500 mM	0.2 ml
10 mM $MnCl_2$	100 mM	10 ml
20 mM DTT	1 M	2 ml
0.5 mM PMSF	100 mM	0.5 ml
DD H_2O		70.7 ml
Protease inhibitor cocktail (stir till dissolved)		1 pellet/50 ml

The solution was centrifuged (Beckman Coulter Centrifuge, JA25.50 rotor, 4°C, 1 hour, 27 000 x *g*) and the supernatant was poured into a beaker and both supernatant and pellet were kept. The supernatant was chilled at 4°C. Ammonium sulphate was slowly added to the supernatant while gently stirring.

$$Ammount\ of\ ammonium\ sulphate\ (g) = 0.25\ x\ volume\ supernatant$$

Once the ammonium sulphate was completely added and dissolved, the sample was left for a minimum of 15 minutes at 4°C. The sample was centrifuged a second time (Beckman Coulter centrifuge, JA25.50 rotor, 4°C, 1 hour, 27 000 x *g*). The supernatant was decanted into another vessel and the supernatant and pellet were kept at 4°C until required.

5.3.4 Purification of Unphosphorylated STAT3β

An IE column (GE Healthcare, Little Chalfont Buckinghamshire, UK) was kept on ice while the ion exchange buffer was prepared (Table 5.2).

Table 5.2: Ion exchange buffer (Salt free)

Component	Amount from Stock for 1000 ml
100 mM Tris pH 8.5 (1 M stock)	100 ml
1 mM EDTA (0.5 M stock)	2 ml
2 mM DTT (1 M stock)	2 ml
DD H_2O	Up to 1000 ml

Varying concentration of ion exchange buffer (100 mL) was prepared using a stock concentration of 5 M NaCl:

0.1 M (2 mL)
0.2 M (4 mL)
0.3 M (6 mL)
0.4 M (8 mL)
1.0 M (20 mL)

The flow on the column was kept to 0.75 – 1.0 mL/min and all fractions and pellet were kept on ice. The column was washed first with 10 mL salt free buffer, followed by 20 mL 1 M NaCl buffer and 20 mL salt free buffer.

The pellet was dissolved in 5 mL salt free buffer and the solution was filtered through a 0.22 µm filter. The filtrate was diluted to 20 mL using double distilled H_2O. Approximately 5 mL of the dissolved protein was passed through the IE column and was collected in a 15 mL tube. To prevent overloading of the column, the fractions were randomly tested for protein using Bio-Rad protein assay. The column was washed with 10 mL salt free buffer which was collected in a 15 mL tube.

Elution was started with 10 mL 0.1 M NaCl IE buffer which was collected in 1.5 mL tubes. Each tube was quickly tested for protein and kept on ice. The column was eluted with 10 mL 0.2 M NaCl IE buffer and collected in a 15 mL tube. The column was eluted with 10 mL 0.3 M NaCl EI buffer and collected in a 15 mL tube. The column was eluted with 10 mL 0.4 M NaCl IE buffer and collected in a 15 mL tube. The column was washed with 20 mL 1 M NaCl buffer followed by 20 mL salt free buffer before continuing onto the next aliquot

5.3.5 1-DE

Standard SDS-PAGE 12% Bis-Tris Gels were prepared for indentification by tryptic in-gel digestion and analysis by mass spectrometry. All solutions were degassed prior to

use. Using mini gel preparations, 15 mL of 12% Bis-Tris was prepared and poured into the mould leaving 2 cm from the top of the glass plates for a resolving gel. Ethanol was poured on top to help level and remove bubbles. The gel was allowed to solidify before a resolving gel was poured on top and allowed to solidify with combs to create wells.

Table 5.3: SDS Page Stacking and Resolving Gel Buffers

Resolving gel Solution (12%)	**Volume (mL)**	**Stacking gel Solution (4%)**	**Volume (mL)**
Protogel	6.0	**Protogel**	0.867
Tris 1 M pH 8.8	6.0	**Tris 0.375 M pH 6.8**	0.676
10 % SDS	0.120	**10% SDS**	0.052
H_2O	2.75	**H_2O**	3.551
TEMED	0.012	**TEMED**	0.0052
Mix			
10 % APS	0120	**10 % APS**	0.052

Using a 5X Tank buffer(1L): 7.55 g Tris HCl
47 g glycine
25 mL 10% SDS.

Tank buffer was prepared by mixing 160 mL 5X Tank buffer and diluting to 800 mL with H_2O.

Sample Buffer (5X) was also prepared: 10 % SDS
10 mM DTT
20% v/v glycerol
0.2 M Tris HCl pH 6.8
0.05% w/v bromophenol blue

Samples were prepared by adding 4 µl of sample buffer (5X) and up to 18 µl sample to a final protein concentration of 25 µg. The sample was heated at 100°C for ten minutes to reduce and denature the proteins. The proteins were alkylated by adding 2 µl of 200 mM iodoacetamide to each sample, mixing and placing in the dark for one hour. The samples were centrifuged prior to pipetting into the gel lanes and run at constant V = 200 V until the bromophenol blue band travelled to the bottom of the gel; between 50-70 minutes.

The gel was removed from the chamber and placed into a clean plastic petri dish with water for a western blot or with Gentaur Instant Blue Stain (Biosera Ltd, Ringmer, East

Sussex, UK) for 15 minutes. The gel was rinsed and the background was removed using H_2O.

5.3.6 Western blotting

Tris-glycine buffer was prepared as described in Table 5.4.

Table 5.4: Tris-glycine buffer preparation for western blotting

	Weight or Volume	Concentration
Tris HCl	2.424 g	25 mM
Glycine	11.52 g	0.19 M
Methanol	160 mL	20%
H_2O	Up to 800 mL	

Nitrocellulose membrane (9.5 cm x 6.5 cm), 3 pieces of 3 mm paper (10 cm x 7 cm) and 4 fibrepads were pre-wet by immersing into 400 ml tris-glycine buffer. The stacking gel was removed and the gel was placed into a petri dish with tris-glycine buffer.

The gel and nitrocellulose were layered into the Invitrogen XCell II™ Blot Module cathode portion starting with 2 blotting pads, followed by a piece of filter paper, the gel, the membrane another piece of filter paper and two more blotting pads. The anode was placed on top and the entirety was then carefully placed into the running chamber. The inner chamber was filled only to the top of the pads with tris-glycine buffer. The outer chamber was filled with chilled water. The gel was blotted for 1 hour at 150 mA, 100V.

Tris-Saline (TS) buffer is prepared as follow (1 L):	1 M Tris pH 7.0	10 mL
	5 M NaCl	30 mL
	H_2O	960 mL

Once completed, the membrane was placed in a petri dish with blocking solution (0.5 g BSA in 50 mL TS buffer) overnight at 4°C. The blocking solution was discarded and was replaced with the primary antibody solution (Primary anti-body (rabbit anti-STAT3 polyclonal K15) – diluted 1:1000 in blocking solution + 0.05 % NP40 (25 µl antibody + 12.5 µl NP40 (Igepal) + 25 ml blocking solution)) and placed on the rocker for 1 hour. The primary antibody was discarded and the membrane was washed with TS buffer and then twice for five minutes with wash solution (Wash solution – TS buffer + 0.05 % NP40) on the rocker. The wash solution was discarded and replaced with the second antibody solution (secondary antibody – (anti-rabbit ECC IgG horseradish peroxidase) diluted 1:1000 in blocking solution + 0.05 % NP40 (25 µl + 12.5 µl NP40 (Igepal) + 25

ml blocking solution). The antibody was discarded and the membrane was washed with TS buffer followed by twice for five minutes with wash solution on the gel rocker.

A SigmaFast DAB tablet and SigmaFast urea H_2O_2 tablet were dissolved in 20 mL H_2O. The mixture was poured over the membrane and as soon as bands started to appear, the solution was drained and the membrane was washed three times with distilled deionised H_2O. The membrane was dried between two sheets of filter paper.

5.3.7 STD-NMR studies on BSA and STAT3β and 1H NMR

Based on the work of Fielding *et al*, BSA was used to test STD-NMR experiment [208-212] using the Bruker Avance 500 NMR equipped with the gradient enable broadband probe[210]. The software used for spectra acquisition and processing were Topspin 2.1 and ICONNMR software. Setup for the saturation transfer difference experiment was as follows: each sample was run to acquire ^1H NMR with water suppression: PROF_1H_ZGPR_01 for 8 scans, followed by STD experiment STDDIFFESGP.Z for 256/1024/2048 scans preceded by 2 dummy scans with an acquisition time of 1.59 seconds, a 90° pulse of 14.5 μseconds, sweep width of 20.7 ppm, relaxation delay of 0.3 seconds, and a temperature of 298K. Protein on-resonance irradiation was performed at −1.1 ppm, and the off-resonance irradiation was set at 114 ppm, where no protein signals were present. Spectra were subtracted internally by phase cycling after every scan using different memory buffers for on- and off-resonance. BSA was prepared at a stock concentration of 5 μM in PBS prepared in D_2O, while naproxen, ibuprofen and salicylic acid were prepared at 1M concentrations in PBS prepared in D_2O. The small molecule solutions were mixed with BSA at 50 μM concentrations. Mixtures of BSA with naproxen and ibuprofen and BSA with naproxen and salicylic acid were also prepared and their NMR spectra acquired. The NMR spectra of BSA in presence of 50 μM STATTIC dissolved in DMSO were also acquired.

STAT3β was dialysed in PBS, and lyophilised. The sample was diluted to 1 mL D_2O to a concentration of 5 μM. Both BSA and STAT3β were prepared and mixed using lower stock concentrations (25 μM and 10 μM) of STATTIC to acquire the NMR spectra.

5.3.8 Solubility studies for STAT3β in mass spectrometry compatible buffers and solvents

STAT3β was tested for its solubility in mass spectrometry compatible buffers at varying concentrations for intact mass analysis. Buffer exchanges were performed as were

precipitation methods using 1:2, 1:5 and 1:10 sample to cold acetone. Samples were left in the freezer (-20°C) overnight before being centrifuged at 4000 x *g*, 20 minutes 4°C. The supernatant was discarded and the pellet was re-dissolved in 100 mM ammonium bicarbonate (ABC). Varying concentrations of ammonium bicarbonate and ammonium acetate (AA) were used for the studies as well as a variety of different molecular weight cut off Vivaspin columns.

Buffer exchanges were performed using 200 mM ABC and using 5000 MWCO spin columns (Sartorius Vivaspin, 30 minutes, 4°C). The columns were filled to the 100 µL mark with ABC and spun. This was repeated twice more and the mixture was then centrifuged down to 10 µL. A second set were prepared with 200 µL 200 mM ABC and concentrated with 30,000 MWCO spin column for 15 minutes before a second 200 µL were added to the spin column and centrifuged for another 15 minutes. A final addition of 200 µL of 200 mM ABC was added to sample and the centrifuge tube was centrifuged a final time for 15 minutes at 12000 x *g*.

Lyophilisation was also used to maintain stability for analysis. pH was also used for stability studies.

5.3.9 Preparation of dialysis tubing and dialysis of STAT3β

Dialysis tubing was boiled for 10 minutes in H_2O followed by 10 minutes in either PBS, 10 mM ammonium bicarbonate or 10 mM ammonium acetate. STAT3β samples were lyophilised accordingly to prepare for NMR and mass spectrometry when needed.

5.3.10 Incubation studies of BSA and STAT3β with inhibitors for mass spectrometry

BSA was prepared at 50 µM in 10 mM ammonium acetate for inhibition studies for mass spectrometry. STAT3β was either dialysed into 10 mM ammonium acetate or 10 mM ammonium bicarbonate, or dialysed and lyophilised from a solution of 10 mM ammonium acetate or 10 mM ammonium bicarbonate. The concentration of STAT3β was tested prior to mass spectrometry analysis. STATTIC was prepared in DMSO and incubated with STAT3β at 0X, 5X, 10X, 15X, and 20X the concentration of STAT3β and BSA. The samples were run on a gel, digested with trypsin and extracted before analysis by LC-MS/MS.

5.3.11 In-gel digestion

After running SDS-Page on control and treated STAT3β and BSA, bands were cut, digested and prepared according to the method stated in previous chapter (section 4.3.13).

5.3.12 Solution digest

Using STAT3β dialysed in PBS, or in 10 mM ammonium acetate to prepare for in solution digest using varying concentrations of STATTIC in DMSO at 0X, 5X, 10X, 50X, 100X, 200X and 500X by incubating for at 37°C for 1 hour before it was digested. Samples were digested by adding 0.04 µg trypsin and digesting overnight at 37°C. The reaction was stopped by adding 10µL of 5% FA before being run on LC-MS/MS.

5.3.13 Bradford Protein Assay using microtiter plates

Using a stock solution of BSA (2.0 mg/mL) and H_2O as a solvent, the following dilutions of BSA were made: 5, 10, 15, 20, 25, 30, 40, 50, 60 µg/µL (Table 5.5). Using a microassay plate, 5 µL of standard or sample was transferred in triplicate onto the plate into separate wells. Reagent was prepared by mixing 1:4 Bio-rad Protein Assay to H_2O. The reagent was filtered using 022 µm filters before use. The diluted protein assay reagent was added to each well to a volume of 200 µL. The plate was shaken gently and read at 595 nm before 1 hour was over.

Table 5.5 BSA calibration curve prepared from 2 mg/mL BSA.

Concentration mg/ml	BSA (µl)	H_2O (µl)	Sample volume (µl)	Reagent volume (ml)
2.0	100	0	5	0.2
1.5	75	25	5	0.2
1.0	50	50	5	0.2
0.75	37.5	62.5	5	0.2
0.5	25	75	5	0.2
0.25	12.5	87.5	5	0.2
0.125	6.25	93.75	5	0.2

Using absorbance vs concentration for BSA, a plot was created and an equation for a straight line was obtained for BSA. Samples were prepared in the same way by adding 5 µL sample to 200 µL reagent. If the absorbance was above that of the BSA, the sample was diluted and run again. The absorbances were then used to determine the approximate concentration using the following equation.

$$\frac{A_{595}}{slope\ of\ curve} \times dilution\ factor = protein\ concentration$$

Where slope of curve is the slope determined for the BSA calibration curve, A_{595} is the absorbance of the sample at 595 nm and dilution factor refers to the amount the sample was diluted, mostly set to 1 referring to no dilution required for analysis. An example of the BSA calibration curve and slope of the line used to calculate protein concentrations can be found in previous chapters 3 and 4.

5.3.14 Calibration of mass spectrometers

Prior to any analysis on the QTOF Global Ultima, the instrument was calibrated according to the method described in (4.3.16)

For intact protein analysis on the QTOF, the system was checked first by infusing 5 pmol/µl BSA to ensure that sensitivity and optimisation was properly set-up.

Calibration of the MALDI-TOF was performed according to the molecular weight range for analysis. For intact protein analysis, the MALDI-TOF was calibrated for a *m/z* 20000- 75000 using BSA as a calibrant (Applied Biosystems).

5.3.15 Intact mass analysis on QTOF

Dialysed or lyophilised BSA and STAT3β samples were prepared and incubated with 100 mM of inhibitors at 37°C for at least 1 hour before analysis. Using Waters QTOF Global Ultima, protein solutions were directly infused into the mass spectrometer for intact mass analysis. Proteins were prepared in 10 mM ammonium bicarbonate or 10 mM ammonium acetate and mixed with 5% FA in 50% acetonitrile 50% H_2O. Using a syringe, the mixture was infused at 0.5 µL/min with a capillary optimised at 2.5 kV, cone 50, RF lens 1 at 50 in positive nano-electrospray mode.

Intact protein masses were deconvoluted using MassLynx 4.1 using mass ranges of 50000 - 75000 for BSA and STATβ, 10000 - 25000 for myoglobin and 2000 - 5000 for the insulin chain B using MaxEnt1 with a resolution of 1.00 Da/channel. Damage model for deconvolution was using a uniform Gaussian with a width at half height of 0.75 Da and default minimum intensity ratios of 33% on both the left and right sides. The deconvolution was iterated to convergence.

5.3.16 Intact mass analysis on MALDI-TOF

STAT3β was attempted using DHB (10 mg/mL in 50:50 ACN/H_2O + 0.1% TFA) and SA (10 mg/mL in 50:50 ACN/H_2O + 0.1% TFA) and α-CA (10 mg/mL in 50:50 ACN/H_2O + 0.1% TFA) by dried droplet, sandwich and layered methods of crystallisation. Samples were analysed using linear positive mode on an AB Sciex Voyager DE-Pro with an accelerating voltage of 25000 V, grid voltage 92%, guide wire 0.3%, extraction delay time 1500 nsec, 100 shots/spectrum.

5.3.17 Capillary LC-MSMS Data Dependent Acquisition

The QTOF Global Ultima system was prepared according to previous chapters for LC-MSMS data dependent acquisition (section 4.3.14).

5.3.18 Mass Spectrometry (MS/MS)

Tandem MS analysis was performed on the QTOF Global Ultima according to methods stated in previous chapters (4.3.15).

5.3.19 Data Processing for MS/MS

Raw data was processed using MassLynx ProteinLynx V4.1 as stated in previous chapters (section 4.3.17).

5.3.20 Protein Lynx Global Server Search[121]

Using the raw data, a processing method, followed by search method were created. The processing method was the same as that was used to create a pkl file:

Mass measure: Smooth by Savitzky Golay 3.00 channel, 2 smooths, centroid peak list using top 80% with minimum peak width at half height = 4. Searching was performed using the sequence for STAT3β. The formatting was modified to be readable using PLGS. A new database search was created with the following parameters:

Enzyme: trypsin	Maximum missed cleavages: 1
Fixed modifications: none	Variable modifications: -Oxidation (Met) -Carbamidomethylation (Cys)
Peptide Tolerance: +/- 0.3 Da	MS/MS tolerance: +/- 0.3 Da
Report top: 200 max	

5.3.21 X!Tandem

A second database was used to search the data. Using the pkl file generated by MassLynx 4.1, the data was searched online using X!Tandem[120] which is run by the GPM (http://h.thegpm.org/tandem/thegpm_tandem.html). Searching was performed using the Human SwissProt database.

Fragment mass error: 0.3	peptide log(e) < -1
	protein log (e) < -1
Protein modifications (variable):	carbamidomethylation (Cys)
	oxidation (Met)
	+210.00 (Cys, Ser, Tyr)
Protein cleavage:	trypsin
Semi-style cleavage:	yes
Spectrum conditioning:	remove redundant yes
Spectrum synthesis:	yes
Quad TOF error:	100 ppm

5.4 Results

The molecular weight for STAT3β was calculated using MassLynx 4.1 and was determined to be 68,098 Da according to the sequence provided. STAT3β was extracted and purified according to the methods optimised in the PPI group at the School of Pharmacy. Upon purification, fractions were checked for protein concentration and some of the fractions were run on SDS-PAGE and as a Western to confirm the presence and purity of STAT3β. (Figure 5.3)

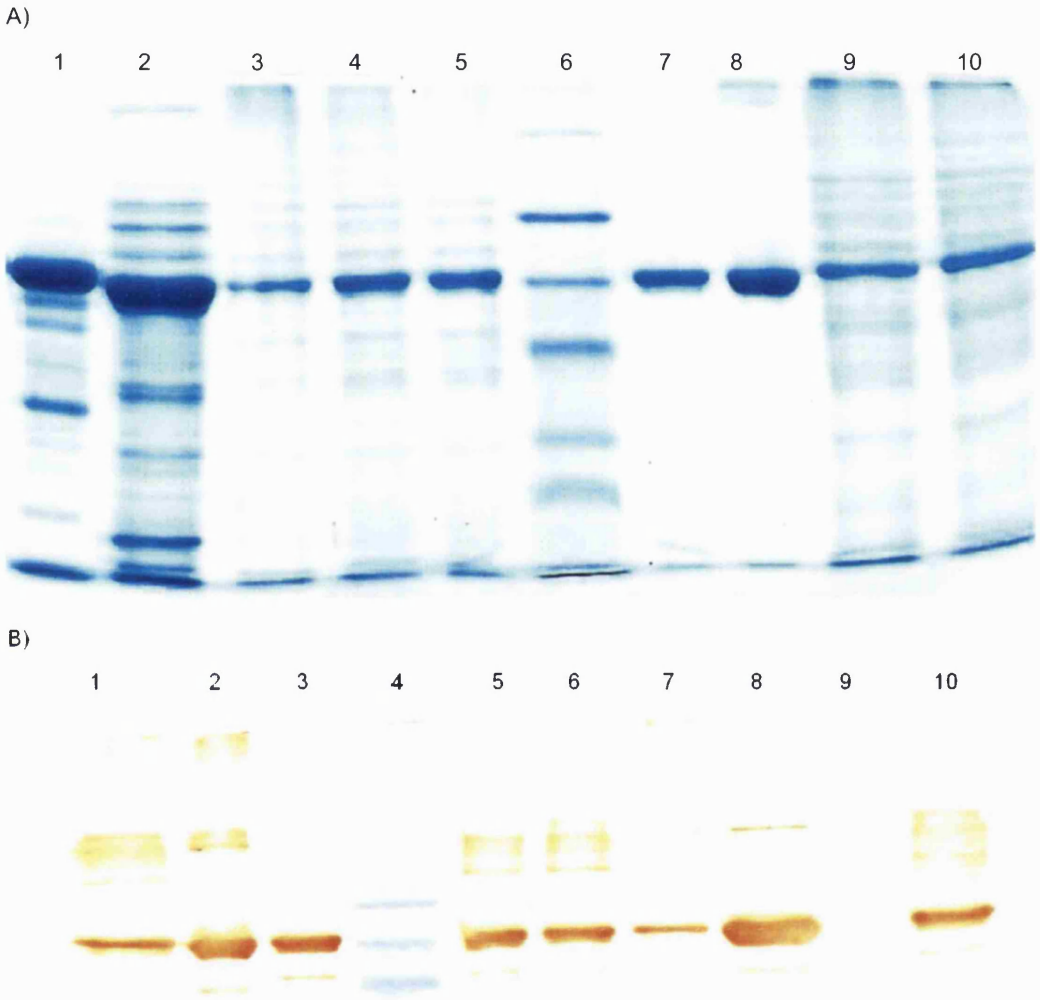


Figure 5.3: SDS-Page (A) and Western (B) of STAT3β purification. A) Lanes: 1-BSA, 2 – SE11-2, 3-purSTAT3B, 4-purSTAT3B, 5 pur STAT3B, 6, MW marker, 7 – STAT3B, 8 – STAT3B, 9 – pSTAT3B, 10 pSTAT3B B) Lanes:1-pSTAT, 2-old STAt3B, 3- new STAT3B, 4-mw marker, 5- pur STAT3B, 6-purSTAT3B, 7-pur STAT3B, 8-SE11-2, 9-BSA, 10 pur STAT3B

The STD experiment for BSA with known non-covalent interacting molecules showed interactions using NMR. The NMR spectra of BSA were initially acquired to test that the pulse sequence was correctly identifying the interactions. ^{1}H 1D NMR spectra were initially acquired to set up the correct parameters of the STD experiment and as a control for all STATTIC experiments (Figure 5.9).

