

Applications of Chemometric and Mass Spectrometric
Methods in the Study of Some Biological Systems

A THESIS PRESENTED FOR THE DEGREE OF DOCTOR OF
PHILOSOPHY IN
THE FACULTY OF SCIENCE
THE UNIVERSITY OF STRATHCLYDE
BY
LIANG ZHENG
(BEng., MSc.)

Department of Pharmaceutical Sciences,
Strathclyde Institute of Biomedical and Pharmaceutical Science,
University of Strathclyde,
27 Taylor Street,
Glasgow, UK, G4 0NR

December, 2009

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright ACTS as qualified by University of Strathclyde Regulation 3.49. Due acknowledgement must always be made of the use of any material contained in, or derived from this thesis.

Abstract

Mass spectrometry in combination with chemometric data processing has secured a central role in the study of biological systems to comprehensively and simultaneously determine the global metabolite levels in whole organisms and their changes as a consequence of stimuli. With unmatched mass resolution, mass accuracy, and detection sensitivity, linear ion trap - Fourier Transform Orbitrap Mass Spectrometry (LTQ-Orbitrap-MS) has the potential for high throughput metabolomic analysis. Chemometric tools like principal component analysis (PCA) and partial least squares (PLS) have provided the capability to extract the most important information from complicated data matrices. In addition the LTQ-Orbitrap is powerful tool for bioanalysis. Five different but linked studies are reported in this thesis. A method based on hydrophilic interaction chromatography (HILIC) in combination with FTMS was used to determine Fe (III) as its EDTA and EDDS complexes in the snail, *Helix aspersa*. It was found that the presence of a complexing agent such as EDTA was necessary to promote uptake of toxic levels of Fe (III) into the snail tissues. A study was carried out on aqueous extracts of different teas in order to develop chromatographic alignment software. Correlation optimised warping (COW) was applied to overcome LC peak shifting from run to run and the better alignment achieved allowed more effective separation of the different teas using multivariate analysis such as PCA, support vector machines (SVM) and random forest (RF). All of the methods were able to discriminate between the different tea samples.

A metabolomics study was carried out on the effects of hypoxia in cells. Zic-Hilic and Cogent-Hydride columns were used in combination with Sieve software. The effect of hypoxia on A378 and HCT cells was studied, hypoxia

induced changes in metabolism were quite different between the two cell lines studied. Another metabolomic study was carried out to compare the metabolism of the protozoa *Trichomonas vaginalis* and *Trichomonas foetus*. There were found to be major differences in the global metabolite profiles between the two closely related organisms. It was observed that biosynthetic pathways such as those for methionine, arginine and sphingolipids were different.

Finally, a reversed normal phase mode LC-MS method based on silica gel was developed and was able to profile 148 of phospholipids in the parasite *Leishmania donovani*. The method produced separations between the major lipid classes phosphatidyl choline, phosphatidyl ethanolamine, phosphatidyl serine and phosphatidyl inositol.

ABBREVIATIONS:

ACC	Amino cyclopropane carboxylate
ANP	Aqueous normal-phase
CART	Classification and regression trees
CE	Capillary electrophoresis
COW	Correlation Optimised Warping
DAGs	Diacylglycerols
EDDS	Ethylenediamine disuccinic acid
EDTA	Ethylenediaminetetraacetic acid
EI	Electron Impact Ionisation
ERM	Empirical risk minimization
FT-ICR	Fourier Transform Ion Cyclotron Resonance
GC	Gas chromatography
Hilic	Hydrophilic interaction chromatography
IPC	Inositol phosphoryl ceramide
LC	Liquid chromatography
MALDI	Matrix Assisted Laser Desorption
MGL	Methionine-gamma- lyase
MST	Mercaptopyruvate sulphur transferase
MTA	Methyl thio adenosine
NAAG	N-acetylaspartylglutamate
NIESI	Negative ion electrospray ionisation
OOB	Out of Bag
OPLS	Orthogonal/Optimised Partial Least Square
PC	Phosphatidyl Choline
PCA	Principle component analysis
PE	Phosphatidyl Ethanolamine
PIESI	Positive ion electrospray ionisation
PIPs	Phosphatidylinositol-phosphates
PLS	Partial Least Square
PS	(Chapter 3) in Picker Shell
PS	Phosphatidyl Serine
RF	Random Forest
ROS	Reactive oxygen species
RPC	Reversed-phase chromatography
SAM	S-adenosylmethionine
SE	Sieve Extractor
SIMCA	Soft independent modelling by class analogy
SRM	Structure risk minimization
SSAT	Spermidine/spermine N1-acetyltransferase
STD	Sexual transmitted disease
TOF	Time of Flight
UPLC	Ultra performance liquid chromatography
VEGF	Vaso-epidermal growth factor

Acknowledgements

I would like to express my sincere gratitude and deep indebtedness to Dr Dave Watson for offering me the opportunity to carry out research and for his enthusiastic encouragement, invaluable advice and the free academic environment throughout the course of the work. From the first day I started until the last minutes of completion, Dave has been spending a lot of efforts and time on supervising me. Also a tremendous appreciation for support from Markus Bieri of Lonza Ltd, Switzerland.

Many appreciations are due to my teachers and collaborators:

RuAngelie Edrada-Ebel for her professional mass spectrometry and chemistry knowledge; Blair Johnston for his advanced mathematical and computer knowledge; Prof. Grahams Coombs, Dr. Ruben T'Kindt, Dr. Gareth Westrop from the parasitology group; Dr. Eyal Gottlieb, Dr. Christian Frezza from Beatson Institute of Cancer Research UK.

Many thanks are due to the PhD students in 308 and masters students in the Department of Pharmaceutical Science: Lynsey, Ahmed, Alex, Abhinav, Ruwida and Vivek.

Finally, but certainly not least, I am deeply indebted to my family for their encouragement and support at all stages and difficulties during my 20 years education.

CONTENTS

ABSTRACT	iii
ABBREVIATIONS	v
ACKNOWLEDGEMENTS	vi

Chapter 1 General Introduction

1.1 Mass Spectrometry in Biology	1
1.1.1 Applications in Biology	1
1.1.2 Mass Spectrometry Based Metabolomics	1
1.2 Mass Spectrometry	1
1.2.1 Electrospray ionisation	2
1.2.2 Electron Impact Ionisation	3
1.2.3 Matrix Assisted Laser Desorption	6
1.3 Ion Separation Methods	7
1.3.1 Magnetic Sector Instruments	7
1.3.2 Quadrupole Instruments	9
1.3.3 Ion Trap Instruments	10
1.3.4 Time of Flight Instruments	12
1.3.5 Fourier Transform Ion Cyclotron Resonance	14
1.3.6 The Orbitrap Mass Spectrometer	15
1.4 Separation Methods	16
1.4.1 Reversed Phase Chromatography	16
1.4.2 Hydrophilic Interaction Chromatography	17
1.4.3 Capillary Gas Chromatography	18
1.4.4 Capillary Electrophoresis	19
1.5 Data Processing	19
1.5.1 SIMCA P and Other Chemometric Software	20
1.5.2 Sieve Software	22
1.6 Metabolomics	23
1.7 Aims and Objectives	26

Chapter 2

Development of a Bioanalytical Method for the determination of iron as its ethylenediaminetetraacetic acid (EDTA) complex and ethylene diamine disuccinic acid derivative (EDDS) complex in the snail *Helix aspersa* by hydrophilic interaction liquid chromatography coupled to Fourier transform electrospray ionisation mass spectrometry

2.1 Introduction	28
2.2 Experimental	31
2.2.1 Chemicals and Materials	31
2.2.2 Microscopy	31

2.2.3 Standards and Sample preparation	32
2.2.4 Feeding and Dissection of Snails	32
2.2.5 Caco 2 Assay System	33
2.2.6 Statistics	35
2.2.7 LC/MS Instrumentation	35
2.3 Results and Discussion	35
2.3.1 Method Validation	35
2.3.2 Studies on Caco-2 Cells	38
2.3.3 Iron Extraction Recovery from Snail Tissues	39
2.3.4 Iron and FeEDTA absorption in Snails	40
2.3.5 General Discussion	44
2.4 Determination of FeEDDS Absorption by Snails	45

Chapter 3

Some Exercises in Data Manipulation Including A Chemometric Study of Chromatograms of Aqueous Extracts of Tea Leaf, the Development of Data Extraction Software and development of a searchable metabolite data base.

3.1 Introduction	48
3.2 Application of COW Followed by Various Chemometric Methods in the Discrimination Between Different Types of Tea	48
3.2.1 Introduction	48
3.2.1.1 Alignment	50
3.2.1.2 Support Vector Machines (SVMs)	51
3.2.1.4 Random Forest (RF)	56
3.2.2 Materials and methods	58
3.2.2.1 Chemicals and Samples	58
3.2.2.2 Instrumentation	58
3.2.3 Results and discussion	59
3.2.3.1 Chromatography	59
3.2.3.2 PCA	61
3.2.3.3 SVM Classification	62
3.2.3.4 Random Forest	67
3.2.4 Conclusion	69
3.3 Development of Pickershell Software for the Extraction of Data from Xcalibur	72
3.3.1 Xcalibur Goggles	72
3.3.2 Picker Shell	72
3.4 Development of Sieve Extractor a Searchable Database for Use with the tables of Accurate Masses Generated By Sieve 1.2	74

Chapter 4

A metabonomic comparison between two cell-lines under normoxic and chronic hypoxic condition

4.1 Introduction	78
4.2 Materials and Methods	79
4.2.1 Chemicals	
4.2.2 Sample Preparation	79
4.2.3 LC-MS Methods	80
4.2.3.1 Separations Carried out on ZICHILIC	80
4.2.3.2 Separations Carried out on a Cogent Diamond Hydride Column with MS Analysis	81
4.3 Results and Discussion	82
4.3.1 Metabolomic Analysis on a ZICHILIC Column in Positive Ion Mode	82

Chapter 5

LC-MS Profiling of Phospholipids in *Leishmania donovani* extracts by Hilic mode

5.1. Introduction	108
5.2 Materials and Methods	112
5.2.1 Materials	112
5.2.2 Chromatography	112
5.2.3 Mobile Phase Composition	113
5.2.4 Mass spectrometry	113
5.2.5 Extraction of Leishmania Samples	114
5.3 Results	114
5.3.1 Characterisation of phospholipids by Infusion ESI-MS	114
5.3.2 Analysis of Standards on the ACE C-4 Column	115
5.3.3 Analysis on a Zic-Hilic column	116
5.3.4 Analysis of the Lipids on a Silica-gel Column	117
5.3.5 Analysis of <i>L.donovani</i> samples	117
5.4 Conclusions	128

Chapter 6

A Metabolomic Study of *Trichomonas vaginalis* and *Trichomonas foetus* Using High Resolution Mass Spectrometry and Hydrophilic Interaction Chromatography

6.1 Introduction	130
6.1.1 <i>T.vaginalis</i> Background	130
6.1.2. Metabolism of <i>T.vaginalis</i>	131
6.2 Materials and Methods	136
6.2.1 Chemicals and reagents	136

6.2.2 Sample Preparation	136
6.2.3 Instrumentation	138
6.2.4 Data Processing	138
6.3 Results and Discussion	138
6.3.1 Normalisation of Data	138
6.3.2 Comparison of Extraction Methods	141
6.3.3 Statistical Analysis	142
6.3.4 Metabolic Pathway Analysis	148
6.3.4.1 Methionine metabolism	148
6.3.4.2 Proline, Arginine and Ornithine Metabolism	151
6.3.4.3 Sphingolipid Metabolism	152
6.3.4.4 Glycine, Serine and Threonine Metabolism	153
6.3.4.5 Variations in nucleotide metabolism	154
6.4 Conclusion	154
Reference	159
Appendix A Metabolomics data-processing Software	171
Appendix 1 MATLAB codes for COW alignment	173
Appendix 2 MS-SQL script for Picker Shell	181
Appendix 3 Macro codes for Sieves Extractor V.10.0	195

Chapter 1 General Introduction

1.1 Mass Spectrometry in Biology

1.1.1 Applications in Biology

Mass spectrometry has wide application in the study of biological systems and has been used such as for the determination of both intact proteins and the primary sequence of proteins (proteomics), for the determination of drugs and their metabolites in biological fluids (bioanalysis) [1-5], for the targeted determination of disease biomarkers in biological fluids and for metabolomic profiling. The current thesis focuses in part on the development of a bioanalytical method using mass spectrometry and then in large part on applications of mass spectrometry in metabolomics and lipidomics with associated development of chemometric and data processing routines.

1.1.2 Mass Spectrometry Based Metabolomics

Trying to define metabolomics and differentiate it from metabonomics usually excites controversy. In this thesis metabolomics is used as a term to describe both untargeted global metabolic profiling i.e. measuring as many metabolites as possible and comparative profiling e.g. comparing two sets of conditions such as healthy and diseased. In the first instance NMR was used as a technique in mammalian metabolomics for measuring health and disease or toxicity and lack of toxicity [6-14] and the first publications on this appeared about ten years ago. Mass spectrometry in the form of gas chromatography mass spectrometry was used from about ten years ago in the field of plant metabolomics [15-18] but is currently less used in mammalian or invertebrate systems. In the past five years the use of liquid chromatography mass spectrometry has grown rapidly in mammalian and invertebrate metabolomics [19-22].

1.2 Mass Spectrometry

Mass spectrometry can be divided into two fundamental processes, ion generation and ion separation. There are three commonly used methods for generating ions and five or six methods for separating ions. Some of these methods are discussed below with particular reference to bioanalysis and metabolomic studies.

1.2.1 Electrospray ionisation

Positive ion electrospray ionisation (PIESI) is the most common ionisation technique which is used in conjunction with liquid chromatography (LC). Since PIEI works well for both low molecular weight and high molecular weight molecules, it has been applied in metabolomics and proteomics work. The technique has been in development for 20 years but the basic working principles have remained basically the same as those seen in Figure 1.1. The analytes and solvents pass through a needle following liquid chromatography. The needle has a high potential difference (with respect to a counter electrode), typically of 3.0-4.5 KV. All molecules inside the needle become charged due to the high voltage. When liquid goes and leaves the needle, it first forms a Taylor cone and slowly bursts away into a fine spray. As the solvent evaporates, the droplets shrink until they reach a limit where the surface tension cannot sustain the charge within the droplet. Thus coulombic explosion occurs and the droplet is ripped apart to form smaller droplets and ultimately gas phase ions. The analytes can be singly or multiply charged. ESI is a soft way of ionisation and very little fragmentation produced and thus it is an important technique in biological research. PIEI is the most sensitive for compounds containing an amine group and such compounds should be detectable to levels below 1 ng/ml. In addition, the technique also works well for amides and, surprisingly, also quite well for phosphates even they didn't undergo volatile. However, PIEI does not work well for polar acids such as Krebs cycle acids or for

neutral sugars, although some fatty acids can be observed in this mode. PIESI produces more efficient ionization when the content of organic solvent in the mobile phase is high since this encourages droplet evaporation and the formation of gas phase ions. In addition it is important to consider ionization suppression under ESI conditions where high levels of environmental contaminants or abundant matrix constituents can negatively affect ionisation capability, accuracy and precision. Thus it is generally best to combine mass spectrometry with a chromatographic step rather than directly infusing samples into the instrument where ion suppressive molecules such as salts are introduced at the same time as the analytes of interest. The spectra produced in PIESI mode are generally quite simple in that they are dominated by a single ion which is usually the protonated molecular ion. However, in global metabolic studies, because of the wide range of concentrations of metabolites analysed, spectra which seem at first sight simple are complicated by the formation of adduct ions with solvents and metals ions such as Na^+ and K^+ , by isotope peaks and by formation of dimers and trimers of the parent compound in the mass spectrometer. Table 1.1 shows common adduct ions which can be observed.

Negative ion electrospray ionisation (NIESI) is generally somewhat less sensitive than PIESI and consequently detects fewer compounds. Some compounds show up both in NIESI and PIESI modes. For example the amino acids can carry either a positive or a negative charge. The main groups of compounds which can be detected exclusively in negative ion mode are neutral sugars and organic acids such as the Krebs cycle acids. The spectra, as in NIESI mode, are quite simple but adducts are formed with solvents and anions such as chloride (seen in table 1.1).

1.2.2 Electron Impact Ionisation (EI)

EI is usually used in conjunction with gas chromatography mass spectrometry (GC-MS) as seen in Figure 1.2. No satisfactory interface combining LC with EI has

been developed. An advantage of EIMS is that the spectra obtained contain many fragments and such fragments can be used to produce a diagnostic fingerprint which can be matched to spectral libraries via mass/intensity correlation and it is simple to prepare a customised library. Figure 1.3 shows a comparison of succinic acid di trimethyl silyl (di TMS) derivative against the spectrum of the di TMS of a succinic acid standard.

Figure1.1 The demonstration of ESI working mechanisms

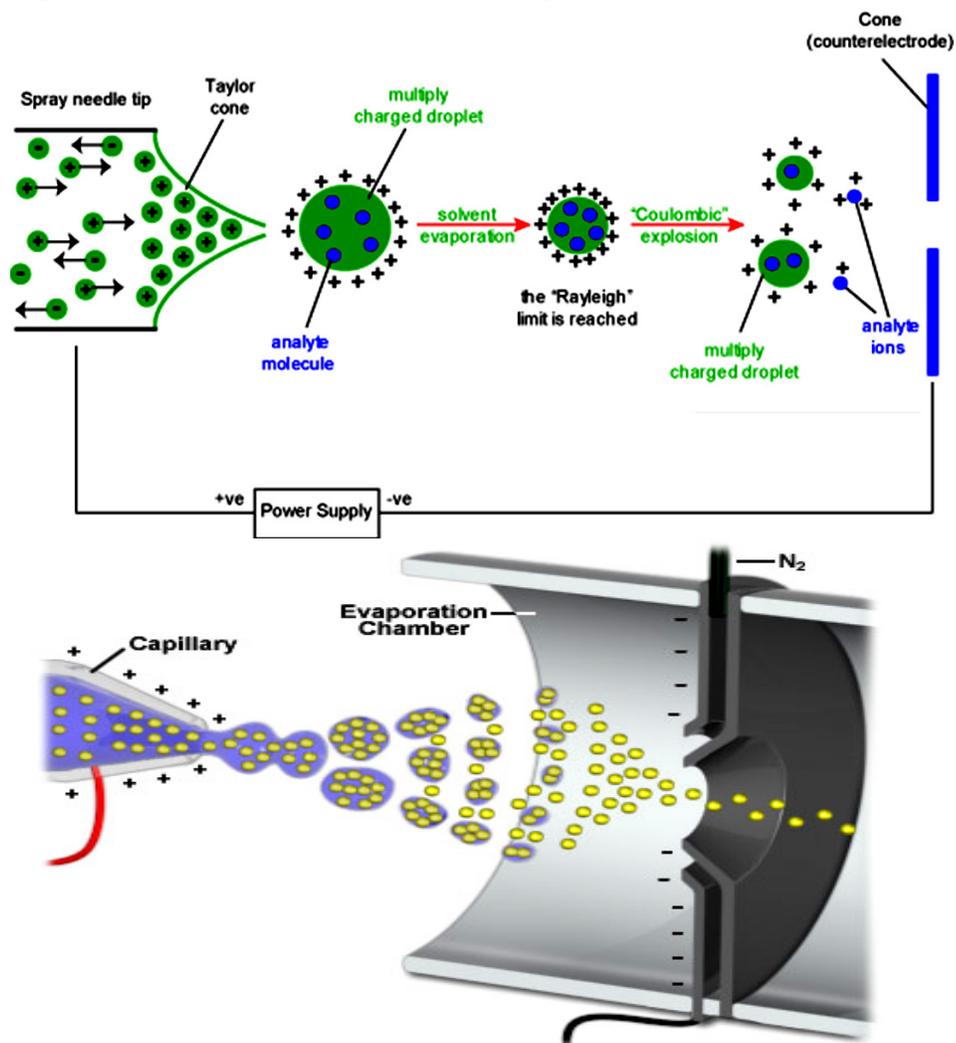
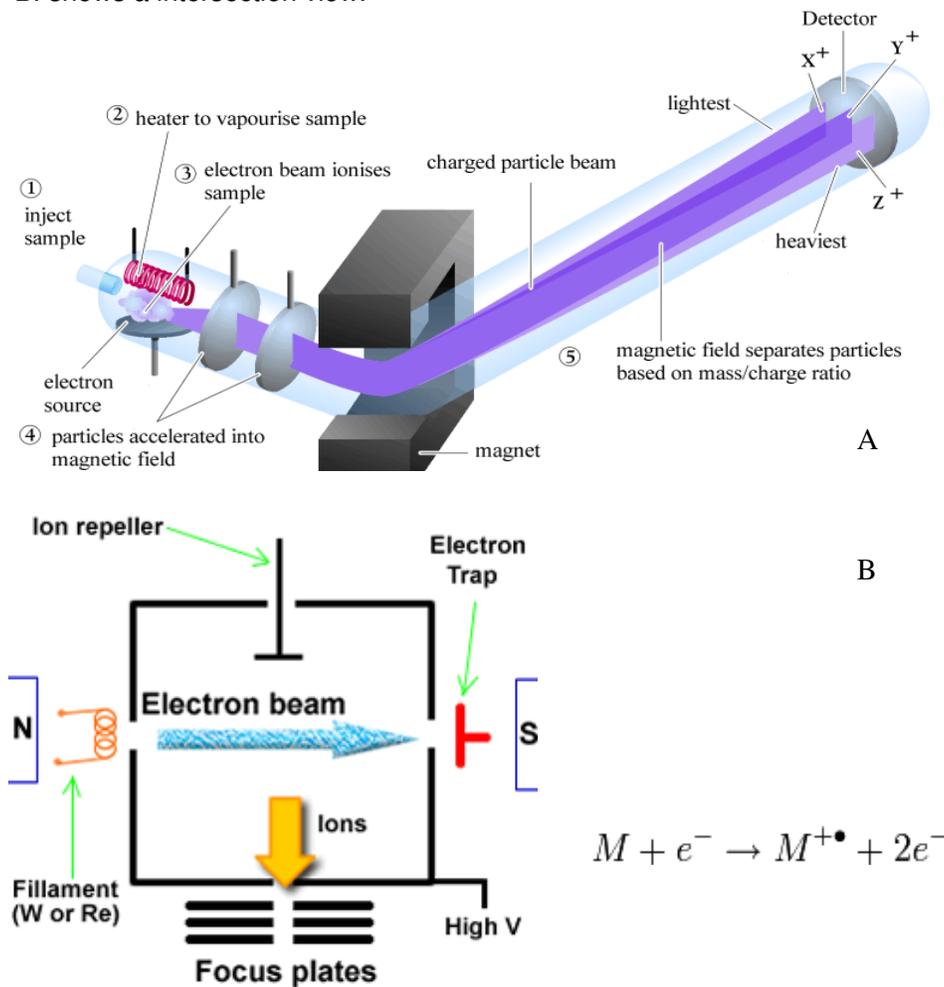


Table 1.1 molecular ions adducts often observed in ESI mass spectra

Ion name	Ion mass	Charge	Ion name	Ion mass	Charge
1. Positive ion mode			2. Negative ion mode		
M+3H	M/3 + 1.0073	3+	M-3H	M/3 - 1.0073	3-
M+2H+Na	M/3 + 8.3346	3+	M-2H	M/2 - 1.0073	2-
M+H+2Na	M/3 + 15.7662	3+	M-H ₂ O-H	M - 19.0184	1-
M+3Na	M/3 + 22.9893	3+	M-H	M - 1.0073	1-
M+2H	M/2 + 1.0073	2+	M+Na-2H	M + 20.9747	1-
M+H+NH ₄	M/2 + 9.5206	2+	M+Cl	M + 34.9694	1-
M+H+Na	M/2 + 11.9983	2+	M+K-2H	M + 36.9486	1-
M+H+K	M/2 + 19.9852	2+	M+FA-H	M + 44.9982	1-
M+ACN+2H	M/2 + 21.5205	2+	M+Hac-H	M + 59.0139	1-
M+2Na	M/2 + 22.9892	2+	M+Br	M + 78.91889	1-
M+2ACN+2H	M/2 + 42.0338	2+	M+TFA-H	M + 112.9856	1-
M+3ACN+2H	M/2 + 62.5471	2+	2M-H	2M - 1.0073	1-
M+H	M + 1.0073	1+	2M+FA-H	2M + 44.9982	1-
M+NH ₄	M + 18.0338	1+	2M+Hac-H	2M + 59.014	1-
M+Na	M + 22.9892	1+	3M-H	3M - 1.0073	1-
M+CH ₃ OH+H	M + 33.0335	1+			
M+K	M + 38.9632	1+			
M+ACN+H	M + 42.0338	1+			
M+2Na-H	M + 44.9712	1+			
M+Propanol+H	M + 61.0653	1+			
M+ACN+Na	M + 64.0158	1+			
M+2K-H	M + 76.9190	1+			
M+DMSO+H	M + 79.0212	1+			
M+2ACN+H	M + 83.0604	1+			
2M+H	2M + 1.0073	1+			
2M+NH ₄	2M + 18.0338	1+			
2M+Na	2M + 22.9892	1+			
2M+K	2M + 38.9632	1+			
2M+ACN+H	2M + 42.0338	1+			
2M+ACN+Na	2M + 64.0158	1+			

Figure 1.2 A demonstration of EI working mechanisms. A. Shows a 3-D view of EI; B. shows a intersection view.



Since EI as a technique has been around for nearly 60 years compared to 20 years for ESI the spectral libraries are extensive and contain over 100,000 spectra, in addition fragmentation patterns are very reproducible in this mode.

1.2.3 Matrix Assisted Laser Desorption (MALDI)

MALDI, or a technique like it, is the missing ionization technique for use in conjunction with liquid chromatographic techniques since it can ionise many molecules which are not ionised under ESI conditions. As seen in Figure 1.4, it is used as a static technique requiring the sample to be deposited on a plate in a light absorbing matrix and ionisation is carried out by using a laser which transfers

energy to the molecules causing them to ionise. There are some molecules which are difficult to ionise under ESI conditions but which ionise efficiently with MALDI. For example highly phosphorylated compounds such as ATP, NADP and acetyl CoA are readily ion suppressed under ESI conditions but ionise well when MALDI is used. Extensively phosphorylated compounds tend to pair strongly with ions commonly available in the biological matrix such as sodium and the tight ion pairs formed do not ionise readily under ESI conditions. Although MALDI does not interface readily with chromatography it is possible to get automated systems which collect LC fractions and spot these onto MALDI plates. MALDI will become increasingly important for the spatial mapping of biomarkers.

1.3 Ion Separation Methods

1.3.1 Magnetic Sector Instruments

The first method used for separating ions in a mass spectrometer was carried out by separation in a magnetic field. As seen in Figure 1.5, the energies of the ions in a magnetic sector instrument are focused in an electrostatic field prior to entering the circular magnet. The ion optics (or electrostatic sectors) accelerates the ions to a kinetic energy given by

$$\text{K.E.} = \frac{1}{2} mv^2 = qV \quad \text{Equation 1.}$$

Where m is the mass of the ion, v is its velocity, q is the charge of the ion and V is the applied voltage of ion optics. The electrostatic field that ions travel acts as an energy-filtering device. The electric sector sorts the ions by their energy. This filtering improves the resolution of the mass spectrometer.

The ions then enter the curved tubes and are deflected by the magnetic field, B :

$$mV^2/r = BqV, \quad \text{Equation 2.}$$

Where r is the radius of curvature.

$$m/q = B^2 r^2 / 2V \quad \text{Equation 3.}$$

In a magnetic field, equation 2 can be rearranged to equation 3.

It shows the m/q ratio of the ions that reach the detector can be varied by changing the magnetic field (B). On the other hand, only ions of mass-to-charge ratio that have equal centrifugal forces will pass through the curvature magnetic sector. In summary, by varying the voltage of the electrostatic field or magnetic field strength of the magnetic-sector analyzer, the individual ion beams are separated spatially and each has a unique radius of curvature according to mass/charge ratio.

Figure 1.3 Comparison of the EI spectrum of the di TMS derivative of succinic acid in a sample (A) against the di TMS of a succinic acid standard in (B).

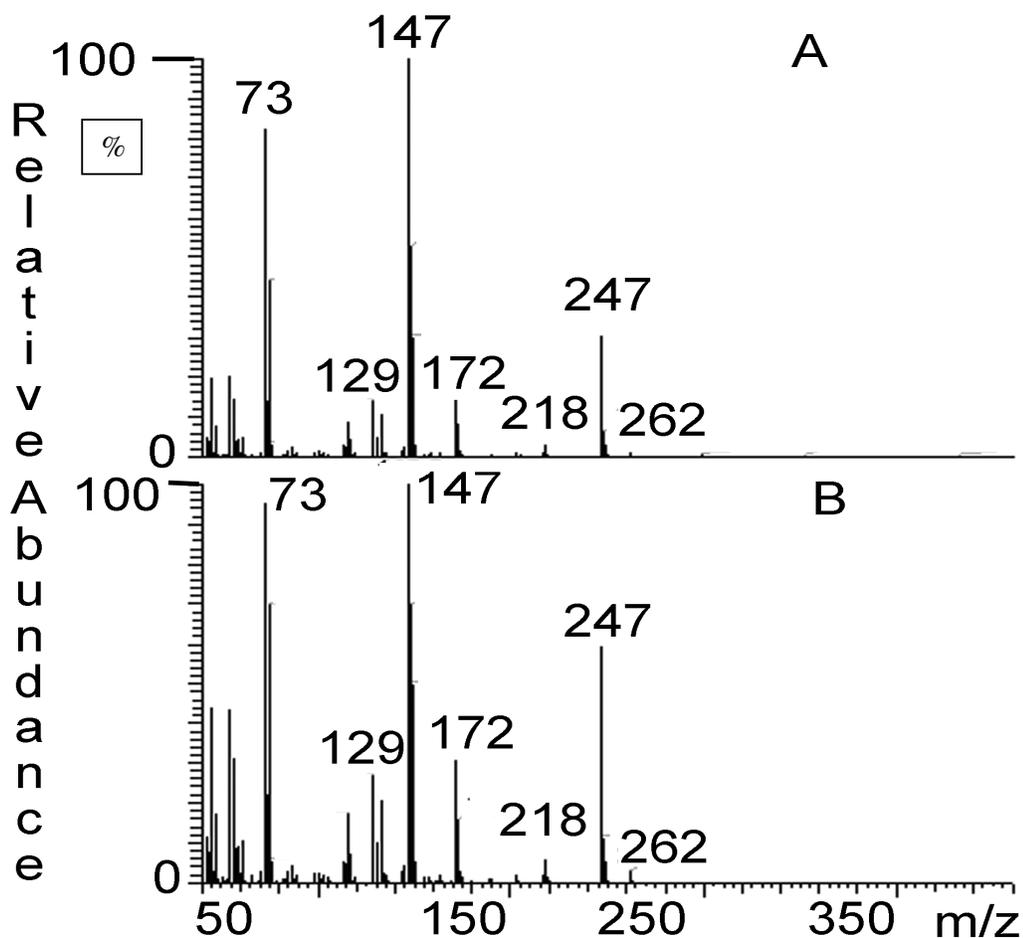
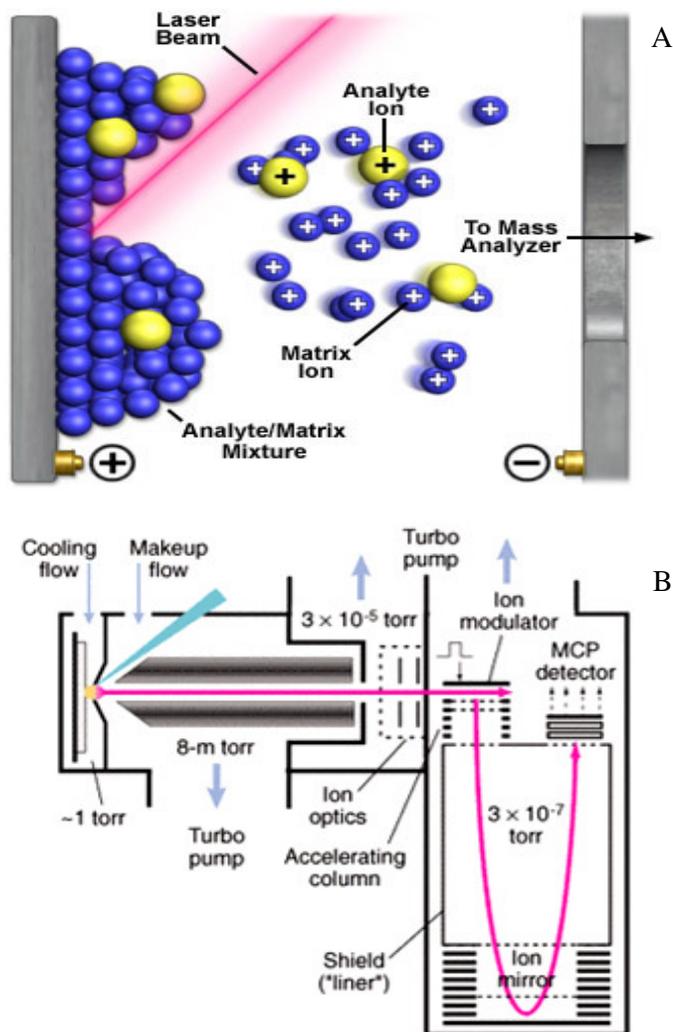


Figure 1.4 The demonstration of MALDI working mechanism and time of flight separation. A. Shows analytes are activated by laser and start to fly with charged ions. B. Shows a diagram in a time-of-flight detector.

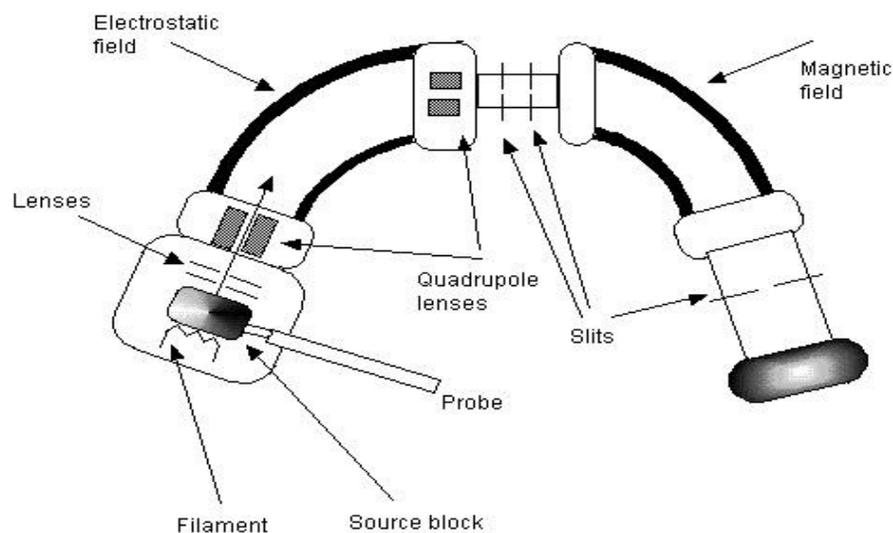


1.3.2 Quadrupole Instruments

Single quadrupole instruments provide entry level mass spectrometry systems and they can provide complex data sets particularly if combined with good chromatography as seen in Figure 1.6. Their main drawback is their limited ability to produce fragmentation information in LC-MS mode and low resolution which is typically 0.5 amu. In GC-MS mode with EI, fragmentation information is available as outlined above which makes single quadrupole separations more information rich in

this mode. In LC-MS tandem MS systems are used to produce fragmentation in conjunction with ESI which only produces limited fragmentation. A tandem system uses three quadrupoles in series, the first quadrupole is used to select an ion of interest, the second quadrupole simply acts as an ion guide and acts as a collision cell where the ion of interest is forced to collide with a heavy gas such as argon or sometimes nitrogen producing fragments and the fragments are then separated by the third quadrupole. Tandem systems deliver the most sensitive analyses in terms of detection limits being about an order of magnitude more sensitive than trap based instruments. They are also widely used in proteomics studies where the output is readily integrated with database searching and also in bioanalytical studies.

Figure 1.5 The demonstration of electrostatic and magnetic sector working mechanisms in a double focusing magnetic sector instrument.



1.3.3 Ion Trap Instruments

Ion trap instruments generally do not deliver the same level of sensitivity as tandem instruments but they can carry out multiple fragmentation experiments in order to provide more detailed structural information for a compound. The classical

ion trap consists of a ring electrode and two hemispherical endcap electrodes as seen in figure 1.7. The ions are introduced into trap by an applied AC voltage on

Figure 1.6 The demonstration of Single Quadrupole working mechanism

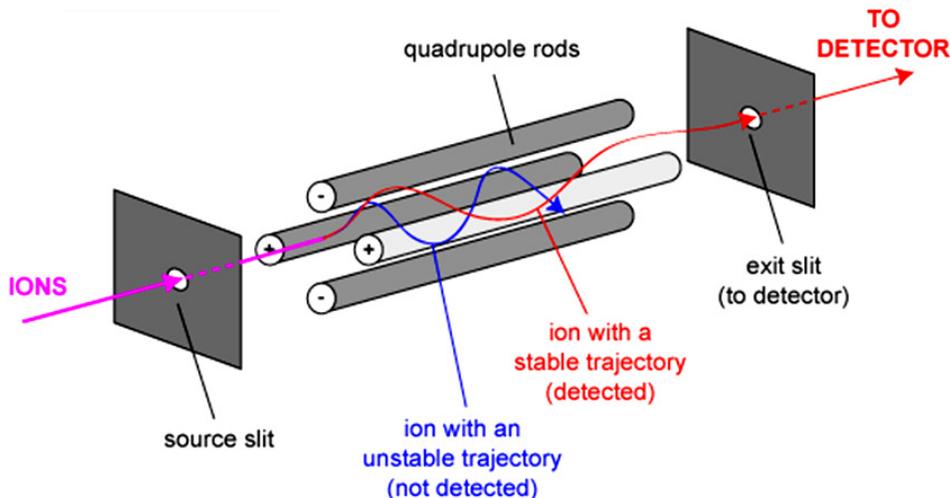
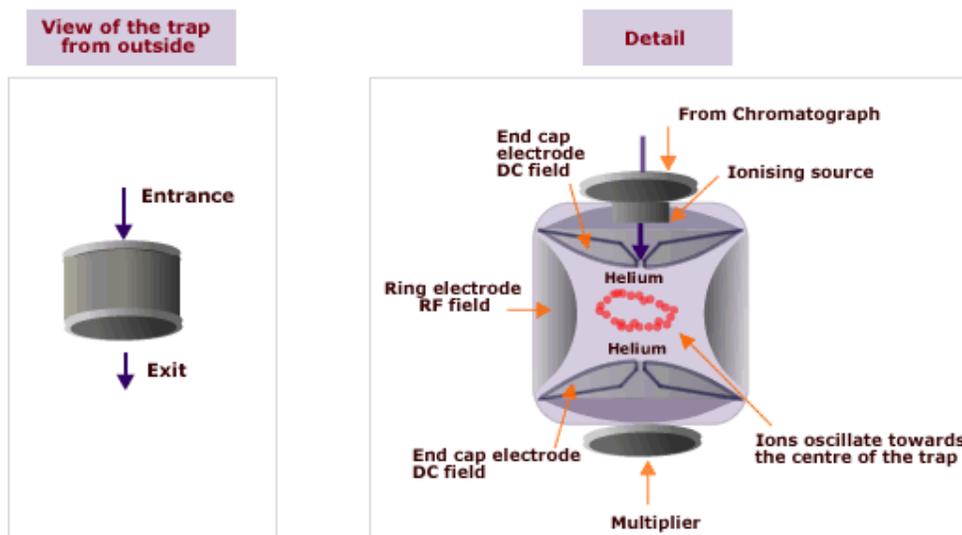


Figure 1.7 The demonstration of Ion trap working mechanism



endcap electrodes and quenched by the helium gas to stabilize ions near the centre of the trap. The helium gas is filled in the ion trap with a pressure of c.a. 1 m torr. both to cool down the temperature of ions and reduce their kinetic energies. Trapped ions can be stably trapped according to their mass to charge ratio, the size

of the ion trap, the oscillating frequency of the fundamental RF and the amplitude of the voltage on the ring electrodes. When the ions of interest are selected by users, the parent ions are excited by changing the potential on the endcap and thus kinetic energy is increased. Other ions will fall outside of the boundaries of the stability, and hit the electrodes causing them to be lost or ejected. Structural information is obtained by application of a low-amplitude AC voltage to make ions faster so that they collide with helium gas and this causes the ions to dissociate and fragment. This process produces random fragmentation ions and ions do not enter the detector because the energy applied is not high enough to eject them. A mass spectrum is generated by sequentially ejecting fragment ions from low m/z to high m/z through applying higher amplitudes of the fundamental RF potential. Ions thus move faster and faster and they go through exit of the trap finally. Fragmentation of the isolated precursor ion can be repeated a number of times and is only limited by the trapping efficiency of the instrument. The MS^n capability is the unique feature of the ion trap and makes structural elucidation capability more powerful than that of triple quadrupole mass spectrometers. The selectivity of MS^n can be used to yield structural information of analytes in the presence of mixtures [23-25].

1.3.4 Time of Flight (TOF) Instruments

Initially TOF was mainly linked to the pulsed ionization technique MALDI or GC. Hybrid instruments combining quadrupole ion focusing with TOF (QTOF) have been developed over the last 15 years. Its working principle is shown in figure 1.8. In this type of instrument TOF separation is compatible with chromatographic systems interfaced to an ESI source. Thus QTOFMS is one of the methods of choice in metabolomic analyses since it can deliver accurate mass data enabling determination of elemental compositions. The technology has been gradually improved with regard to the ability of instruments to deliver a wide dynamic range.

Linear dynamic ranges are around 10^3 and resolving power of around 3 ppm on most instruments. There is a physical limitation to the resolving power of the instrument which is that greater resolution depends on longer flight tubes. Bruker have taken the technology to its limit with the production of their Maxis instrument [26, 27]. Its long extended flight tube allows it to deliver comparable mass accuracy to Fourier transform instruments and its ETD MS/MS capability has been widely used for top-down protein identification.

Figure 1.8 A demonstration of the Q-TOF working mechanism

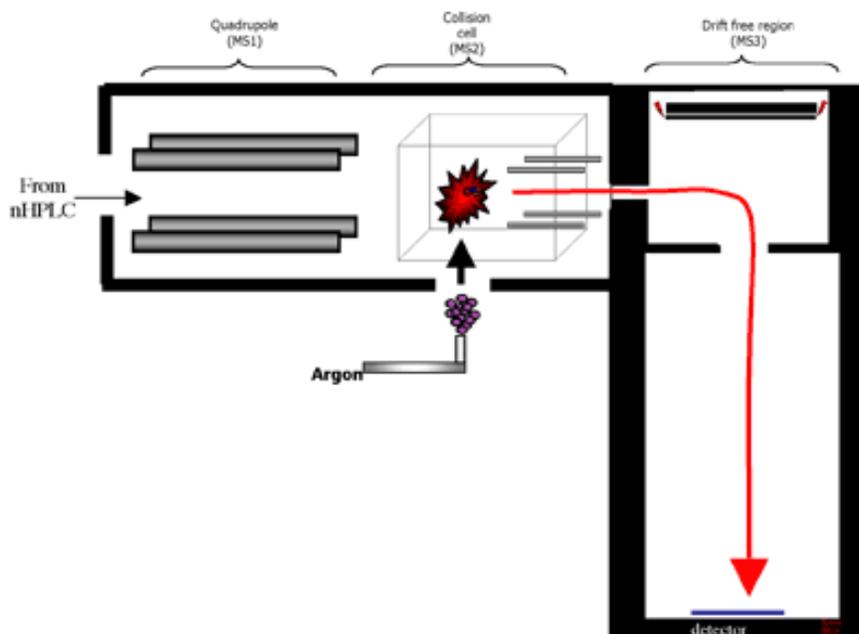
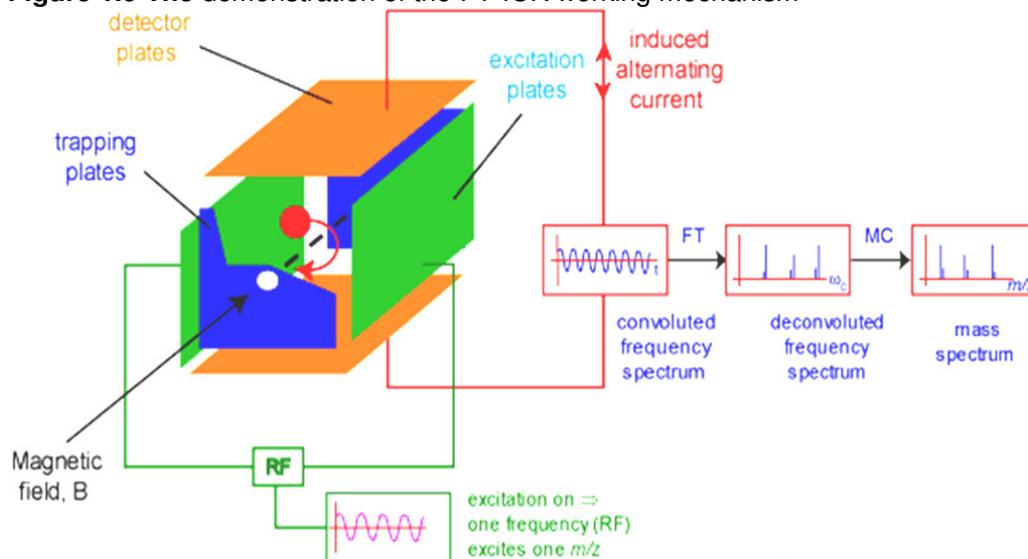


Figure 1.9 The demonstration of the FT-ICR working mechanism



1.3.5 Fourier Transform Ion Cyclotron Resonance (FT-ICR)

FT-ICR offers the highest resolving power and accuracy of all MS instruments to date and although FT-ICR offers superb resolution and the ion-ion interactions in FT-ICR decrease the dynamic range of measurement and thus inevitably metabolite coverage. FT-ICR requires a longer time to make accurate mass measurements and is thus less compatible with chromatographic methods than other ion separation methods [28-30]. Its working principle is based on the circular movement of charged particles in a strong magnetic field (cyclotron movement). The cyclotron frequency depends directly on the mass-to-charge ratio of the ions. The periodic movement of ion packets is recorded and converted to a frequency spectrum with a Fourier transform, which is finally converted to a mass spectrum after calibration. Figure 1.9 illustrates the ion separation process used in FT-ICR MS.

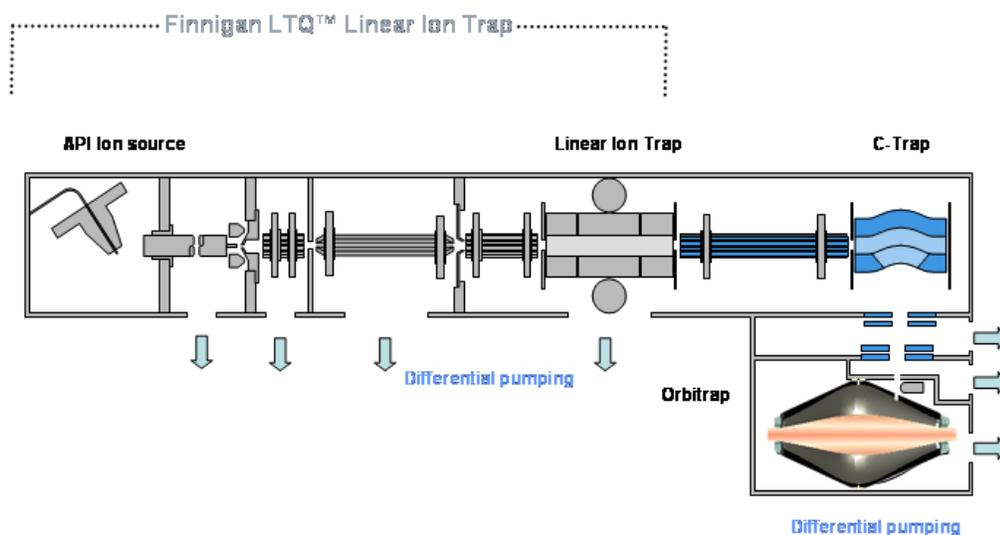
Unlike TOF, Quadrupole and magnetic separations which destroy the ions during analysis, FT-ICR is a non-destructive method that both allow users to run further experiments on the ions and permits longer data collection times, thus increasing sensitivity. In theory the longer the detection time, the better the certainty and

precision of the measurement. Thermo and Bruker each have their own FTMS for metabolomics and proteomics study that Thermo's LTQ-FT, for instance, is a hybrid linear ion trap-FTMS, and Bruker's APEX-Qq is a triple quadrupole (Qq front end) mass analyzer interfaced to an FT-MS [31]. In 2009, Bruker launched the cutting edge Solarix system. It is equipped with the world's highest field FTMS magnet which provides record-breaking analytical power for the most challenging complex samples [32].

1.3.6 The Orbitrap Mass Spectrometer

In the Orbitrap the ion separation technology (figure 1.10) is based on trapping the ions injected into the trap between an outer barrel-like electrode and an inner

Figure 1.10. The demonstration of LTQ-Orbitrap working mechanism



spindle-like electrode [33]. The ions exhibit angular, radial and axial oscillations at frequencies all of which are mass-dependent [34]. The m/z ratio is measured by an image current generated by the axial oscillation which is completely independent of energy and of the spatial spread of the ions. The instrument produces fast measurement of accurate masses and is thus compatible with chromatography.

The instrument is very sensitive and capable of measuring analytes at a concentration of < 1 ng/ml [35]. Compared to FT-ICR, Orbitrap has several advantages: they are less expensive and have lower maintenance costs; they do not have a huge magnetic sector and thus a complicated cooling system is not necessary. It also has some of the limitations of an FT-ICR instrument in that space-charge effects limit the number of ions that can be delivered to the trap and thus the dynamic range, however, dynamic ranges approaching 10^5 can be achieved. The Orbitrap is now available in four configurations: the Exactive which can measure accurate masses but has no fragmentation capability, the Discovery which can measure accurate masses and produce fragmentation information, the Orbitrap XL which can deliver accurate mass and both high and low energy fragmentation [36-39]. LTQ Orbitrap Velos was launched in 2009. It is equipped with the new technology of an S-lens replacing the old tube lens, a dual pressure linear ion trap and rapid scanning cycle time of Orbitrap Velos. LTQ Orbitrap Velos is believed to be the fastest and the most precise accurate mass instrument in the world [40].

1.4 Separation Methods

Some early work in mass spectrometry based metabolomics used direct infusion of samples into ion cyclotron resonance mass spectrometers. However, a chromatographic step is generally important in order to minimize the risk of ion suppression effects and also in order to be able to distinguish between isomers. Certainly separation methods are always used in bioanalytical studies since in a biological matrix there is an abundance of interfering compounds.