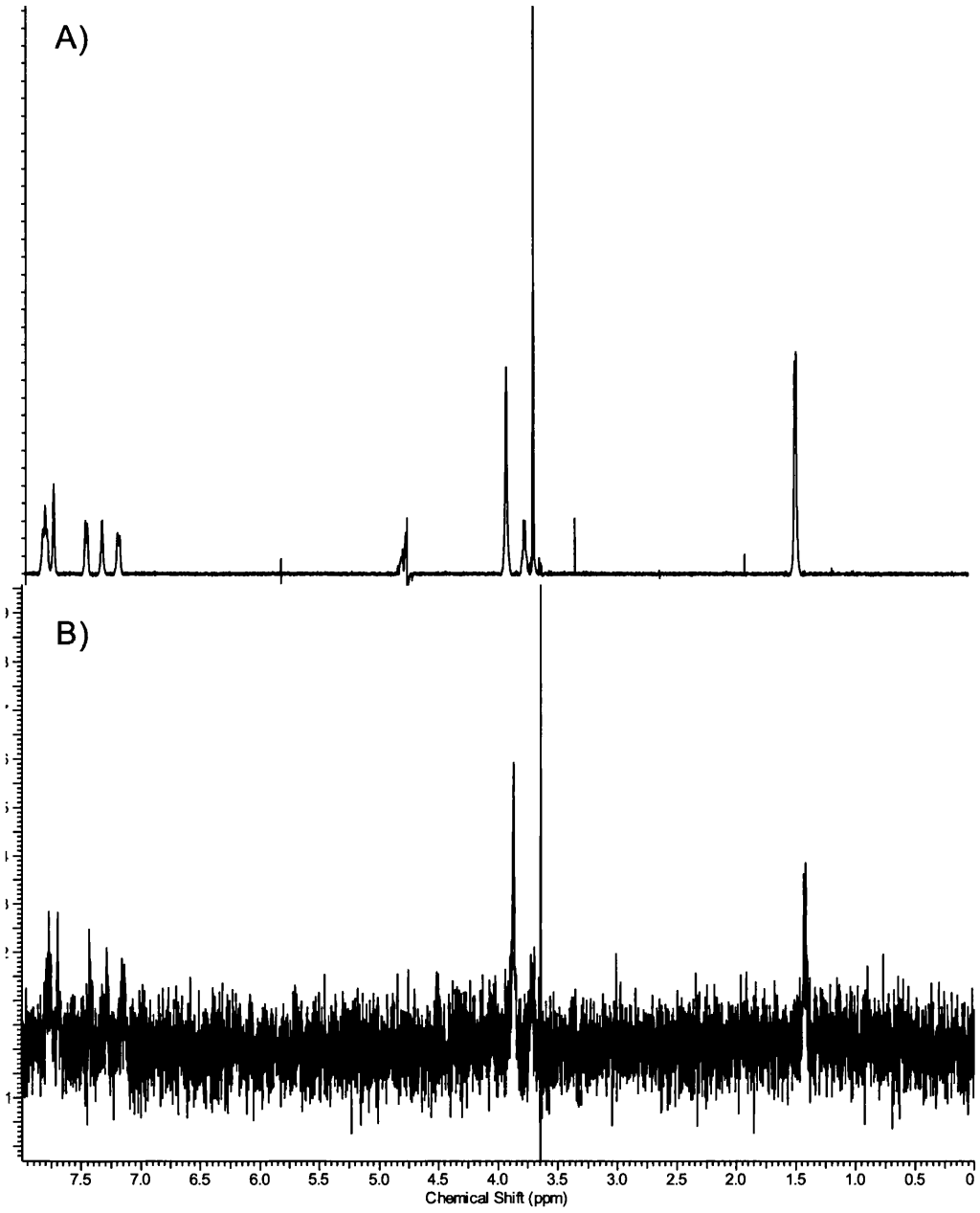


Figure 5.4: STD experiment of BSA with naproxen. A) ^{1}H NMR of naproxen, B) STD experiment with BSA and naproxen showing non-covalent interaction of BSA with naproxen.

BSA was run using naproxen as a non-covalent inhibitor showing that non-covalent interactions can be seen using the STD-NMR experiment (Figure 5.4). The same was performed using ibuprofen (Figure 5.5). The signal to noise ratio in these spectra indicated that ibuprofen is a weaker binder than the naproxen.

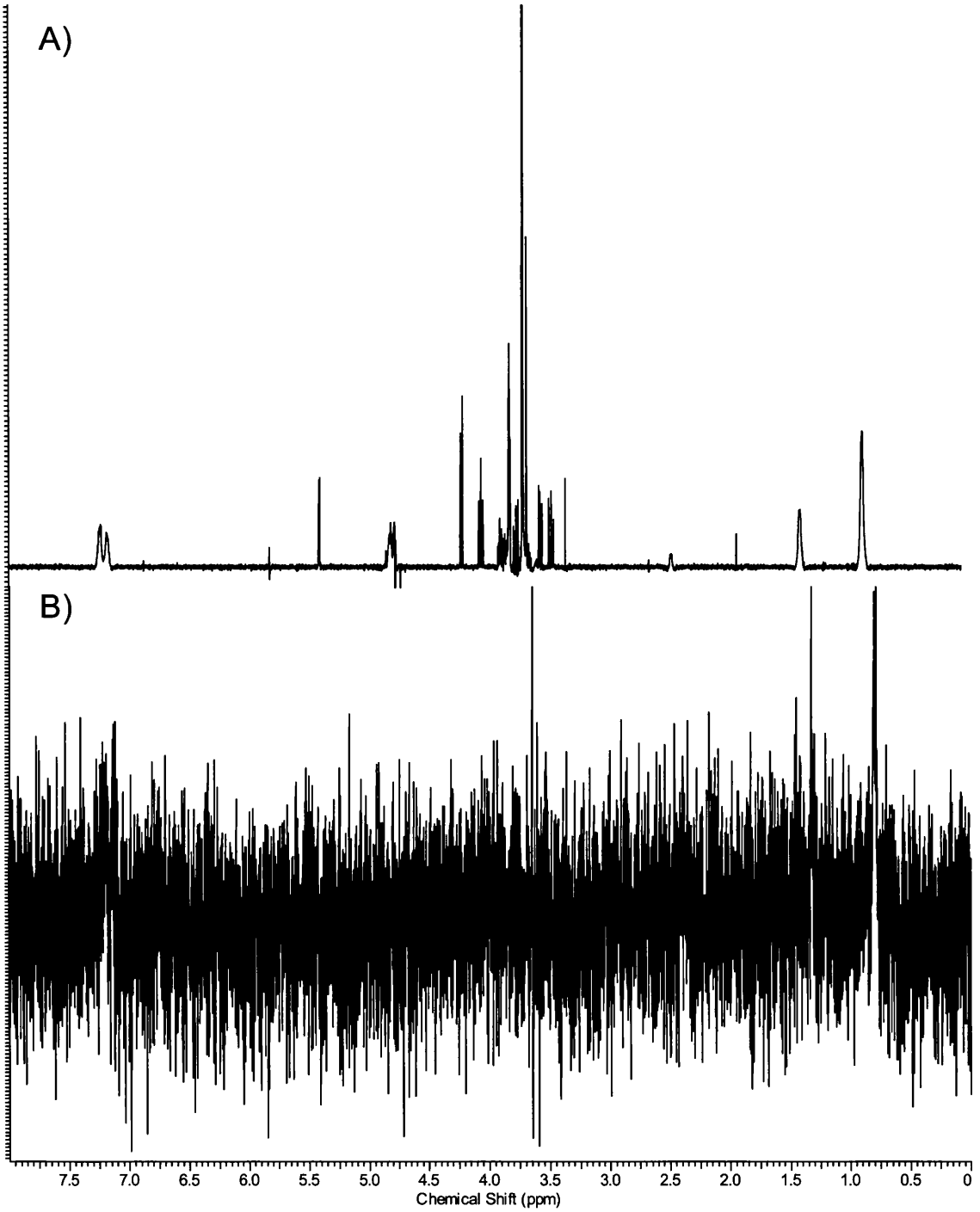


Figure 5.5: STD-NMR experiment of BSA and ibuprofen. A) ^{1}H NMR of ibuprofen, B) STD-NMR of BSA and ibuprofen.

A final component, salicylic acid was also tested with BSA, which proved to be the strongest binder (Figure 5.6).

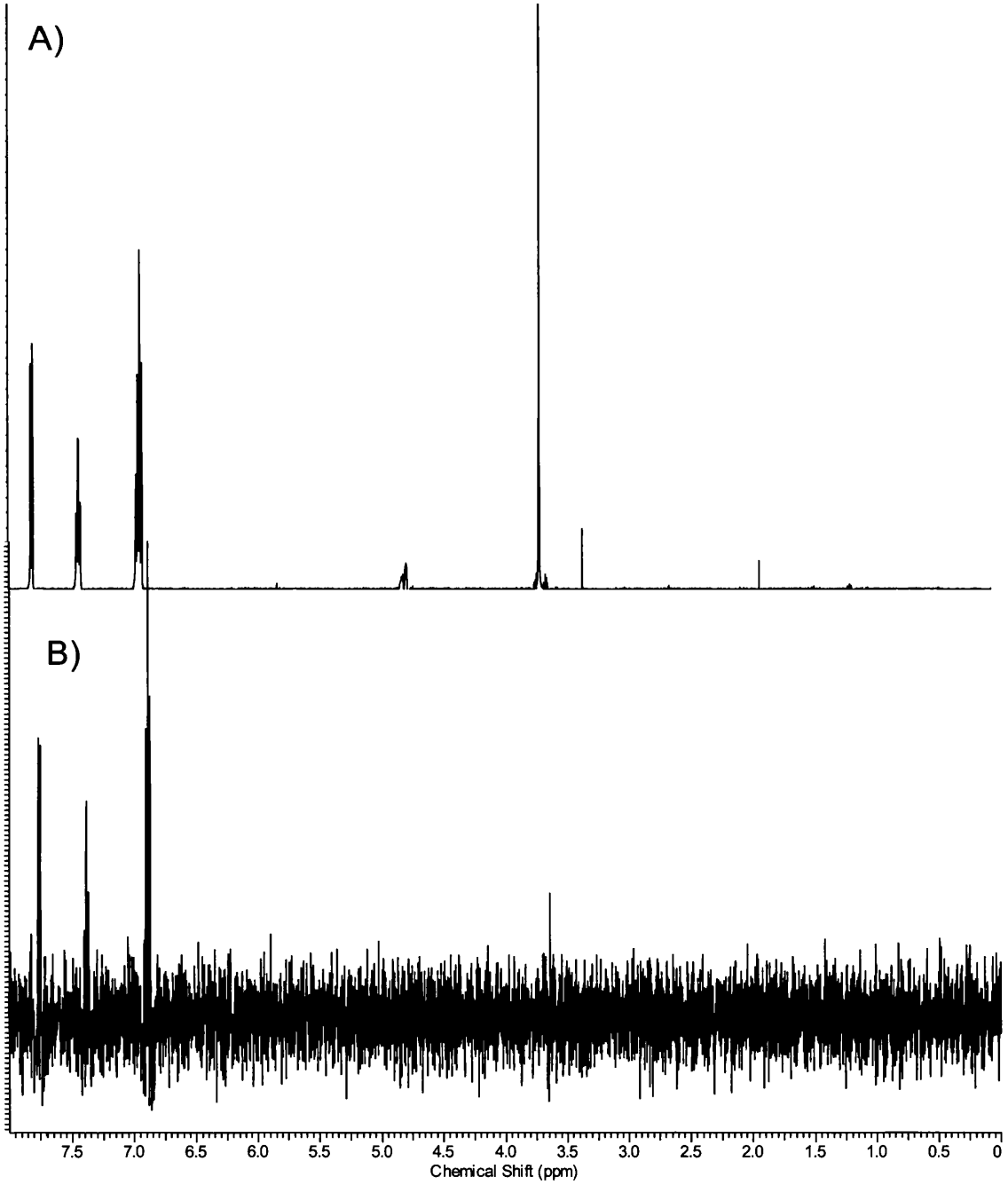


Figure 5.6: STD-NMR experiment of BSA and salicylic acid. A) ^{1}H NMR of salicylic acid, B) STD experiment of BSA and salicylic acid.

Using the three molecules which are known to interact with BSA, naproxen and ibuprofen were mixed together and the same STD-NMR experiment was also performed. The peaks of salicylic acid had highest intensity due to the strongest binding.

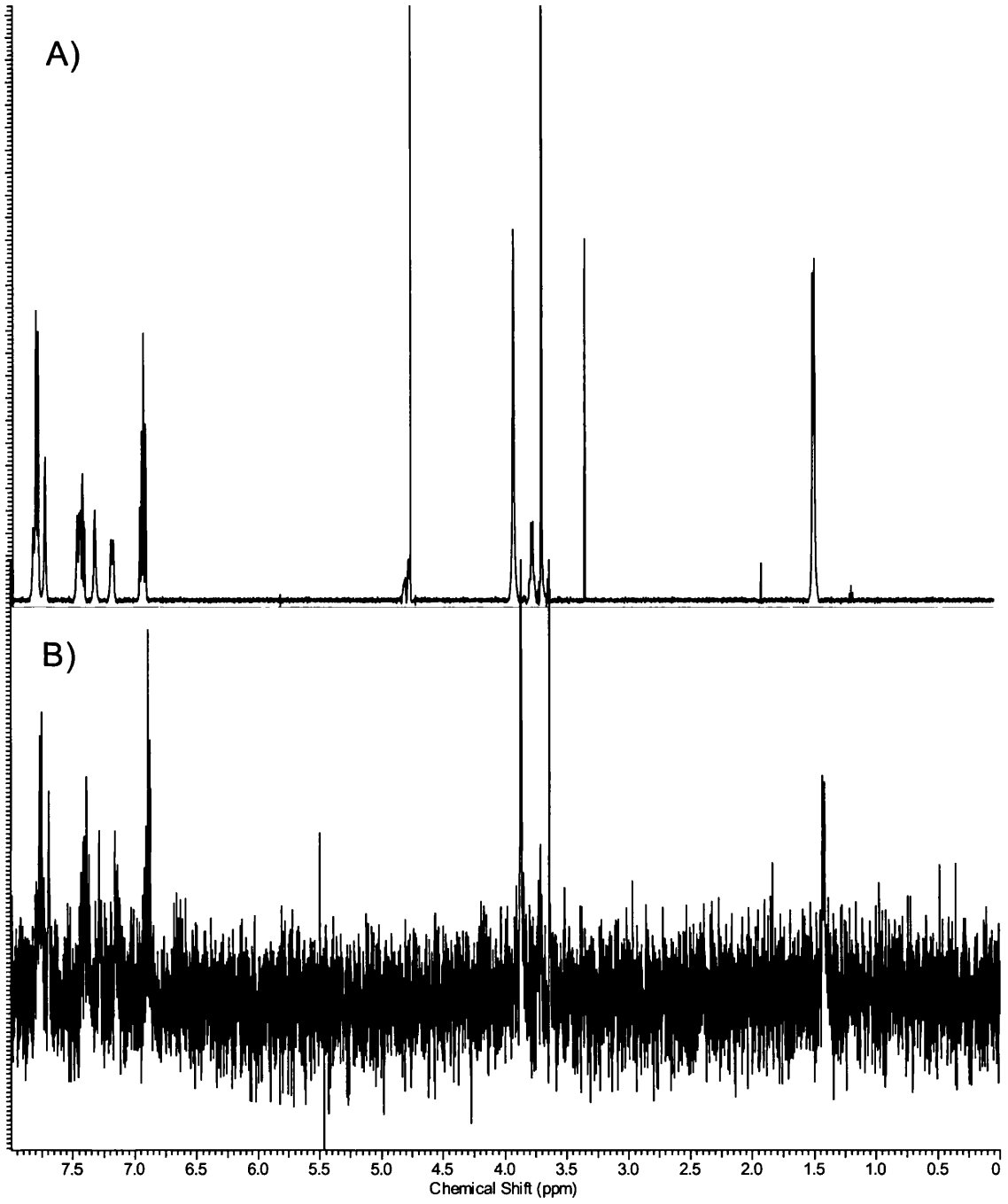


Figure 5.7: STD-NMR of BSA with naproxen and salicylic acid. A) ^{1}H-NMR of naproxen and salicylic acid. B) STD experiment showing naproxen peaks.

Competitive and non-competitive inhibition were both observed using BSA as the target. Non-competitive binding was observed for two sites on BSA with naproxen and salicylic acid, showing both molecules binding to BSA (Figure 5.7). Competitive inhibition was seen using naproxen and ibuprofen with BSA, where only peaks of naproxen were observed in the STD experiment (Figure 5.8).

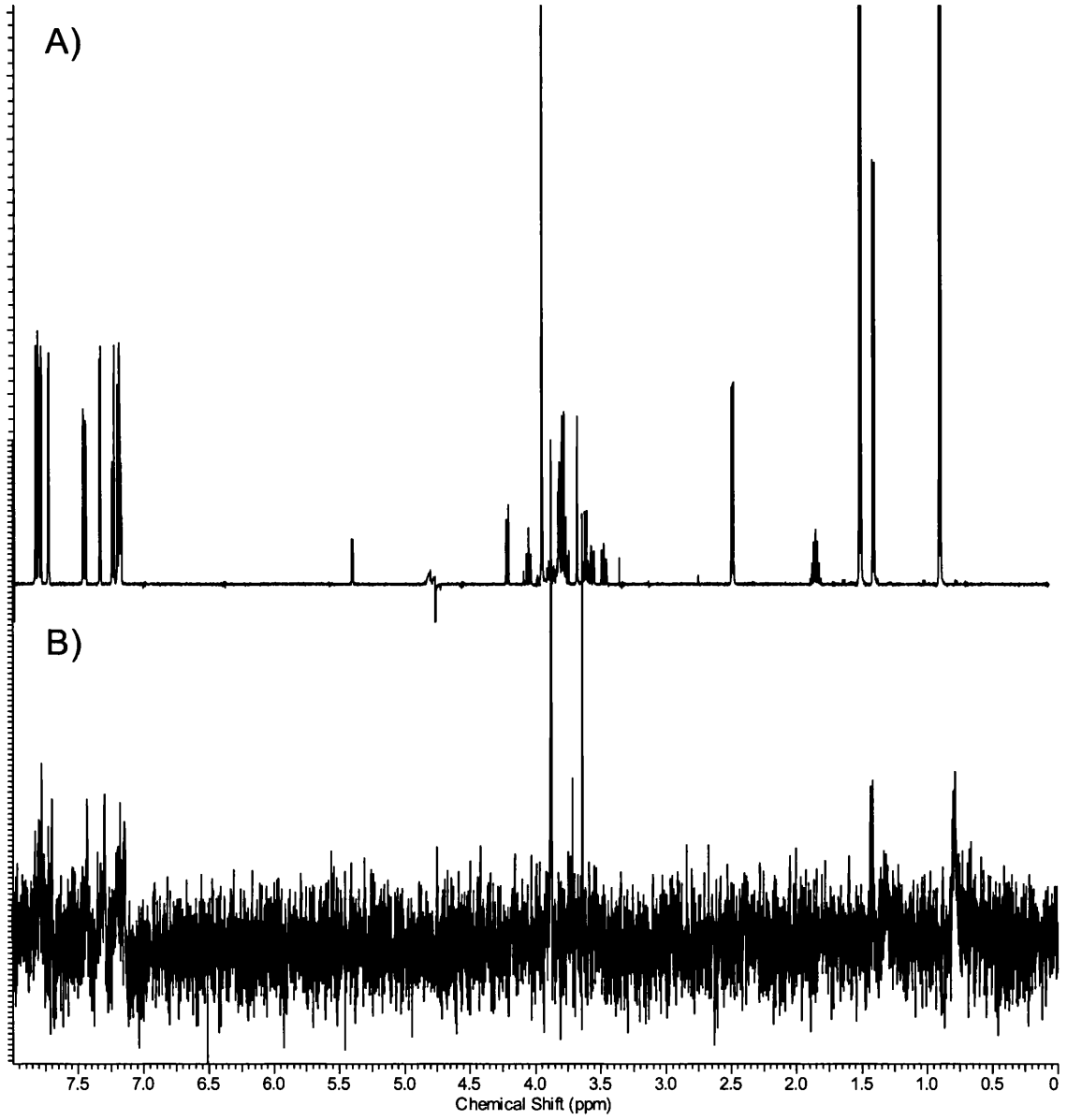


Figure 5.8: STD-NMR of BSA with naproxen and ibuprofen. A) ^{1}H NMR of naproxen and ibuprofen. B) STD-NMR experiment showing interaction of both with BSA.