1.4.1 Reversed Phase Chromatography

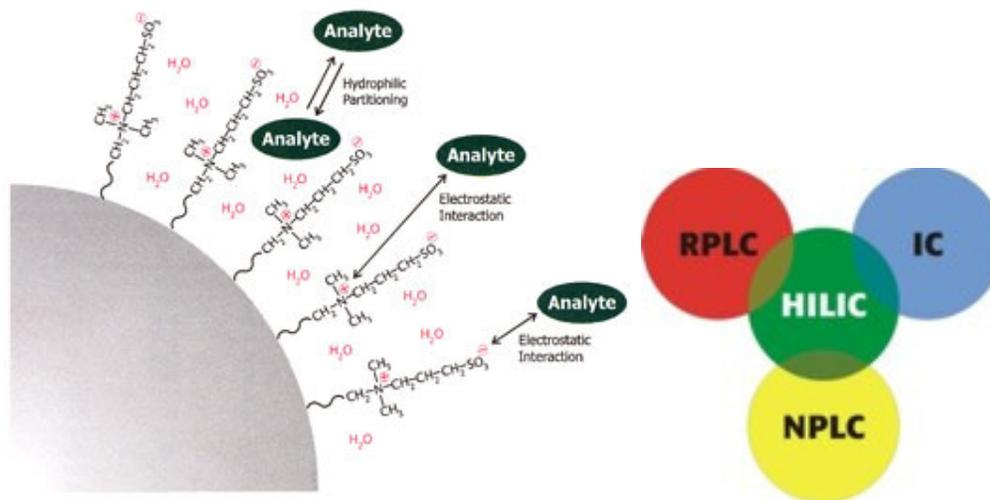
Reversed-phase chromatography (RPC) is often used for analysis of metabolomic samples. It is very useful for lipophilic compounds and is the technique of choice for the analysis of most drugs in biological matrices. There is a

disadvantage to using RPC with biological samples because lipophilic compounds such as phospholipids and polymeric residues from plastic ware tend accumulate on-column and elute in subsequent runs causing interference and ion suppression, unless a washout programme with high levels of organic solvent is included in the run. In metabolic profiling, the resolving power of RPC can be increased by using ultra performance liquid chromatography (UPLC) [41-44]. UPLC offers high chromatographic efficiency: a typical 1.7 μ m particle size column delivers about 2.5 times the efficiency of a 5 μ m particle column. The flat van Deemter plot obtained with low particle size means that very high mobile phase flow rates can be used without compromising efficiency. While fast separations can be obtained by using UPLC, because of the high back pressures the number of plates per column is limited i.e. the pressure required to pump solvent through long columns is high. The highest number of theoretical plates (peak capacity) could be obtained by connecting several 5 μ m columns in series [45], the disadvantage of this approach is that run times are relatively long, however, in biomarker discovery, high throughput analyses are less important than accurate identification. UPLC is perhaps more appropriate for rapid screening post-biomarker discovery. Many biomolecules are not well retained in RPC, and thus polar amino acids such as glycine and alanine will elute in the void volume of most columns - as will sugars, small peptides, pyrimidines and polyamines. Elution in the void volume is not desirable because no true chromatographic information is available and also there is the risk of ion suppression since the inorganic salts present in the biological matrix also elute in the void. Some of these problems can be solved by using hydrophilic interaction chromatography (HILIC).

1.4.2 Hydrophilic Interaction Chromatography (HILIC)

HILIC provides an alternative to reversed-phase chromatography. This option has been around for years in the form of aminopropyl columns which are commonly used in the analysis of sugars. Figure 1.11 illustrates the principle behind HILIC, on the ZICHILIC phase from Sequant, which involves the use of the water layer associated with column surface as a pseudo stationary phase. The zwitterion coating in theory is charge neutral thus minimising ion exchange interactions with the analyte and the chain length separating the positive and negative charges of the ZICHILIC column is optimised so that the charges on the column surface can fold round and interact to neutralise each other. However, undoubtedly some ion exchange interactions occur. The column is used with a high organic solvent content in the mobile phase which favours ionisation under ESI conditions and is also beneficial in reducing the diffusion related mass transfer terms in the van Deemter equation thus increasing column efficiency. There are an increasing number of HILIC columns on the market or under development [46-49].

Figure 1.11 The chemistry of a Zic-Hilic column.



1.4.3 Capillary Gas Chromatography (GC)

Capillary GC offers very high separation efficiency and typical efficiencies for a standard 30-metre GC capillary column are ca 140,000 plates compared to

efficiencies of around 20,000 plates for a 15 cm HPLC column packed with 3 μ m particles. This equates to a resolving power for the GC column which is 2.6 times that of the HPLC column. The drawback of GC is that the sample has to be volatile and not all compounds are volatile. However, a wide range of compounds are volatile once they are converted to their trimethylsilyl (TMS) derivatives including: Krebs cycle acids, amino acids, sugars and sugar phosphates and steroids. There is a choice of capillary GC columns but generally non-polar columns such as DB-1 (based on methyl silicone) or weakly polar columns such as DB-5 are used. An advantage of GC is that peak resolution can be simply controlled by the temperature programme used for the GC oven. Thus GC-MS methods are completely complementary to LC-MS methods and allow analysis of certain sets of metabolites such as Krebs cycle acids with a high degree of confidence.

1.4.4 Capillary Electrophoresis

Capillary electrophoresis (CE) interfaced with mass spectrometry is a popular method of analysis for metabolomic samples with over 100 publications employing it within the last five years, 50 of these falling within the past year. It has a similar or even better resolving power than capillary GC and thus is very good for separating isomeric compounds and it also allows injection of extremely small volumes. Recently the direct analysis of aliquots (6 nL) of the contents of a single cell using CEMS was reported [50]. Since the volumes which can be loaded onto CE columns are only very small it has suffered from a lack of sensitivity, however, the stability of the interfaces between CE and MS have improved allowing much more sensitive analyses. Sensitivity in CE can also be improved by sweeping and stacking methods which utilise the ability of CE to focus analytes injected in large volumes on the basis of differential migration rates. The simplest stacking method is to inject the sample in an electrolyte of low conductivity compared to the running buffer so that

the ions in the sample experience a relatively high potential difference compared to the background electrolyte and focus at the leading edge of the sample plug. A recent paper on metabolomics used isotachopheresis to focus a plug of sample ions between plugs of formic acid and ammonium hydroxide. In simple terms, an individual analyte, which has to be pH sensitive, is squeezed into a narrow band since the formic acid and ammonia force it in opposite directions until all its molecules are uniformly charged [51] .

1.5 Data Processing

The greatest amount of time in metabolomic experiments is devoted to data processing. The first metabolomics experiments were carried out by using NMR and the standard way presenting differences between sample groups was by use of principal components analysis (PCA) in which multidimensional data was reduced often to two dimensions.

1.5.1 SIMCA P and Other Chemometric Software

SIMCA (soft independent modelling by class analogy) is a chemometric method for unsupervised / supervised classification of data [52]. Common features of each class are extracted by principle components (PCs) to represent whole model. Samples which are described by spectra or chromatograms are mapped onto a much lower dimensional subspace for classification. The classifications are judged depending on class membership by the orthogonal projection distance between an unknown sample and the PC model of each class. If an unknown sample is assigned to one class, it will score a high probability. Otherwise, the residual variance of the sample exceeding the upper limit of model will be an outlier. SIMCA is sensitive to the quality of the data sets. There are diagnostics to assess the quality of the data, such as the modelling power or the discriminatory power. The modelling power describes how well a variable helps the principal components to

model variation, and discriminatory power describes how well the variable helps the principal components to classify the samples in the data set.

There are several popular SIMCA softwares including SIMCA-P, Statistica, and Unscrambler which can be used in life sciences research. Statistica and Unscrambler are professional mathematical software not only for mathematics but also for statistics. SIMCA-P is particularly designed for chemometrics, which focuses on principal components or a generalised extension (projection to latent structures, or PLS), in contiguous time series [53-56].

PCA is a relatively simple modelling procedure which is carried out without supervising the datasets. It is an orthogonal linear transformation that calculates a series of Eigen vectors to replace the higher dimensional data matrix. The results of a PCA are usually discussed in terms of component scores and loadings. PCA is a non-parametric analysis and the result is unique and independent of any hypothesis regarding data probability distribution. No prior knowledge therefore can be incorporated and a PCA model often incurs loss of information. Also it is an unsupervised method, no parameters can indicate whether or not the model is fit or not for unknown samples.

Partial least squares (PLS) or PLS- discriminant analysis (DA) is a supervised method. It models the relationship between two matrices X and Y, i.e., the data set obtained from chromatograms or spectral and class affiliation matrix. A PLS model tries to find the multidimensional direction in the X space that explains the maximum multidimensional variance direction in the Y space (It determines a set of latent variables corresponding to principal components in PCA, but explains as much as possible of the covariance between the two matrices). PLS-regression is suitable when the matrix of predictors has more variables than observations, and when there is multicollinearity among X values. The output of PLS is the score matrix that can

be plotted in a similar manner to PCA, and predictor matrices, and which estimate class affiliation. The comparison of particular vectors of the predictor matrix with respondent vectors of the target matrix shows the correctness or incorrectness of a particular sample classification.

Conventional PLS uses X to construct a model of Y , where the objective is to predict Y from X . The orthogonal/optimised PLS (OPLS) method is a recent modification of the PLS method, which is designed to handle variation in X that is orthogonal or uncorrelated to Y . OPLS separates the systematic variation in X into two parts, one that is predictive of Y and one that is orthogonal to Y . This partitioning of the X -data provides improved model interpretability and does not change the predictive power because the data matrix is modelled separately from the variation common to the X and Y matrices.

In order to use chemometric software the data has to be put into a form which the software can model. In the case of a chromatogram or a one dimensional NMR spectrum the data is sliced and binned into sections of a defined width e.g. 0.01 ppm in NMR or 0.1 minutes when considering a chromatogram. A particular problem with chromatographic data is to align the chromatograms so that any retention time drift from run to run is eliminated. Chemometric approaches are generally less useful with high mass spectrometry data since the dimensions (peaks) present in a sample run can be >20,000 and PCA does not really assist in uncovering the components which have changed because the data in the loadings plots provided by the mass spectrometry data are complex.

1.5.2 Sieve Software

Most commercial data processing software aims to plot extracted ion chromatograms across the full scan range of the mass spectrometer. In order to do this a bin width for the ions is set e.g. 0.02 amu and extracted ion chromatograms in

these bins are compared between two sample groups. Differences in the plots are highlighted and the exact masses of the ions producing the differences can be determined. This is best observed in figure 1.12 where the software has observed a difference in the intensity of the peak for S-adenosylmethionine (SAM) between trypanosomes cultured with glucose as their carbon source and trypanosomes cultured with proline as their carbon source. The red plots for the ion at m/z 399.14447 show the levels of SAM in 5 glucose grown samples and the blue plots the levels of SAM in the 5 proline grown samples. The software gives a ratio for the mean peak areas and a P value. The P value is affected by how well the chromatograms align and with poor alignment the P value will not be accurate. In figure 1.12 the alignment of the chromatographic peaks from different runs is good but not perfect. With Sieve software the data has to be carefully checked in order to ensure that no artifacts are produced by poor peak alignment. Another task in data processing is to identify metabolites from their exact masses. Automated data base searching is incorporated in Sieve software but the output is not user friendly. It is possible to write a macro in Excel to search against compound lists such as the Kegg database. In practice there are many compounds which yield clear chromatographic peaks where their exact mass does not correspond to compounds in any database thus the metabolome databases are far from complete.

In addition to Sieve, there are many other softwares for metabolomics data processing such as Masslynx from Waters. The software list is in shown table 1.2 in Appendix A.

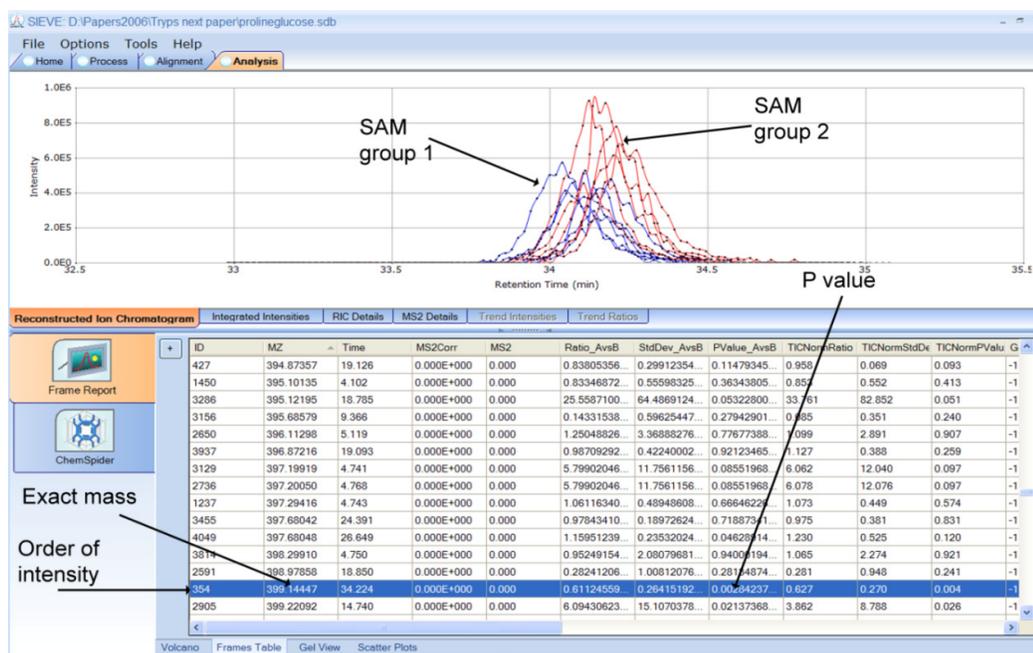
1.6 Metabolomics

Metabolomics is an emerging branch of “omics” dealing with metabolites. Along with transcriptomics and proteomics, metabolomics sets up a bridge between genotype and phenotype, as shown in figure 1.13. Apart from high molecular weight molecules such as nucleic acids, proteins, lipids, small molecular weight compounds

like sugars, amino acids and nucleotides are the main metabolites which undergo changes in living systems under pathophysiological conditions such as environment, drug response, nutritional intervention, genetic manipulation, or in response to toxicity [57].

In 1971, Horning and Mamer undertook the first metabolomics experiment using mass spectrometry [58, 59]. In the same year Pauling and Robinson published the first paper in this field [60]. The term metabolomics was not used until 1998 when Oliver *et al* introduced “metabolome” in their publication studying systematic functional analysis of the yeast genome [61]. Nicholson *et al* gave the study of metabonomics a clear definition as “the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification” [62] in 1998. It is also defined as the comprehensive and simultaneous systematic determination of metabolite levels in whole organisms and their changes over time as a consequence of stimuli, such as diet, life style, environment, genetic effects and pharmaceutical interventions, both beneficial and adverse [63].

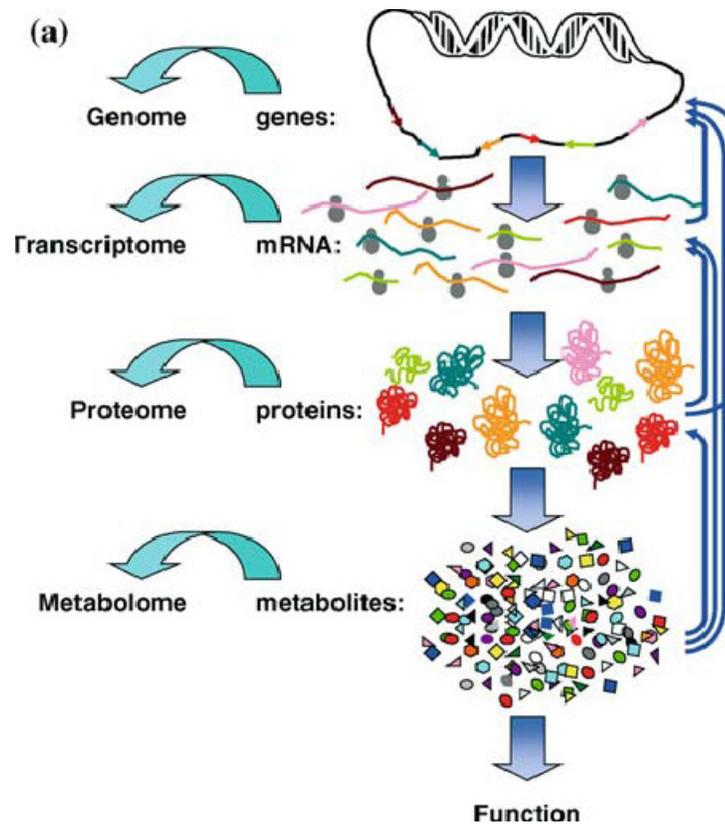
Figure 1.12 Application of Sieve 1.2 software to uncovering differences between two sample groups.



The samples studied in metabolomics include biofluids like serum, urine, bile or seminal fluid, cell cultures, tissue extracts or microorganisms [63-69]. In the agricultural field, plants have been modified genetically or environmentally in order to improve diet, health and nutrition. This causes the biological chemistry to be altered and the newly produced metabolism can be effectively monitored by means of a metabolomics study. However, a few difficulties need to be solved carefully such as biodiversity in the distribution of the plant ingredients, differences between plant parts, different soils or different seasons [70-72].

In the clinical field, metabolomics has been used in drug discovery, drug safety, drug development, toxicology and pathophysiology of diseases. Current studies in the field developed slowly because there are some difficulties such as the inaccessibility of biofluids of cerebrospinal fluid, biological variability in the the dilution of urine and difficulties of experimental design in the long-term ethical considerations. Furthermore, discontinuity in the studies also might delay the collection of experimental samples and cause statistical problems [73-75].

Figure 1.13 Demonstration of the roles of genomics, transcriptomics, proteomics and metabolomics



In the microbial field, the application of metabolomics focuses on the understanding of genetics and physiology of microbes in order to improve human health, drug effects and the environment. It is usually easy to control the microbial culture and to repeat sampling over a short period of time. However, the homogeneity of the samples is difficult because of variations in cell counts, variability of the cells, e.g. some are wild and some are resistant and the matrix effect of the components of the culture medium. Moreover, the metabolic quenching during sample preparation should be validated very well in order to ensure the reproducibility and accuracy of the analysis [76].

1.7 Aims and Objectives

The aim of the thesis was to apply ESI ion trap mass spectrometry methods, in conjunction with chromatographic separation, to the acquisition of data which could then be modeled using chemometric methods and other data processing methods, to a range of bioanalytical, metabolomic and lipidomics problems.

Chapter 2

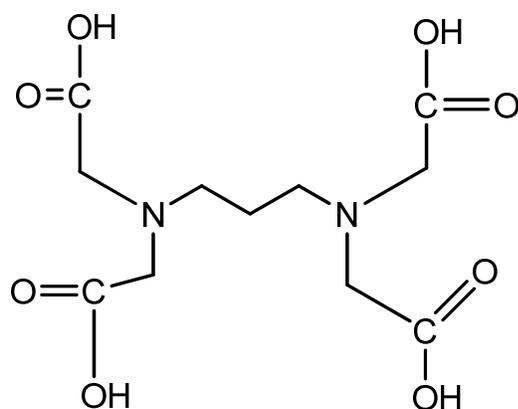
Development of a Bioanalytical Method for the determination of iron as its ethylenediaminetetraacetic acid (EDTA) complex and ethylene diamine disuccinic acid derivative (EDDS) complex in the snail *Helix aspersa* by hydrophilic interaction liquid chromatography coupled to Fourier transform electrospray ionisation mass spectrometry

2.1 Introduction

The project was presented and funded by Lonza Ltd. who were interested in getting a better understanding of the mechanism of action of a slug pellet produced by one of their competitors. They were keen to publish the findings and the proposed work presented a considerable technical challenge since it involved both mass spectrometry and intricate biological dissection and thus it was quite suitable for part of a PhD project. Iron (III) phosphate containing pellets are used as molluscicides but relatively little is known about the mechanism of action of these pellets although iron has been observed to be deposited in digestive gland and body wall of the mollusc [77]. It has been reported that the ability of metals to act as molluscicides was improved when they were formulated as organic complexes [78] and in 1995 a patent was taken out on the use of ferric EDTA and related complexes as molluscicides [79]. Recently it has been observed that earthworms are also very sensitive to molluscicides based on FePO_4 [80]. The aim of the current study was to develop a method for the determination of iron in the organs of *Helix aspersa* as a model for what may occur in other molluscs. A number of studies have been conducted to determine whether or not the uptake of iron via the gut is more effective when the iron is administered as a complex such as FeEDTA compared with administration of a simple salt such as FeSO_4 . In one study [81] it was found

that iron was accumulated to a lesser extent in liver and spleen of rat when FeEDTA was administered than when FeSO₄ was administered but other organs were not examined. Zhu *et al* [82] found that the penetration of iron through Caco-2 cells, used as a model for the cells of the gut, was less effective in the case of Fe (III) EDTA compared with FeSO₄ and FeCl₃. However, an *in vivo* study by the same group found that NaFeEDTA produced different tissue distributions of iron compared to FeSO₄ in the rat [83]. Levels of iron in the kidney were 83% higher when rats were fed NaFeEDTA and levels in the liver were 53% lower when compared with feeding FeSO₄.

Figure 2.1. The chemical structure of EDTA



Among a large number of chelating reagents for formation of complexes with metals, particularly iron, EDTA is considered to be very effective. EDTA can normally be represented as H₄Y (Figure 2.1), in the fully protonated form because of four acidic groups. The pK_a values of four corresponding hydrogens are pK_{a1}=0.26, pK_{a2}=0.96, pK_{a3}=2.60; pK_{a4}=2.76 [84]. The hexadentate ligand has an unusual chelating power with metals forming 1:1 complexes with many metals. Of all the important metals, Fe³⁺ has the highest complexation stability constant (logK of 25.1),

followed by Cu (logK 18.8), Zn (logK16.3), Co²⁺ (logK16.2) and so on [85]. In aqueous solution EDTA and Fe(III) forms an octahedral complex [FeY]⁻ exhibiting a negative charge.

The determination of Fe(III) as its FeEDTA complex has been carried out by a number of researchers. Owens et al. [86] studied EDTA complexes by capillary electrophoresis (CE) with UV detection. Complexes of various metal ions were formed prior to analysis including Fe³⁺, Ca²⁺, Co²⁺, Cu²⁺, Zn²⁺, Cd²⁺, and Pb²⁺. The authors reached a detection limit of [FeY]⁻ at 4 µg·mL⁻¹. Blatny et al. studied metal ions in water with CE; EDTA was added to water to form a complex which was then detected at 254 nm [87]. The limit of detection for Fe was 13 ng·mL⁻¹. Sillanpaa et al. [88] performed ion pair HPLC to measure EDTA complexes with ferric ions. They used a buffer solution containing the ion-pairing agent cetrimide as the mobile phase. The complex was detected at 260 nm and the limit detection was 0.5 µg·mL⁻¹. A similar ion pair method was used by Lucena et al. with tetrabutylammonium chloride as ion pair reagent and UV detection at 280 nm [89]. Chen et al. [90] developed an analysis of [FeY]⁻ by CE coupled with electrospray ionization-mass spectrometry. The authors reached a detection of limit at 4 ng·mL⁻¹ for [FeY]⁻ in the negative ion mode. Quintana et al. [91] used liquid chromatography coupled with ESI-MS to detect the EDTA-Fe (III) complex. They added the volatile ion-pairing reagent tributylamine to the mobile phase. A phenyl-hexyl column was used and the detection of limit was 1 ng·mL⁻¹ with a 50 µL injection volume. Álvarez-Fernández et al used an ESI-time of flight method to determine a number of Fe(III) chelates used in fertilisers in various agricultural matrices. The used reverse phase chromatography in the negative ion mode, high sensitivity was achieved through the use of ⁵⁷Fe-labelled internal standards and the LOQ for Fe (III) EDTA achieved was 328 pmol/ml (114.1 ng/ml) [92] Inductively Coupled plasma-mass spectrometry

(ICP-MS) has been also used to determine Fe. Yeh et al. [93] use ICP-MS to speciate V, Cr and Fe with CE separation, the limit of detection was $10 \text{ ng}\cdot\text{mL}^{-1}$ for $[\text{FeY}]^-$. Xuan *et al* [94] introduced a ZIC-HILIC column to separate and identify the different phytosiderophores of the mugineic acid family as their metal complexes. Various metals (Zn, Cu, Ni, and Fe) were added to 2'-deoxymugincic acid (DMA) and nicotinamine to form complexes. They used a mobile phase consisting of NH_4AC and CH_3CN at $\text{pH}=7.3$. When authors checked the stability of the complexes during separations using EDTA, they found EDTA-Fe (III) peaks $[\text{FeY}]^-$ at 21.7 mins with m/z 344.0. However, they had did not show chromatograms or validate the method.

In the current work a HILIC chromatography method interfaced with FT-ESIMS was developed for the determination of uptake of iron (III) into tissues of the snail following feeding of a commercially available molluscicide based on iron (III).

2.2 Experimental

2.2.1 Chemicals and Materials

All chemicals used were analytical grade. EDTA disodium salt dihydrate, EDDS trisodium salt solution, iron chloride hexahydrate were purchased from Sigma-Aldrich. HPLC grade acetonitrile and formic acid were from VWR International, UK. For all eluents and standards preparations, sample preparations, deionised water was provided by a water purification system (Milli-Q system, Millipore). Snails (*Helix aspersa*) were purchased from Blades Biological Systems (Kent, UK). Ferramol slug pellets (W.Neudorff, Germany) were purchased locally. Iron (III) phosphate pellets (1% w/w) were provided by Lonza Ltd and contained iron (III) phosphate without addition of EDTA.

2.2.2 Microscopy

All snails were dissected under a stereo-binocular microscope from Brunel, U.K. with magnification power objectives x 4 and an eyepiece x 10. A micro-dissection

tool kit was used for dissecting the snails. The dissected tissue was homogenised using a hand held homogenizer (LabGen 7B, USA).

2.2.3 Standards and Sample preparation

FeEDTA:

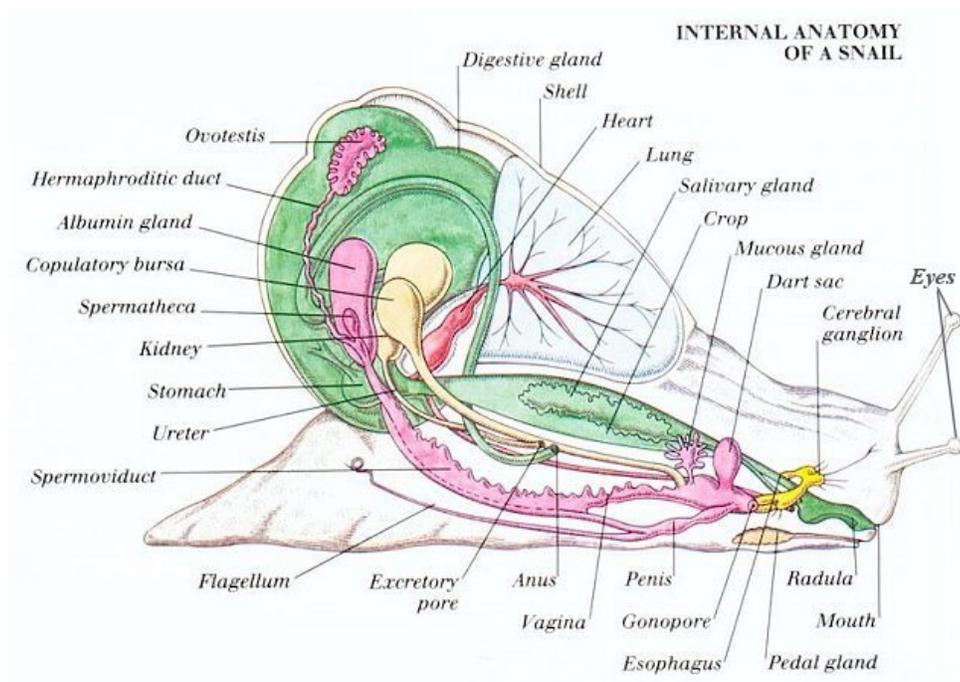
$\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$ (ca 77.7 mg MW=270.3 g·mol⁻¹) and $\text{C}_{10}\text{H}_{14}\text{O}_8\text{N}_2\text{Na}_2 \cdot 2\text{H}_2\text{O}$ (ca 106.9 mg MW=372.2 g·mol⁻¹) were weighed out and both dissolved in 100mL of water in the same volumetric flask, thus preparing solution A with FeEDTA (referred to hereafter as the FeEDTA complex standard) at a concentration of 1mg·mL⁻¹. Solution A was diluted x1000 with acetonitrile/0.1% v/v formic acid (70:30) producing a stock solution of FeEDTA complex standard at a concentration of 1 µg·mL⁻¹. From this solution calibration series was prepared with 2 ng·mL⁻¹, 10ng·mL⁻¹, 50ng·mL⁻¹, 200ng·mL⁻¹ and 500ng·mL⁻¹, 1000ng·mL⁻¹, 5000 and 10000ng·mL⁻¹ of FeEDTA.

2.2.4 Feeding and Dissection of Snails

The snails were fed manually with pellets which contained either FePO_4 or commercial pellets containing FeEDTA and kept in an aquarium. Once they had died (after 3 days), dissections were carried under the microscope. Firstly, the outer shells were removed by breaking bits of shell away with forceps taking care not to damage the soft tissue. The anatomy diagram of snails is shown in Figure 2.2. Pieces of the shell and any other surface contaminations were removed by washing the body in water (3 x 500mL). The mantle skirt was located and cut away; another cut was made transversely from the pneumostome to the left side of the mantle along the posterior edge of skirt. The body was cut posteriorly along the left side of the mantle to the posterior end of the lung. Thus the mantle was laid aside and the atrium, aorta, ventricle and kidney were removed from the lung roof. The dart sac in the body cavity was also removed. The kidney, heart and dart sac were placed into 5 mL vials and extracted with either 1mL of 0.1 % v/v formic acid in water in order to

determine FeEDTA in the tissues after feeding with FeEDTA pellets or 1 ml of 0.1% v/v formic acid in water containing $10 \mu\text{g}\cdot\text{mL}^{-1}$ of EDTA where total iron in the tissues was compared between snails on FeEDTA pellets, FePO_4 pellets

Figure 2.2. The anatomy diagram of snails.



and a control group not fed on pellets. The tissue was then homogenised with a hand held homogeniser. A drop of 1% v/v Triton solution was added to the vials and the samples were left sonicate for one hour at 40 °C. Acetonitrile (1.5 mL) was then added to each vial and all vials were vortexed for 5 mins. The solutions were filtered using a syringe filter and transferred to HPLC autosampler vials.

2.2.5 Caco 2 Assay System

Caco-2 cells (HTB-37) were obtained from the ATCC. It is a continuous line of heterogeneous human epithelial colorectal adenocarcinoma cells. Permeability studies were carried out using the BIOCOAT HTS Caco-2 assay system purchased from Becton Dickinson Labware Europe. This system provides an *in vitro* intestinal

model for the transport of drugs and natural compounds. Consisting of a fibrillar collagen coated 24 well insert plate and a multiwell feeder tray for culturing cell monolayers; combined with basal seeding medium, Entero-STIM a serum free medium containing butyric acid to induce cell differentiation and MITO+Serum extender containing the hormones, growth factors and metabolites required for serum free cell culture.

A suspension of Caco-2 cells at a density of 4×10^5 cells/ml was prepared in basal medium supplemented with MITO+ serum extender, 500 μ l of which was pipetted into each of the 24 collagen inserts to give a seeding density of 2×10^5 cells per insert. A volume of the same medium (35 ml) was pipetted into the Multiwell feeder tray and the system was incubated at 37°C, 5% CO₂ and 100% humidity for 24 hours. The medium was then removed from the feeder tray and from each insert taking care not to damage the cell monolayer. 500 μ l of Entero-STIM medium supplemented with MITO+ Serum extender was added to the interior of each insert and 35ml of the same medium was added to the feeder tray. The system was then incubated under the conditions previously described for 48 hours to allow formation of the differentiated enterocyte- like monolayer. The Multiwell insert plate was removed from the feeder tray and placed on a 24 well plate. The Entero-STIM medium was carefully removed from each insert and the cell monolayers were washed with PBS. Prior to permeability testing Caco-2 monolayer integrity was confirmed by means of Lucifer yellow rejection. Rejection values of 98.9% were obtained which indicated well formed cell monolayers. Permeability of EDTA and FePO₄ made up in PBS (pH 7.4) was tested at concentrations of 10, 50 and 100 μ M in duplicate. The test solutions were added in a volume of 300 μ l to the inside of the inserts and 1000 μ l PBS was added to each well of the receiving 24 well plate. The test plate was incubated at 37°C, 5% CO₂, 100% humidity for 120 minutes, with

intermittent shaking at 50rpm. The insert plate was then removed, and the solutions in the receiving 24 well plate were analysed by LCMS to determine % permeability.

2.2.6 Statistics

The one way and two ways ANOVA with replication was performed on the results obtained for the analysis of Iron (III) as FeEDTA in the snail organs.

2.2.7 LC/MS Instrumentation

The HPLC system consisted of Finnigan Surveyor MS Pump Plus; an online Finnigan Surveyor degasser; Finnigan Surveyor MS Autosampler Plus; Finnigan Surveyor PDA Plus detector (Thermo Finnigan, Massachusetts, USA). A ZICHILIC column was used (50mm X 4.6 mm, 5 μ m) purchased from HiChrom Ltd, Reading, U.K. It was fitted with a ZICHILIC guard column (20mm X 2.1 mm, 5 μ m) from the same supplier. The determination of EDTA-Fe (III) was carried out using a mixture of 0.1% v/v formic acid in water with acetonitrile (30:70) as the mobile phase at room temperature. The run time was 18 mins with a flow rate of 0.5 mL \cdot min⁻¹ and an injection volume of 10 μ L. For the separation of other metal EDTA complexes, the eluent was changed to 0.1% v/v formic acid in water: CH₃CN (40:60) and parameters remained the same. The FT-ESIMS instrument was a Finnigan LTQ Orbitrap (Thermo Finnigan, Massachusetts, USA) operated in the negative ion mode. The capillary spray voltage was set to 4.5 KV. The temperature of the heated capillary was maintained at 250 °C. Nitrogen gas flow was kept at a flow rate of 41 and auxiliary gas at 28, the tube lens voltage was -100V. The scan range was set to 100-800 amu at the beginning when confirmation of the peaks of [FeY]⁻, then set to scan the narrow range of 343-345 amu to monitor FeEDTA.

2.3 Results and Discussion

2.3.1. Method Validation

The measurements of the EDTA-Fe (III) complex were optimised in several steps. The various LC conditions and mass spectrometer conditions were optimised and the mass spectrometer was tuned to the molecular ion for the EDTA complex using standard solutions at a concentration of $1\mu\text{g}\cdot\text{mL}^{-1}$. The ZICHILIC column is particularly effective type of HILIC column based on a zwitterionic stationary phase; it gives very stable chromatography when used with simple aqueous modifiers such as 0.1% formic acid. The mass spectrum of the $[\text{FeY}]^-$ is shown in Figure 2.3. The spectrum shows the three major ions. The peak m/z 344.00 (343.9931) is from the $[\text{FeY}]^-$ and the lower spectrum shows the MS^2 obtained at a normalised collision energy of 35 V. The peaks m/z 299.9931, m/z 256.0435 correspond to successive losses of the carboxylate group starting from the parent ion $[\text{FeY}]^-$. These fragmentations confirmed the ion at m/z 344.0 to be due to the $[\text{FeY}]^-$. The specificity of the LC-MS method was demonstrated to be free from interference by running an EDTA solution blank which can be compared with the peak for FeEDTA obtained at the LOD (Figure 2.4).

The Intra-day instrument precision of the method was evaluated using two concentrations of the FeEDTA complex standard which was spiked into homogenised snail tissue. Nine consecutive runs were carried out. The relative standard deviation (RSD.) values obtained are shown in Table 2.1 and in both cases were $ca \pm 0.5\%$.

The Inter-day precision was evaluated on day 1 and day 2 with spiked solution concentrations of at $12.5\text{ ng}\cdot\text{mL}^{-1}$ and at $5000\text{ ng}\cdot\text{mL}^{-1}$. Six injections were carried on each day. The means and RSD values were similar for day 1 and day 2.

The linearity of the method was obtained for the concentration range 2-10000 $\text{ng}\cdot\text{mL}^{-1}$. The calibration curves of the peak area as function of concentration were linear, with a correlation coefficient of 0.9998. Because the samples were

determined during several months, the calibration curves were prepared freshly each time. The R.S.D. values of the line slopes (K), intercept (B) and correlation coefficient R^2 shown in table 2.2 indicate good precision for the linearity over the time period.

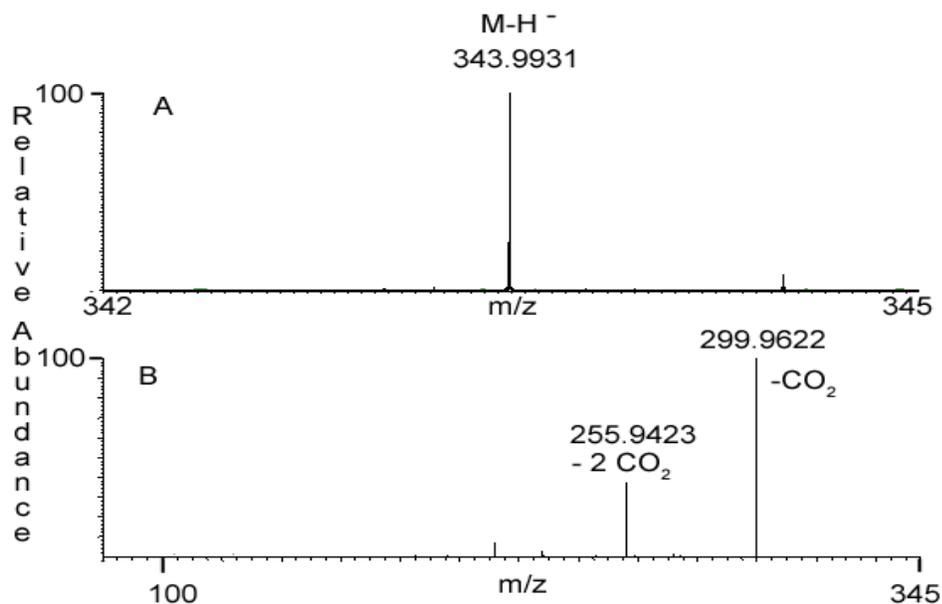
Table 2.1 Inter- and Intra-day precision for FeEDTA solution spiked into samples of homogenised snail tissue filtered then analysed by FT-ESIMS.

Inter-day variation		
	[FeY]⁻=12.5	[FeY]⁻=5000ng/mL
Mean	348145	99281296
S.D.	2216	491096
R.S.D. (%)	0.636	0.495
Intra-day variation		
Day 1.		
Mean	347759.3	99248153
S.D.	2048.5	548116.7
R.S.D. (%)	0.589	0.557
Day2.		
Mean	347807.5	99254116
S.D.	1984.4	603225
R.S.D. (%)	0.571	0.608

Table 2.2 Calibration curve statistics for slope, intercept and R^2 (n=3) for calibration in the range 2-10000 ng.mL⁻¹

No.	Slope (K)	Intercept (B)	R²
1	16697	92555	0.9988
2	16726	78522	0.9990
3	17889	85554	0.9998
Mean	17104	85544	0.9992
R.S.D.	3.976	8.202	0.053

Figure 2.3 A representative mass spectrum of FeEDTA obtained in negative ESI mode. B MS2 of FeEDTA at 35V collision energy

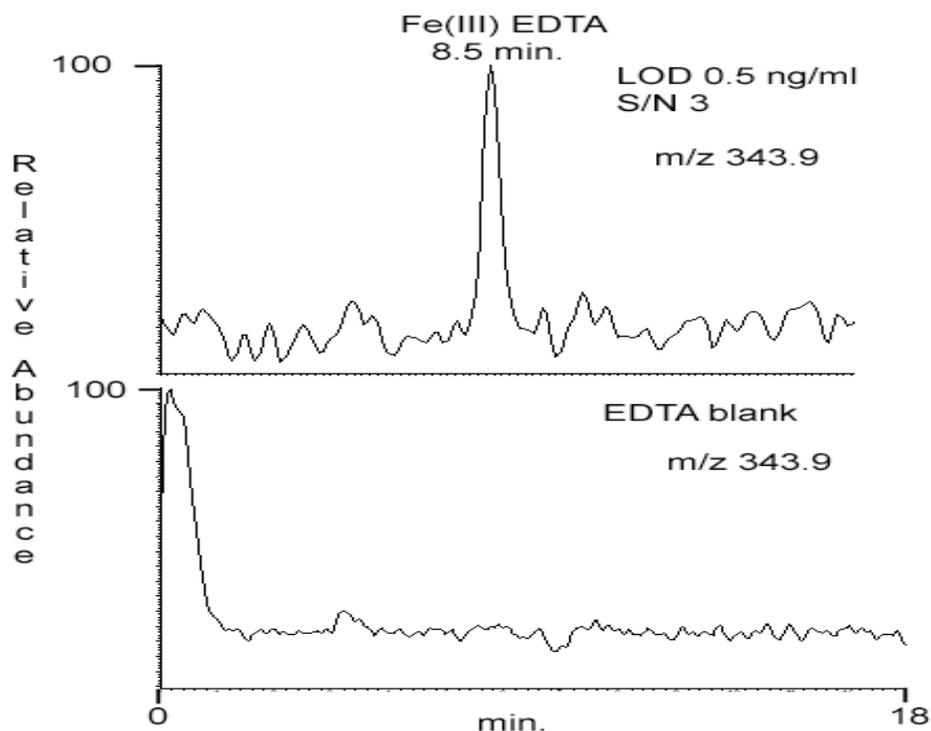


The observed detection limit was about $0.5 \text{ ng}\cdot\text{mL}^{-1}$ for a $20 \mu\text{L}$ injection volume ($S/N=3$) (figure 2.4). Once the method had been validated the Ferramol slug pellets were analysed for their FeEDTA content and it was found to be $1.08\% \text{ w/w} \pm 0.42\% \text{ RSD}$ ($n=3$). Although the active ingredient in the pellets according to the label claim is FePO_4 it is obvious, because of the high stability constant for FeEDTA, that this complex will form as soon as the pellets become wet since the pellets obviously also contain EDTA.

2.3.2. Studies on Caco-2 Cells

Caco 2 cells provide a model for the gut absorption. The permeability of the FeEDTA complex standard was tested on Caco 2 Cells. The data in table 2.3 indicates that the FeEDTA complex is absorbed intact through a Caco-II cell monolayer

Figure 2.4 A Selected ion chromatogram of a 0.5 ng/mL [FeY]⁻ standard solution (20 μL injection). **B** Reagent blank EDTA in water.



2.3.3 Iron Extraction Recovery from Snail Tissues

Aliquots of Fe(III) phosphate solutions equivalent to 10 ng mL⁻¹, 20 ng mL⁻¹, 35 ng mL⁻¹, 175 ng mL⁻¹ of FeEDTA were added to 1 ml aliquots of solution containing homogenised snail tissue and EDTA was added into the solution which was heated at 40 °C and sonicated for 1 hour. As indicated in Table 2.4 the recoveries were around 96%. This demonstrated that the free Fe (III) could be recovered from the snail tissues as its FeEDTA complex reproducibly.

Table 2.3 The permeability of FeEDTA complex through Caco-II cell monolayers, the complex solution was dissolved in PBS at pH=7.4 and applied to the surface of the monolayer. (n=3)

Feed solution level (ng mL ⁻¹)	FeEDTA (ng mL ⁻¹)
3.44X10 ⁷	430
1.72X10 ⁷	210
3.44X10 ⁶	130

Table 2.4 Extraction recovery studies for Fe (III). Fe (III) was spiked into snail tissue, EDTA was then added and the solution was filtered and then analysed, in comparison with an unextracted FeEDTA standard at the same concentration. (n=2)

Nominal Concentration of FeEDTA	Recovery of FeEDTA
ng mL ⁻¹	
10	96.1%
20	96.2%
35	96.0%
175	96.1%
Mean	96.1%
R.S.D.	0.08%

Table 2.5 Statistical evaluation of the results on feeding FeEDTA pellets to *H.aspersa* (snails stayed in the tank together with the pellets) in comparison with *H.aspersa* which were not fed pellets. *p*-Value given for 95% confidence. In this case extraction was carried out with 0.1% formic acid.

No.	Kidney		Heart (ng/organ)	
	Feed	Control	Feed	Control
1	34786	6	4365	0
2	66702	394	3654	0
3	9381	151	5623	0
4	23596	234	2168	0
5	4587	1132	1260	0
<i>p</i>-Value		0.035		0.002

2.3.4 Iron and FeEDTA absorption in Snails

The snails were fed on lettuce and carrots until they were removed to a separate tank for feeding with pellets. They were either fed with FePO₄ (1% w/w) pellets or the commercial pellets, which contained FePO₄ (1% w/w) according to the label claim along with a quantity of EDTA which was not specified on the label. Five snails

at a time were transferred to a separate tank and commercial FeEDTA or FePO₄ only pellets were introduced into the tank for feeding. After 3 days the snails were removed from the tanks and dissected immediately to acquire the hearts, kidneys and dart sacs. Table 2.5 shows that the quantities of FeEDTA in heart and kidney after feeding with the commercial slug pellets are significantly higher than the control group with *p* value below 0.05. In this experiment extraction was carried out with 0.1% formic acid alone thus there did appear to be contamination of the kidneys in the control group with FeEDTA since no FeEDTA had been fed to this group. When dissecting the dead snails, it was observed that debris from the pellets adhered to the body and shells of the snails. Thus it was hard to eliminate the possibility that the FeEDTA might be absorbed via the skin or contaminate the organs and also the instruments during dissection. Although the snails were rinsed with 0.5 litre of deionised water 3 times the sticky pellets were not easily removed. Finally, the snails were fed manually with pellets softened with deionised water until they stopped eating. The snails fed on the FeEDTA pellets were fed 4-5 pellets and died three days after this single feeding. The snails fed on the FePO₄ pellets were fed 3-4 pellets on three occasions over 3 days since they continued to eat the pellets and suffered no apparent ill-effects. In addition the organs were removed from unfed snails, drowned in deionised water as a control. In this experiment the total iron in the tissues was measured by extracting the organs with deionised water containing EDTA. Table 2.6 and figure 2.5 show the iron accumulated in snail organs after manual feeding with FePO₄ pellets (1% w/w), FeEDTA pellets (1 % w/w) and after feeding just with a normal diet.

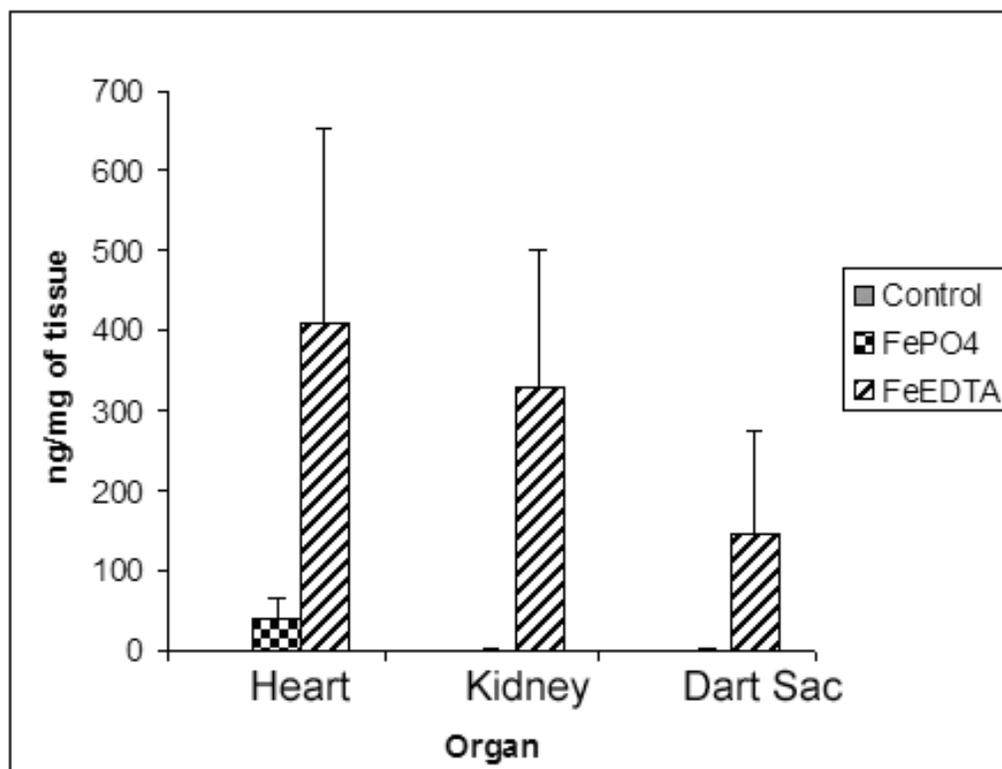
Figure 2.5 Accumulation of FeEDTA in the organs of *H.aspersa*

Table 2.6 Results of FeEDTA distribution in organs from feedings to three groups of snails: control (no pellets fed); FePO₄ (Pellets containing FePO₄ alone); FeEDTA (Pellets containing FePO₄ with EDTA). In this case 0.1% v/v formic acid in water containing 10 µg·mL⁻¹ of EDTA was used to extract all the samples thus determining total iron in the tissues.

Feed Group.*	Heart	Kidney	Dart Sac (ng/mg)
0.117±0.219	0.014±0.021	0.028±0.053	0.117±0.219
41.352±18.080	0.685±0.679	0.982±1.345	41.352±18.080
409.834±196.370	327.131±139.151	144.439±103.975	409.834±196.370

* N=5, $\mu = X \pm S.D.$

Table 2.7 The two way ANOVA with replication (N=5) on the accumulation of Fe(III)/FeEDTA in organs after three feedings to *H.aspersa*

Source of						
Variation	SS	d.f.	MS	F	p-value	F crit
Organs	2091.43	2	1045.71	5.14	0.01	3.26
Feedings	21808.98	2	10904.49	53.6	<0.001	3.26
Interaction	2940.83	4	735.20	2.11	0.06	2.63
Within	7321.04	36	203.36			
Total	34162.28	44				

Negligible amounts of iron were found in the snails which had not been fed with pellets and very large differences in iron accumulation in the organs were observed between the snails fed FePO₄ pellets and those fed FeEDTA pellets. Snails fed with the 1% w/w FePO₄ did not suffer any ill or toxic effects and could be maintained on this diet indefinitely. Thus it would appear that the presence of EDTA in the formulation along with iron (III) is necessary for the absorption of toxic levels of iron. Very high levels of iron were found in hearts and kidney when snails were fed with FeEDTA compared with much lower levels in snails fed with FePO₄, as indicated in Table 2.6. Two Way ANOVA with replication (N=5) was performed on the results. A 95% confidence interval was applied to determine if the variance was statistically significant. From the data in Table 2.7 the type of feeding had a significant effect (with $p < 0.001$) which means the feedings significantly influence the uptake of Fe or FeEDTA. The snails' organs were differently distributed with complex the significance of which is revealed by $p = 0.01$. Interactions between the type of feedings and organs were not found to be significant ($p = 0.06$) which means the effect of feedings is independent of the organs. This indicates that the different feedings do not affect the distribution of iron. In addition, considering the FeEDTA

and FePO_4 feeding only, the former significantly increased the absorption of Fe or FeEDTA in snails.