STAT3β was dialysed in PBS, then lyophilised before being dissolved in D_2O. The protein concentration was determined and set to 5 μM. STATTIC was added at 50 μM in DMSO to STAT3β causing immediate precipitation of both molecule (red) and protein (white). The same was attempted using BSA, which resulted in the same red and white precipitate forming. Using a lower concentration of STATTIC (25 μM) resulted in a fairly stable solution with BSA. The mixture was immediately run as a STD-NMR experiment (Figure 5.9).

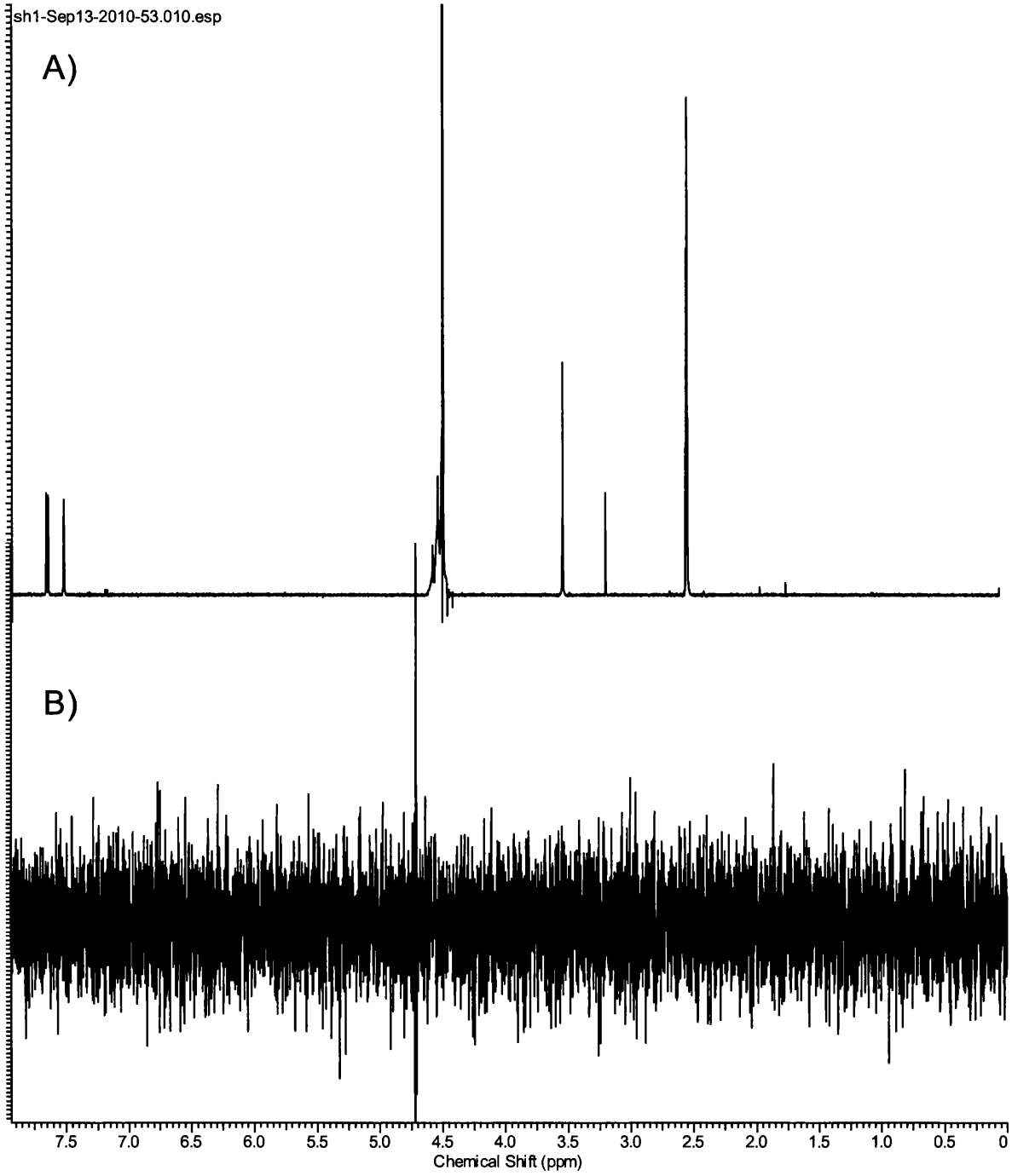


Figure 5.9: STD-NMR of BSA and STATTIC. A) ^{1}H NMR of STATTIC, B) STD-NMR of BSA and STATTIC.

The result from the STD (Figure 5.9B) experiment showed that there is no STATTIC seen in the experiment when compared to the ^{1}H NMR (Figure 5.9A). The lack of magnetisation transfer from BSA to ligand suggests that STATTIC is either not interacting with BSA or is covalently binding to BSA.

The STD experiments of STAT3β in the presence of STATTIC were attempted in the same fashion as BSA however, the protein continued to precipitate out of solution when STATTIC was added. STAT3β was mixed with DMSO slowly to see how soluble

it was in the aqueous/DMSO mixture. Using a total volume of 1 mL for these attempts, it was determined that up to 500 µL DMSO did not change solubility of STAT3β and did not cause it to precipitate, suggesting that the STATTIC molecule was causing precipitation at the concentrations needed to carry out NMR experiments.

The standard purification of STAT3β in the PPI laboratory resulted in STAT3β in a PBS, Tris or Hepes buffer containing DTT and EDTA. MALDI-TOF analysis of such samples was attempted, however was quite unsuccessful (Figure 5.10).

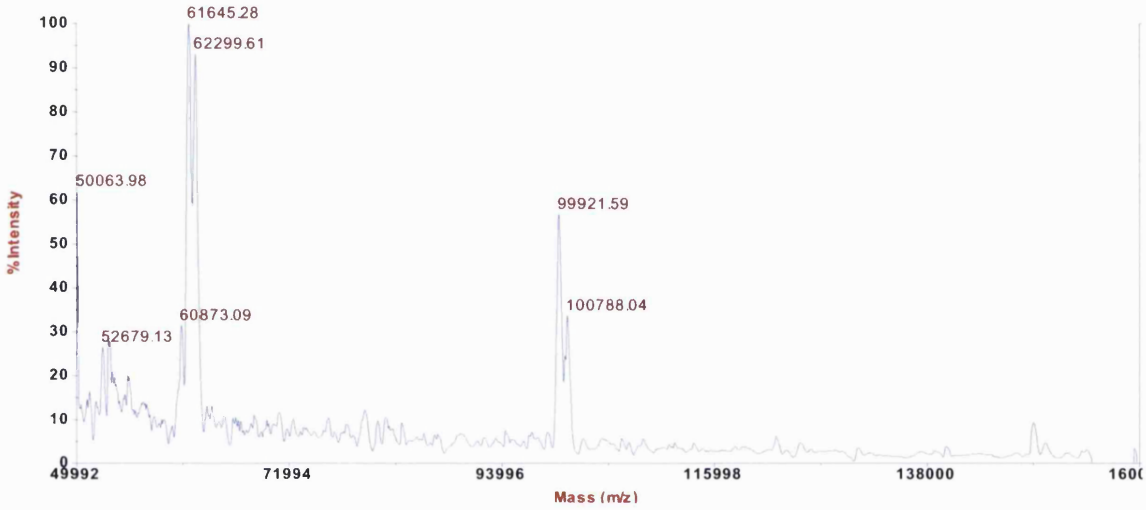


Figure 5.10: STAT3β pellet, extracted but not purified and analysed by MALDI-TOF

Initial buffer exchanges were performed using 5 kDa and 30 kDa MWCO centrifuge columns. Attempts were made to get STAT3β into 200 mM ABC, however when analysed on the MALDI-TOF, nothing was seen in the 68 kDa *m/z* range using the 5kDa MWCO centrifuge tubes and only a very small peak was seen using the 30 kDa MWCO centrifuge tubes (Figure 5.11).

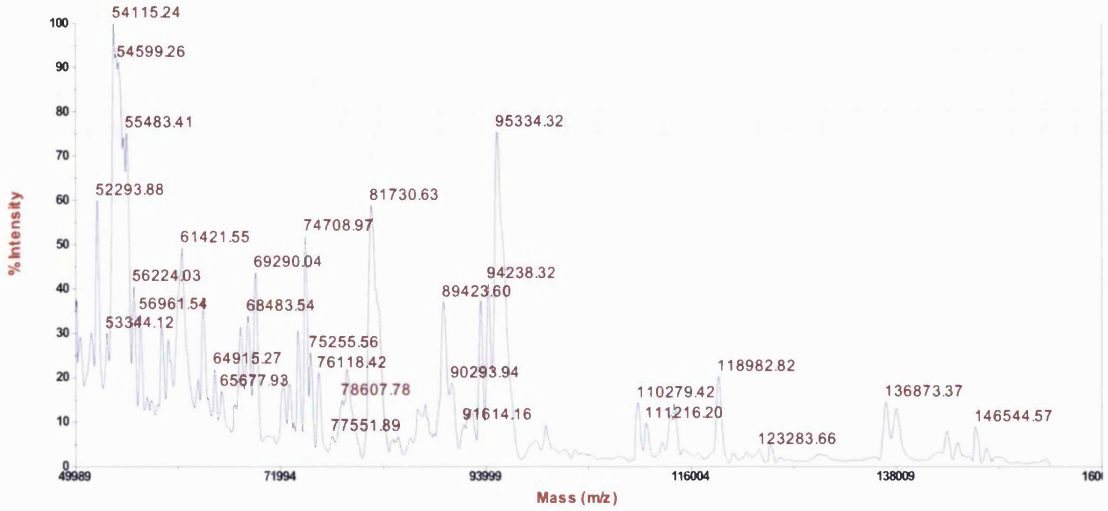


Figure 5.11: MALDI-TOF spectrum of STAT3β concentrated and desalted into 100 mM ammonium bicarbonate with a 30,000 MWCO centrifuge column.

The spectra for 30 kDa MWCO analysis showed a very low signal which cannot be confirmed. In another attempt to buffer exchange, 200 µL of STAT3β in PBS was precipitated using 1000 µL cold acetone (Figure 5.12). The mixture was left at -20°C overnight before being centrifuged at 12000 x *g* for 10 minutes. The supernatant was discarded and 50 µL of 200 mM ABC was added to dissolve the pellet.

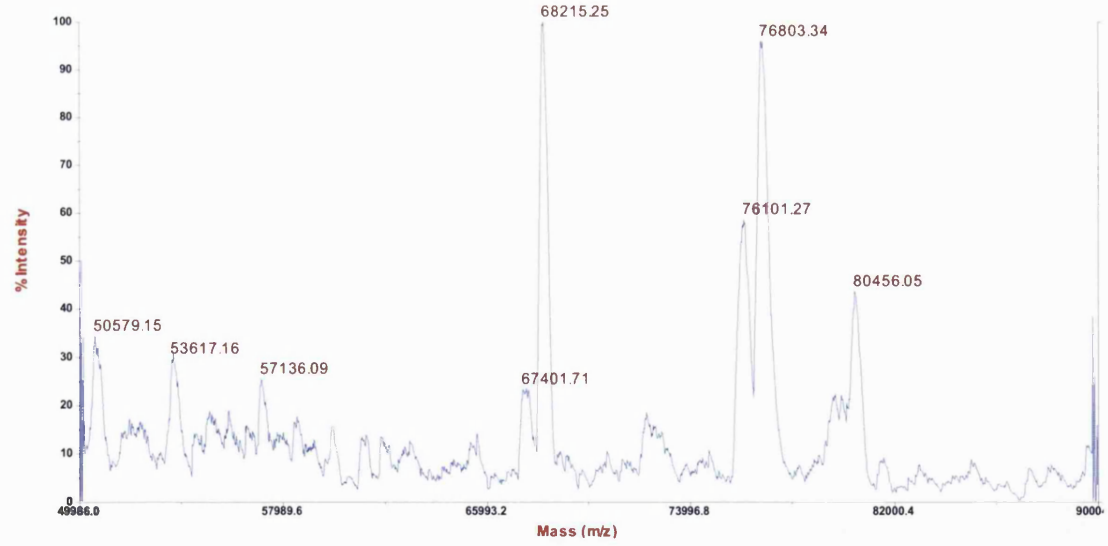


Figure 5.12: MALDI spectrum of cold acetone precipitation 1:5 with STAT3β. Redissolved in 100 mM ammonium bicarbonate.

The STAT3β in PBS (Figure 5.13) and in ABC were spotted onto the MALDI target plate and analysed. The same was attempted using 200 mM ammonium acetate. A small peak potentially at the mass of STAT3β was seen in the precipitated samples, however the peak size was quite small, indicating that the sample had not completely re-dissolved.

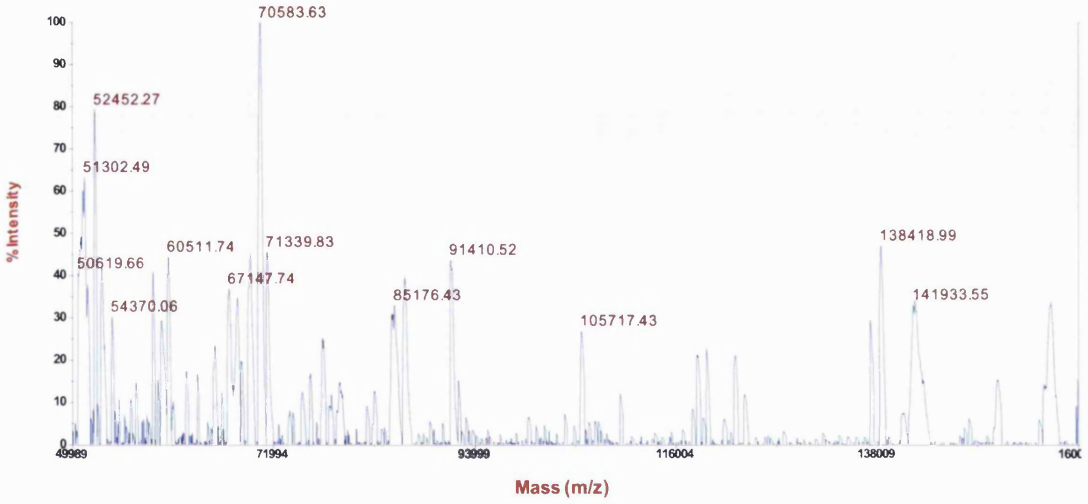


Figure 5.13: STAT3β in PBS analysed by MALDI-TOF

Using a pellet of unpurified STAT3β, a test was setup to see how concentrations of ammonium acetate and ABC as well as pH would to affect solubility of STAT3β (Table 5.6).

Table 5.6: pH study on solubility of STAT3β

Buffer	pH	Pellet dissolved (mg)	MALDI-TOF	MALDI-TOF (with STATTIC)
Tris	7.4	6	X	X
H_2O	7	4	X	68 803
ABC 200 mM	8.5	6	X	X
ABC 200 mM	7	4	X	X
AA 200 mM	6.5	5	68 223	68 131
AA 200 mM	5.5	6	X	68 848
ABC 50 mM	8.5	7	X	X
ABC 50 mM	7	4	X	X
AA 50 mM	6.5	6	X	X
AA 50 mM	5.5	4	68 314	68 445

Analysis of STAT3β in varying concentrations of ammonium acetate or ABC and varying the pH, indicated that ammonium acetate gave better analysis than ABC at both 200 mM and 50 mM ABC.

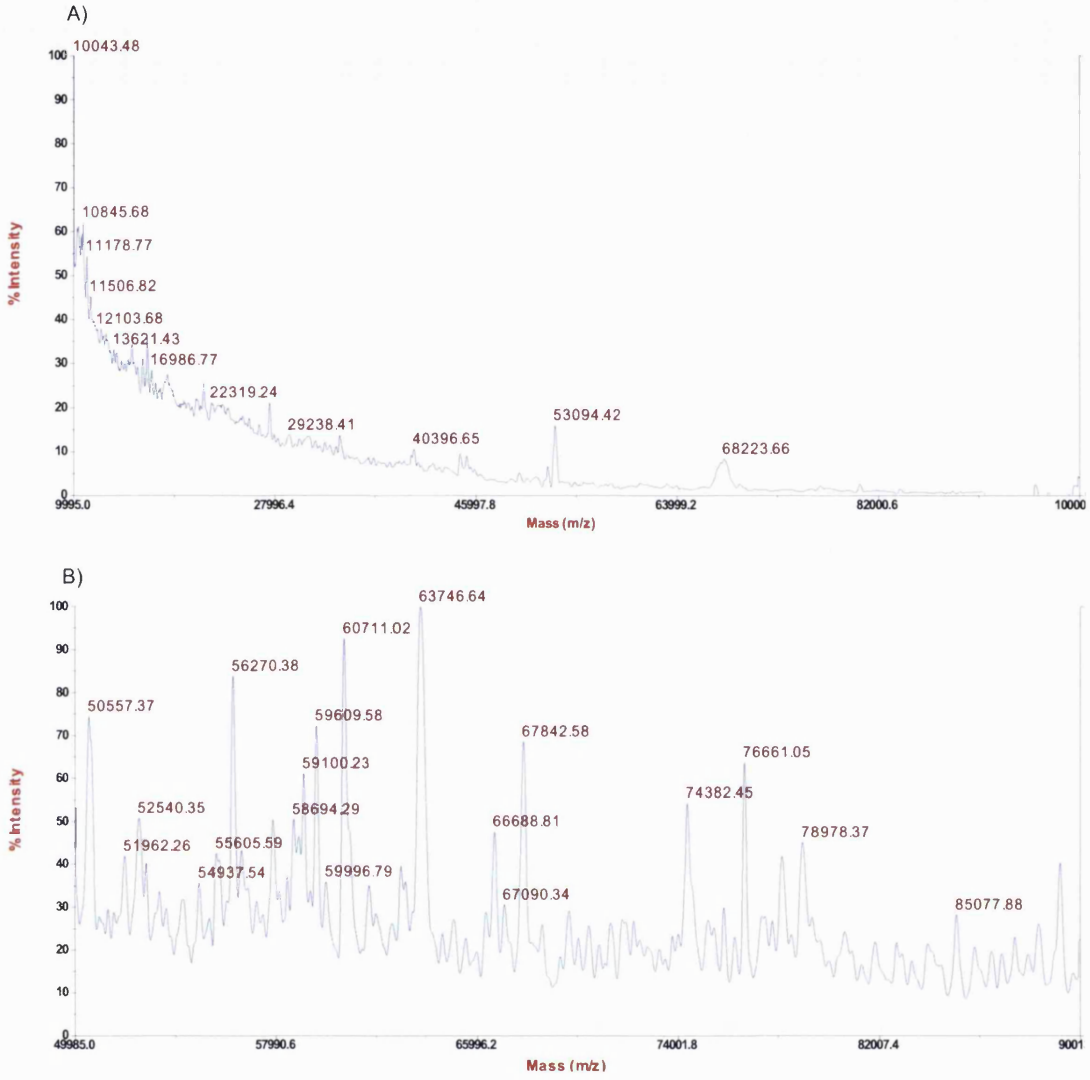


Figure 5.14: MALDI spectra of pellet of STAT3β dissolved in A) 200 mM ammonium acetate and in B) 50 mM ammonium acetate with STATTIC

The mixtures were then analysed using MALDI-TOF, resulting in *m/z* at 68 314 for STAT3β in 50 mM ammonium acetate (Figure 5.14B) and *m/z* 68 223 in 200 mM ammonium acetate (Figure 5.14A). In both cases, the signal was low and more than just STAT3β was seen, suggesting that the sample was not actually as pure as the gel would suggest.

A fresh batch of STAT3β was dialysed into PBS and prepared for in-gel and in-solution digestion. STAT3β protein concentration was calculated (0.33 mg/mL) and STATTIC in DMSO was added in concentrations of 0, 1X, 2X, 5X, 10X, 15X and 20X. BSA was prepared in the same concentration and was run using 0X, 5X and 20 X. The samples were incubated at 37°C for one hour before preparing for SDS-PAGE and in-gel digest (Figure 5.15).

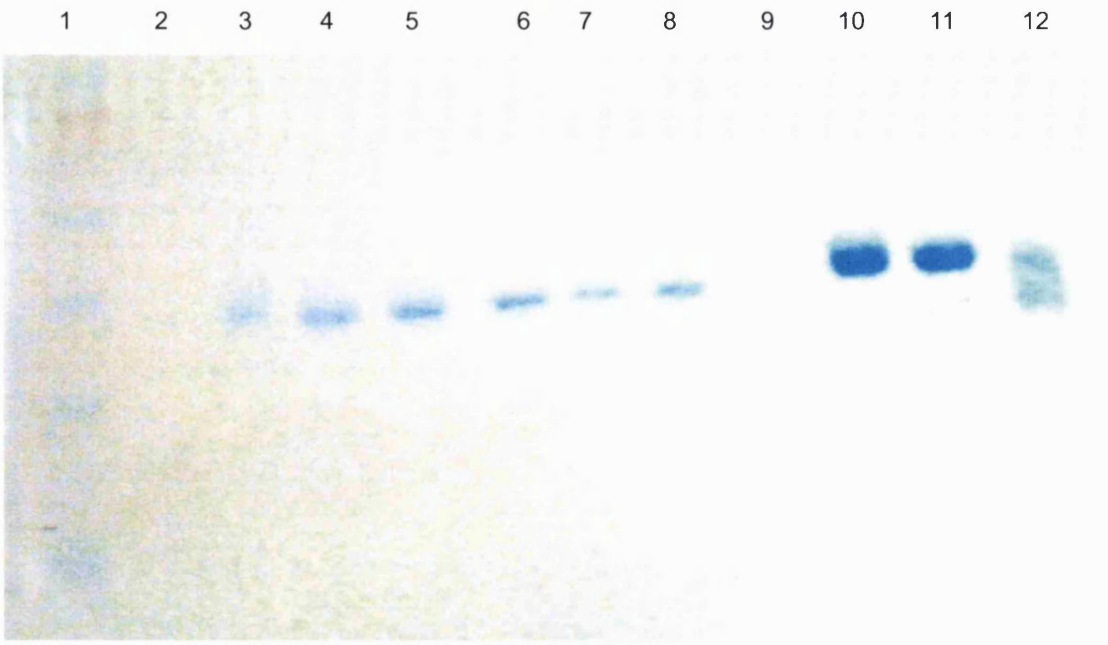


Figure 5.15: Incubation of STAT3β and BSA with STATTIC. Lanes: 1- molecular weight marker, 3- STAT3β, 4 – 1X STATTIC, 5- 2X STATTIC, 3-5X STATTIC, 4-10X STATTIC, 5 – 15X STATTIC, 6-20X STATTIC, 8- BSA, 9-BSA/5X STATTIC, 10- BSA-20X STATTIC.