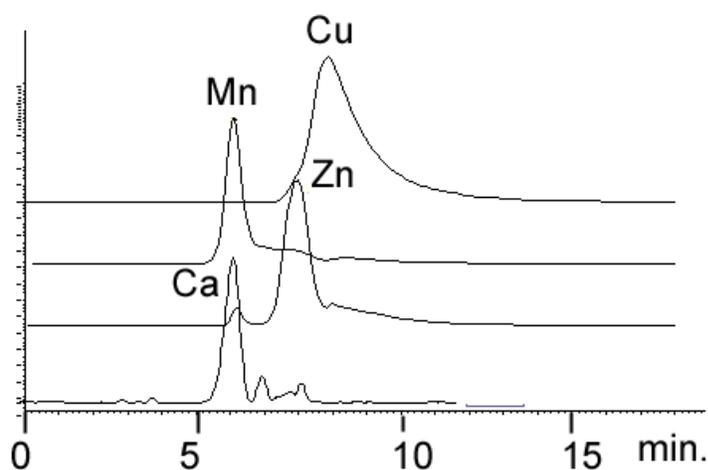
2.3.5 General Discussion

It was found by previous workers that Fe (III) in the form of FeEDTA did not pass through a Caco-2 monolayer any better than Fe (III) as simple iron salts [26]. As seen from the data in table 2.6 the Fe(III) was not evenly distributed between heart, kidney and the reproductive system. The hearts always had the highest levels of iron, with lower amounts in the kidney and the smallest amount in reproductive system. Interestingly, the heart and kidneys are both located in the lung cavity and quite adjacent with each other, while the dart sac is close to the stomach in the body cavity. Accumulation of FeEDTA in the kidney reflects the observations of Zhu and Miller with respect to rat kidney [83]. The exact mechanism of iron toxicity in the snails is not known but given that the natural levels of iron in snail tissue appear to be very low and that snails use copper to carry oxygen in their haemolymph rather than iron, they may be very sensitive to the presence of iron in their tissues. Since the FeEDTA is present at very high concentrations in the tissues it may be that, under physiological conditions, where the stability of the FeEDTA complex could change, essential metals such as copper could become complexed by the EDTA.

It was not possible to use the LC-MS method to distinguish between iron(II) and iron(III), in theory the iron(II) EDTA complex should be doubly charged but there was no evidence of this when the iron(II) EDTA complex was analysed. It may be that iron (II) becomes oxidised to iron (III) during the electrospray process. In addition in theory the FT-MS system should be able to measure a difference in mass of one electron but again it was not possible to confidently distinguish between iron (II) EDTA and iron (III) EDTA on the basis of mass when the two complexes were mixed. In fact it might have been expected that in order to maintain a single negative

charge the iron (II)EDTA complex would have carried an extra proton and thus would have had a m/z value of 345. The LC-MS analysis of other metals was also observed to be possible and the EDTA complexes of Cu^{2+} , Zn^{2+} , Ca^{2+} and Mn^{2+} could be determined by the current method. Figure 2.6 shows extracted ion chromatograms for these complexes which give good peak shapes apart from the Cu^{2+} complex. Thus method could provide a sensitive general method for metal determination although would not compete with ICP-MS for sensitivity.

Figure 2.6 TIC of Cu, Mn, Zn, Ca after complexation with EDTA run ZICHILIC column with 0.1% v/v formic acid in water: CH_3CN (40:60).



2.4 Determination of FeEDDS Absorption by Snails

EDDS is a structural isomer of EDTA and has two chiral centres [95]. Like EDTA, EDDS forms stable hexadentate chelates with transition metals. However, EDDS is readily bio-degradable and became an increasingly widely applied chelator in the environment. The readily biodegradable form is the S,S form which is based on the naturally occurring amino acid L-aspartic acid. Although, compared to EDTA, EDDS has the lower stability constant for chelating metal, with the useful range roughly $3 < \text{pH}_{\text{EDDS}} < 9$ and $2 < \text{pH}_{\text{EDTA}} < 11$, its range is sufficient for most applications [96, 97].

As the metal-based ingested poisons, snails must eat and absorb compounds within their digestive tract in sufficiently high levels to cause a pesticidal effect.

According to Puritch, compared with EDTA, EDDS delivers Fe more freely and disperses it throughout the mollusc body more quickly. This accelerates absorption of the simple metal compound from the mollusc digestive tract into the internal organs. As a result, it causes the rapid, irreversible destruction of the cellular integrity which prevents them from continuing their feeding, eventually leading to death [98].

Most of the work focused on the determination of absorption of FeEDTA by snails, in the mean time, there was also an interest in the determination of iron(III) levels in snails following feeding of pellets containing Fe(III)EDDS complex.

The complex has identical molecular weight and mass spectrum to Fe(III)EDTA although there is a difference in retention time of the ZICHILIC column between the complexes. Essentially the same methodology as was used above was followed. The complex was both able to pass through CACO cells (table 2.8) and caused ion accumulation in the organs of the snail (table 2.9).

Table 2.8 The permeability of EDDS-Fe (III) complex on Caco-II cells, the complex solution was dissolved in PBS with pH=7.4.

Feed solution level	EDDS-Fe (µg/mL)
1.72×10^7	0.52
3.44×10^6	0.27
3.44×10^5	0.10

Table 2.9 Results for FeEDDS distribution in organs resulting from feeding with pellets containing FeEDDS (N=5, $\mu=X \pm S.D$)

(ng/mg)	Heart	Kidney	Dart.Sac
Mean	398.5822	327.2797	85.9813
S.D.	44.6632	61.8410	15.1938
Range	398.58±44.66	327.28±61.84	85.98±15.19

Chapter 3

Some Exercises in Data Manipulation Including A Chemometric Study of Chromatograms of Aqueous Extracts of Tea Leaf, the Development of Data Extraction Software and development of a searchable metabolite data base.

3.1 Introduction

During the early stages of the metabolomic work described later in the thesis there was no completely satisfactory software for handling the data derived from FT-MS analyses. Thus several strategies were tried as a move towards producing in house software. Some progress was made but eventually ThermoElectron offered an improvement in their Sieve software which solved many of the data manipulation problems faced.

3.2 Application of Correlation Optimised Warping Followed by Various Chemometric Methods in the Discrimination Between Different Types of Tea

3.2.1 Introduction

Tea is an infusion of the leaves of *Camellia sinensis*; different types of tea are produced according to the treatment of the leaves after they have been picked. Tea contains a variety of constituents including gallic acid, epigallocatechin gallate, epigallocatechin, catechin, rutin, caffeine, theobromine and phenylalanine. A number of methods have been used to profile the components of tea including: gas chromatographic analysis of volatile fractions of green tea [99]; capillary electrophoresis to characterise amino acids in tea leaves [100]; near infra-red spectrophotometry to identify green, black and Oolong tea [101]; nuclear magnetic resonance (NMR) to profile black tea which was then correlated to organoleptic

properties [102] and HPLC which has been used by a number of research groups to profile the different chemical components of tea [103-107].

Comparison of chromatographic fingerprints by chemometric approaches is regarded as a powerful tool since it helps to highlight chemically relevant information and patterns present within the data [108]. Prior to chemometric processing of chromatograms, data pre-processing must be considered since it not only eliminates signal noise, which would otherwise impede analysis, but can also be used to align peaks to eliminate retention time drift from run to run. Peak shifting is a common occurrence in LC due to minor changes in mobile phase composition, temperature and flow variation. Nielsen *et al.* [109] developed correlation optimization warping (COW) for use in their chromatographic analysis of fungal extracts which proved to be an effective method in the current study when dealing with retention time shifts. Robust alignment is required to improve the quality of chromatography data but also in order to determine which type of data analysis should be considered [110]. The study of alignment algorithms has increasingly become a separate subject in this field [111-113].

Principle component analysis (PCA) is used to discriminate between sample categories by reducing the dimensions of the variables in order to simplify subsequent analysis. Loading plots show which areas of the chromatograms are significant and contribute to the largest variance in the data and are thus useful for sample classification. Support vector machines (SVMs) and random forest (RF) classifications are relatively new pattern recognition and regression methods. They are not comparable to traditional chemometrics, which has a full theoretical foundation in statistics [114]. However, their advantages include good predictive capability [115-120] and the balancing all variables in case of overfitting [121]. Unlike other classical chemometrics methods, SVMs and RF fix the classification

decision on structure risk minimization (SRM) rather than empirical risk minimization (ERM). SVMs solve classification problems by finding a hyperplane with a maximal margin and uses support vectors to represent it. When a dataset is noisy and has cross-linked interactions, SVMs projects the dataset to a higher dimension and transforms it so that it is linearly separable. Recently, SVMs has been used in phytochemical, protein and cancer classifications [118, 122, 123]. Wu *et al.* studied image textures for sorting of tea categories with SVMs [124]. RF achieves a classification by constructing a series of decision trees [115] and takes input variables down all trees in order to optimise classification. RF places no limit on the dimensionality of the data. The so called 'out-of-bag' error estimate can enable classification to be cross-validated internally and thus increase the accuracy of prediction and resistance to noise [125-128]. In 2008, Grimm *et al.* applied RF to the mapping of organic carbon concentration in soil [129]. Perdiguero-Alonso *et al* used RF to differentiate fish populations using parasites as biological tags [130]. These methodologies have not yet been explored in the field of tea profiling by liquid chromatography, and the analysis of teas represents a useful challenge since their chromatographic profiles are very similar. Such methods could for example have wider applications in the quality control of herbal medicines.

3.2.1.1 Alignment

COW aligns chromatograms by means of sectional linear stretching and compression, which shifts the peaks of one profile to better correlate with that of a reference profile. The profile to align, P , and the reference profile, T , are first divided into a number of subsections, N . Each section, i , in P is warped to meet the i th section in the reference profile (T) sequentially: the i th section is stretched or compressed by shifting the position of the end-point by a limited number of lengths, s , called slack parameters. Each section end-point can be shifted by $-s$ to $+s$ points.

For example, if s equals 1, three possible shifts are possible: -1 (compressed by 1), 0, or +1 (stretched by 1). If stretched, the length of the section is linearly interpolated to the gap in P . Once one a section has been aligned, the next section in P and T will be treated using the same approach described above. The end point of the previous section becomes the new start point of the second section in P . Thus the end point of the second section in P is again shifted from $-s$ to s , and interpolated to meet the length in T . The selection of the s parameter is usually guided by observed peak shifts and by the section length. The slack parameter, s , should be a large enough value to ensure satisfactory flexibility when warping and the number of data points in a section minus s should be larger than 1. To measure the warping effect, an objective function, F , is constructed as a cumulative sum of the correlation coefficients of all sections. Only the highest value of the correlation coefficient is stored for each section. A more detailed explanation of COW can be found in a number of references [109, 131, 132]. In the current study, COW was implemented with MATLAB 7.0 (Mathworks, Natick, MA, USA) using a core code supplied by Giorgio Tomasi [133]. Our changed version is shown in Appendix 1.

3.2.1.2 Support Vector Machines (SVMs)

SVMs are a set of supervised learning methods used for regression and prediction; it is a linear classifier. The original algorithm was proposed by Vapnik and Chervonenkis in the Vapnik-Chervonenkis theory published in 1963 [118, 119]. Noisy and highly dimensional datasets present a challenge to SVMs. In 1992, Boser, Guyon and Vapnik suggested a way to create non-linear classifiers by applying the kernel to maximum-margin hyperplanes [134]. In this theory, every dot product is replaced by a non-linear kernel function in the transformed feature space. Here, a basic description of SVMs is presented, and the technique has been described in more detail elsewhere [135, 136].

The classic SVMs deal with two-class problems, in which the data are separated with a hyperplane defined by a number of support vectors. There are many hyperplanes that can classify the data, and the best effort is in finding the maximum separation (margin) between the two classes. Take a two dimensional situation for example: the action of SVMs is shown in Figure 3.1. The classes of data are given symbols - circles (class A) and squares (class B). First, the SVMs set up an appropriate hyperplane so that the distance between the space margin boundary and the data point is maximal. Then the boundary is placed in the middle of the margin, the nearest data points which are laid on the margin are used to define the classifier, called support vectors (SVs). Once the SVs are determined, SV will represent all the necessary information of that classifier.

In more detail, suppose a class problem with l training samples, $\{(x_i, y_i) | i=1, 2, 3, \dots, l\}$, where $x \in R^n$, a n -dimensional vector, $y \in \{-1, +1\}$, is a class ID. Thus, the boundary is expressed as:

$$\{(x_i, y_i) | y = \omega \cdot x + \beta, \omega \in R^n, \beta \in R\} \quad (1)$$

Considering the classification, the equation for A and B can be as:

$$\omega \cdot x + \beta \leq -1, \text{ class B (Square)} \quad (2)$$

$$\omega \cdot x + \beta \geq +1, \text{ class A (Circle)} \quad (3)$$

Clearly, SVs correspond to the extreme of the data for the class. In that the equation is equal to +1 or -1. The function of sign can represent all data which lies between -1 and +1; the decision function is used to classify any point in either class A and B,

$$F(x) = \text{sign}((\omega \cdot x) + \beta) \quad (4)$$

Considering equation (2) and (3), the margin width is calculated as $\frac{2}{|\omega|}$. The broader margin will give out a smaller $|\omega|$. SVMs is trying to find an optimal

hyperplane to achieve the maximal margin, so the minimum $\|\omega\|$ will be represented as:

$$\Delta(\omega) = \min \{ \frac{1}{2} \omega^T \omega \} \text{ or } \Delta(\omega) = \min \{ \frac{1}{2} \|\omega\|^2 \} \quad (5)$$

$$\text{Subject to } y_i (\omega \cdot X + \beta) \geq 1 \quad (y_i = -1 \text{ or } +1)$$

Figure 3.1 SVMs Classification. The solid line and dashed line denotes the hyperplane and margins. Squares and circles denote the negative and the positive training samples. The green arrows in margin denote the support vectors.

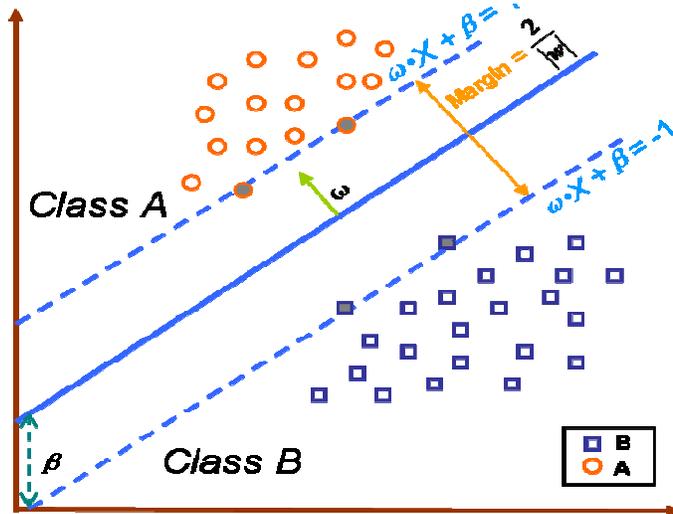
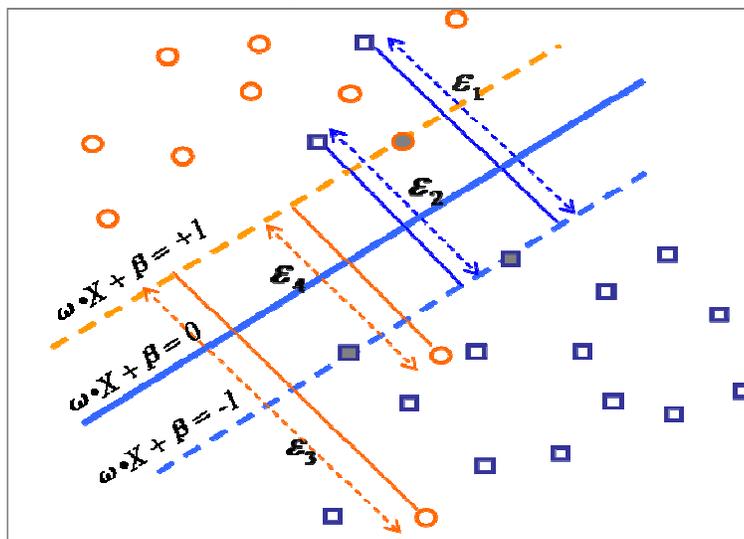


Figure 3.2 Application of the slack parameter. The separated points start to merge. Some blue squares go to red and red circles go to blue. ξ_i is the slack parameter.



Equation (5) is a common calculation for simple linear dataset. As the square (or circle) point gradually merged to the opposite side of class, as seen in Figure 3.2, the previous calculation results in misclassification and fault prediction. In 1995, Cortes and Vapnik proposed a soft margin idea by introducing a slack variable, ξ_i , $i = 1, 2, \dots, l$ which measures the degree of misclassification of data. The new formula incorporating slack variables will be:

$$\Delta(\omega) = \min_{\omega, \beta, \xi} \left[\frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \xi_i \right] \quad (6)$$

$$y_i (\omega \cdot x_i + \beta) \geq 1 - \xi_i \quad \text{and} \quad \xi_i \geq 0 \quad \text{for all } i$$

Where C is called the penalty coefficient, which is used to control the tradeoff between minimizing the training error and model complexity. The constraint, new ω can be optimised as:

$$\omega = \sum y_i \alpha_i x_i \quad (7)$$

Where, α_i is called the Lagrange multiplier. Apply equation (7) to (4); the decision function will be as:

$$f(x) = \text{sign} \left(\sum_{i,j} \alpha_i x_i y_i \alpha_j x_j y_j + \beta \right) \quad (8)$$

Where, $j = 1, 2 \dots l$. As the dataset complexity continue the linear boundary in input spaces might not be enough to separate them properly, shown in Figure 3.3a and 3.3b. It is thought whether to create a hyperplane that allows linear separation in the higher dimension (had to use curved surface in lower-dimensional input space). In SVMs, It is solved by a transformation function $\Phi(x)$ that converts the data from an input space to feature space:

$$p = \Phi(x) \quad (9)$$

Fig. 3.3c shows the transformation from the input space to the feature space. The non-linear boundary has been transformed into a linear boundary in feature space. New classification appears.

A kernel function is used to perform this transformation. Two advantages of a kernel include reducing the computation load and retaining the effect of higher-dimensional transformation. The kernel function $K(x_i, x_j)$ is defined as follows:

$$K(x_i, y_i) = \Phi(x_i) \cdot \Phi(x_j) \quad (10)$$

Where, Φ is a function to project the data into feature spaces. The more popular kernel functions now are the radial basic function (RBF), polynomial, sigmoid kernel function, etc., as follows:

$$\text{RBF} \quad K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (11)$$

$$\text{Polynomial} \quad K(x_i, x_j) = (1 + x_i x_j)^\sigma \quad (12)$$

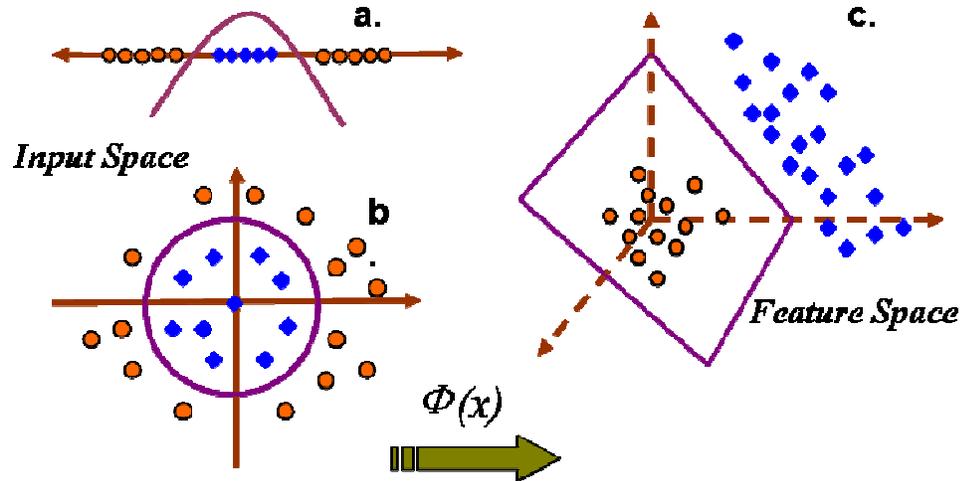
$$\text{Sigmoid} \quad K(x_i, x_j) = \tanh(\alpha x_i x_j + \nu)^\sigma \quad (13)$$

Where, parameter σ of the kernel that defines implicitly the non-linear mapping from input space to feature space. Finally, consider the use of a kernel to substitute for input space functions, The SVMs will be changed as equation 14. Then apply equations (11) (12) or (13) to (8).

$$f(x) = \text{sign}\left(\sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, y_j) + \beta\right) \quad (14)$$

Where $K(x_i, y_j)$ can be RBF, polynomial or sigmoid kernels. SVMs in this study were implemented in STATISTICA 7.0.

Figure 3.3 The Transformation of input to a higher dimension by SVMs. a. Non-linearly separable variables in one dimension, b. Non-linearly separable variables in two-dimensions c. Projection to feature space and classification.



3.2.1.4 Random Forest (RF)

RF is another advanced method of machine learning. The classification is achieved by constructing an ensemble of randomised classification and regression trees (CART) [126]. For a given training dataset, $A = \{(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)\}$, where $(x_i, = 1, 2, \dots, n, \text{ is a variable or vector and } y_i \text{ is its corresponding property or class label})$, the basic RF algorithm is presented as follows:

Each training set is drawn with replacement from the original dataset A . Bootstrapping allows replacement, so that some of samples will be repeated in the sample, while others will be “left out” of the sample. The “left out” samples constitute the “Out-of-bag (OOB)” which has, for example, one-third, of samples in A which are used later to get a running unbiased estimate of the classification error as trees are added to the forest and variable importance.

For each bootstrap sample, a tree is grown. m variables (m_{try}) are selected at random from all n variables ($m_{try} \leq n$) and the best split of all m_{try} is used at each node. Each tree is grown to the largest extent (until no further splitting is possible)

and no pruning of the trees occurs. Each tree is constructed on the bootstrap sample. The OOB samples are not used and therefore regarded as a test set to provide an unbiased estimate of the prediction accuracy. Each OOB sample is put down the constructed trees to get a classification. A test set classification is formed. At the end of the run, take k to be the class which got most of the “votes” every time sample n was OOB. The proportion of times that k is not the true class of n averaged over all samples is the OOB error estimate.

RF has the ability to rank the variable importance. For each tree grown in the forest, put down the OOB and count the number of votes cast for the correct class. Permute the value of variable m in OOB randomly and put these samples down the tree. Count the number of votes for the correct class in the variable- m -permuted OOB data. Again count the number of votes for the correct class in the untouched OOB data. Subtracting the two counts and averaging this number over all trees in the forest is the raw importance score for variable m . Finally importance score will be computed depending on the correlations between trees.

RF has several advantages over other statistical modelling methods [126]. Its variables can be both continuous and categorical. As a large number of trees are induced and averaged during the run, RF can produce the low bias and low variation results but highly accurate classification and good prediction. Since RF can make OOB error estimates which test the classifications by vote on a small number of samples this further strengthens the model. Additionally, RF has no apparent limits on the dataset complexity and therefore can handle much higher dimension datasets than other learning methods. However, the entire computation algorithm is over-complicated and hard to interpret and it is often called a “Black Box” approach [137]. The interaction between response and predictor cannot be monitored and thus the mechanism of classification cannot be readily explained.

3.2.2 Materials and methods

3.2.2.1 Chemicals and Samples

HPLC grade methanol and water were obtained from VWR (Lutterworth, UK). Tea bags were purchased locally. AnalaR grade formic acid (98%) was obtained from BDH-Merck (Dorset, U.K.).

3.2.2.2 Instrumentation

Six commercial teas Assam, Ceylon, Darjeeling, English Breakfast, Green and Decaffeinated tea were examined. One tea bag was extracted with 100 mL of boiling water and allowed to stand for an hour. This was repeated with five tea bags for each tea sample. A volume of the sample was filtered and then 0.8 mL of sample was mixed with 0.2 mL of methanol. The tea sample was run on a Kromasil phenyl column: 150 x 4.6 x 5 μm particle size (Hichrom, Reading U.K.). The mobile phase used was A: 0.1% v/v formic acid in water and B: methanol. The gradient programme was: 0-10 min 10% B; 45 min 25% B; 60 min 80% B. The columns were re-equilibrated for 10 minutes between runs. The flow rate was set at 0.4 mL/min and detection was carried out at 280 nm. A ThermoSeparations P2000 pump was used and the data was acquired using Chromquest software. Chromatographic data were exported as ASCII files producing 36000 data points. The data sets were reduced by profile binning every 60 data points. Normalisation is by dividing each of intensity by sum of all intensities. 2-Point Baseline correction is by plotting a connecting line between beginning and end points, then subtracting it from the trace. These operations were achieved by writing a macro for Microsoft Excel to automate them.