As each lane was loaded with the same amount of protein, it appears that the protein band gets thinner, however, the gel appears to be slanting upwards which does not allow for a comparison of molecular weight.

Samples were digested and analysed by LC-MS/MS. STAT3β was identified using PLGS and by X!Tandem. PLGS was unable to take into account 210.99 as modification by covalent binding of STATTIC but was able to use the STAT3β sequence for searching (Appendix B). In order to look for a +210.99 addition, X!Tandem was used to search the data using Human SwissProt database (Table 5.7).

Table 5.7: X!Tandem results from database searches for gel digest of STAT3β and BSA incubated with STATTIC

Sample		pI	MW (kDa)	%	Description	Modification
STAT3β	**sp|P40763|**	**5.94**	**88**	**34**	**STAT3**	**No**
STAT3β 1:1	sp|P40763|	5.94	88	29	STAT3	No
STAT3β 1:5	**sp|P40763|**	**5.94**	**88**	**38**	**STAT3**	**No**
STAT3β 1:20	sp|P40763|	5.9	88	33	STAT3	No
BSA	**sp|ALBU_BOVIN|**	**5.8**	**69.2**	**31**	**BSA**	**No**
BSA 1:5	sp|ALBU_BOVIN|	5.82	69.2	29	BSA	C288
BSA 1:20	**sp|ALBU_BOVIN|**	**5.8**	**69.2**	**31**	**BSA**	**C288**

Initial digest of STAT3β did not show any binding of STATTIC while BSA showed binding at residue C_{288}.

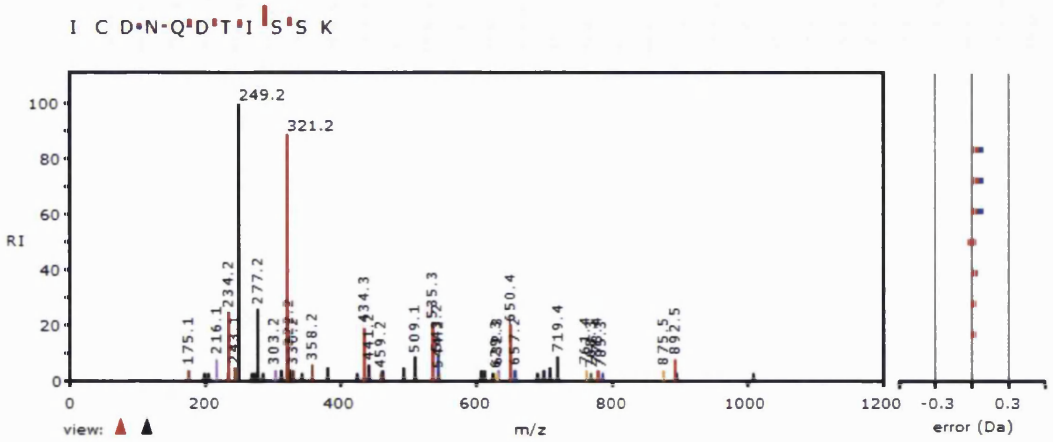


Figure 5.16: Fragmentation of a BSA peptide with +210.99 bound to the cysteine.

The same was repeated in PBS and AA using higher concentrations of STATTIC: 0X, 5X, 10X, 50X 100X 200X, and 500X in DMSO and MeOH, at 1hr and 24 hr incubations at 37°C. The samples were run on LC-MS/MS and searched initially using PLGS and then searched by X!Tandem. Again, PLGS could not show results for binding of STATTIC to STAT3β (Appendix B), while X!Tandem showed binding of STATTIC to potentially be a the C_{367} residue of STAT3.

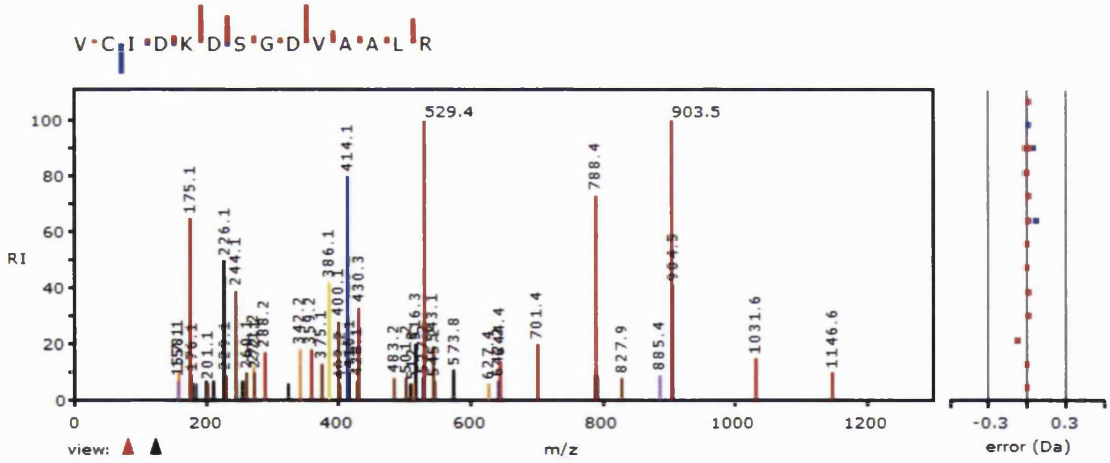


Figure 5.17: STAT3β peptide suggesting binding of STATTIC to C_{367}.

A comparison of the parts of the protein identified for the mixed buffer results for 1hr and 24hr incubations is shown in the Table 5.8.

Table 5.8: Percent coverage of regions of STAT3β with 1 hour incubation of STATTIC prior to solution digest.

			DNA specificity		SH2 Domain	
	4-helix bundle		B-barrel	Connector		pTyr
Amino acid	1	320	465	585	688	722
DMSO 1hr AA						
0aa	3.33		1.87	1.73	4	0
5aa	3.07		3.33	1.87	4	0
10aa	1.6		3.07	1.73	5.47	0
50aa	3.33		1.87	1.73	5.47	0
100aa	1.47		1.87	1.73	3.33	0
200aa	13.47		5.73	0	4.4	0
500aa	1.6		2.13	1.73	6.93	0
PBS						
0pbs	10.27		0	1.87	11.07	0
5pbs	6.13		5.33	1.73	13.47	2.93
10pbs	4.8		2.27	1.87	5.2	0
50pbs	10.27		5.47	1.87	5.6	0
100pbs	6.93		4.13	1.73	5.2	0
200pbs	9.07		3.6	1.73	6	0
500pbs	9.47		4.4	0	2.13	0
MeOH AA						
0aa	0		0	3.6	2.1	0
5aa	1.87		1.87	1.73	4	0
10aa	6.8		1.87	0	4	0
50aa	3.33		1.87	1.73	4	0
100aa	0		1.87	0	4	0
200aa	1.47		1.87	1.87	3.6	0
500aa	0		2.13	1.73	4	0
PBS						
0pbs	12.13		2.27	1.87	8.13	0
5pbs	10.67		4.13	1.73	2.8	0
10pbs	7.73		3.47	1.73	5.33	0
50pbs	12		0	0	0	0
100pbs	12		3.6	1.73	0	0
200pbs	13.33		3.6	2.5	3.6	0
500pbs	11.2		3.6	2.4	6.53	0

From the data observed after 1hr incubation, the ammonium acetate (AA) mixtures were observed to give more consistent results and indicated that STATTIC can possibly bind to a sequence containing a C_{367}. The peptide $[M+2H]^{2+}$ 836.8635 was not in the control samples (Table 5.9), indicating that the the C_{367}, was the amino acid modified by STATTIC. While the ammonium acetate incubation and digest did not give the same amount of coverage as the PBS samples, it gives consistent results, as the

digest of the methanol samples give a less consistent result for all the samples. Identification of the protein was spread across the entire protein, although no peptides were identified in the pY region of the SH2 domain (Table 5.8). The modified cysteine residue was found within the β-barrel region of the protein and was only found in the samples incubated with STATTIC and not in the controls. An expanded summary of individual peptides identified can be found in Appendix B.

Table 5.9:STAT3β incubations with STATTIC in DMSO 1 hr

AA	% (measured)	unique	MW	Accession	Description	Modification
STAT	9.9	8	88	sp|P40763|	STAT3	No
STAT-5	13	13	88	sp|P40763|	STAT3	C_{367}
STAT-10	12	10	88	sp|P40763|	STAT3	C_{367}
STAT-50	11	10	88	sp|P40763|	STAT3	C_{367}
STAT-100	12	7	88	sp|P40763|	STAT3	C_{367}
STAT-200	36	34	88	sp|P40763|	STAT3	C_{367}
STAT-500	16	16	88	sp|P40763|	STAT3	C_{367}
PBS						
STAT	27	23	88	sp|P40763|	STAT3	No
STAT-5	29	27	88	sp|P40763|	STAT3	C_{367}
STAT-10	21	18	88	sp|STAT3_HUMAN|	STAT3	No
STAT-50	29	24	88	sp|P40763|	STAT3	C_{367}
STAT-100	28	24	88	sp|P40763|	STAT3	C_{367}
STAT-200	24	17	88	sp|P40763|	STAT3	No
STAT-500	22	15	88	sp|P40763|	STAT3	C_{367}

The same was repeated with a 24 hour incubation to allow for possibly more complete modification of the protein.

Table 5.10: Percent coverage of regions of STAT3β with 24hr incubation with STATTIC prior to solution digest.

		DNA specificity		SH2 Domain	
	4-helix bundle	B-barrel	Connector		pTyr
	1 320		465 585	688	722
DMSO 24hr AA					
0aa	0	0	1.73	4.27	0
5aa	3.33	2.13	0	3.33	0
10aa	5.07	6.13	4.4	4.8	0
50aa	4.93	3.2	2.13	1.87	0
100aa	5.06	3.07	2.13	4	0
200aa	1.6	4.4	4	3.2	0
500aa	11.2	5.73	0	1.87	0
PBS					
0pbs	5.2	1.87	2.27	3.33	0
5pbs	12.67	4.13	2.13	5.07	0
10pbs	10	7.87	1.73	8.8	0
50pbs	11.6	2.13	0	7.73	0
100pbs	13.73	5.47	1.73	1.87	0
200pbs	7.2	4.13	1.73	1.87	0
500pbs	5.47	0	0	1.87	0
MeOH AA					
0aa	1.6	1.73	0	1.87	0
5aa	1.6	5.47	1.73	4	0
10aa	1.6	0	0	1.87	0
50aa	3.07	3.47	2.13	1.87	0
100aa	1.6	1.87	2.13	1.87	0
200aa	1.6	2.13	2.13	1.87	0
500aa	1.6	2.13	2.13	1.87	0
PBS					
0pbs	3.07	2.27	2.13	3.33	0
5pbs	3.6	5.47	2.13	5.2	0
10pbs	8.93	3.47	2.13	3.33	0
50pbs	5.87	2.27	2.13	1.87	0
100pbs	0	0	0	0	0
200pbs	2	4.13	2.13	1.87	0
500pbs	0	0	0	0	0

Data from the 24 hour incubation with STATTIC gave far more variable results and lower coverage of the protein, which suggested that the peptides or the protein may still be precipitating in solution the longer they are kept with STATTIC. In the 24 hour samples, the same C_{367} was identified by X!Tandem as a potential place for modification by STATTIC. Neither the peptide, nor the modification were seen in the control samples which were incubated with DMSO and not STATTIC. The coverage of

the protein compared to the 1 hour incubation was mostly the same (Table 5.9,Table 5.11) however the identification of the modified cysteine residue was far less complete over a longer period of time. A full list of peptides found in the samples can be found in Appendix B.

Table 5.11: STAT3β incubations with STATTIC in MeOH

AA	% (measured)	total	MW	Accession	Description	Modification
STAT	6.8	7	88	sp|P40763|	STAT3	No
STAT-5	12	15	88	sp|P40763|	STAT3	C_{367}
STAT-10	14	13	88	sp|P40763|	STAT3	C_{367}
STAT-50	14	15	88	sp|P40763|	STAT3	C_{367}
STAT-100	11	12	88	sp|STAT3_HUMAN|	STAT3	C_{367}
STAT-200	12	10	88	sp|P40763|	STAT3	C_{367}
STAT-500	14	16	88	sp|P40763|	STAT3	C_{367}
PBS						
STAT	23	28	88	sp|STAT3_HUMAN|	STAT3	No
STAT-5	24	25	88	sp|P40763|	STAT3	C_{367}
STAT-10	22	16	88	sp|P40763|	STAT3	No
STAT-50	32	38	88	sp|P40763|	STAT3	No
STAT-100	21	21	88	sp|P40763|	STAT3	No
STAT-200	28	29	88	sp|STAT3_HUMAN|	STAT3	No
STAT-500	34	38	88	sp|STAT3_HUMAN|	STAT3	No

Using STAT3βTC, a truncated STAT3 human protein used for modelling studies[196], the location of the cysteine residue and peptide were observed in Figure 5.18. The 3-dimensional structure used for the modelling was prepared by Dr. Shozeb Haider by using X-ray crystallography data and creating a 'snap shot' view of the protein using homology modelling of the sequence. The resulting structure was used for modelling dynamic simulations carried out on Amber v10. For the purposes of this study, the protein and ligand were optimized using Grid/Glue 22 for binding and the resulting image was then imported into DS Visualizer 2.5.

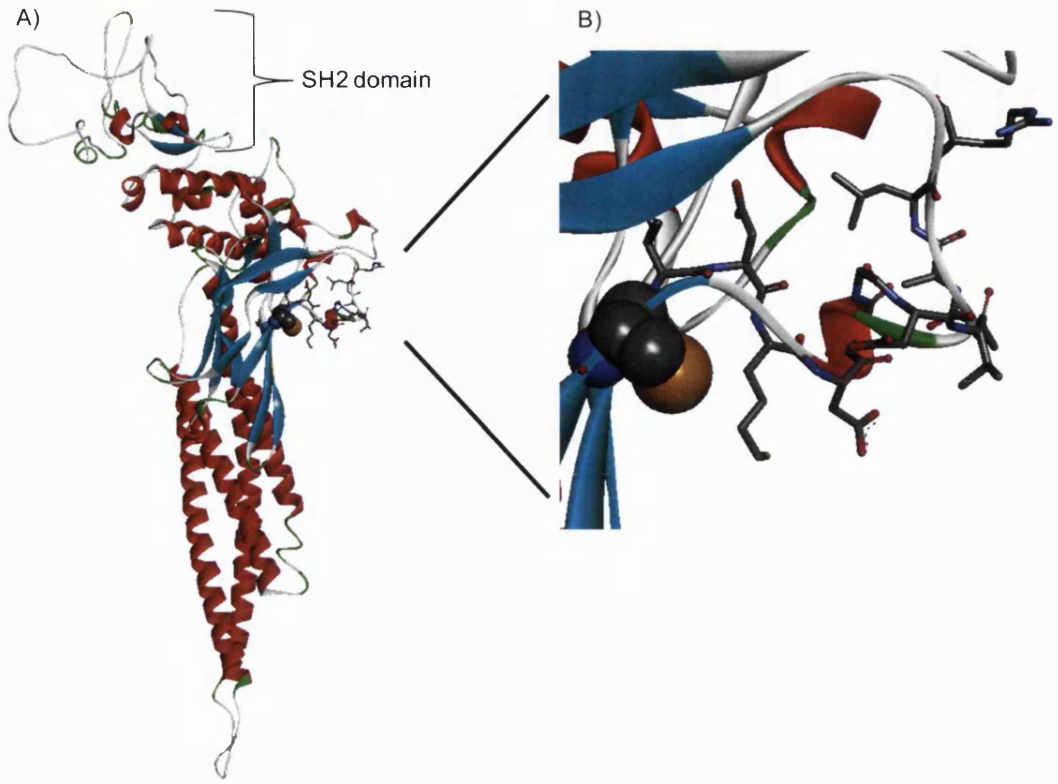


Figure 5.18: STAT3βTC. A) Full protein showing the SH2 domain. B) Expanded view showing the cysteine residue in CPK and the rest of the peptide identified in stick using DS Visualizer 2.5.

The location of the cysteine residue is on the beta barrel on the opposite side of the protein to the SH2 domain. An overview of the general sequence coverage on STAT3β with and without STATTIC showed that areas of digest from the control (Figure 5.19A, blue), treated (blue and green) and modified peptide (red) give no identification of the SH2 domain (Figure 5.17) nor in many of the helices which appear to be interacting with each other.

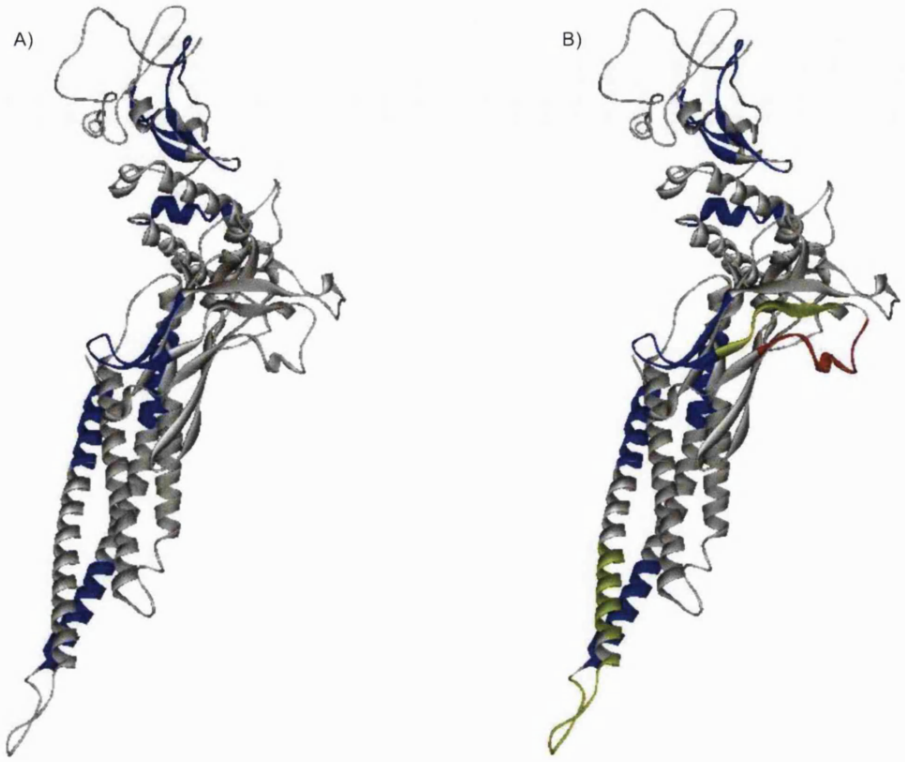


Figure 5.19: Digest coverage of areas of STAT3β. A) Control STAT3β with no STATTIC. Coverage of peptides in blue. B) STAT3β incubated with STATTIC. Average peptides identified with similar peptides to the control in blue, new areas in green and modified in red. Viewed using DS Visualizer 2.5.

While one modified cysteine residue was identified, it was uncertain if multiple modifications were present as the coverage of the protein was incomplete. Getting an intact mass would allow the number of covalently bound STATTIC molecules to be identified and to confirm the digest data.

STAT3β was dialysed in PBS and in 100 mM ABC. For each buffer, MALDI-TOF protein analysis was attempted as some initial spectra of STAT3β in PBS samples showed potential for intact mass analysis. A sample of 100 mM ABC was also lyophilised and re-suspended in H_2O. Samples were centrifuged prior to preparation for MALDI-TOF analysis. MALDI-TOF samples were prepared using α-CA, SA, or DHB by dried droplet, sandwich and layer methods. The summary of all experiments is shown in Table 5.12, where only two spectra had a peak of intact mass of STAT3β.

Table 5.12: Results from matrices and sample preparation techniques

Sample	Method	Matrix	Mixed	Result
100 mM ABC	Dried droplet	α-CA	2:1	X
			1:1	X
			1:5	X
	Layer			X
	Sandwich			X
	Dried droplet	DHB	2:1	X
			1:1	X
			1:5	X
	Layer			X
	Sandwich			YES
	Dried droplet	SA	2:1	X
			1:1	X
			1:5	X
	Layer			X
	Sandwich			X
PBS	Dried droplet	α-CA	2:1	X
			1:1	X
			1:5	X
	Layer			X
	Sandwich			X
	Dried droplet	DHB	2:1	X
			1:1	X
			1:5	X
	Layer			X
	Sandwich			X
	Dried droplet	SA	2:1	X
			1:1	X
			1:5	X
	Layer			X
	Sandwich			YES
Lyophilised	Dried droplet	α-CA	2:1	X
			1:1	X
			1:5	X
	Layer			X
	Sandwich			X
	Dried droplet	DHB	2:1	X
			1:1	X
			1:5	X
	Layer			X
	Sandwich			YES
	Dried droplet	SA	2:1	YES
			1:1	YES
			1:5	X
	Layer			X
	Sandwich			YES

The lyophilised sample gave better promise with respect to analysis in comparison to the other prepared samples. SA gave better results than any of the other matrices attempted.