LC-MS data were acquired using a LTQ Orbitrap instrument (Thermo Fisher Scientific, Hemel Hempstead, U.K.) set at 30000 resolutions. Sample analysis was carried out under negative ion mode. The mass scanning range was m/z 50-1500

and the capillary temperature was 250°C. The HPLC conditions used were the same as those used for HPLC UV. The mass axis of the instrument was externally calibrated according to the manufacturer's instructions just before commencing the experiment,

3.2.3 Results and discussion

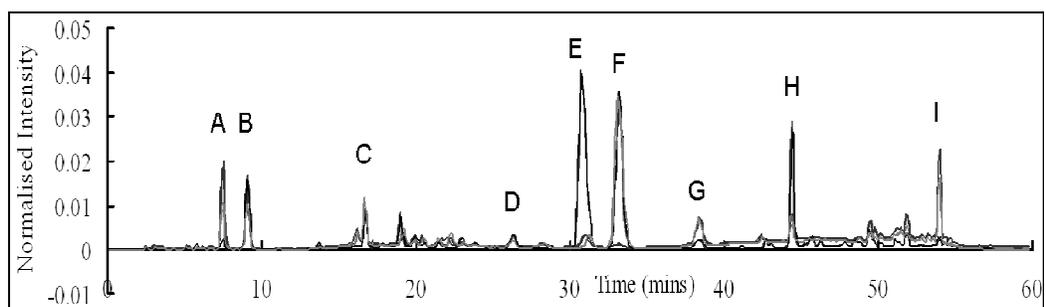
3.2.3.1 Chromatography

The LC-FTMS analysis allowed tentative identification of several compounds according to their elemental composition including: gallic acid (A); galloyl quinic acid (B); catechin (C); epigallocatechin (D); caffeine (E); *p*-coumaryl quinic acid (F); epicatechin-gallate (G) rutin (H); kaempferol-3-rutinoside (I). Elemental matches were all within 5 ppm of the predicted elemental composition.

A representative HPLC chromatogram is shown in figure 3.4. The chromatographic region between 5-57 minutes was modelled because it contained almost all of the chromatographic peaks.

COW was tested on the tea samples and various values were tried for *N* and *s*. The optimum values of *N* and *s* were found to be 40 (length=15) and 5 respectively, in order to obtain the best warping. Figure 3.5 shows the chromatograms of the 30 tea samples before and after alignment. The largest peak shifts were in the middle of the chromatograms, for example, the *p*-coumaryl quinic acid peak at *ca.* 33

Figure 3.4 A representative chromatogram of aligned tea extracts. Gallic acid (A); galloyl quinic acid (B); catechin (C); epigallocatechin (D); caffeine (E); *p*-coumaryl quinic acid (F) , epicatechin-gallate (G) rutin (H); kaempferol-3-rutinoside (I).



minutes. Figure 6 shows a detailed view of the alignment carried out for this peak. The consequence of alignment was further evaluated as shown in Figure 3.7 in terms of peak height, the width at base, the width at 5% of peak height and asymmetry factor. The low variations for these parameters indicated the peak shape and peak areas did not undergo distortion and therefore peak features were preserved.

Figure 3.5 The effect of chromatogram alignment by COW: a. Before. b: After

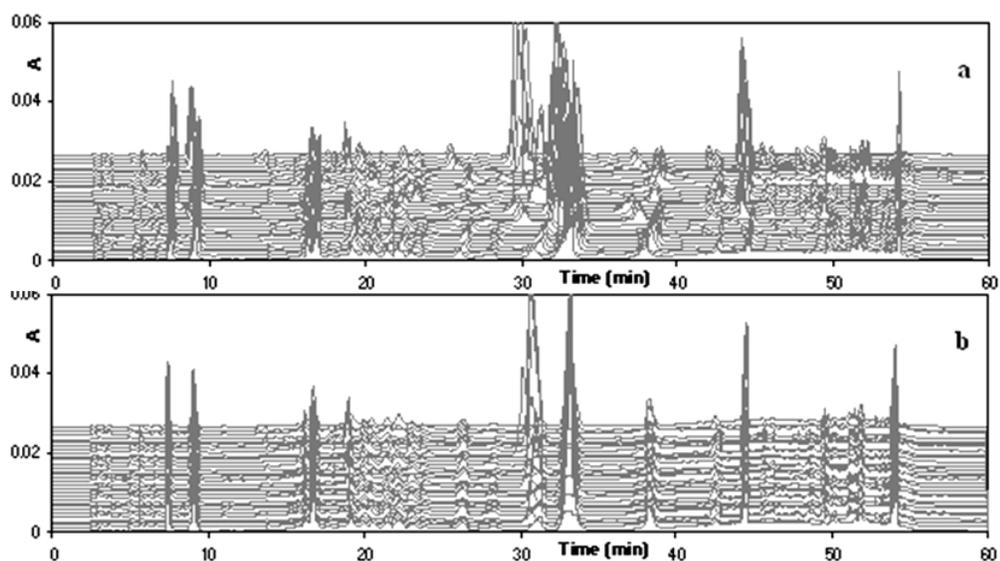


Figure 3.6 A detailed view of 4-*p*-coumaryl quinic acid peak alignment a.unaligned peaks. b. The peaks after alignment.

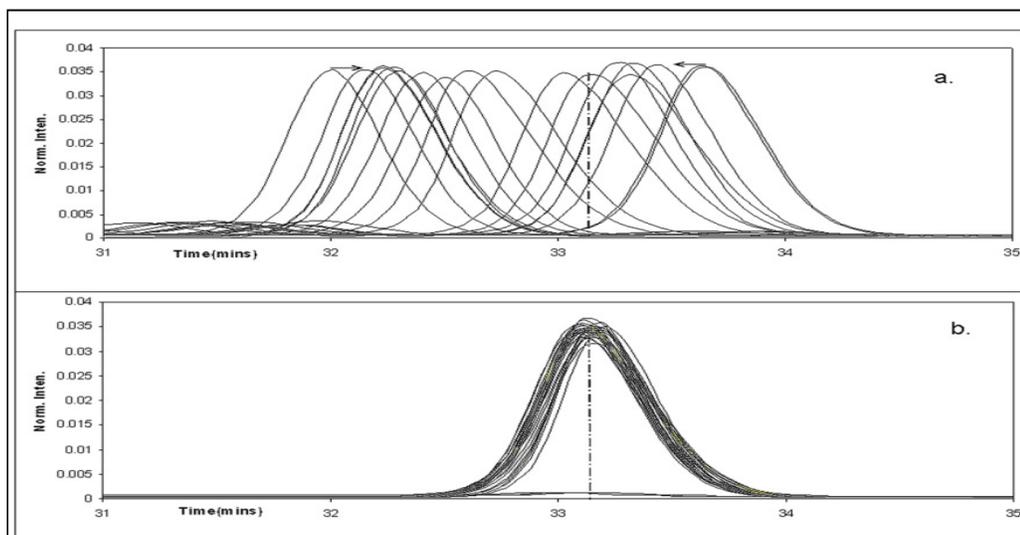
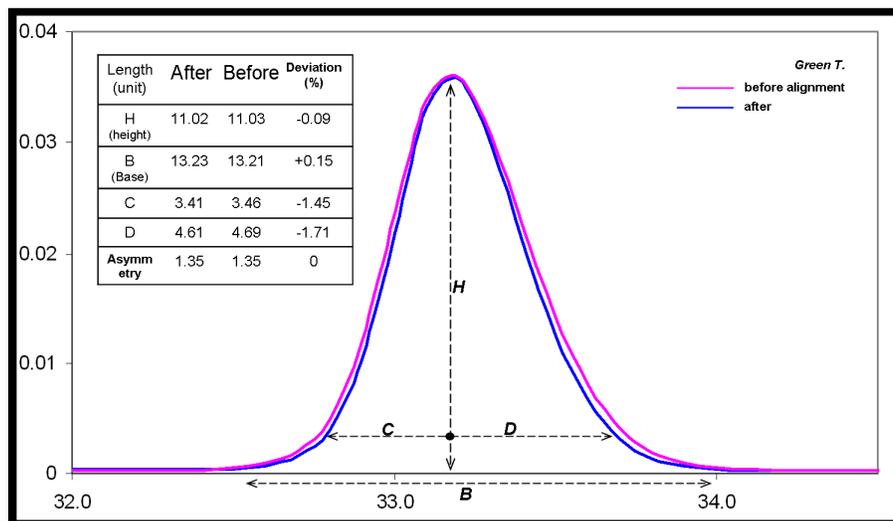


Figure 3.7 A plot of peak shape and area after alignment. H, peak height; B, peak base; C, the left part of peak width at 5% peak height; D, the right part of peak width at 5% peak height. The measurements are shown in the inset table.



3.2.3.2 PCA

All datasets were imported to SIMCA P and used to make a PCA model. A larger value for the cumulative variance (78% up to PC3 and consistent for all components) was found after alignment, as shown in Figure 3.8, which highlights the good fitting of the COW model and the benefits of alignment. Figure 3.9 shows the effect of signal treatment on the score scatter plot. The score plots obtained before processing were very different to those obtained afterwards. This difference was due to reduction of the chromatographic time shifts, instrument noise/baseline drift and variations in sample extraction. It can be observed that discrimination between the six types of tea was following COW was improved. Before warping, discrimination of the Darjeeling, Assam and English Breakfast teas was not possible. After warping, the score plot revealed clear groupings despite the samples having similar chromatograms. The decaffeinated tea and green tea are well separated from the rest of the teas. Ceylon tea is in the middle, quite close to the English Breakfast variety. A PCA model based on all observations cannot be used

to classify individual tea types. Instead, separate PCA models (one for each class) must be constructed. To determine the class limits, the significance levels for the distance to model (DModX) parameter can be tweaked. Having a significance level of 0.05 will capture 95% of all objects that belong to a class, whereas 0.01 will capture 99%. However, the lower the significance level, the more prone the model will be to positively classify objects that do not belong to the class. The PCA score plot however does not provide a model in which samples can be classified on the basis of the probability of them belonging to a particular class. A simple way of modeling PCA data is via the use of Cooman's plot which sets a critical value of 0.05 or 0.01 below which misclassification is possible. Table 3.1 shows the predictions of tea class according to PCA modelling. Ceylon, Darjeeling, Assam, decaffeinated and green teas are predicted correctly. English Breakfast tea is predicted by the model correctly when the critical value is set e.g. 0.01 while still having a good distance to the other teas.

The loadings plot explains which compounds discriminate between the teas. As can be seen from Figure 3.10, the right-hand-side cluster features variables which correspond to the composition of the chemicals abundant in the decaffeinated tea. These are mainly three compounds A, G, I. They are: gallic acid, epicatechin-3-gallate and kaempferol-3-rutinoside. Important discriminatory peaks in green tea include galloyl quinic acid, caffeine and rutin. Assam tea had catechin as an important discriminating compound. There are no obvious feature peaks observed for the other teas, which means they have an average amount of marker compounds.

3.2.3.3 SVM Classification

SVMs was used to identify the six categories of tea by using Statistica to produce the model. Although SVMs usually solves two-class problems, it can reconstruct and

transform multi-class datasets to a two-class model [138]. In this study of tea, 24 out of 30 samples were randomly selected as the training set, the rest were used as the test set. As mentioned earlier, various kernel functions can be used in the SVMs equations. In order to obtain the best model accuracy, three kernels were used with Statistica. Their parameters including cross-validation were used to improve performance. Table 3.2 shows the descriptive statistics for the three kernels. The

Figure 3.8 The cumulative explained % Variance vs. PCs; Curves: ▲(After alignment) ●(Before alignment).

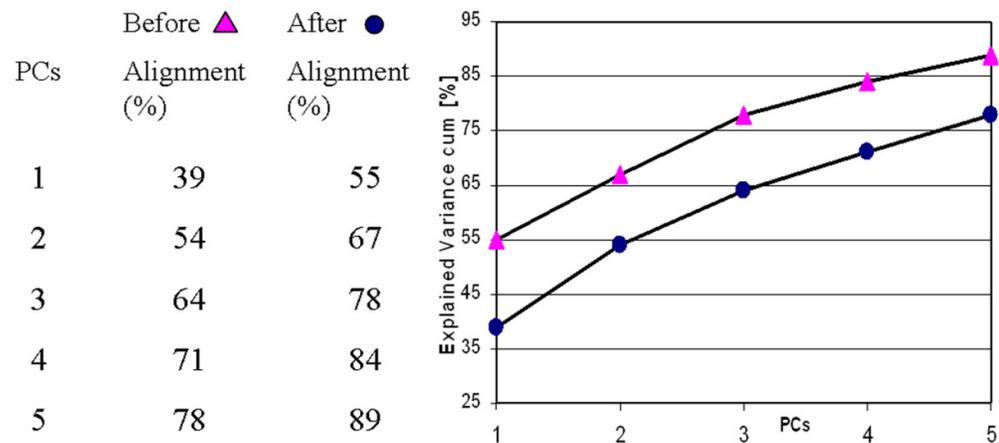
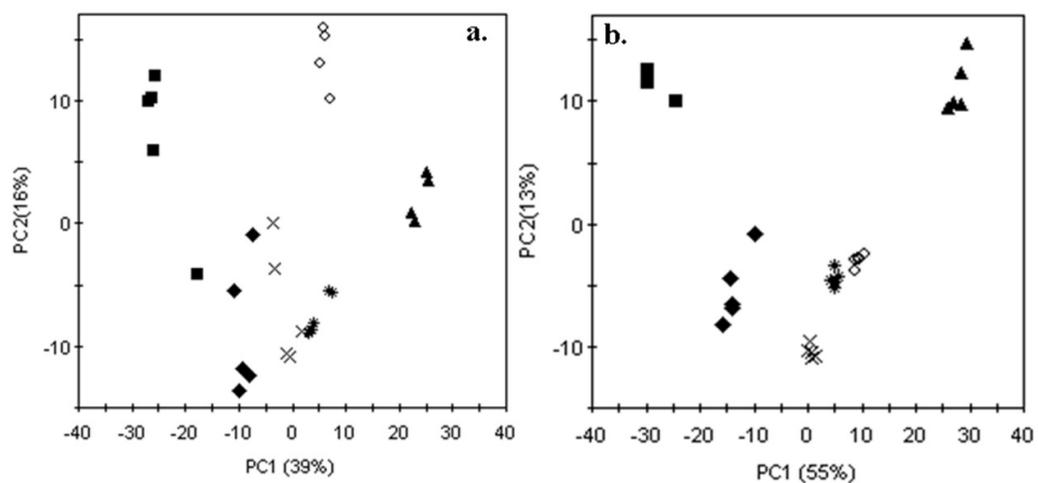


Figure 3.9 The PCA Score Plots (PC1-PC2); a. Before alignment. b. After alignment ▲(Decaffeinated), ■(Green), ◆(Darjeeling), ◇(Ceylon), *(English Breakfast), x(Assam).



polynomial kernel possessed a cross validation accuracy of 16.67 %, while the RBF had an accuracy of 100%. The RBF used 23 support vectors to model the six tea classes while the polynomial, and sigmoid kernels used 24. For the classification and prediction accuracy, the RBF got 100% correct. Thus RBF was not only the best classifier but also the most effective one, which used least number of support vectors and achieved the best results. Similar trends also could be found in the individual prediction results. Table 3.3 shows the prediction results on the test samples. For example in the classification of Assam, polynomial gave the lowest value of 33% and RBF got the highest of 60%, the sigmoid kernel yielded a probability of 42 %. RBF thus exhibited the best predictive ability. However, it was observed that when the polynomial and sigmoid kernels were used to predict English Breakfast and Ceylon tea, that the probability had decreased to 19% and 25% respectively. This might be explained in two ways. Ceylon and English Breakfast tea are very close according to the PCA score plots and therefore their chromatograms are very similar. In addition, the polynomial and sigmoid kernels are less useful for classifying the groups if the number of variables is large. There were 24 samples in the training set, and 23-24 SVs were used in the model. This model is not persuasive and indicated that SVMs could not reduce the number of variables if the dimensions were too high. Thus, PCA was considered as a method for reducing the variables. Six PCs were extracted containing 98.8% of the variation. This new dataset was used in SVMs to do the analysis again. Table 3.3b shows that seven SVs were created in the new model and RBF still classified correctly. After extraction of the PCs, Assam got the lowest value of 27% by polynomial and the highest of 66% by RBF. Ceylon and English Breakfast tea got 19% and 13% by polynomial, 52% and 50% by RBF. The decreased prediction might be explained by the fact that Ceylon and English Breakfast tea are close together in the PCA model.

By using six PCs, the number of SVs was decreased to 7 from 23, while the results did not undergo too much change.

Table 3.1 Classification of teas according to Cooman's plot (D-Crit0.05)

	M1.PModX	M2.PModX	M3.PModX	M4.PModX	M5.PModX	M6.PModX
Sample	PS	PS	PS	PS	PS	PS
	+ [2]	+ [2]	+ [2]	+ [2]	+ [2]	+ [2]
	Assam	English Br. Tea	Ceylon	Darjeeling	Decaff.	Green
Green_T	0	0	0	0.002	0	0.112
English Br._T	0.002	0.043	0.001	0.002	0	0
Ceylon_T	0.001	0.001	0.084	0.002	0	0
Darjeeling _T	0	0.001	0.001	0.100	0	0
Decaff._T	0	0	0	0.001	0.091	0
Assam_T	0.176	0.002	0.001	0.001	0	0

Figure 3.1 PCA Loadings plot (PC1:PC2) with location of the compounds A-H, for their names refer to figure 4 legend.

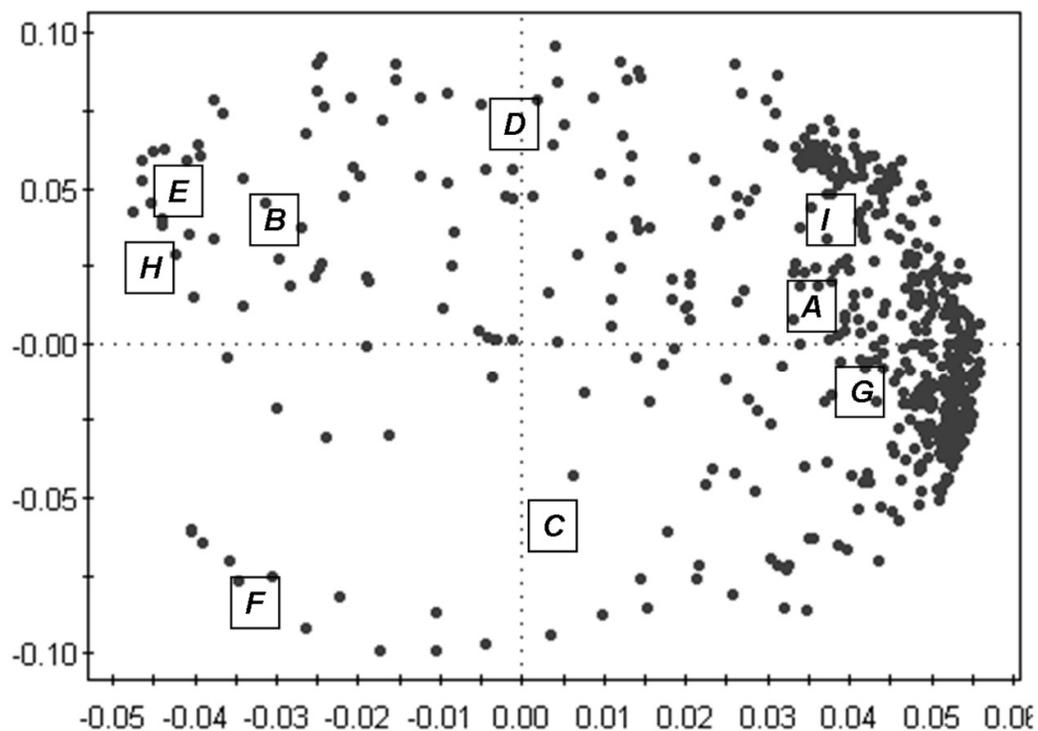


Table 3.2 Overall classifications for the tea samples according to SVMs with three different kernel functions using full variables or 6 PCs.

	No. of SVs	Cross-validation accuracy [%]	Class-accuracy (Training)	Class-accuracy (Test)
a. Full variables				
Polynomial	24	16.67	88.12	83.33
RBF	23	100	100	100
Sigmoid	24	83.33	95.13	83.33

b. Six PCs				
Polynomial	7	33.33	88.12	66.67
RBF	7	100	100	100
Sigmoid	7	66.67	88.12	83.33

Table 3.3 The probability of classification of tea samples using three different kernels with full variables or with 6 PCs.

	Assam [%]	Darjeeling [%]	Decaffeinated [%]	English Breakfast [%]	Ceylon [%]	Green [%]
a. Full variables						
Polynomial	33	33	33	19	19	33
RBF	60	60	60	60	60	60
Sigmoid	42	42	42	25	31	42
b. Six PCs						
Polynomial	27	19	31	13	13	21
RBF	66	66	66	50	52	66
Sigmoid	38	38	33	13	19	27

3.2.3.4 Random Forest (RF)

The tea datasets were analysed by using RF. Statistica provides several parameters to optimise classification and prediction performance. All retention time bins were defined as continuous predictors and the tea identities were defined as categorical predictors. Misclassification costs were set to be equal for all categories. The bootstrap splitting ratio was defined as 0.2 in order to make a constant proportion with SVMs. The number of predictors was important and was optimised by Statistica, this is helpful for large datasets. Statistica uses Breiman's method and ten predictors were set [127]. The maximum number of trees was set to be 150. The

stopping condition parameters were defined as five for the minimum number of child nodes, 100 for the maximum number of nodes and a minimum 5% for the decrease in training estimate error after one iteration. The stopping conditions were used to terminate the iteration of the RF model at the point where no more useful trees were generated. Table 3.4 shows the results of the RF model performance. Green and decaffeinated tea gave 95% probability for correct prediction. This was much better prediction than that given by SVMs. Darjeeling was correctly classified with 92.5% probability. Ceylon and English Breakfast tea were predicted correctly with 85% and 90% accuracy since there was some probability of misclassification of English Breakfast as Ceylon or Assam and Ceylon as Assam. The results indicated that RF gave correct answer for each class, however, the high number of variables and the small size of the samples also pose a question to RF whether or not the model is correct. However, according to Breimann [127, 139] the number of variables used in the RF approach is not critical. The PCA extraction data was used again to study the effect of reducing the number of variables. The results in table 4 show that green tea gets a probability of 62.8% as the highest and Ceylon gets 31.3% as the lowest. The predictions are correct for each class when using six PCs even though the probabilities are lower. There is nothing in the literature to suggest that the number of variables should be restricted in RF models. RF uses variable importance to describe the weight of the different predictors. Table 3.5 shows the top 20 dominating variables; some of these were not identified. Of the identified variables the regions of the peaks for caffeine, rutin, galloyl quinic acid and gallic acid were important. Even the categorical predictor, class name, was regarded as a weighted predictor. However not all of them were helpful in interpreting the RF model. Overall it was confirmed that caffeine, rutin, galloyl quinic acid and gallic acid were very relevant descriptors for differentiating green, Darjeeling and decaffeinated teas.

3.2.4. Conclusion

Chromatography based chemometrics was performed in order to distinguish between different teas. This represented a significant challenge since the differences in the tea chromatograms were small. Signal treatment which included COW, binning and normalisation proved to be effective for removing effects that did not contribute to the classification. This then allowed better classification by PCA. Although PCA gives a classification of the teas it does not give a model where the classification can be assigned a probability. PCA, SVMs and RF were comparatively evaluated as methods for processing the tea chromatograms for classification and prediction of the different tea samples. Cooman's plots derived from the PCA model all teas correctly when a significance threshold of 0.01 was used. In the case of SVMs, RBF using seven support vectors and produced the best accuracy, thus RBF proved to be the best kernel. RF exhibits the better predictive power than SVMs in classifying the tea extracts in the test set. Several components which were identified by MS were found to be important in classifying the teas. These compounds agreed with the loadings observed in the PCA model. In this study, SIMCA-P and Statistica were used. SIMCA was very suitable for multivariate regression with good visual output; while Statistica had much more comprehensive statistical power.

Table 3.4 Probability of classification of different tea samples according to RF modeling using full variables or 6 PCs.

Test set \ Training set	Assam [%]	Darjeeling [%]	Decaff. [%]	English Br. [%]	Ceylon. [%]	Green. [%]
a. Full Variables						
Assam	75	5	0	10	10	0
Darjeel.	5	92.5	0	0	2	0.5
Decaff.	0	0	95	0	5	0
English Br.	5	0	2.5	90	0	2.5
Ceylon	10	0	0	5	85	0
Green	0	0	0	0	5	95
b. Six PCs						
Assam	0.020	0.111	0.050	0.191	0.000	0.628
Darjeel.	0.172	0.290	0.250	0.098	0.090	0.100
Decaff.	0.160	0.230	0.313	0.110	0.090	0.097
English Br.	0.130	0.090	0.060	0.580	0.090	0.050
Ceylon	0.110	0.222	0.133	0.031	0.414	0.090
Green	0.480	0.280	0.172	0.067	0.000	0.000

Table 3.5 Important chromatographic regions for classification of the tea samples according to RF modelling (uk=unknown). * No peaks observed in this region

Variable	Variable Rank	Importance	Identity
472	100	1.00	uk *
537	82	0.82	uk
305	74	0.74	caffeine
311	69	0.69	caffeine
318	68	0.68	caffeine
112	68	0.68	uk
347	67	0.67	uk
284	64	0.64	uk
442	63	0.63	rutin
367	63	0.63	uk
447	62	0.62	rutin
140	62	0.62	uk
2	61	0.61	tea identity
424	59	0.59	uk
92	55	0.55	galloyl quinic acid
181	54	0.54	uk
392	54	0.54	uk
73	54	0.54	gallic acid
225	53	0.53	uk
83	52	0.52	uk

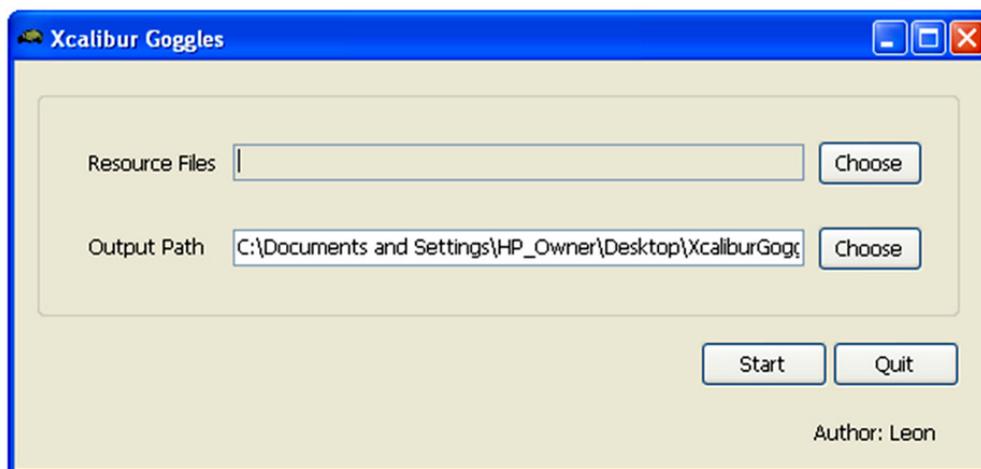
3.3 Development of Pickershell Software for the Extraction of Data from Xcalibur

3.3.1 Xcalibur Goggles

In the initial stages of the metabolomic studies software was written to process Xcalibur data. The improvement in Sieve software meant that development was stopped. However, the software has the potential to offer some advantages over Sieve although further development would be a PhD project on its own.