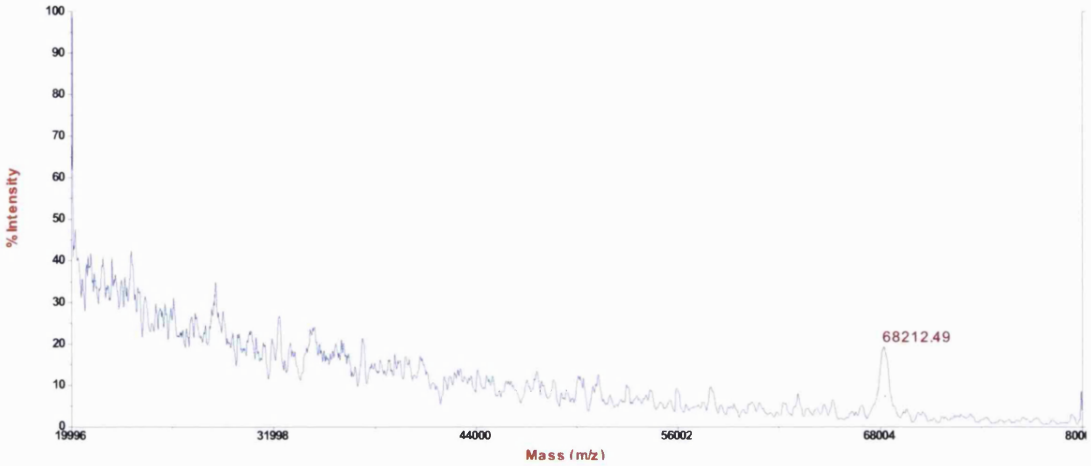


Figure 5.20: STAT3β analysis on the MALDI-TOF using 2:1 sample: SA using the dried droplet method.

Analysis of the sample kept in the fridge overnight using SA and the dried droplet method yielded no results, suggesting that the lyophilised sample was less stable than the regular mixture and that mass spectrometry analysis needs to be done immediately or the signal diminishes.

To ensure that the modification by STATTIC can be detected, the STAT3β and BSA samples were incubated with STATTIC and their spectra acquired using MALDI-TOF and QTOF. STAT3β was dialysed against 10 mM ammonium acetate while BSA (5μM) was prepared in 10 mM ammonium acetate. Both BSA and STAT3β were mixed with 0.1 % FA 50:50 ACN/H_2O and infused on the QTOF, while the MALDI samples were mixed with SA. The aliquot was incubated at 37°C for 1 hour with a final concentration of 100 mM STATTIC dissolved in DMSO. Parameters and optimisation for the QTOF were performed using BSA followed by BSA incubated with STATTIC.

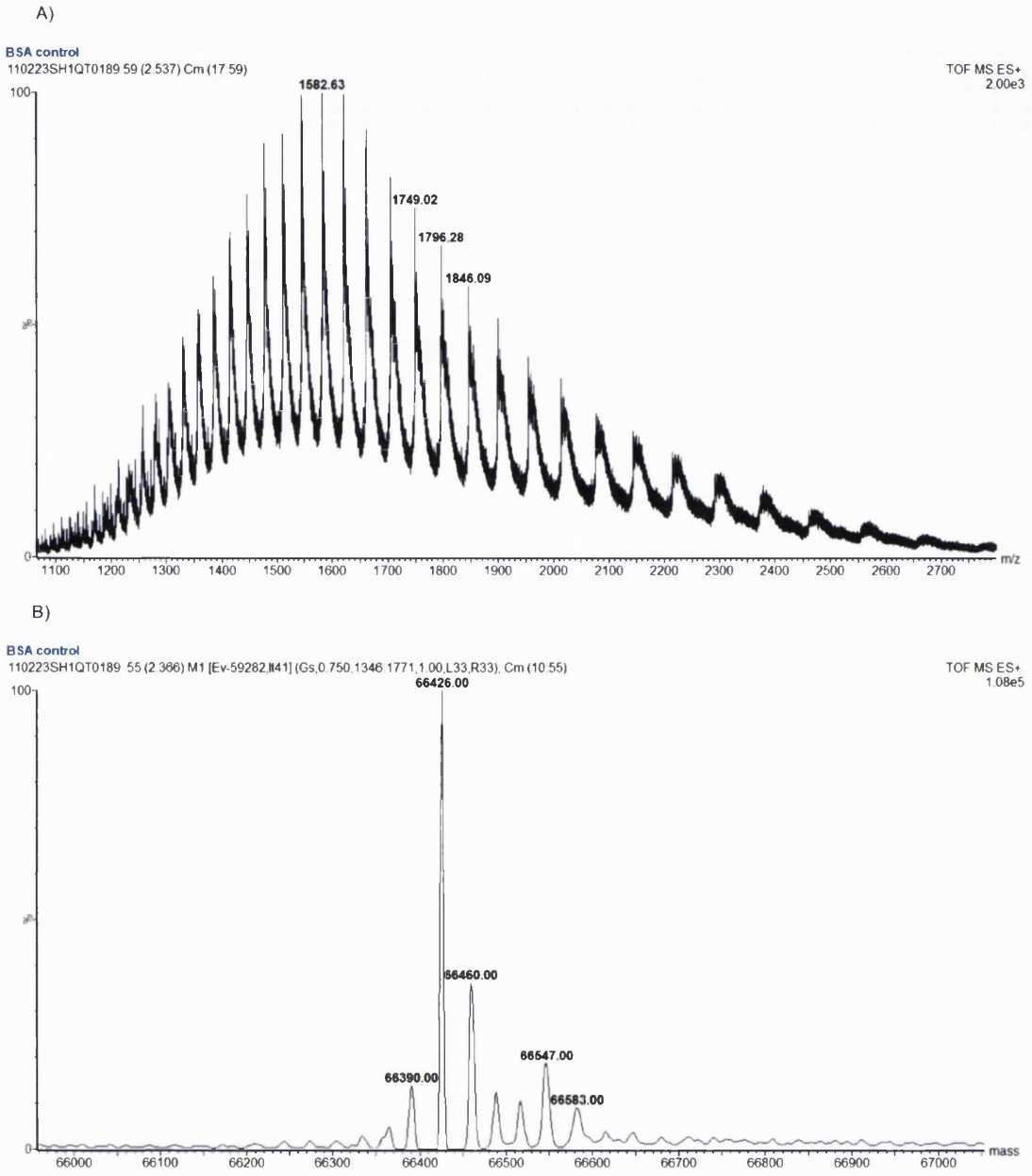


Figure 5.21: BSA analysis with STATTIC on the QTOF. A) BSA electrospray spectrum. B) BSA deconvoluted. (Continued on the next page)

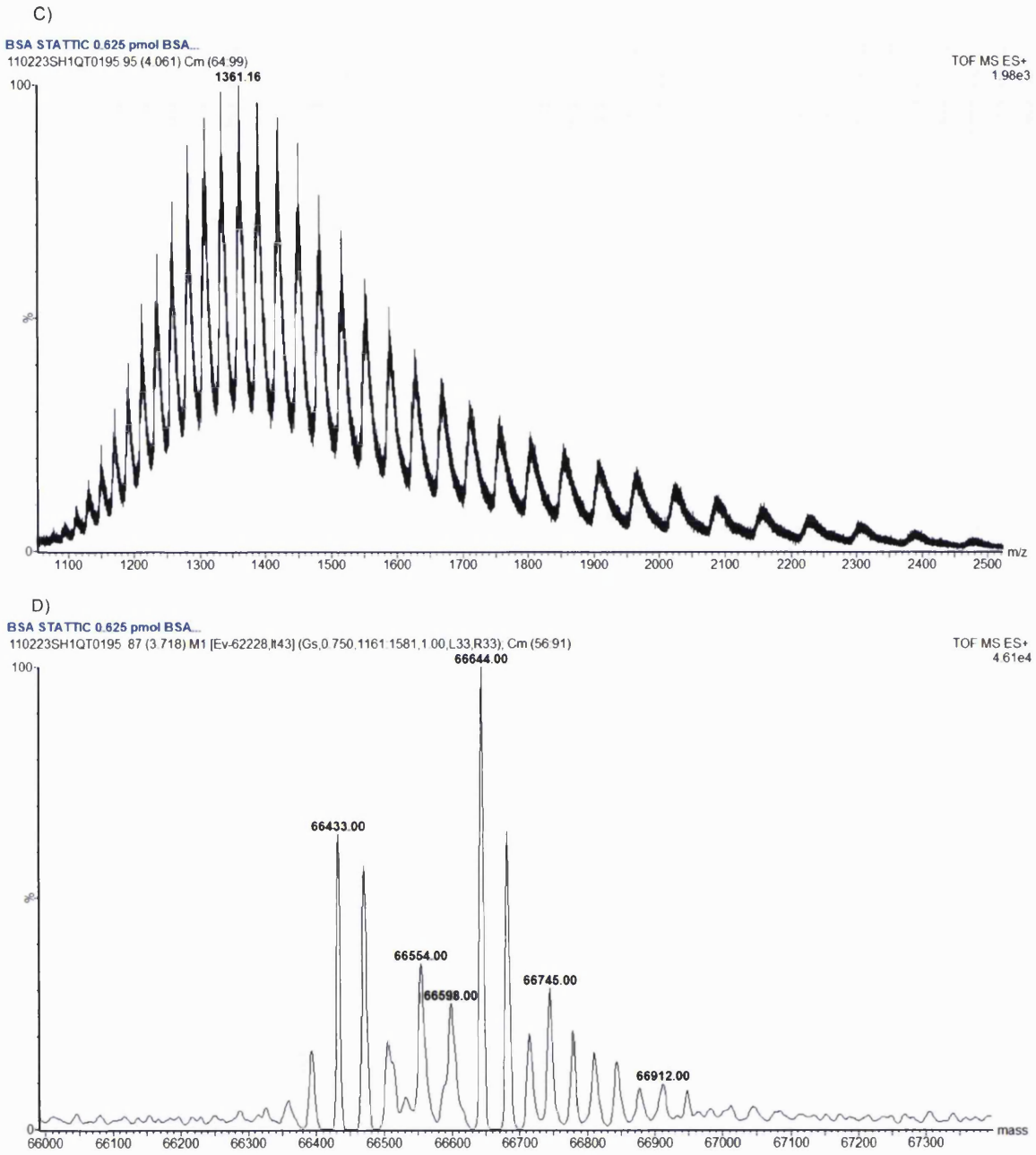


Figure 5.21: BSA analysis with STATTIC on the QTOF. C) BSA and STATTIC electrospray. D) BSA and STATTIC deconvoluted.

BSA was visible and the deconvolution of BSA and BSA with STATTIC showed differences in the spectrum with an addition of only one 210.99 in the spectrum with BSA and STATTIC. It is however seen that BSA without STATTIC was also present in the mixture (Figure 5.21D) showing incomplete modification of the protein at the cysteine residue.

Continuing with STATTIC in DMSO, STAT3β was purified and dialysed into 10 mM AA and the resulting protein was run on the QTOF.

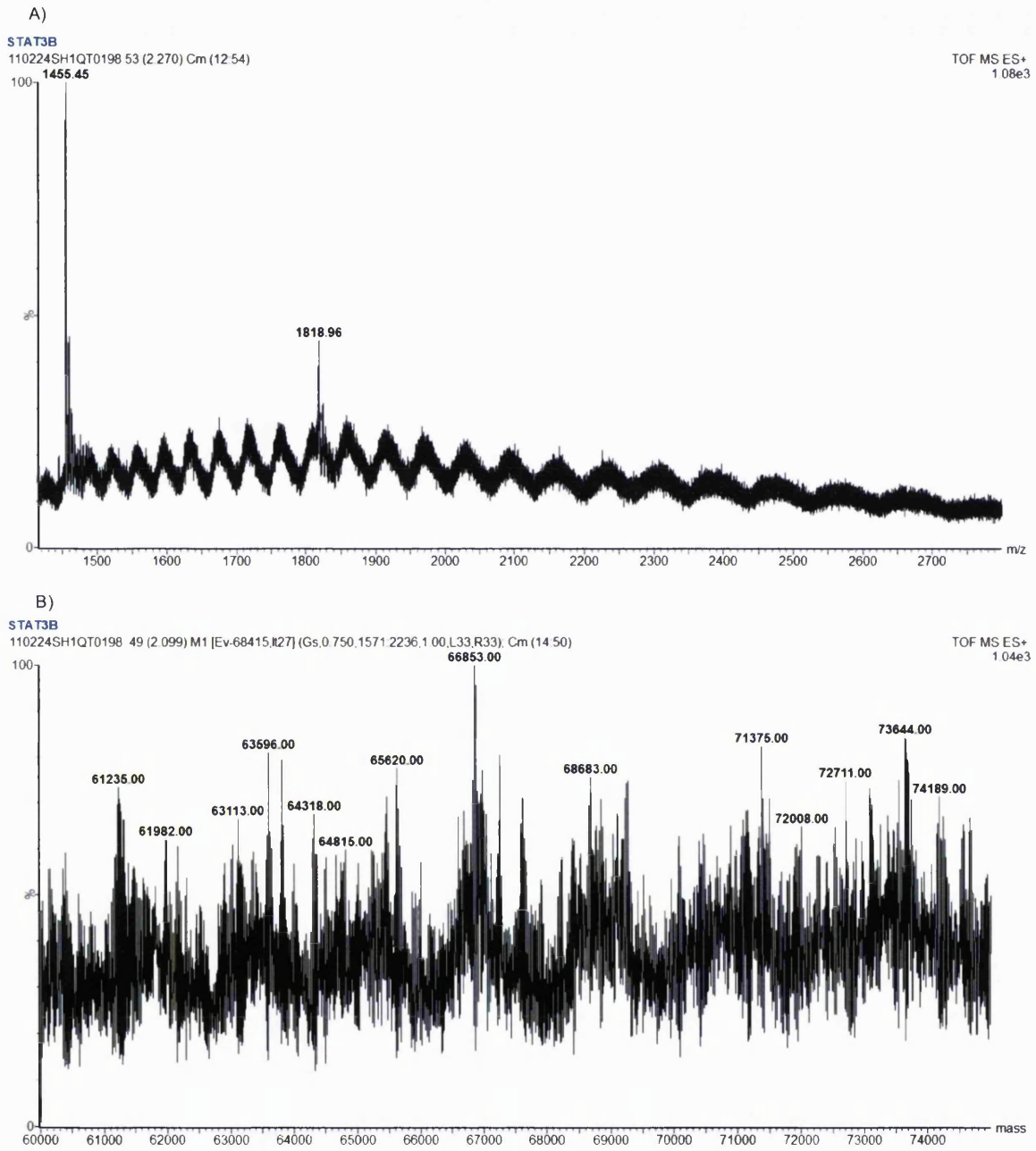


Figure 5.22: STAT3β analysis on QTOF. A) ESI spectrum and B) deconvoluted spectrum of STAT3β.

The spectrum of STAT3β intact on the QTOF was seen to be very low and did not deconvolute into one peak. Addition of STATTIC to the protein caused precipitation of the protein and resulted in no spectra.

BSA and STAT3β were attempted with and without STATTIC on the MALDI-TOF using SA dried droplet method.

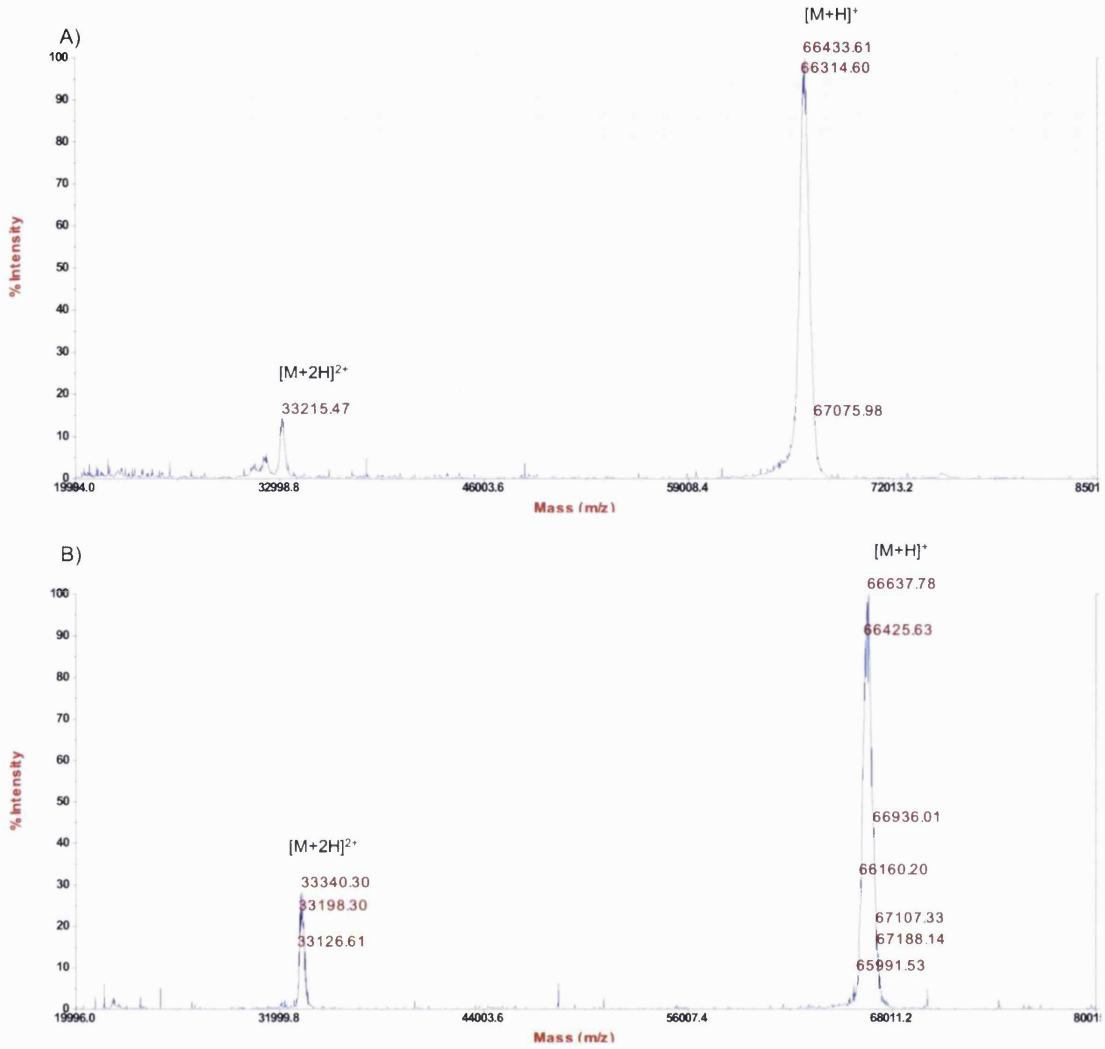


Figure 5.23: MALDI-TOF analysis of BSA A) BSA, B) BSA in the presence of STATTIC. $[M+H]^+$ corresponds to the protonated BSA sample while $[M+2H]^{2+}$ corresponds to the doubly charged form of the protein.

BSA was visible in the presence and the absence of STATTIC with a mass shift of +238 which is less than 0.1 % mass difference. The same was repeated on a separate day to confirm the mass shift seen on the BSA and to attempt to see STAT3β with STATTIC.

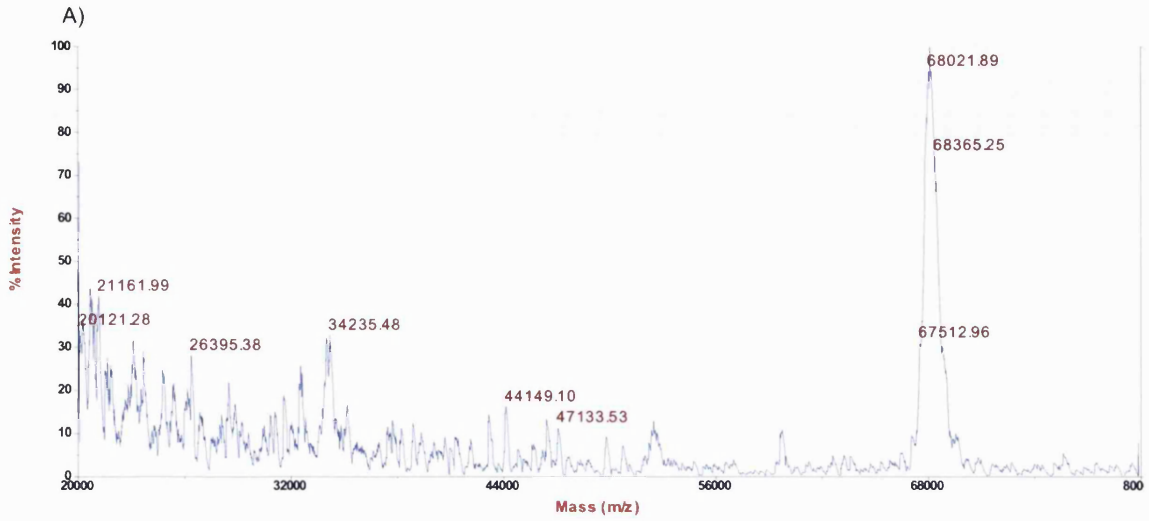


Figure 5.24: MALDI-TOF Analysis of STAT3β A) STAT3β, B) STAT3β in the presence of STATTIC.

STAT3β was visible on the MALDI-TOF, however once incubated with the STATTIC molecule, precipitated out, preventing any analysis on the mass spectrometer. Myoglobin and insulin (oxidised chain) were also incubated using the same conditions with STATTIC and infused on the QTOF.

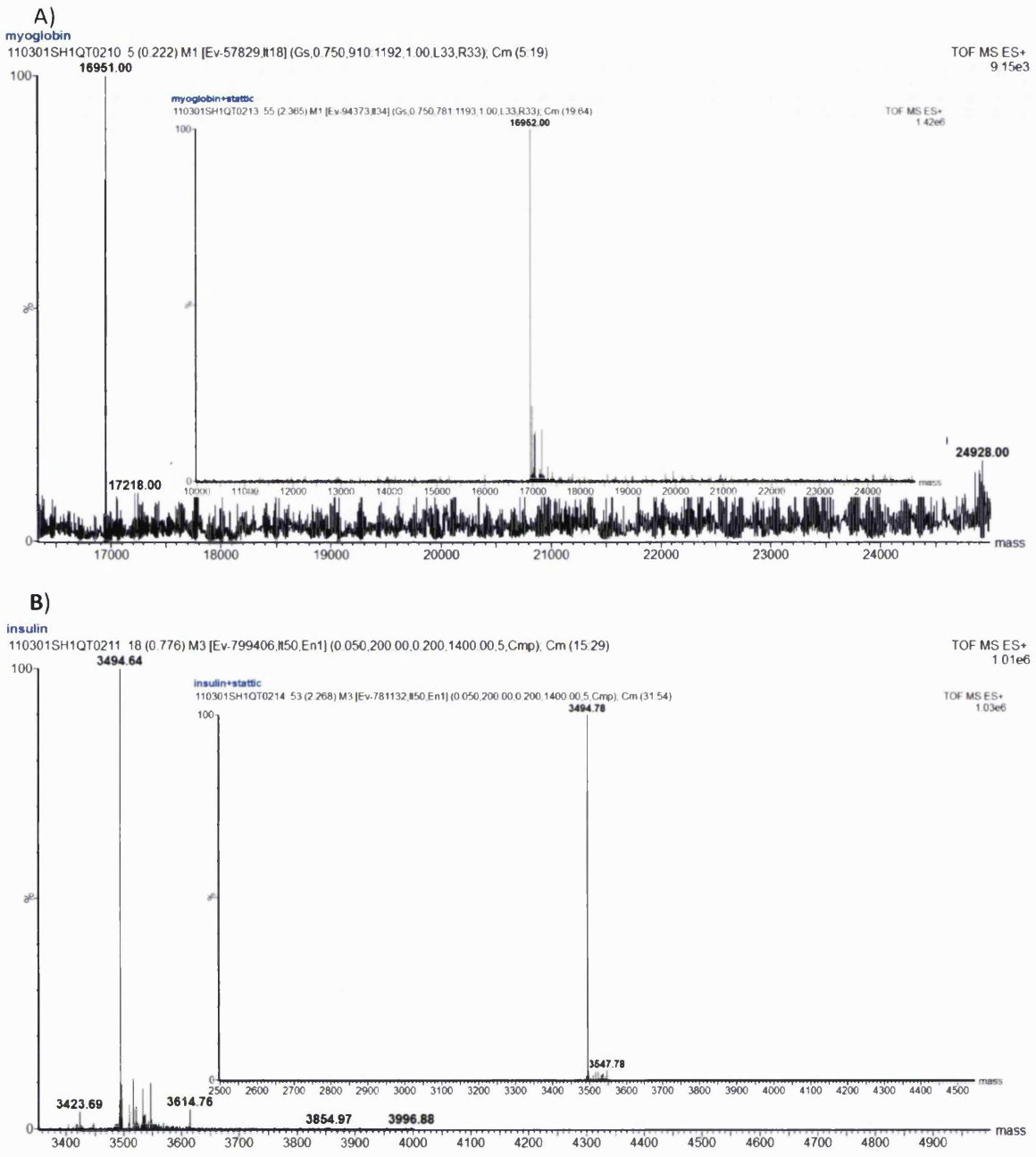


Figure 5.25: Incubation of Myoglobin and Insulin with STATTIC. A) Myoglobin deconvoluted and with STATTIC (inset). B) Insulin deconvoluted and with STATTIC (inset).

Both myoglobin and insulin did not show any sign of STATTIC binding.

5.5 Discussion

Using STAT3β, a truncated version of STAT3 composed of residues 122-722 of the intact protein, studies were undertaken to determine the binding site of STATTIC, a small molecule which was found to bind to STAT3β. In particular, it was of interest to establish whether the STATTIC molecule would specifically bind to the SH2 domain of the protein. The mostly likely targets of STATTIC would be the nucleophilic residues: cysteine, followed by serine and tyrosine using the proposed mechanism (Figure 5.2). Previously published studies on STAT3 showed that STATTIC killed or inhibited STAT3 dependent cancer cell lines.[196]

STD or saturation transfer difference spectroscopy is an NMR technique that is used to monitor large molecule - ligand interactions. The large molecule, generally protein, is initially screened by NMR to identify a region that contains only resonances from the protein itself. The resonance of the protein is selectively saturated in this region, and general loss of magnetisation occurs through the spin diffusion through the entire protein. If a ligand weakly interacts with the protein, the intermolecular transfer of magnetisation from the protein onto ligand protons will occur via spatial spin diffusion. The resulting STD spectrum shows only resonances that have experienced saturation, such as the peaks of the ligand. Typically the protein is maintained at a low concentration while the ligand is at a far higher concentration, eliminating the need to filter out the peaks of the protein.[213]

Saturation transfer dissociation NMR was carried out first on BSA with ibuprofen/naproxen/salicylic to test the NMR method and ensure that the method was working properly. NMR STD experiments were performed to determine if the small molecule-protein interaction is a non-covalent interaction. If the interactions are non-covalent, then the peaks of a small molecule will become visible during the experiment as the protein transfers magnetisation to the small molecule, causing it to become visible. If peaks are not seen then it is likely that there is no interaction or there is strong interaction with a possible formation of a covalent bond. BSA was seen to interact with all three molecules (Figure 5.4, Figure 5.5, Figure 5.6) as well as with mixtures of the three molecules showing both non-competitive inhibition with naproxen and salicylic acid and competitive inhibition between naproxen and ibuprofen – with naproxen being visible (Figure 5.7, Figure 5.8).

BSA when incubated with STATTIC showed no peaks visible in STD NMR spectrum suggesting that there is either no interaction or covalent bond formation (Figure 5.9)

while STAT3β constantly precipitated out of solution upon the addition of STATTIC in sufficient concentration for NMR analysis.

STAT3β proved to be difficult to analyse by mass spectrometry methods.[196] It was hoped that the addition of STATTIC would help with stabilising the protein to allow for MS spectrum acquisition and analysis. If not, the addition of STATTIC followed by digestion would allow for identification of the peptide(s) to which the STATTIC should be bound.

STAT3β was initially analysed on the MALDI-TOF in PBS, Hepes and Tris buffers (Figure 5.10). Previous attempts at protein analysis in these buffers gave good results and with the instability of STAT3β in solution it was deemed worthwhile to attempt analysis in non-volatile buffer conditions. Previously published attempts at mass spectrometry of STAT3 have been unsuccessful.[196] An attempt was made to take the pellet of extracted STAT3β and dissolve it in a variety of mass spectrometry compatible buffers at different pHs and concentrations (Table 5.6), to determine which, if any, would be optimal for solubility and for intact mass. In most cases, nothing at all was seen which may have been due to the complexity of the mixture, however all showed precipitate forming after a day, suggesting that the stability of the STAT3β is very low in both ammonium bicarbonate and ammonium acetate. Attempts to concentrate purified STAT3β or to buffer exchange using Vivaspin or microcons was unsuccessful as no MALDI spectrum for STAT3β was obtained (Figure 5.11). The instability of STAT3β was seen in the previous buffer study through precipitation of the protein suggesting that it is quite likely that the STAT3β precipitated onto the membrane when a certain concentration was reached or when the PBS was exchanged to ammonium acetate or ammonium bicarbonate.

The same buffer mixtures were used for incubation of STATTIC with STAT3β. The hope was that STATTIC might stabilise the STAT3β protein. This, however was not the case and in most instances it was observed that STATTIC increased the speed at which STAT3β precipitated out of solution. With the current lack of success to maintain STAT3β in solution, purified STAT3β was dialysed from PBS into 10 mM ammonium bicarbonate overnight at 4°C. The protein was removed from the dialysis tubing and immediately analysed on the MALDI-TOF, resulting in no intact mass analysis.

STAT3β dialysed in PBS or ammonium acetate and a portion was incubated with STATTIC at concentrations of 0, 1X, 2X, 5X, 10X, 15X and 20X the concentration of the protein. The resulting mixture was digested in solution and in-gel with trypsin at

37°C. The resulting digests were run on LC-MS/MS database and searching was performed using PLGS and X!Tandem. While PLGS was able to search for the specific sequence of STAT3β, it was unable to take into account the the addition of STATTIC (+211, +422 etc), while X!Tandem allowed for additional modifications to be added to the search. BSA was used as a control and was also incubated with STATTIC, digested, run on LC-MSMS and searched. The search for BSA resulted in a positive identification of BSA as well as identifying a peptide mass with an addition of 211, the mass of the STATTIC molecule. This peptide was found with good amino acid coverage confirming it as a peptide from BSA (Table 5.7). BSA was used as a random control as it had been used to prepare the NMR-STD experiments. While previously published work on STATTIC found specificity in its action towards STAT3[196], the belief is that the specificity of STATTIC is not absolute for STAT3, confirmed by the addition of STATTIC to a C_{288} residue in BSA. A second set of incubation studies found the STATTIC molecule bound to the C_{288} residue of BSA as well as to a C_{367} residue in STAT3β (Figure 5.16, Figure 5.17). The peptide to which it was bound, was not identified in control samples of STAT3β.

A 3-D protein model of STAT3βTC – a modified and truncated version of STAT3β – was used to identify the location of the C_{367} residue (Figure 5.18). The assumption was that in order for inhibition of STAT3β dimerisation to occur, STATTIC would mostly likely have to bind somewhere in or near the SH2 domain – which is the region of the protein which interacts with another monomer of STAT3β to dimerise. Review of the peptide digest data, showed that no peptides were identified from the SH2 domain (Figure 5.19) resulting in an incomplete view of the protein. While one binding side has been identified, there is a possibility of a second binding site that has not yet been identified.

Work was continued on identifying an intact mass of STAT3β with STATTIC as a means of confirming the digest results obtained for BSA and STAT3β. A newly purified STAT3β was dialysed in 100 mM ammonium bicarbonate and attempts were made using α-CA, SA and DHB for MALDI-TOF analysis. Dried droplet, sandwich and layer methods were attempted with both. The same was attempted with incubations of STATTIC at 37°C. Although the signal to noise is very low, there is some signal seen with dried droplet with SA, but not with any of the others (Table 5.12). Using different buffers for solubilizing STAT3β as well as different pHs showed that the best and most consistent results were obtained from incubating in 10 mM AA and with a 1hr incubation at 37°C. STAT3β was dialysed against 10 mM ammonium acetate and was analysed on the QTOF and the MALDI-TOF. MALDI-TOF (Figure 5.20), gave results

for intact mass analysis, while the QTOF (Figure 5.22) did not provide any results. The sample was re-analysed by MALDI-TOF and QTOF the following day, however no signal was obtained on either instrument, suggesting that the stability of the protein again was very short. Samples were checked for protein concentration after centrifugation and while protein was still present in solution no spectra were obtainable on either instrument. Another sample of STAT3β was prepared with 100 μM STATTIC in DMSO and incubated at 37°C. This was then analysed using MALDI-TOF (Figure 5.24) and QTOF. STAT3β was observed to precipitate at any STATTIC concentrations and no spectrum for STAT3β on the QTOF was obtained (Figure 5.24). BSA was treated as and used as a control for incubation studies and using both MALDI-TOF and QTOF, showed a mass difference of 211 between the control and the treated BSA samples, corresponding to one STATTIC molecule (Figure 5.21). BSA and STATTIC were confirmed at low (pmol) concentration of BSA. Attempts to remove the STATTIC molecule by increasing capillary, cone or collision energy proved ineffective. STAT3β incubations with STATTIC resulted in almost an immediate precipitation of the protein.

Although BSA was expected to be a negative control, that is STATTIC was not expected to bind, digestion studies revealed a single binding site on the BSA protein at a cysteine residue. STATTIC was then incubated using horse heart myoglobin – which does not contain any cysteine residues, and insulin chain B oxidised – which contains a modified cysteine residue, to see if STATTIC would bind to any other nucleophilic amino acid when cysteine residues are not present (Figure 5.25). Incubations of these with STATTIC resulted in no binding using the same buffers and concentrations used with the BSA model suggesting that while STATTIC's specificity for STAT3β is not absolute as was suggested, it does have some specificity as it binds to BSA, but not to myoglobin or insulin and appears to be specific towards cysteine residues.

5.6 Conclusion and Future work

Despite major difficulties in maintaining STAT3β in solution, with or without STATTIC, a potential binding site was found consistently in most of the digested samples. C_{367} seems to be highly reactive to STATTIC and was the only residue identified with a +210.99 modification, although the lack of peptides identified in the SH2 domain does not preclude multiple binding sites. Identification of a binding site confirms what Schust *et al* [196] believed, but could not confirm in their own mass spectrometry attempts. Two other methods were used to identify STAT3β and STATTIC to confirm covalent binding. The first method was used to exclude non-covalent binding through the use of STD-NMR experiments which are able to monitor non-covalent interactions. The second method involved protein-ligand analysis by MALDI-TOF and QTOF for covalent additions to the mass of the protein. MALDI-TOF intact mass was obtained for the STAT3β protein. While a specific binding site was found for STAT3β, it was also found that STATTIC binds to BSA although not to myoglobin or insulin, suggesting that while there is specificity, it is not necessarily only to STAT3β, and not necessarily to the SH2 domain which was the proposed location to prevent dimerisation of STAT3.

Future work could include comparing the binding areas of STAT3β and BSA to determine if there is something in the location that makes it more specific for addition, as well, to test other members of the STAT family to determine if they have the same reactivity and if that particular region of the protein is modified in any other member of the STAT family. As mass spectrometry analysis of the SH2 domain has proven to be difficult, creating a truncated protein consisting only of the SH2 domain would allow for further incubation studies and to potentially allow for identification of more STATTIC binding sites. Site directed mutagenesis of this particular cysteine residue would determine the significance of the results by mass spectrometry or if the binding is inconsequential to the STATTIC activity.

6 Summary and General Discussion

The last three decades have seen a phenomenal increase in the use of mass spectrometry, in many cases thanks to the soft ionisation techniques: ESI and MALDI. This has allowed for the analysis of biological molecules such as high molecular weight proteins as well as small metabolites using both their intact masses and their product ion masses for identification. Its use continues to increase with newer and more sensitive instrumentation such as FT-ICR and the Orbitrap as well as new techniques for fragmentation that allows for the identification of post-translational modifications such as ECD and ETD.

From small molecule analysis of oxysterols in plasma using an ion-trap, it can be seen that even some of the older technologies can still be used for complex questions. Oxysterols which are mainly neutral, were extracted from plasma, derivatised and separated from other metabolites in an attempt to look at 24S-hydroxycholesterol a hydroxycholesterol synthesised solely in the brain and 27-hydroxycholesterol, a hydroxysterol commonly found in plasma. The two co-elute so their peak areas and heights were combined and compared to that of 3β-hydroxycholest-5-en-27-oic acid. Ratios of the two were considered a practical means for quantitative analysis as absolute values of 24S-hydroxycholesterol would vary from person to person, however the ratio of this to other metabolites should be the same. This analysis would be useful for targeting neurological issues related to cholesterol in the brain such as Alzheimers, allowing for an early prognosis. The steps for extraction are not simple and are time-consuming as the cholesterol needs to be separated from the hydroxysterols to prevent autoxidation resulting in a change of hydroxysterol analysis. The derivatisation required, while simple also requires purification before the analysis can be completed using LC-MS and selective reaction monitoring to increase sensitivity. This method for hydroxysterol analysis proved to be very robust and reproducible, which has been transferred to an Orbitrap, leading to more sensitive analysis and identification of more than just the hydroxysterols in the samples.[4]

The 24S-hydroxysterol, is only produced in the brain and as a consequence, the protein cytochrome P450 46A1 was searched for as an important protein in brain. Screening involved separating the brain into three fractions: cytosolic, microsomal and mitochondrial fractions. In the process of searching for proteins involved in the synthesis and metabolism of cholesterol 1838 individual proteins were identified from replicate runs, with only 13, less than 1% of the total proteins identified from steroidal pathways, receptors or transporters. P450 46A1 was identified, but not reproducibly.

The use of 1D-gel electrophoresis for total protein identification gave a higher number of identifications when compared to other techniques used by the HUPO brain pilot study. While the analysis on the QTOF for these low abundant proteins was difficult, it was considered to be the separation and not the instrument that caused the problem. Separation such as 2D-LC would probably give better separation by allowing for more peptides to be identified. Another method of separation would be using a solution digest – followed by 2D-LC which would prevent peptide loss that may be seen in the SDS-PAGE extraction.

Using the same method of separation above, 1-DE followed by LC-MS/MS on the QTOF for data dependent acquisition was considered an appropriate starting point for the comparative analysis of control and UV treated *Anabaena variabilis*, which upon UV irradiation forms a UV absorbing compound shinorine. While shinorine is readily seen using UV, it was not well seen using mass spectrometry and quantification was performed using analytical HPLC-UV as it was far more sensitive than the mass spectrometers. Once shinorine production was confirmed at 72 hours, control and treated cyanobacteria were split into soluble and insoluble fractions and run on 1-DE. From the 1-DE, differences were clearly visible and the bands were cut, digested, extracted and run using data dependent acquisition and analysed using X!Tandem, PLGS and MASCOT. While X!Tandem and PLGS were compared to each other, the use of NCBI database by X!Tandem for *Anabaena variabilis* made it difficult to do a direct comparison to PLGS as the redundant database caused multiple identifications of the same protein. Nevertheless, some reductases and other proteins of interest were found. MASCOT was run and homology searches were performed to identify proteins of interest, resulting in particular proteins, from the shikimate pathway being identified, which further suggested that this pathway has a part to play in the synthesis of shinorine as evidenced by inhibition studies. Further separation of proteins or a more targeted approach to their identification would allow for more identification by reducing the complexity of the samples. Despite the fact that this pathway induces the production of a small molecule, the proteins themselves appear to be at a low concentration and are still masked by other higher abundant proteins.

Protein interactions were also studied using STAT3β, a truncated form of STAT3, a human oncoprotein involved in the activation of transcription factor 3. Previously published work suggested that STATTIC a small molecule was a selective inhibitor of STAT3 and it was assumed that this inhibition was through covalent modification of the SH2 domain – the domain that allows for dimerisation of STAT3 with another STAT3 monomer. Intact mass analysis was unable to determine the number of STATTIC

molecules however in-gel and in-solution digest showed identification of a binding site on a cysteine residue that was however not in the SH2 domain. While no peptides were identified in the SH2 domain, it could not be ruled out that a second binding site may also occur within this domain. A cysteine binding site was also identified on BSA confirming that the specificity of STATTIC isn't just for STAT3β, which suggests that further testing and analyses are required.

While mass spectrometry has proven to be a very powerful technique, it also has limitations. It has been shown to identify proteins of interest in screening methods, but in these cases, it requires good sample preparation prior to analysis to reduce the complexity. It has also been shown to be able to quantitatively analyse small molecules and has been able to identify covalent binding of a small molecule to a protein, showing its ability to answer many different questions related to biological systems. However, quantification of a small neutral UV absorbing molecule proved to be more effective by UV-HPLC and identification of difficult proteins was still more easily obtained by western blot in the case of STAT3β. As well, the SH2 domain, a domain of interest has proven itself to be difficult to analyse by mass spectrometry – in intact form or in digested form. Mass spectrometry while a valuable too, when used in tandem with other biochemical and molecular biology techniques, such as Western blots, NMR, SDS-PAGE, HPLC and fluorescence assays, can give valuable information and as the instrumentation and techniques improve, mass spectrometry will continue to improve in its ability to help us understand complex biological systems.

References

1. Simpson, R.J., *Proteins and proteomics : a laboratory manual.* 2003, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
2. Walther, T.C. and M. Mann, *Mass spectrometry-based proteomics in cell biology.* The Journal of Cell Biology, 2010. **190**(4): p. 491-500.
3. Lapolla, A., D. Fedele, and P. Traldi, *Diabetes and mass spectrometry.* Diabetes/Metabolism Research and Reviews, 2001. **17**(2): p. 99-112.
4. Griffiths, W.J., et al., *Discovering Oxysterols in Plasma: A Window on the Metabolome.* Journal of Proteome Research, 2008. **7**(8): p. 3602-3612.
5. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics.* Nature, 2003. **422**(6928): p. 198-207.
6. Joshi, S., et al., *Oncoproteomics.* Clinica Chimica Acta, 2011. **412**(3-4): p. 217-226.
7. Griffiths, W.J., et al., *Targeted Metabolomics for Biomarker Discovery.* Angewandte Chemie International Edition, 2010. **49**(32): p. 5426-5445.
8. Chace, D.H., *Mass spectrometry in newborn and metabolic screening: historical perspective and future directions.* Journal of Mass Spectrometry, 2009. **44**(2): p. 163-170.
9. Drexler, D., M. Reily, and P. Shipkova, *Advances in mass spectrometry applied to pharmaceutical metabolomics.* Analytical and Bioanalytical Chemistry, 2011. **399**(8): p. 2645-2653.
10. de Hoffmann, E. and V. Stroobant, *Mass spectrometry : principles and applications.* 2007, Chichester, England; Hoboken, NJ: J. Wiley.
11. Siuzdak, G., *The expanding role of mass spectrometry in biotechnology.* 2003, San Diego, Calif.: MCC Press.
12. March, R.E., *Quadrupole ion traps.* Mass Spectrometry Reviews, 2009. **28**(6): p. 961-989.
13. Lane, C., S, *Mass spectrometry-based proteomics in the life sciences.* Cellular and Molecular Life Sciences, 2005. **62**(7-8): p. 848-869.
14. Fenn, J.B., *Electrospray ionization mass spectrometry: How it all began.* J Biomol Tech, 2002. **13**(3): p. 101-118.
15. Douglas, D.J., A.J. Frank, and D. Mao, *Linear ion traps in mass spectrometry.* Mass Spectrometry Reviews, 2005. **24**(1): p. 1-29.
16. Gaskell, S.J., *Electrospray: Principles and Practice.* Journal of Mass Spectrometry, 1997. **32**(7): p. 677-688.
17. Yamashita, M. and J.B. Fenn, *Electrospray ion source. Another variation on the free-jet theme.* J. Phys. Chem., 1984. **88**(20): p. 4451-4459.
18. Yamashita, M. and J.B. Fenn, *Negative ion production with the electrospray ion source.* The Journal of Physical Chemistry, 1984. **88**(20): p. 4671-4675.
19. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.
20. Wilm, M.S. and M. Mann, *Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last?* International Journal of Mass Spectrometry and Ion Processes, 1994. **136**(2-3): p. 167-180.
21. Kebarle, P., *A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry.* Journal of Mass Spectrometry, 2000. **35**(7): p. 804-817.
22. Karas, M. and F. Hillenkamp, *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons.* Analytical Chemistry, 1988. **60**(20): p. 2299-2301.
23. Tanaka, K., et al., *Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry.* Rapid Communications in Mass Spectrometry, 1988. **2**(8): p. 151-153.
24. Taneja, S., et al., *Plasma peptidome profiling of acute hepatitis E patients by MALDI-TOF/TOF.* Proteome Science, 2011. **9**: p. 13.