The small software Xcalibur Goggles was used to export the files from Xcalibur text files. The raw text file from the instrument has many explanatory text words, like: micro scans, scan header, scan number etc. Their presence makes the chromatograms very easy to understand but difficult to process by computer. Xcalibur Goggles was written in open SQL script and helped to efficiently pick up the useful information from the raw files and assemble the whole spectral data in standard format. This allowed Picker Shell to import the standardized files and start the data handling straightforward. Its operation window is shown in figure 3.11.

Figure 3.11 The interface of Xcalibur Goggles



3.3.2 Picker Shell

Picker Shell (PS) is automated software for metabolomic data analysis. It performs differential analysis of sample populations, comparing LC/MS spectra from control and treatment samples to find differentially expressed features. It has an embedded the database to identify the metabolites through masses by accurate measurement. Picker shell works as shown in the flow chart in Figure 3.12 and has two modes: i. Mass spectral mode; ii. LC chromatograms mode.

i. Mass spectral mode: When a LCMS raw file was acquired from LTQ-Orbitrap, the full mass list was exported as text file to represent the whole metabolite profile. PS first starts the scan of full list and sort the m/z values from low to high. For each individual mass profile, PS filters the ions by intensity threshold A (normally 2000) or alternatively by S/N (normally = 3). The ions below the predefined cut-off value will be cleared to produce a new list. PS then starts to calculate each ppm between every two consecutive masses and merge the ions if the mass difference is below a predefined ppm (normally = 3) and make their intensities integrated. The first step includes the above actions, in which a single file was processed and new mass list was generated. In these new files, masses became discrete and background ions are removed.

Next step is to compare two groups of samples (control against treatment). PS puts all files into a three dimensional array: m/z , intensity and file name. It again sorts the m/z ascending by m/z and calculates the ppm. If the ppm is below the predefined value (normally =3), PS will regard the m/z as the identical ions in each file and start to calculate the p-values, ratios and normalise their intensities.

The third step is to give names to metabolites. PS calculate the ppm of the scanned m/z and theoretical m/z . The name will be transferred to ions only if the ppm is below the predefined value C (normally 3 ppm).

ii. LC chromatogram mode:

In this mode, files imported to PS will have retention times (RT). Files cannot be exported from Xcalibur but from Xcalibur goggles. This mode is very similar to mode above except that the additional parameters RT. So, the intra-file and inter-file processing need to consider not only mass but also retention time, which means two criteria need satisfaction. Peak index is used to calculate the distortion of peak shapes, which measures of the peaks to an isosceles triangle. The closer the shape to a triangle, the closer number assigned is to one. On the other hand, if the number is in very low, the chromatographic shape will be more in the baseline or spike form; if the number high, more in a real peak shape. As a result, by using the peak index, the chromatograms will be filtered. However, under LC chromatogram modes, chromatographic peak shifting quite often happened. Thus peak alignment is a necessary way to overcome the problem. The function of peak alignment could be achieved in Matlab as described above in section 3.2. However, this was not implemented in PS since advances in Sieve software led to the abandonment of this line of research.

The Picker Shell program was written in Open SQL script. The program can be easily executed without compiling, decoding and can be easily modified through codes. Their codes are shown in Appendix 2.

3.4 Development of Sieve Extractor a Searchable Database for Use with the tables of Accurate Masses Generated By Sieve 1.2

Sieve Extractor (SE) was created to help automate the identification of metabolites by exact masses. It can easily manipulate the Sieve results in Excel spreadsheets. The software was written as an Excel Macro (Microsoft 2007) and can be easily modified if requirements changed. Two parameters are used in the

Sieve Extractor, “Mass Width” and “R.T.Width”. After the Sieve results in the form of a frames table are copied and pasted into Sieve Extractor. The exact masses obtained from the Sieve analysis are compared with the m/z values in the database.

Figure 3.12 Picker Shell Flow Chart

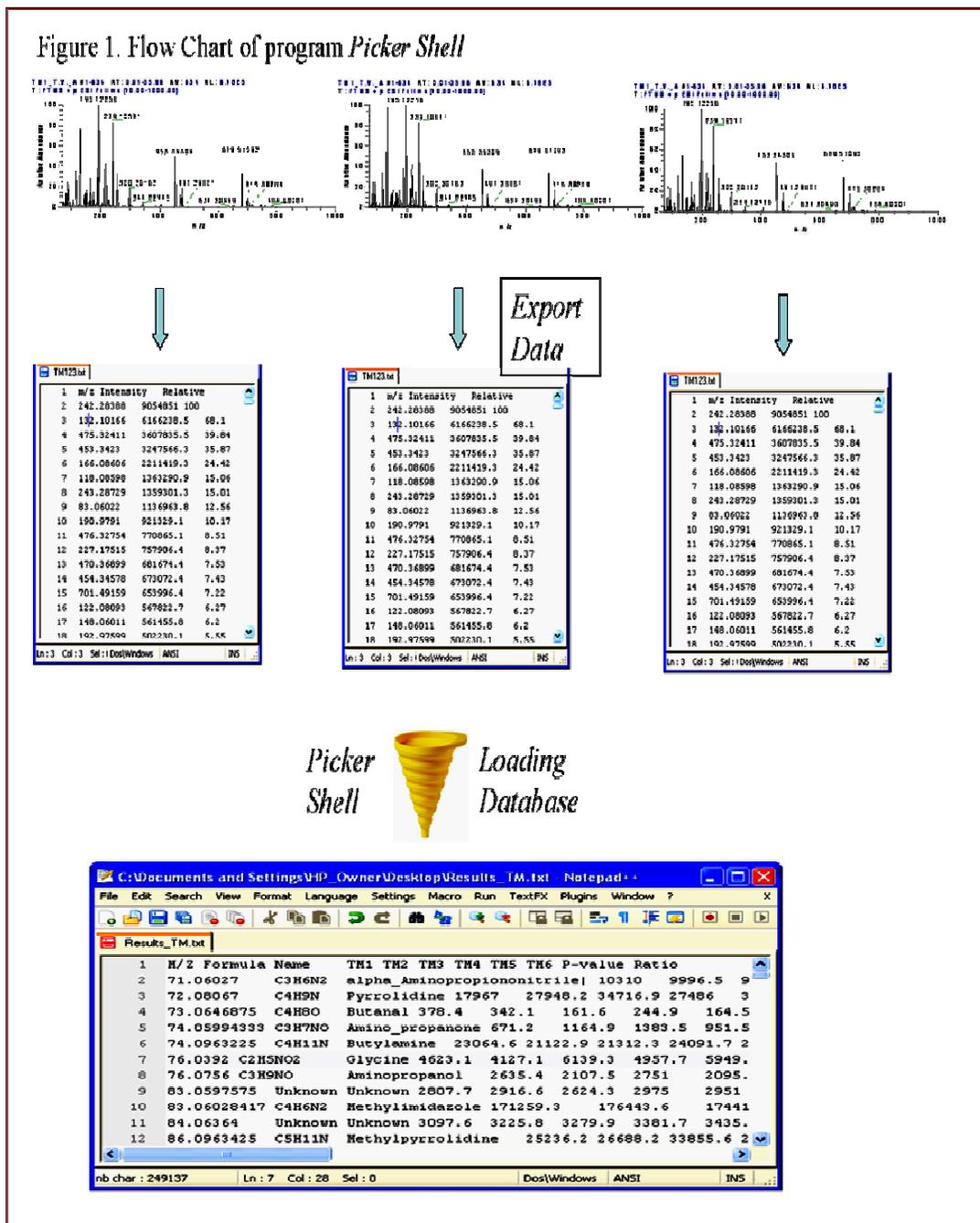
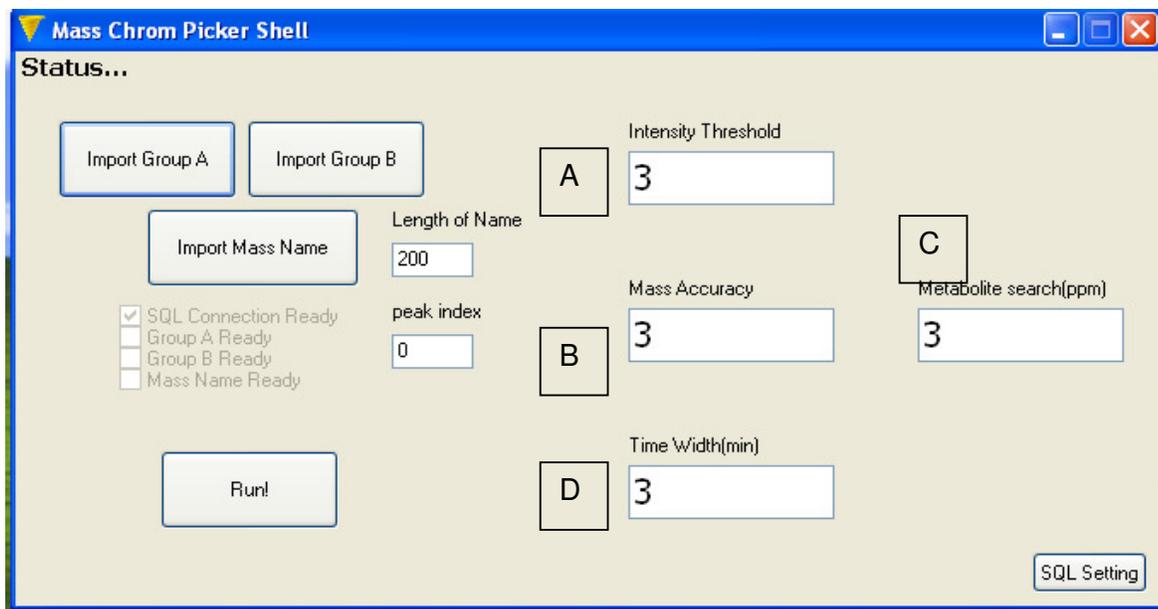


Figure 3.13 The interface of Picker Shell

A window for allowable mass deviation can be set, usually ± 3 ppm and addition a retention time window can be defined for known reference metabolites run under the same chromatographic conditions as the sample. In practise the retention time window is less often used since only a fraction (98 metabolites have been run under standard conditions on ZICHILIC) of the metabolites present in the data bases are available as standards. The software transfers the metabolites' names and the linked pathway names into two two separate columns. Finally, the software returns the number of total hits and the number of metabolites found in each pathway. The current version of the Sieve Extractor 10.0 database was composed of Kegg database (<http://www.genome.jp/kegg/>) and Metlin (<http://metlin.scripps.edu/>) database. The interface of Sieve Extractor is shown in figure 3.14. The Visual Basic coding can be seen in Appendix 3.

Figure 3.14 The Interface of Sieve Extractor.

Microsoft Excel - Sieves_Extractor 8.0 gamma_christian

File Edit View Insert Format Tools Data Window Help

Type a question for help

Go to Office Live | Open | Save

K22

	E	F	G	H	I	J	K	L	M	N	O
1	Pathway	Pathway cod	RT				R.T. Width	20.00	mins		
2	Propn	65	RT.							1 Glutamate metabolism	Glu
3	Propn	77	RT.							2 Purine biosynthesis	Purine
4	Lipid	65	RT.				Mass Width	3.00	ppm	3 Methionine metabolism	Met
5	Mesn	81	RT.							4 Leucine/Valine degradation	Leu/Val(deg)
6	Lipid	65	RT.							4 Leucine/Valine biosynthesis	Leu/Val(syn)
7	AminoPhos_GF	65_70	RT.							5 Histidine metabolism	His
8	btAla_Propn	14_77	RT.							6 Arginine/Proline metabolism	Arg/Pro
9	Butan	33	RT.							7 Tyrosine/Phenylalanine metabolism	Tyr/Phe
10	btAla_Cnamin	14_67	RT.							8 Tryptophan metabolism	Trp
11	Pyruv	32	RT.							9 Pyrimidine metabolism	Pyrimidine
12	Butan	33	RT.							10 Cysteine metabolism	Cys
13	btAla	14	RT.							11 Serine/Glycine/Threonine metabolism	Ser/Gly/Thr
14	Lipid	65	RT.				Total	14	Hits	12 Nicotinamide metabolism	Nicotinamide
15	Fructo_Pyruv_L	32_57_65	RT.							13 Lysine degradation	Lys(deg)
16	Nicotinamide_E	12_77_79	RT.							13 Lysine biosynthesis	Lys(syn)
17	Butan	33	RT.							14 beta-Alanine metabolism	btAla
18	btAla_Urea	14_17	RT.							15 Alanine/Aspartate metabolism	Ala/Asp
19	Porphy	85	RT.							16 Taurine metabolism	Taurine
20	Mesn	81	RT.							17 Urea cycle	Urea
21	Glyoxy_Hex	62_76	RT.							18 Glutathione biosynthesis	GSH
22	Pyruv_Lipid	32_65	RT.							19 Sulphur metabolism	Sul
23	Benz	64	RT.							20 Thiamine metabolism	Thia
24	btAla_Urea	14_17	RT.							21 Riboflavin metabolism	Ribfv
25	Butan	33	RT.							22 Vitamin B6 metabolism	VB6
26	Chloacry	78	RT.							23 Pantothenate	Pantt
27	Glyoxy	76	RT.							24 Folate	Fola
28	Propn_Pyruv_G	29_32_77	RT.							25 Biotin	Biot
29	Propn_btAla	14_77	RT.							26 Ubiquinone	Ubin

Metabolites / Sieves / Hit

Ready

Chapter 4

A metabonomic comparison between two cell-lines under normoxic and chronic hypoxic conditions.

4.1 Introduction

Cell cultures provide useful systems for developing mass spectrometric based methods for metabolomic profiling [140]. While human metabolomics is necessarily observational, cell cultures provide a more controlled system for exploring the impact of drug treatments or environmental changes. The goal of metabolomics is to provide new hypotheses or add further support to existing hypotheses. These hypotheses then can be tested in more detail and the results of further tests may involve further metabolomic analyses, the measurement of targeted markers identified during metabolomic experiments or the deployment of other 'omic' techniques such as proteomics or transcriptomics in support of new hypothesis. Hypoxia is an important factor in the development of solid tumours. The cells at the centre of the tumour are adapted to grow under hypoxic conditions because of the paucity of the blood supply to the tumour. Hypoxic cells are more resistant to radio- and chemotherapy and hypoxic cells are more malignant. Hypoxia inducible factor HIF is central to the adaptive response to hypoxia [141]. The protein is a heterodimer consisting of HIF-1 α and HIF-1 β and the two subunits combine to promote the transcription of the genes involved in the adaptive response to hypoxia. Under normoxic conditions HIF-1 α is rapidly degraded by the action of proline hydroxylase but under hypoxic conditions it is stabilised and able to combine with HIF-1 β and trigger transcription leading to the expression of a number of enzymes/proteins. Up-regulation of the following proteins occurs via the action of HIF: carbonic anhydrase XI, glucose transporters which service the tumour's increased need for glycolysis and VEGF (vaso-epidermal growth factor). HIF

regulates the change in metabolism that occurs in response to reduced O₂ availability through the transcriptional activation of key genes encoding metabolic enzymes, including: LDHA (Lactate dehydrogenase A), which converts pyruvate to lactate [142] and *PDK1* (pyruvate dehydrogenase kinase 1), which inactivates the enzyme responsible for conversion of pyruvate to acetyl-CoA, thereby shunting pyruvate away from the mitochondria [143, 144] (figure 4.1). It has been known for more than 50 years that very low oxygen levels protect cells from death caused by irradiation. Until recently, however, there has been no definitive proof that hypoxia in human tumours contributes to radiotherapy treatment failure. It is known that hypoxia in solid tumours is not only a major problem for radiation therapy but also leads to resistance to most anticancer drugs and appears to accelerate malignant progression and increase metastasis. However, efforts to overcome the problem of hypoxia have had only limited success. The recent development of new drugs that are nontoxic until they are activated in the hypoxic cell opens a new era [145].

The aim of the current study was to examine the global metabolic effects of hypoxia on two different cell lines in order to see if a model could be produced reflecting the hypoxia present in solid tumours.

4.2 Materials and Methods

4.2.1 Chemicals

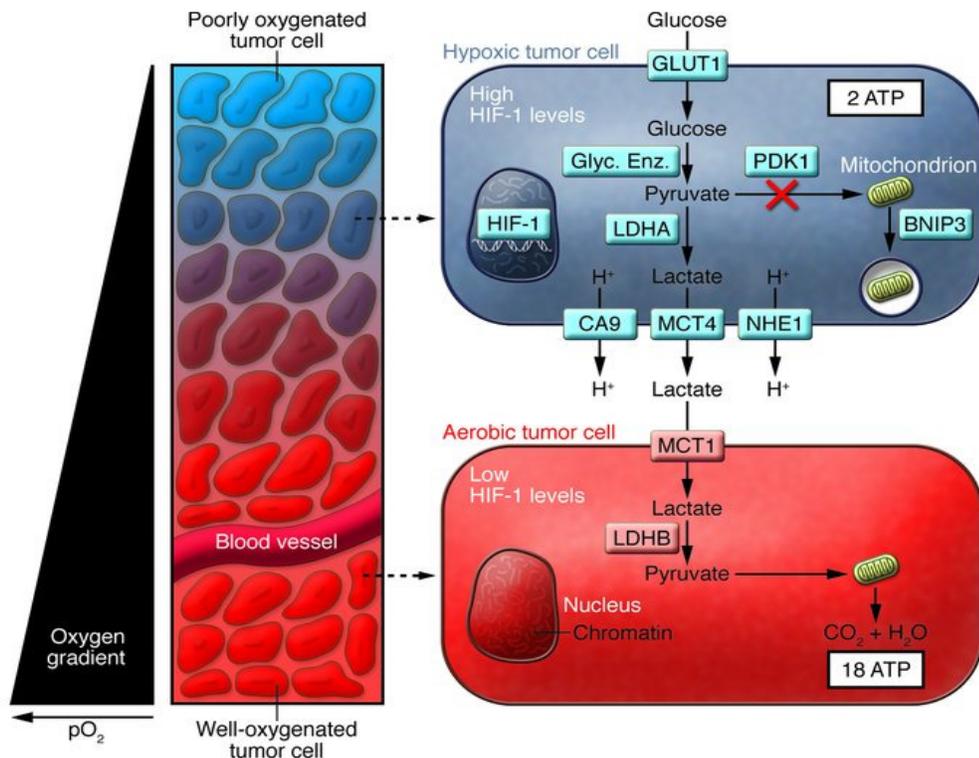
HPLC grade acetonitrile and water were obtained from VWR International Ltd (Lutterworth, U.K.). AnalaR grade formic acid (98%) was obtained from BDH-Merck (Dorset, U.K.).

4.2.2 Sample Preparation

A375 and HCT116 series of cells, were used for the study. The hypoxic samples, were generated from cells which were kept for 36 hours at 1% oxygen in a hypoxic chamber. In order to harvest the cells, they were washed once with ice cold PBS

and scraped with an ice-cold solution of 80% Methanol to reach the concentration of 1×10^6 cells per mL. In order to count the cells on the day of experiment a "counter dish" was used and plated under the same time as the extracted samples. After

Figure 4.1 Metabolic changes in hypoxic tumours.



scraping the cells were immediately centrifuged at $0^\circ C$ for 15 minutes. The supernatant was kept at $-80^\circ C$ until the mass spectroscopy analysis.

4.2.3 LC-MS Methods

4.2.3.1 Separations Carried out on ZICHILIC

LC-MS data were acquired using a Finnigan LTQ Orbitrap instrument (Thermo Fisher Scientific, Hemel Hempstead, U.K.) set at 30000 resolution. The instrument was tuned at least once a week with the manufacturers recommend tuning solution to ensure that mass accuracy remained $< \pm 2$ ppm. Sample analysis was carried out under positive ion mode. The mass scanning range was m/z 50-1200, while the capillary temperature was $250^\circ C$ and the sheath and auxiliary gas flow rates were

50 and 10 respectively. The LC-MS system (controlled by Xcalibur version 2.0, Thermo Fisher Corporation) was run in binary gradient mode. Solvent A was 0.1 % v/v formic acid/water and solvent B was acetonitrile containing 0.1% v/v formic acid; the flow rate was 0.3 ml/min. A ZIC-HILIC column 5 μm 150 \times 4.6 mm (HiChrom, Reading, U.K.) was used for all analyses. The gradient was as follows: 90% B (0 min.) - 40% B (16 min.) - 20% B at 18 min – 20% B at 28 min-90% at 33 min-90% B at 36 min. The injection volume was 10 μl . Mass measurement was externally calibrated according to the manufacturer's instructions just before commencing the experiment, and was internally calibrated by lock masses (83.06037 for acetonitrile, 195.03765 for caffeine and 391.28429 for plasticizer).

4.2.3.2 Separations Carried out on a Cogent Diamond Hydride Column with MS Analysis

The LC system consisted of a Dionnex Ultimate 3000 instrument with a Ultimate 3000 autosampler (UK). The Cogent Diamond Hydride column (150mm X 4.6 mm i.d. 4 μm particle size) was from MicroSolv Tech (NJ, USA).

The mobile phase conditions were as follows:

A: 0.02M ammonia formate in water (pH=6.2). B: acetonitrile/water/0.02 M ammonia formate (90/10 by volume). The flow rate was 400 $\mu\text{L}/\text{min}$ under isocratic conditions with 50% of A: 50% B for 15 mins.

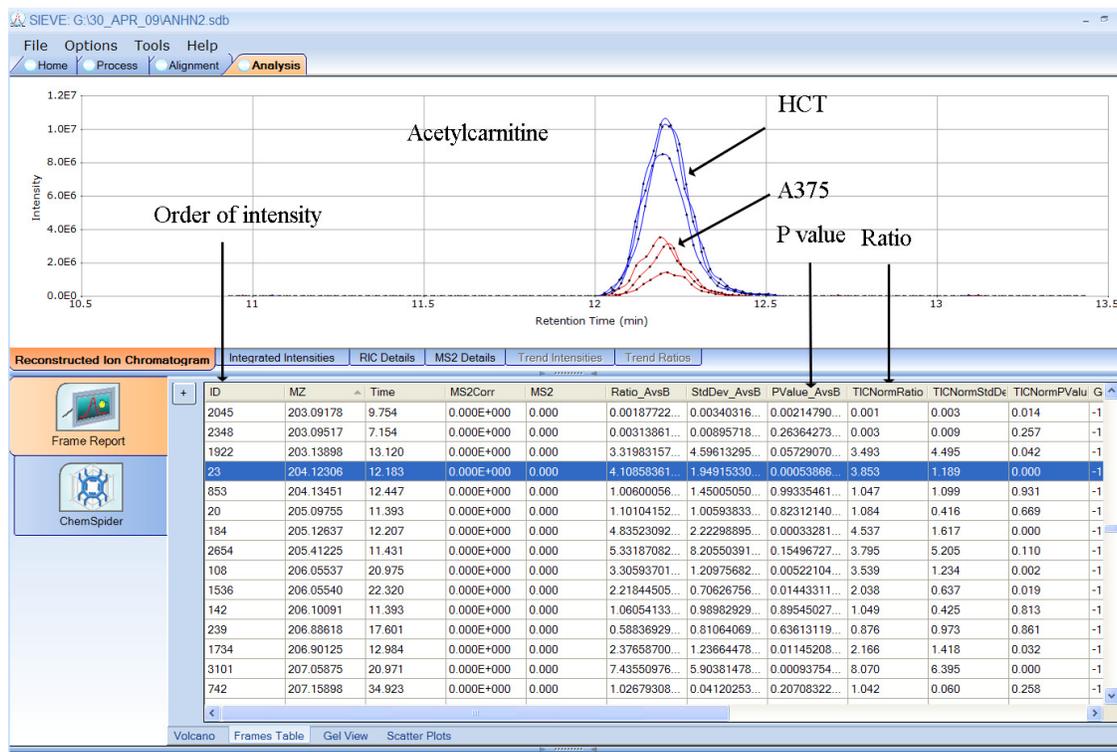
The LC system described above was coupled online to an Exactive mass spectrometer (Thermo fisher Scientific, Germany) equipped with an electrospray ionization (ESI) source. The ion spray voltage was set to 4.5 KV in the negative ion mode. The capillary was set to 25 V. The heated capillary temperature was set to 280 $^{\circ}\text{C}$ and tube lens was set to -40 V. Data acquisition was controlled with Xcalibur 2.1.0 (Thermo Fisher). All acids were measured by a full scan from 50- 300 amu with a resolution of 25,000.