25. Xiao, D., et al., *Analysis of the urinary peptidome associated with Helicobacter pylori infection.* World Journal of Gastroenterology, 2011. **17**(5): p. 618-624.
26. Mann, M. and G. Talbo, *Developments in matrix-assisted laser desorption/ionization peptide mass spectrometry.* Current Opinion in Biotechnology, 1996. **7**(1): p. 11-19.
27. Griffin, P.R., et al., *Direct database searching with MALDI-PSD spectra of peptides.* Rapid Communications in Mass Spectrometry, 1995. **9**(15): p. 1546-1551.
28. Perkins, J.R., et al., *Application of electrospray mass spectrometry and matix-assisted laser desorption ionization time-of-flight mass spectrometry for molecular weight assignment of peptides in complex mixtures.* Journal of The American Society for Mass Spectrometry, 1993. **4**(8): p. 670-684.
29. Rasmussen, H.H., et al., *Identification of transformation sensitive proteins recorded in human two-dimensional gel protein databases by mass spectrometric peptide mapping alone and in combination with microsequencing.* Electrophoresis 1994. **15**(1): p. 406-416.
30. Zaluzec, E.J., D.A. Gage, and J.T. Watson, *Matrix-Assisted Laser Desorption Ionization Mass Spectrometry: Applications in Peptide and Protein Characterization.* Protein Expression and Purification, 1995. **6**(2): p. 109-123.
31. Zhu, Y.F., et al., *Revisit of MALDI for small proteins.* Rapid Communications in Mass Spectrometry, 1995. **9**(13): p. 1315-1320.
32. Glocker, M.O., et al., *Characterization of Specific Noncovalent Protein Complexes by UV Matrix-assisted Laser Desorption Ionization Mass Spectrometry.* Journal of Mass Spectrometry, 1996. **31**(11): p. 1221-1227.
33. Loo, J.A., et al., *High sensitivity mass spectrometric methods for obtaining intact molecular weights from gel-separated proteins.* Electrophoresis, 1999. **20**(4-5): p. 743-748.
34. Puchades, M., et al., *Analysis of intact proteins from cerebrospinal fluid by matrix-assisted laser desorption/ionization mass spectrometry after two-dimensional liquid-phase electrophoresis.* Rapid Communications in Mass Spectrometry, 1999. **13**(24): p. 2450-2455.
35. Vaidyanathan, S., et al., *Sample preparation in matrix-assisted laser desorption/ionization mass spectrometry of whole bacterial cells and the detection of high mass (>20 kDa) proteins.* Rapid Communications in Mass Spectrometry, 2002. **16**(13): p. 1276-1286.
36. Macht, M., et al., *"Affinity-proteomics": direct protein identification from biological material using mass spectrometric epitope mapping.* Analytical and Bioanalytical Chemistry, 2004. **378**(4): p. 1102-1111.
37. Lin, M., et al., *Intact protein analysis by matrix-assisted laser desorption/ionization tandem time-of-flight mass spectrometry.* Rapid Communications in Mass Spectrometry, 2003. **17**(16): p. 1809-1814.
38. Demirev, P.A., et al., *Bioinformatics and Mass Spectrometry for Microorganism Identification:Proteome-Wide Post-Translational Modifications and Database Search Algorithms for Characterization of Intact H. pylori.* Analytical Chemistry, 2001. **73**(19): p. 4566-4573.
39. Lochnit, G. and R. Geyer, *Carbohydrate Structure Analysis of Batroxobin, a Thrombin-Like Serine Protease from Bothrops moojeni Venom.* European Journal of Biochemistry, 1995. **228**(3): p. 805-816.
40. Remoortere, A., et al., *MALDI imaging and profiling MS of higher mass proteins from tissue.* Journal of The American Society for Mass Spectrometry, 2010. **21**(11): p. 1922-1929.
41. Fenselau, C. and P.A. Demirev, *Characterization of intact microorganisms by MALDI mass spectrometry.* Mass Spectrometry Reviews, 2001. **20**(4): p. 157-171.
42. Grey, A.C., et al., *MALDI Imaging Mass Spectrometry of Integral Membrane Proteins from Ocular Lens and Retinal Tissue* Journal of Proteome Research, 2009. **8**(7): p. 3278-3283.

43. Schwamborn, K. and R.M. Caprioli, *Molecular imaging by mass spectrometry - looking beyond classical histology.* Nat Rev Cancer, 2010. **10**(9): p. 639-646.
44. Schwamborn, K. and R.M. Caprioli, *MALDI Imaging Mass Spectrometry - Painting Molecular Pictures.* Molecular Oncology, 2010. **4**(6): p. 529-538.
45. Zhang, J. and R. Zenobi, *Matrix-dependent cationization in MALDI mass spectrometry.* Journal of Mass Spectrometry, 2004. **39**(7): p. 808-816.
46. Stults, J.T., *Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS).* Current Opinion in Structural Biology, 1995. **5**(5): p. 691-698.
47. Glückmann, M. and M. Karas, *The initial ion velocity and its dependence on matrix, analyte and preparation method in ultraviolet matrix-assisted laser desorption/ionization.* Journal of Mass Spectrometry, 1999. **34**(5): p. 467-477.
48. Gabelica, V., E. Schulz, and M. Karas, *Internal energy build-up in matrix-assisted laser desorption/ionization.* Journal of Mass Spectrometry, 2004. **39**(6): p. 579-593.
49. Chang, W.C., et al., *Matrix-assisted laser desorption/ionization (MALDI) mechanism revisited.* Analytica Chimica Acta, 2007. **582**(1): p. 1-9.
50. Vestal, M.L., *Modern MALDI time-of-flight mass spectrometry.* Journal of Mass Spectrometry, 2009. **44**(3): p. 303-317.
51. Chen, X., J. Carroll, and R. Beavis, *Near-ultraviolet-induced matrix-assisted laser desorption/ionization as a function of wavelength.* Journal of The American Society for Mass Spectrometry, 1998. **9**(9): p. 885-891.
52. Kaufmann, R., *Matrix-assisted laser desorption ionization (MALDI) mass spectrometry: a novel analytical tool in molecular biology and biotechnology.* Journal of Biotechnology, 1995. **41**(2-3): p. 155-175.
53. Vestal, M. and P. Juhasz, *Resolution and mass accuracy in matrix-assisted laser desorption ionization-time-of-flight.* Journal of The American Society for Mass Spectrometry, 1998. **9**(9): p. 892-911.
54. Amster, I.J., *Fourier Transform Mass Spectrometry.* Journal of Mass Spectrometry, 1996. **31**(12): p. 1325-1337.
55. Marshall, A.G., C.L. Hendrickson, and G.S. Jackson, *Fourier transform ion cyclotron resonance mass spectrometry: A primer.* Mass Spectrometry Reviews, 1998. **17**(1): p. 1-35.
56. Hu, Q., et al., *The Orbitrap: a new mass spectrometer.* Journal of Mass Spectrometry, 2005. **40**(4): p. 430-443.
57. Makarov, A. and M. Scigelova, *Coupling liquid chromatography to Orbitrap mass spectrometry.* Journal of Chromatography A, 2010. **1217**(25): p. 3938-3945.
58. Perry, R.H., R.G. Cooks, and R.J. Noll, *Orbitrap mass spectrometry: Instrumentation, ion motion and applications.* Mass Spectrometry Reviews, 2008. **27**(6): p. 661-699.
59. March, R.E., *An Introduction to Quadrupole Ion Trap Mass Spectrometry.* Journal of Mass Spectrometry, 1997. **32**(4): p. 351-369.
60. Hopfgartner, G., et al., *Triple quadrupole linear ion trap mass spectrometer for the analysis of small molecules and macromolecules.* Journal of Mass Spectrometry, 2004. **39**(8): p. 845-855.
61. Douglas, D.J., *Linear quadrupoles in mass spectrometry.* Mass Spectrometry Reviews, 2009. **28**(6): p. 937-960.
62. Stephens, W., *Proceedings of the American Physical Society.* Physical Review, 1946. **69**(11-12): p. 674.
63. Dunphy, J.C. and K.L. Busch, *Imaging analysis and selected sequence monitoring of small peptides using planar chromatography/secondary ion mass spectrometry.* Biomed Environ Mass Spectrom, 1988. **17**(5): p. 405-10.
64. Karas, M.c., *Time-of-Flight Mass Spectrometer with Improved Resolution, W. C. Wiley and I. H. McLaren, (Reprinted from Rev. Sci. Instrum., 26, 1150 (1955)).* Journal of Mass Spectrometry, 1997. **32**(1): p. 1-11.

65. Chernushevich, I.V., A.V. Loboda, and B.A. Thomson, *An introduction to quadrupole–time-of-flight mass spectrometry.* Journal of Mass Spectrometry, 2001. **36**(8): p. 849-865.
66. Abian, J., *The coupling of gas and liquid chromatography with mass spectrometry †.* Journal of Mass Spectrometry, 1999. **34**(3): p. 157-168.
67. Cubbon, S., et al., *Metabolomic applications of HILIC–LC–MS.* Mass Spectrometry Reviews, 2009. **29**(5): p. 671-684.
68. Karu, K., et al., *Liquid chromatography-mass spectrometry utilizing multi-stage fragmentation for the identification of oxysterols.* Journal of Lipid Research, 2007. **48**(4): p. 976-987.
69. Shukla, A.K. and J.H. Futrell, *Tandem mass spectrometry: dissociation of ions by collisional activation.* Journal of Mass Spectrometry, 2000. **35**(9): p. 1069-1090.
70. Johnson, R.S., et al., *Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine.* Analytical Chemistry, 1987. **59**(21): p. 2621-2625.
71. Herron, W.J., D.E. Goeringer, and S.A. McLuckey, *Gas-Phase Electron Transfer Reactions from Multiply-Charged Anions to Rare Gas Cations.* J. Am. Chem. Soc., 1995. **117**(46): p. 11555-11562.
72. Lubec, G., K. Krapfenbauer, and M. Fountoulakis, *Proteomics in brain research: potentials and limitations.* Progress in Neurobiology, 2003. **69**(3): p. 193-211.
73. Krapfenbauer, K., M. Fountoulakis, and G. Lubec, *A rat brain protein expression map including cytosolic and enriched mitochondrial and microsomal fractions.* Electrophoresis, 2003. **24**(11): p. 1847-1870.
74. Boguski, M.S. and M.W. McIntosh, *Biomedical informatics for proteomics.* Nature, 2003. **422**(6928): p. 233-237.
75. Yates III, J.R., et al., *Proteomics of organelles and large cellular structures.* Nat Rev Mol Cell Biol, 2005. **6**(9): p. 702-714.
76. Gustafsson, J.O.R., et al., *MALDI Imaging Mass Spectrometry (MALDI-IMS)-Application of Spatial Proteomics for Ovarian Cancer Classification and Diagnosis.* International Journal of Molecular Sciences, 2011. **12**(1): p. 773-794.
77. Bowers, J.J., et al., *Protein identification via ion-trap collision-induced dissociation and examination of low-mass product ions.* Journal of Mass Spectrometry, 2008. **43**(1): p. 23-34.
78. Surinova, S., et al., *On the Development of Plasma Protein Biomarkers.* Journal of Proteome Research, 2011. **10**(1): p. 5-16.
79. Katagiri, T., et al., *Proteomic analysis of proteins expressing in regions of rat brain by a combination of SDS-PAGE with nano-liquid chromatography-quadrupole-time of flight tandem mass spectrometry.* Proteome Science, 2010. **8**: p. 12.
80. Dumont, D., et al., *Gel-free analysis of the human brain proteome: Application of liquid chromatography and mass spectrometry on biopsy and autopsy samples.* Proteomics, 2006. **6**(18): p. 4967-4977.
81. De Iuliis, A., et al., *A proteomic approach in the study of an animal model of Parkinson's disease.* Clinica Chimica Acta, 2005. **357**(2): p. 202-209.
82. Debora, D., et al., *Characterization of mature rat oligodendrocytes: a proteomic approach.* Journal of Neurochemistry, 2007. **102**(2): p. 562-576.
83. Fountoulakis, M., et al., *Rat brain proteins: Two-dimensional protein database and variations in the expression level.* Electrophoresis, 1999. **20**(18): p. 3572-3579.
84. He, S., et al., *Proteomic analysis and comparison of the biopsy and autopsy specimen of human brain temporal lobe.* Proteomics, 2006. **6**(18): p. 4987-4996.
85. Shapiro, A.L., E. Viñuela, and J. V. Maizel, *Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels.* Biochemical and Biophysical Research Communications, 1967. **28**(5): p. 815-820.

86. Weber, K. and M. Osborn, *The Reliability of Molecular Weight Determinations by Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis.* Journal of Biological Chemistry, 1969. **244**(16): p. 4406-4412.
87. Rabilloud, T., et al., *Two-dimensional gel electrophoresis in proteomics: Past, present and future.* Journal of Proteomics, 2010. **73**(11): p. 2064-2077.
88. Fountoulakis, M. and J.-F. Juranville, *Enrichment of low-abundance brain proteins by preparative electrophoresis.* Analytical Biochemistry, 2003. **313**(2): p. 267-282.
89. Boyd-Kimball, D., et al., *Proteomic identification of proteins specifically oxidized by intracerebral injection of amyloid [beta]-peptide (1-42) into rat brain: Implications for Alzheimer's disease.* Neuroscience, 2005. **132**(2): p. 313-324.
90. Carrette, O., et al., *Age-related proteome analysis of the mouse brain: a 2-DE study.* PROTEOMICS, 2006. **6**(18): p. 4940-4949.
91. Föcking, M., et al., *2-D DIGE as a quantitative tool for investigating the HUPO Brain Proteome Project mouse series.* Proteomics, 2006. **6**(18): p. 4914-4931.
92. Fountoulakis, M., et al., *Differences in protein level between neonatal and adult brain.* Electrophoresis, 2000. **21**(3): p. 673-678.
93. Griffin, T.J. and R. Aebersold, *Advances in Proteome Analysis by Mass Spectrometry.* Journal of Biological Chemistry, 2001. **276**(49): p. 45497-45500.
94. Seefeldt, I., et al., *Evaluation of 2-DE protein patterns from pre- and postnatal stages of the mouse brain.* Proteomics, 2006. **6**(18): p. 4932-4939.
95. Delahunty, C.M. and J.R. Yates, *MudPIT: multidimensional protein identification technology.* Biotechniques, 2007. **43**(5): p. 563-.
96. Adachi, J., et al., *The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins.* Genome Biology, 2006. **7**(9): p. R80.
97. Granvogl, B., M. Ploscher, and L.A. Eichacker, *Sample preparation by in-gel digestion for mass spectrometry-based proteomics.* Analytical and Bioanalytical Chemistry, 2007. **389**: p. 991-1002.
98. Olsen, J.V., S.-E. Ong, and M. Mann, *Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues.* Molecular & Cellular Proteomics, 2004. **3**(6): p. 608-614.
99. Elliott, M.H., et al., *Current trends in quantitative proteomics.* Journal of Mass Spectrometry, 2009. **44**(12): p. 1637-1660.
100. Marouga, R., S. David, and E. Hawkins, *The development of the DIGE system: 2D fluorescence difference gel analysis technology.* Analytical and Bioanalytical Chemistry, 2005. **382**(3): p. 669-678.
101. Imai, K., A. Koshiyama, and K. Nakata, *Towards clinical proteomics analysis.* Biomedical Chromatography, 2010. **25**(1-2): p. 59-64.
102. Stewart, II, T. Thomson, and D. Figeys, *O-18 Labeling: a tool for proteomics.* Rapid Communications in Mass Spectrometry, 2001. **15**(24): p. 2456-2465.
103. Gerber, S.A., et al., *Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(12): p. 6940-6945.
104. Seibert, C., et al., *Multiple-Approaches to the Identification and Quantification of Cytochromes P450 in Human Liver Tissue by Mass Spectrometry.* Journal of Proteome Research, 2009. **8**(4): p. 1672-1681.
105. Zinzalla, G. and D.E. Thurston, *Targeting protein-protein interactions for therapeutic intervention: a challenge for the future.* Future Medicinal Chemistry, 2009. **1**(1): p. 65-93.
106. Dowsey, A.W., et al., *Image analysis tools and emerging algorithms for expression proteomics.* Proteomics, 2010. **10**(23): p. 4226-4257.
107. Jonker, N., et al., *Recent developments in protein–ligand affinity mass spectrometry.* Analytical and Bioanalytical Chemistry, 2011. **399**(8): p. 2669-2681.

108. Lee, D.Y., B.P. Bowen, and T.R. Northen, *Mass spectrometry-based metabolomics, analysis of metabolite-protein interactions, and imaging.* Biotechniques, 2010. **49**(2): p. 557-565.
109. Alvarez-Manilla, G., et al., *Tools for Glycoproteomic Analysis: Size Exclusion Chromatography Facilitates Identification of Tryptic Glycopeptides with N-linked Glycosylation Sites.* Journal of Proteome Research, 2006. **5**(3): p. 701-708.
110. Wang, Y., et al., *Membrane Glycoproteins Associated with Breast Tumor Cell Progression Identified by a Lectin Affinity Approach.* Journal of Proteome Research, 2008. **7**(10): p. 4313-4325.
111. Mikhailov, V. and H. Cooper, *Activated ion electron capture dissociation (AI ECD) of proteins: Synchronization of infrared and electron irradiation with ion magnetron motion.* Journal of The American Society for Mass Spectrometry, 2009. **20**(5): p. 763-771.
112. Mikhailov, V.A., J. Iniesta, and H.J. Cooper, *Top-Down Mass Analysis of Protein Tyrosine Nitration: Comparison of Electron Capture Dissociation with "Slow-Heating" Tandem Mass Spectrometry Methods.* Analytical Chemistry, 2010. **82**(17): p. 7283-7292.
113. Stokes, A.A., et al., *Top-down protein sequencing by CID and ECD using desorption electrospray ionisation (DESI) and high-field FTICR mass spectrometry.* International Journal of Mass Spectrometry, 2010. **289**(1): p. 54-57.
114. Tsybin, Y.O., et al., *Peptide and protein characterization by high-rate electron capture dissociation Fourier transform ion cyclotron resonance mass spectrometry.* Journal of Mass Spectrometry, 2004. **39**(7): p. 719-729.
115. Runz, H., et al., *Inhibition of Intracellular Cholesterol Transport Alters Presenilin Localization and Amyloid Precursor Protein Processing in Neuronal Cells.* J. Neurosci., 2002. **22**(5): p. 1679-1689.
116. Wiesner, J., T. Premsler, and A. Sickmann, *Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications.* Proteomics, 2008. **8**(21): p. 4466-4483.
117. Grønborg, M., et al., *A Mass Spectrometry-based Proteomic Approach for Identification of Serine/Threonine-phosphorylated Proteins by Enrichment with Phospho-specific Antibodies.* Molecular & Cellular Proteomics, 2002. **1**(7): p. 517-527.
118. Wang, J., et al., *Phosphopeptide detection using automated online IMAC-capillary LC-ESI-MS/MS.* PROTEOMICS, 2006. **6**(2): p. 404-411.
119. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-67.
120. Craig, R. and R.C. Beavis, *TANDEM: matching proteins with tandem mass spectra.* Bioinformatics, 2004. **20**(9): p. 1466-1467.
121. *ProteinLynx Global Server Version 2.2.5 User's Guide*. Revision A ed, ed. W. Corporation. 2006, Milford MA: Waters Corporation. 432.
122. Chamrad, D.C., et al., *Gaining knowledge from previously unexplained spectra-application of the PTM-Explorer software to detect PTM in HUPO BPP MS/MS data.* PROTEOMICS, 2006. **6**(18): p. 5048-5058.
123. Griffiths, W.J., ed. *Metabolomics, Metabonomics and Metabolite Profiling*. RSC Biomolecular Sciences, ed. P.S.N. (Chairman). 2008, Royal Society of Chemistry: London. 323.
124. Holmes, E., I.D. Wilson, and J.K. Nicholson, *Metabolic Phenotyping in Health and Disease.* Cell, 2008. **134**(5): p. 714-717.
125. Atherton, H.J., et al., *Metabolomics of the interaction between PPAR-[alpha] and age in the PPAR-[alpha]-null mouse.* Mol Syst Biol, 2009. **5**.
126. Liu, S., J. Sjövall, and W.J. Griffiths, *Neurosteroids in Rat Brain: Extraction, Isolation, and Analysis by Nanoscale Liquid Chromatography-Electrospray Mass Spectrometry.* Analytical Chemistry, 2003. **75**(21): p. 5835-5846.

127. Lakshmanan, V., K.Y. Rhee, and J.P. Daily, *Metabolomics and malaria biology.* Molecular and Biochemical Parasitology, 2011. **175**(2): p. 104-111.
128. Tiller, P.R., et al., *High-throughput, accurate mass liquid chromatography/tandem mass spectrometry on a quadrupole time-of-flight system as a 'first-line' approach for metabolite identification studies.* Rapid Communications in Mass Spectrometry, 2008. **22**(7): p. 1053-1061.
129. Babiker, A. and U. Diczfalusy, *Transport of side-chain oxidized oxysterols in the human circulation.* Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism, 1998. **1392**(2-3): p. 333-339.
130. Iuliano, L., et al., *Measurement of oxysterols and [alpha]-tocopherol in plasma and tissue samples as indices of oxidant stress status.* Analytical Biochemistry, 2003. **312**(2): p. 217-223.
131. Sjövall, J., *Fifty years with bile acids and steroids in health and disease.* Lipids, 2004. **39**(8): p. 703-722.
132. Wikoff, W.R., et al., *Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites.* Proceedings of the National Academy of Sciences, 2009. **106**(10): p. 3698-3703.
133. Lövgren-Sandblom, A., et al., *Novel LC-MS/MS method for assay of 7[alpha]-hydroxy-4-cholesten-3-one in human plasma: Evidence for a significant extrahepatic metabolism.* Journal of Chromatography B, 2007. **856**(1-2): p. 15-19.
134. Calvano, C.D., C.G. Zambonin, and O.N. Jensen, *Assessment of lectin and HILIC based enrichment protocols for characterization of serum glycoproteins by mass spectrometry.* Journal of Proteomics, 2008. **71**(3): p. 304-317.
135. Eneroth, P., K. Hellström, and R. Ryhage, *Identification and quantification of neutral fecal steroids by gas-liquid chromatography and mass spectrometry: studies of human excretion during two dietary regimens.* Journal of Lipid Research, 1964. **5**(2): p. 245-262.
136. Hooper, N.K. and J.H. Law, *Mass spectrometry of derivatives of cyclopropene fatty acids.* Journal of Lipid Research, 1968. **9**(2): p. 270-275.
137. Pauling, L., et al., *Quantitative Analysis of Urine Vapor and Breath by Gas-Liquid Partition Chromatography.* Proceedings of the National Academy of Sciences, 1971. **68**(10): p. 2374-2376.
138. Griffiths, W.J., et al., *Derivatisation for the characterisation of neutral oxosteroids by electrospray and matrix-assisted laser desorption/ionisation tandem mass spectrometry: the Girard P derivative.* Rapid Communications in Mass Spectrometry, 2003. **17**(9): p. 924-935.
139. Dzeletovic, S., et al., *Determination of Cholesterol Oxidation Products in Human Plasma by Isotope Dilution-Mass Spectrometry.* Analytical Biochemistry, 1995. **225**(1): p. 73-80.
140. Gowda, G.A.N., et al., *Quantitative Analysis of Blood Plasma Metabolites Using Isotope Enhanced NMR Methods.* Analytical Chemistry, 2010. **82**(21): p. 8983-8990.
141. Lütjohann, D., et al., *Plasma 24S-hydroxycholesterol (cerebrosterol) is increased in Alzheimer and vascular demented patients.* Journal of Lipid Research, 2000. **41**(2): p. 195-198.
142. Leoni, V., et al., *Changes in human plasma levels of the brain specific oxysterol 24S-hydroxycholesterol during progression of multiple sclerosis.* Neuroscience Letters, 2002. **331**(3): p. 163-166.
143. Burkard, I., K.M. Rentsch, and A. von Eckardstein, *Determination of 24S- and 27-hydroxycholesterol in plasma by high-performance liquid chromatography-mass spectrometry.* Journal of Lipid Research, 2004. **45**(4): p. 776-781.
144. Lütjohann, D., *Cholesterol metabolism in the brain: importance of 24S-hydroxylation.* Acta Neurologica Scandinavica, 2006. **114**: p. 33-42.
145. Wang, Y., et al., *Matrix-Assisted Laser Desorption/Ionization High-Energy Collision-Induced Dissociation of Steroids: Analysis of Oxysterols in Rat Brain.* Analytical Chemistry, 2005. **78**(1): p. 164-173.

146. Wang, Y. and W.J. Griffiths, *Capillary liquid chromatography combined with tandem mass spectrometry for the study of neurosteroids and oxysterols in brain.* Neurochemistry International, 2008. **52**(4-5): p. 506-521.
147. Karu, K., *The analysis of oxysterols by capillary liquid chromatography tandem mass spectrometry*, in *Pharmaceutical and Biological Chemistry.* 2009, University of London, School of Pharmacy: London. p. 321.
148. Hamacher, M., et al., *HUPO Brain Proteome Project: Summary of the pilot phase and introduction of a comprehensive data reprocessing strategy.* Proteomics, 2006. **6**(18): p. 4890-4898.
149. Stühler, K., et al., *Pilot study of the Human Proteome Organisation Brain Proteome Project: Applying different 2-DE techniques to monitor proteomic changes during murine brain development.* Proteomics, 2006. **6**(18): p. 4899-4913.
150. Stephan, C., et al., *HUPO Brain Proteome Project Pilot Studies: Bioinformatics at Work.* PROTEOMICS, 2005. **5**(11): p. 2716-2717.
151. Reidegeld, K.A., et al., *The power of cooperative investigation: Summary and comparison of the HUPO Brain Proteome Project pilot study results.* Proteomics, 2006. **6**(18): p. 4997-5014.
152. Stephan, C., et al., *Automated reprocessing pipeline for searching heterogeneous mass spectrometric data of the HUPO Brain Proteome Project pilot phase.* PROTEOMICS, 2006. **6**(18): p. 5015-5029.
153. Mueller, M., et al., *Functional annotation of proteins identified in human brain during the HUPO Brain Proteome Project pilot study.* Proteomics, 2006. **6**(18): p. 5059-5075.
154. Martens, L., et al., *A comparison of the HUPO Brain Proteome Project pilot with other proteomics studies.* PROTEOMICS, 2006. **6**(18): p. 5076-5086.
155. Dowsey, A.W., et al., *Examination of 2-DE in the Human Proteome Organisation Brain Proteome Project pilot studies with the new RAIN gel matching technique.* Proteomics, 2006. **6**(18): p. 5030-5047.
156. Park, Y.M., et al., *Profiling human brain proteome by multi-dimensional separations coupled with MS.* Proteomics, 2006. **6**(18): p. 4978-4986.
157. Fountoulakis, M., et al., *Two-dimensional database of mouse liver proteins: Changes in hepatic protein levels following treatment with acetaminophen or its nontoxic regioisomer 3-acetamidophenol.* Electrophoresis, 2000. **21**(11): p. 2148-2161.
158. Witt, A.E., et al., *Functional Proteomics Approach to Investigate the Biological Activities of cDNAs Implicated in Breast Cancer.* Journal of Proteome Research, 2006. **5**(3): p. 599-610.
159. Kim, W.S., et al., *Role of ABCG1 and ABCA1 in Regulation of Neuronal Cholesterol Efflux to Apolipoprotein E Discs and Suppression of Amyloid-β Peptide Generation.* Journal of Biological Chemistry, 2007. **282**(5): p. 2851-2861.
160. Waelsch, H., W.M. Sperry, and V.A. Stoyanoff, *A study on the synthesis and deposition of lipids in brain and other tissues with deuterium as an indicator.* Journal of Biological Chemistry, 1940. **135**(1): p. 291-296.
161. Björkhem, I. and S. Meaney, *Brain Cholesterol: Long Secret Life Behind a Barrier.* Arterioscler Thromb Vasc Biol, 2004. **24**(5): p. 806-815.
162. Meaney, S., et al., *Evidence that the major oxysterols in human circulation originate from distinct pools of cholesterol: a stable isotope study.* J. Lipid Res., 2001. **42**(1): p. 70-78.
163. Seth, G., R.S. McIvor, and W.-S. Hu, *17[beta]-Hydroxysteroid dehydrogenase type 7 (Hsd17b7) reverts cholesterol auxotrophy in NS0 cells.* Journal of Biotechnology, 2006. **121**(2): p. 241-252.
164. Voet, D., V.J. G., and C.W. Pratt, *Chapter 9 Lipids*, in *Fundamentals of Biochemistry.* 1999, John Wiley & Sons Inc.: New York, NY. p. 219-231.

165. Dietschy, J.M. and S.D. Turley, *Thematic review series: Brain Lipids. Cholesterol metabolism in the central nervous system during early development and in the mature animal.* J. Lipid Res., 2004. **45**(8): p. 1375-1397.
166. Raffai, R.L. and K.H. Weisgraber, *Cholesterol: from heart attacks to Alzheimer's disease.* J. Lipid Res., 2003. **44**(8): p. 1423-1430.
167. Hedlund, E., J.A. Gustafsson, and M. Warner, *Cytochrome P450 in the brain; a review.* Curr Drug Metab, 2001. **2**(3): p. 245-63.
168. Dumont, D., et al., *Characterization of mature rat oligodendrocytes: a proteomic approach.* J Neurochem, 2007. **102**(2): p. 562-76.
169. Bhagwat, S.V., M.R. Boyd, and V. Ravindranath, *Rat brain cytochrome P450. Reassessment of monooxygenase activities and cytochrome P450 levels.* Drug Metab Dispos, 1995. **23**(6): p. 651-4.
170. de Gruijl, F. and J. Leun, *Environment and health: 3. Ozone depletion and ultraviolet radiation.* CMAJ, 2000. **163**(7): p. 851-855.
171. Gallagher, R.P. and T.K. Lee, *Adverse effects of ultraviolet radiation: A brief review.* Progress in Biophysics and Molecular Biology, 2006. **92**(1): p. 119-131.
172. Norval, M., et al., *The effects on human health from stratospheric ozone depletion and its interactions with climate change.* Photochemical & Photobiological Sciences, 2007. **6**(3): p. 232-251.
173. Shick, J.M. and W.C. Dunlap, *Mycosporine-like amino acids and related gadusols: Biosynthesis, Accumulation, and UV-Protective Functions in Aquatic Organisms.* Annual Review of Physiology, 2002. **64**(1): p. 223-262.
174. Shick, J.M., *Ultraviolet Stress*, in *Encyclopedia of Tidepools and Rocky Shores*, M.W. Denny and S.D. Gaines, Editors. 2007, University of California Press: Los Angeles and Berkeley. p. 604-609.
175. Booth, C.R. and J.H. Morrow, *The Penetration of UV into Natural Waters.* Photochemistry and Photobiology, 1997. **65**(2): p. 254-257.
176. Barrios-Llerena, M.E., et al., *Shotgun proteomics of cyanobacteria-applications of experimental and data-mining techniques.* Briefings in Functional Genomics & Proteomics, 2006. **5**(2): p. 121-132.
177. Favrebonvin, J., N. Arpin, and C. Brevard, *STRUCTURE OF MYCOSPORINE (P-310).* Canadian Journal of Chemistry-Revue Canadienne De Chimie, 1976. **54**(7): p. 1105-1113.
178. Portwich, A. and F. Garcia-Pichel, *Biosynthetic pathway of mycosporines (mycosporine-like amino acids) in the cyanobacterium Chlorogloeopsis sp strain PCC 6912.* Phycologia, 2003. **42**(4): p. 384-392.
179. Starcevic, A., et al., *Gene Expression in the Scleractinian Acropora microphthalma Exposed to High Solar Irradiance Reveals Elements of Photoprotection and Coral Bleaching.* Plos One, 2008. **5**(11): p. 10.
180. Balskus, E.P. and C.T. Walsh, *The Genetic and Molecular Basis for Sunscreen Biosynthesis in Cyanobacteria.* Science, 2010. **329**(5999): p. 1653-1656.
181. Schmidt, E.W., *An Enzymatic Route to Sunscreens.* Chembiochem. **12**(3): p. 363-365.
182. Happe, T., K. Schutz, and H. Bohme, *Transcriptional and Mutational Analysis of the Uptake Hydrogenase of the Filamentous Cyanobacterium Anabaena variabilis ATCC 29413.* J. Bacteriol., 2000. **182**(6): p. 1624-1631.
183. Weyman, P.D., B. Pratte, and T. Thiel, *Transcription of hupSL in Anabaena variabilis ATCC 29413 Is Regulated by NtcA and Not by Hydrogen.* Appl. Environ. Microbiol., 2008. **74**(7): p. 2103-2110.
184. Singh, S., et al., *Role of various growth media on shinorine (mycosporine-like amino acid) concentration and photosynthetic yield in Anabaena variabilis PCC 7937.* World Journal of Microbiology and Biotechnology, 2008. **24**(12): p. 3111-3115.
185. Bandaranayake, W.M., *Mycosporines: are they nature's sunscreens?* Natural Product Reports, 1998. **15**(2): p. 159-172.

186. Shick, J.M., et al., *Inhibition of the shikimate pathway blocks UVB radiation-stimulated accumulation of mycosporine-like amino acids (MAAs) in the coral stylophora pistillata.* Photochemistry and Photobiology, 1999. **69**: p. 93S-93S.
187. Singh, S.P., et al., *Effects of Abiotic Stressors on Synthesis of the Mycosporine-like Amino Acid Shinorine in the Cyanobacterium Anabaena variabilis PCC 7937.* Photochemistry and Photobiology, 2008. **84**(6): p. 1500-1505.
188. Keller, A., et al., *Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search.* Analytical Chemistry, 2002. **74**(20): p. 5383-5392.
189. Nesvizhskii, A.I., et al., *A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry.* Analytical Chemistry, 2003. **75**(17): p. 4646-4658.
190. Cardozo, K.H.M., et al., *Fragmentation of mycosporine-like amino acids by hydrogen/deuterium exchange and electrospray ionisation tandem mass spectrometry.* Rapid Communications in Mass Spectrometry, 2006. **20**(2): p. 253-258.
191. Whitehead, K. and J.I. Hedges, *Electrospray ionization tandem mass spectrometric and electron impact mass spectrometric characterization of mycosporine-like amino acids.* Rapid Communications in Mass Spectrometry, 2003. **17**(18): p. 2133-2138.
192. Ehling-Schulz, M., et al., *The UV-B stimulon of the terrestrial cyanobacterium Nostoc commune comprises early shock proteins and late acclimation proteins.* Molecular Microbiology, 2002. **46**(3): p. 827-843.
193. Shick, J.M., et al., *Ultraviolet-B radiation stimulates shikimate pathway-dependent accumulation of mycosporine-like amino acids in the coral Stylophora pistillata despite decreases in its population of symbiotic dinoflagellates.* Limnology and Oceanography, 1999. **44**(7): p. 1667-1682.
194. Dunlap, W., et al., *A Microtiter Plate Assay for Screening Antioxidant Activity in Extracts of Marine Organisms.* Marine Biotechnology, 2003. **5**(3): p. 294-301.
195. Oyamada, C., et al., *Mycosporine-Like Amino Acids Extracted from Scallop (Patinopecten yessoensis) Ovaries: UV Protection and Growth Stimulation Activities on Human Cells.* Marine Biotechnology, 2008. **10**(2): p. 141-150.
196. Schust, J., et al., *Stattic: A small-molecule inhibitor of STAT3 activation and dimerization.* Chemistry & Biology, 2006. **13**(11): p. 1235-1242.
197. Fletcher, S., J. Turkson, and P.T. Gunning, *Molecular Approaches towards the Inhibition of the Signal Transducer and Activator of Transcription 3 (Stat3) Protein.* ChemMedChem, 2008. **3**(8): p. 1159-1168.
198. Siddiquee, K., et al., *Selective chemical probe inhibitor of Stat3, identified through structure-based virtual screening, induces antitumor activity.* 2007. **104**(18): p. 7391-7396.
199. Bromberg, J. and J.E. Darnell, *The role of STATs in transcriptional control and their impact on cellular function.* Oncogene, 2000. **19**(21): p. 2468-2473.
200. Yu, H. and R. Jove, *The STATs of cancer-new molecular targets come of age.* Nat Rev Cancer, 2004. **4**(2): p. 97-105.
201. Levy, D.E. and C.-k. Lee, *What does Stat3 do?* The Journal of Clinical Investigation, 2002. **109**(9): p. 1143-1148.
202. Becker, S., B. Groner, and C.W. Muller, *Three-dimensional structure of the Stat3 beta homodimer bound to DNA.* Nature, 1998. **394**(6689): p. 145-151.
203. Darnell, J.E., *STATs and gene regulation.* Science, 1997. **277**(5332): p. 1630-1635.
204. Buettner, R., L.B. Mora, and R. Jove, *Activated STAT signaling in human tumors provides novel molecular targets for therapeutic intervention.* Clinical Cancer Research, 2002. **8**(4): p. 945-954.
205. Bowman, T., et al., *STATs in oncogenesis.* Oncogene, 2000. **19**(21): p. 2474-2488.
206. Darnell, J.E., *Validating Stat3 in cancer therapy.* Nature Medicine, 2005. **11**(6): p. 595-596.

207. Turkson, J., et al., *Novel peptidomimetic inhibitors of signal transducer and activator of transcription 3 dimerization and biological activity.* Molecular Cancer Therapeutics, 2004. **3**(3): p. 261-269.
208. Angulo, J., P.M. Enríquez-Navas, and P.M. Nieto, *Ligand–Receptor Binding Affinities from Saturation Transfer Difference (STD) NMR Spectroscopy: The Binding Isotherm of STD Initial Growth Rates.* Chemistry – A European Journal, 2010. **16**(26): p. 7803-7812.
209. Fielding, L., *NMR methods for the determination of protein-ligand dissociation constants.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2007. **51**(4): p. 219-242.
210. Fielding, L., S. Rutherford, and D. Fletcher, *Determination of protein–ligand binding affinity by NMR: observations from serum albumin model systems.* Magnetic Resonance in Chemistry, 2005. **43**(6): p. 463-470.
211. Wang, Y.-S., D. Liu, and D.F. Wyss, *Competition STD NMR for the detection of high-affinity ligands and NMR-based screening.* Magnetic Resonance in Chemistry, 2004. **42**(6): p. 485-489.
212. Xia, Y., et al., *Clean STD-NMR spectrum for improved detection of ligand-protein interactions at low concentration of protein.* Magnetic Resonance in Chemistry, 2010. **48**(12): p. 918-924.
213. Mayer, M. and B. Meyer, *Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy.* Angewandte Chemie International Edition, 1999. **38**(12): p. 1784-1788.

Appendix A Publications

Publications resulting from the work in this thesis

Griffiths, W.J., et al., *Discovering Oxysterols in Plasma: A Window on the Metabolome.* Journal of Proteome Research, 2008. **7**(8): p. 3602-3612.

Conference contributions resulting from the work in the thesis

Heidelberger, S., *Structural Chemistry Approaches for the Identification of Potential Interaction Sites of Ligands to STAT3.* **Oral presentation** for the School of Pharmacy Research Day 2010.

Heidelberger, S., Karu, K., Wang, Y., Griffiths, W. *Proteomic Strategies for the Analysis of Proteins Involved in Steroid Synthesis, Metabolism and Transport.* **Poster submission** for the School of Pharmacy Research Day, 2006, London, UK.

Heidelberger, S., Karu, K., Wang, Y., Griffiths, W. *Proteomic Strategies for the Analysis of Proteins Involved in Steroid Synthesis, Metabolism and Transport.* **Abstract submission** for the International Mass Spectrometry Conference, 2006, Prague.

Other Publications

Mkrtchyan, H., et al., *Purification, characterisation and identification of acidocin LCHV, an antimicrobial peptide produced by Lactobacillus acidophilus n.v. Er 317/402 strain Narine.* International Journal of Antimicrobial Agents, 2010. **35**(3): p. 255-260.

Antonow, D., et al., *Structure-Activity Relationships of Monomeric C2-Aryl Pyrrolo[2,1-c][1,4]benzodiazepine (PBD) Antitumor Agents.* Journal of Medicinal Chemistry, 2010. **53**(7): p. 2927-2941.

Castaldo, G., et al., *Proposed Arrangement of Proteins Forming a Bacterial Type II Polyketide Synthase.* Chemistry & Biology, 2008. **15**(11): p. 1156-1165.

Balan, S., et al., *Site-Specific PEGylation of Protein Disulfide Bonds Using a Three-Carbon Bridge.* Bioconjugate Chemistry, 2007. **18**(1): p. 61-76.

Shaunak, S., et al., *Site-specific PEGylation of native disulfide bonds in therapeutic proteins.* Nat Chem Biol, 2006. **2**(6): p. 312-313.

Khan, M.A., et al., *Analysis of derivatised steroids by matrix-assisted laser desorption/ionisation and post-source decay mass spectrometry.* Steroids, 2006. **71**(1): p. 42-53.

Appendix B: Accompanying CD

Chapter 2: Steroid data summary of areas and heights for 6 samples

Chapter 3: FINAL: Identification and sorting of proteins: Steroid proteins of interest
Master list of proteins
One hit wonders
Mitochondrial proteins
Microsomal proteins
Cytosolic proteins

Chapter 4: Comparison of X!Tandem and PLGS data completed
PLGS data from all bands
X!Tandem data from all bands
Mascot sample cyanobacteria individual files
Mascot sample report for cyanobacteria
Mascot cyanobacteria probability

Chapter 5: Data protein location PLGS
Data protein location X!Tandem
X!Tandem – PLGS Gel DMSO
X!Tandem – PLGS Gel MeOH