4.3 Results and Discussion

4.3.1 Metabolomic Analysis on a ZICHILIC Column in Positive Ion Mode

Samples were extracted on two occasions three months apart and run in random order with batches of three normoxic and three hypoxic cell cultures of each cell line. The files were loaded into Sieve software in order to highlight the differences between the two cell lines in advance of considering the effects of hypoxia on each individual cell line. Figure 4.1 shows the type of output which can be observed from Sieve. Sieve works by aligning total ion chromatograms for the data files being compared and then binning extracted ion chromatograms for a narrow mass width e.g. 0.02 amu extracted from the aligned data. The series of extracted ion chromatograms is then compared and any differences between feature peaks are compared. Figure 4.1 shows the Sieve output for comparison between A375 and HCT cells. The figure highlights a comparison of the extracted ion chromatograms for acetyl carnitine between the two samples Acetyl carnitine was found to be higher in the HCT cell line. The first column reports an id number which indicates the order of intensity in descending order from the most intense peak in the samples. In general below an id of 500, peaks are getting very noisy, however, 500 cannot be reliably used as a cut off point since sometimes some of the low intensity peaks are due to genuine compounds. The data table from Sieve was then loaded into the Excel macro Sieve Extractor and compared against a mass list of 6000 compounds. Comparison of normoxic HCT cells and hypoxic using Sieve software indicated that about 160 metabolites could be tentatively identified in the two cell lines (table 4.1) and many of these were at similar levels in both cell lines indicated by non-significant P values and, as a rule of thumb, ratios between 0.5 and 2. The list in table excludes most of the lipid fraction which which would add many more compounds to the table. A similar list has been recently reported [146]. The

Figure 4.1 Sieve comparisons between acetyl carnitine in HCT and A375 cells.



compounds in table 4.1 have not been fully validated and there is no automated method for excluding spurious identification due to isotope peaks, adducts, dimers and trimers and fragment ions. The best strategy for this is to target the compounds which vary significantly for further investigation. Thus although the metabolites observed using the HILIC-Orbitrap combination are not the full set of all possible metabolites, the list was complete enough to observe some interesting variations between cell lines and between hypoxic and normoxic conditions. To illustrate a point one of the most relatively abundant compounds in the A375 cells is apparently Co-enzyme M7, according to the Kegg database, which is a co-enzyme in folate biosynthesis. Inspection of the extracted ion chromatograms (figure 4.2) indicates that, although this is a low intensity metabolite, there is indeed a difference between the two cell lines regarding this compound. Inspection of the spectrum for the peak does not initially look promising (figure 4.3) since the peak is present in a region of

the chromatogram where there are a large amount of sodium formate cluster ions. However, there is an ion at 484.055 in amongst the cluster ions, which, since it is quite a high mass compound gives a mass match to a range of different compositions as can be seen in table 4.2.

Table 4.1 Metabolite list obtained from comparing HCT against A 373 cells

Compounds	MZ	Time	HCT/A375	PValue_AvsB
NAAG	305.0982	8.1	0	0.000204
Coenzyme M 7-mercaptoheptanoylthreonine-phosphate heterodisulfide	484.0542	18.1	0.064	0.005619
thiamine	265.1119	32.8	0.11	0.004291
GPC	258.1102	18.1	0.19	0.001815
N-acetyl aspartate	176.0554	7.6	0.32	0.001462
adenosine	268.1042	12.0	0.36	0.003873
Deisopropylhydroxyatrazine	156.088	15.5	0.57	0.137446
S-Adenosyl-L-methionine+	399.1447	34.0	0.63	0.012731
Uridine	245.077	11.3	0.75	0.084233
Pantothenate	220.1181	6.7	0.76	0.064525
Urocanate	139.0502	14.6	0.78	0.010569
hydroxymethylpropanitrile	86.06013	15.2	0.85	0.379297
Urocanate	139.0503	13.3	0.86	0.004847
gamma-N-methylaminobutyrate	118.0862	25.5	0.86	0.051835
phosphoethanolamineR	216.0633	16.5	0.87	0.120682
Phosphoric acid	98.98412	16.7	0.88	0.156707
L-Leucine	132.1019	25.8	0.93	0.67533
Oleamide	282.2792	4.8	0.93	0.542333
Urocanate	139.0502	10.8	0.94	0.171143
L-Leucine	132.1019	27.1	0.94	0.781238
gamma-N-methylaminobutyrate	118.0862	36.6	0.94	0.257013
gamma-N-methylaminobutyrate	118.0862	20.96	0.94	0.496387

L-Cysteine	122.027	14.9	0.94	0.898611
1H-Imidazole-4-ethanamine	112.0869	15.1	0.94	0.007948
5-Aminopentanamide	117.1021	14.4	0.96	0.269851
gamma-N-methylaminobutyrate	118.0862	34.4	0.96	0.261775
1H-Imidazole-4-ethanamine	112.0869	13.6	0.97	0.098606
L-Leucine	132.1019	28.8	0.97	0.906263
L-Aspartate	134.0448	15.9	0.97	0.777534
L-Leucine	132.1018	23.9	0.97	0.895622
5-Aminopentanamide	117.1021	16.0	0.98	0.824188
Amino-proline	131.0815	37.4	0.99	0.997364
gamma-N-methylaminobutyrate	118.0862	12.3	1.00	0.772443
Urocanate	139.0502	9.5	1.01	0.479795
Butanoic acid	89.05968	15.8	1.02	0.615416
L-Leucine	132.1018	33.5	1.03	0.841078
(S)(+)-Allantoin	159.0514	19.7	1.05	0.858942
gamma-N-methylaminobutyrate	118.0862	37.8	1.05	0.188996
Glutamic acid	148.0604	15.3	1.05	0.138201
L-Leucine	132.1019	22.2	1.05	0.759821
(S)(+)-Allantoin	159.0514	18.4	1.06	0.829684
Urocanate	139.0502	5.8	1.06	0.172007
Butanoic acid	89.05968	36.8	1.07	0.012223
2-Aminoacrylate	88.03928	16.9	1.10	0.04916
GSH	308.0913	14.5	1.10	0.109454
Cys-Gly	179.0485	14.5	1.10	0.180557
L-Leucine	132.1019	20.8	1.11	0.594157
Amino-proline	131.0815	35.7	1.12	0.000692
gamma-N-methylaminobutyrate	118.0862	32.6	1.13	0.147121
Triethanolamine	150.1125	15.6	1.13	0.582174
acetylcholine 1+	146.1176	10.9	1.14	0.838844
Choline	104.1069	15.5	1.14	0.552969
Butanoic acid	89.05964	7.29	1.14	0.001344

L-Serine	106.0498	16.8	1.14	0.12759
C17 sphinganine	288.2898	9.1	1.25	0.750112
L-Ornithine	133.0972	15.8	1.26	0.198164
Glutamate 5-semialdehyde	132.0655	15.2	1.29	0.027692
Spermidine	146.1652	25.0	1.30	0.750365
L-Phenylalanine	166.0863	12.7	1.32	0.111891
L-Leucine	132.1019	36.0	1.33	0.167702
Alanine	90.05496	15.3	1.35	0.042858
L-Leucine	132.102	19.4	1.36	0.341285
Glycerone phosphate	171.0057	39.4	1.38	0.175028
Piperidine	84.08074	27.49	1.40	0.403754
Aminoacetone	74.06001	27.4	1.41	0.376263
L-Histidine	156.0767	25.0	1.41	0.104915
L-Ornithine	133.0971	27.8	1.42	0.370218
Pyroline-4-hydroxy-2-carboxylate	130.0498	15.2	1.44	0.021393
Creatine	132.0767	14.9	1.45	0.005569
L-Tyrosine	182.0812	13.8	1.45	0.126884
L-Leucine	132.1018	12.0	1.45	0.131993
Pyroline-4-hydroxy-2-carboxylate	130.0498	16.4	1.45	0.039929
Phenylpyruvate	165.0546	13.8	1.47	0.12025
L-Phenylalanine	166.0862	11.4	1.48	0.122101
L-Leucine	132.1019	13.3	1.48	0.069345
gamma-N-methylaminobutyrate	118.0862	13.8	1.50	0.111888
2-Phenylacetamide	136.0757	13.8	1.50	0.124026
L-Threonine	120.0655	15.5	1.51	0.039278
Indole acrylic acid	188.0706	12.1	1.52	0.155447
L-Tryptophan	205.0972	12.1	1.53	0.148194
Methylhypoxanthine	151.0617	12.9	1.55	0.082354
tetrahydrothiophene carboxylic acid	133.0318	12.9	1.57	0.078145
L-Methionine	150.0584	12.9	1.57	0.069623
trans-Cinnamate	149.0596	11.4	1.57	0.075231

O-Acetyl-L-serine	148.0604	8.3	1.57	0.001033
gamma-N-methylaminobutyrate	118.0862	16.4	1.61	0.034542
Benzoate	123.044	13.7	1.61	0.074524
gamma-N-methylaminobutyrate	118.0863	15.1	1.63	0.066855
Choline phosphate+	184.0734	22.7	1.66	0.000952
diaminooxopentanoic acid(Msc)	147.0764	16.4	1.67	0.006469
Spermidine	146.1652	22.8	1.69	0.610086
L-Histidine	156.0768	13.9	1.71	0.009105
methylproline	130.0863	27.4	1.73	0.276634
N-Acetylornithine	175.1077	15.2	1.74	0.30856
Pyridoxine	170.0812	13.7	1.76	0.222368
L-Lysine	147.1128	27.4	1.78	0.261542
(Z)-4-Hydroxyphenylacetaldehyde-oxime	152.0706	13.8	1.79	0.258221
L-Arginine	175.119	27.0	1.79	0.239995
5,6-Dihydrothymine	129.0658	15.0	1.86	0.017595
Pyrroline-4-hydroxy-2-carboxylate	130.0498	9.9	1.91	0.066228
Putrescine	89.10731	35.8	1.91	0.0154
Taurine	126.0219	15.5	1.97	0.016635
L-Citrulline	176.103	17.1	1.99	0.161118
Pyrroline-4-hydroxy-2-carboxylate	130.0499	13.09	2.03	0.063192
Pyrroline-4-hydroxy-2-carboxylate	130.0498	8.6	2.03	0.090757
Phosphocreatine	212.0432	15.3	2.06	0.002098
L-Leucine	132.1019	16.4	2.07	0.055535
Pyrroline-4-hydroxy-2-carboxylate	130.0499	7.4	2.14	0.176561
Folate	442.147	9.8	2.17	0.252529
Diethanolamine	106.0862	15.6	2.21	0.231715
Creatinine	114.0662	15.2	2.23	0.103426
Phytosphingolipid	304.2846	9.0	2.25	0.437105
C19 sphinganine	316.3211	8.9	2.30	0.323316
2-Methyl-3-hydroxy-5-	182.0453	13.8	2.30	0.184716

formylpyridine-4-carboxylate				
Creatinine	114.0662	16.5	2.42	0.066631
Hydroxyindoline	134.06	13.7	2.43	0.272685
Hydroxyvaline	134.0811	14.9	2.48	0.016064
L-Methionine S-oxide	166.0533	12.9	2.49	0.069841
L-Leucine	132.1019	14.5	2.51	0.075204
Nicotinamide	123.0553	8.76	2.51	0.187018
Pirbuterol	241.1546	14.0	2.60	0.191048
Glycine	76.03934	16.0	2.62	0.000137
Dodecyldimethylamine oxide	230.248	8.9	2.68	0.282813
Carnitine	162.1124	15.0	2.70	0.018301
Spermidine	146.1652	19.9	2.70	0.50381
(R)-Pantoate	149.0806	16.4	2.74	0.026035
Indolylmethylthiohydroximate	207.0587	21.4	2.76	0.002902
acetylcholine	146.1176	14.4	2.78	0.002193
GABA	104.0706	15.2	3.32	0.000149
valerylcarnitine	246.1701	10.4	3.56	0.009086
(S)(+)-Allantoin	159.0513	13.0	3.59	0.201341
L-Asparagine	133.0608	16.8	3.72	4.41E-08
Propionyl carnitine	218.1388	11.7	3.87	0.002987
Butyrylcarnitine	232.1545	11.0	3.98	0.00462
Riboflavin	377.1457	7.6	3.99	0.245665
O-Acetylcarnitine	204.1231	12.8	4.28	0.002143
Lue-Pro	229.1547	14.0	4.57	0.200991
butanedione	87.04408	15.2	4.64	0.003303
5-Formiminotetrahydrofolate	459.1726	9.0	4.79	0.000105
Proline Betaine	144.1019	14.4	5.31	0.020025
N-Acetyl-L-glutamate	190.071	7.2	6.08	0.000121
2-amino-3-(2-amino-3-hydroxy-3-oxo-propyl)disulfanyl-propanoic acid	241.0312	20.9	6.79	0.346524

1-Aminopropan-2-ol	76.07565	14.4	6.84	0.154017
2,3,4,5-Tetrahydrodipicolinate	172.0604	7.2	7.16	6.34E-05
N-acetylspermidine	188.1758	34.8	8.83	0.00088
L-Serine	106.0498	8.3	9.18	5.30E-05
N-Acetylputrescine	131.1178	15.4	12.23	0.005407
1,1,3-tris(ethoxymethyl)urea	235.1655	14.8	12.70	0.100467
N-Methylethanolamine phosphate	156.042	17.2	14.15	0.445065
NAD+	664.1166	17.1	16.95	0.354933
O-Acetyl-L-homoserine	162.0761	13.0	22.87	0.014088
Acetylaminocarnitine	203.1393	14.6	25.88	0.005941
S-Adenosyl-L-methionine+	399.1442	32.2	46.35	0.003675
thiamine(+)	265.1118	31.0	1122.28	0.059582

Figure 4.2 Comparison between extracted ion traces for co-enzyme M7 for A 375 cells (top) vs HCT cells.

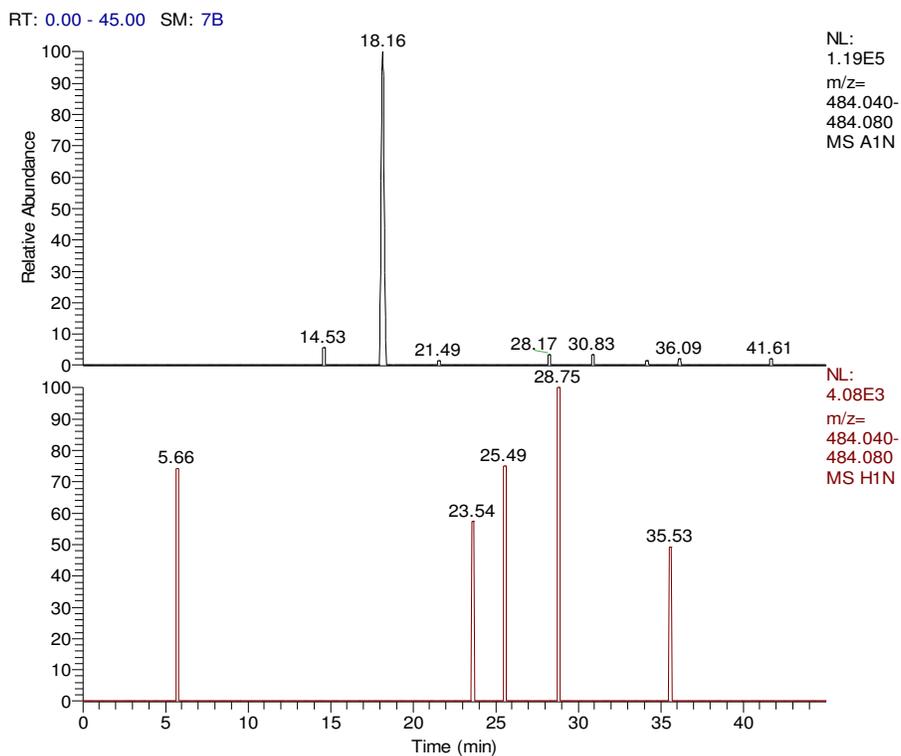
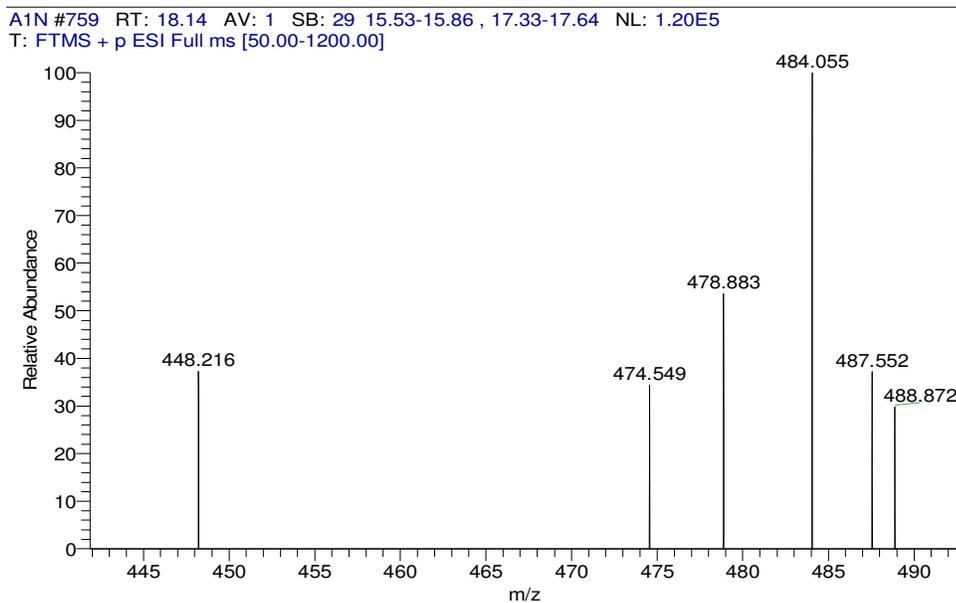
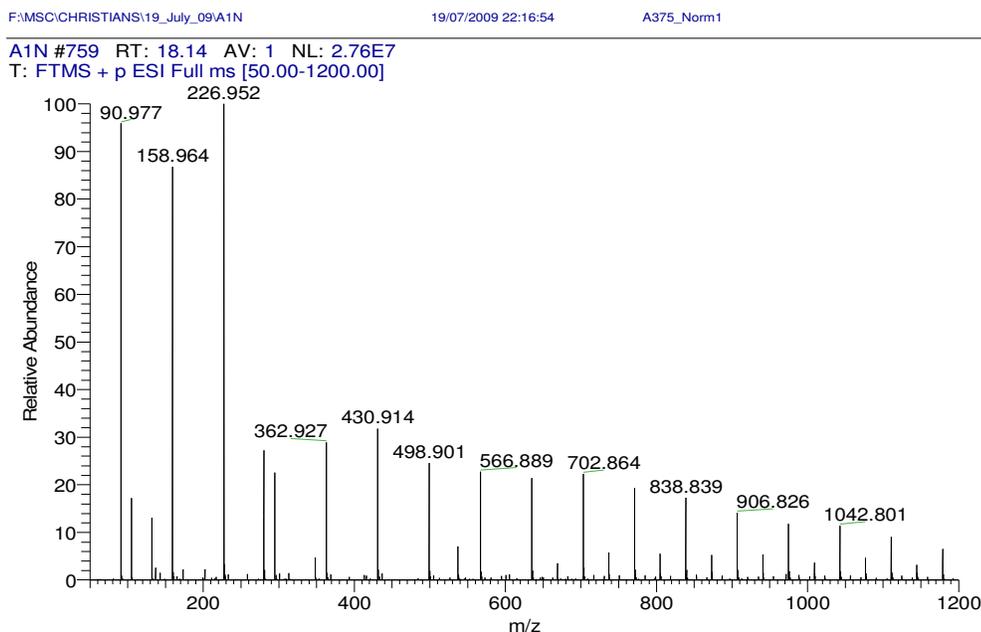


Figure 4.3 Sodium formate cluster ions with expanded region corresponding to Coenzyme M7



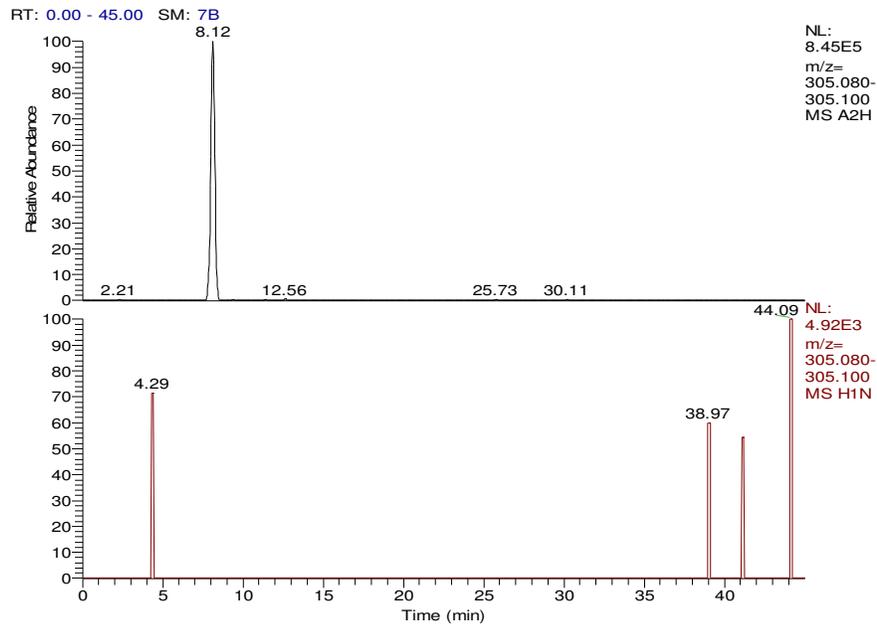
The formula corresponding to co-enzyme M is 17th on the list, however, all the other formula in the obviously do not represent compounds in the Kegg data base. Some of the formulae can be discounted immediately 1, 5, 7, 8, 12, 13 and 14 cannot occur since there have to be at least four oxygens for each phosphorus or at

least seven for two phosphorus atoms in biological systems. In addition 2, 5, 6, 10, 16 are unlikely because of their degree of unsaturation would require an extensive series of rings. Thus it is possible that the ion does belong to co-enzyme M7 although in the absence of standard fragmentation studies would be necessary to confirm this and even these might not yield an answer because the starting ion for fragmentation is weak.

Table 4.2 Elemental matches for ion at m/z 484.05

m/z	Δ (ppm)	RDB	Composition
1. 484.05	0.08	19.0	C28 H21 P S3
2. 484.05	0.12	21.0	C25 H12 O9 N2
3. 484.05	-0.19	6.0	C17 H26 O8 P2 S2
4. 484.05	-0.33	11.0	C18 H20 O6 N4 S3
5. 484.05	0.46	16.0	C20 H17 O3 N6 P S2
6. 484.05	-0.52	16.5	C21 H15 O9 N3 P
7. 484.05	-0.56	14.5	C24 H24 N P2 S3
8. 484.05	-0.89	25.0	C28 H13 O N4 P S
9. 484.05	-0.97	6.5	C14 H23 O6 N5 P S3
10. 484.05	1.10	20.5	C24 H14 O3 N5 S2
11. 484.05	-1.16	12.0	C17 H18 O9 N4 P2
12. 484.05	1.23	15.5	C23 H20 O5 N P2 S
13. 484.05	-1.53	20.5	C24 H16 O N5 P2 S
14. 484.05	-1.61	2.0	C10 H26 O6 N6 P2 S3
16. 484.05	-1.67	20.0	C26 H16 O4 N2 S2
17. 484.05	1.80	1.5	C13 H27 O10 N P S3

Looking at the data in table 4.1 co-enzyme M7 is not of immediate interest since it does not immediately link into sets of related metabolites and is a low intensity peak. There were numerous significant differences between the two cell lines examined. The number of elevated metabolites in A375 compared to HCT cells was smaller and the most marked difference was in N-acetylaspartylglutamate (NAAG). This compound was absent from the HCT cells as can be seen in figure 4.4. The mass spectrum of NAAG is shown in figure 4.5. The mass spectrum shows the molecular ion for NAAG a 305.098 and an additional dominant peak at m/z 123.055 which is due to nicotinamide which, since it is also an amide, co-elutes from the ZICHILIC column. The ion at m/z 148.06 is a source fragmentation ion which corresponds to glutamic acid generated by fragmentation of NAAG, this indicates the the position of acetylation of the dipeptide must be on the aspartate and that glutamic acid is at the C-terminus. Glutamic acid itself elutes at 15.1 minutes since it is an amine and thus more strongly retained on the ZICHILIC column. The identity of of NAAG was confirmed by its co-elution with a standard and which gave the same fragmentation pattern under MS^2 conditions yielding predominantly a fragment ion due to glutamic acid. NAAG is very abundant in brain tissue and is a neuroprotective peptide [147]. The ACT375 cell-line derives from melanoma which in turn originates from melanocytes which in turn derive from neural crest cells. Thus there is a rather tenuous link between these cells and neurological tissue. Linked to high NAAG levels are elevated levels of its precursor N-acetyl aspartate (table 4.1). NAAG plays a central role in neurological function and is released following nerve depolarisation stimulating glucose transport during repolarisation [148, 149]. Adenine is also elevated in the A375 cells (figure 4.6). Linked to this are elevated levels of adenosine (figure 4.7) which again reflects the origin of these cells in neurological tissue, adenosine is abundant in the brain [150-152].

Figure 4.4 NAAG in A375 cells (top) and HCT cells.**Figure 4.5** Mass spectrum of NAAG including dominant co-eluting ion from nicotinamide.

A2H #332-347 RT: 7.93-8.29 AV: 16 SB: 104 6.37-7.62, 8.86-10.05 NL: 5.90E5
T: FTMS + p ESI Full ms [50.00-1200.00]

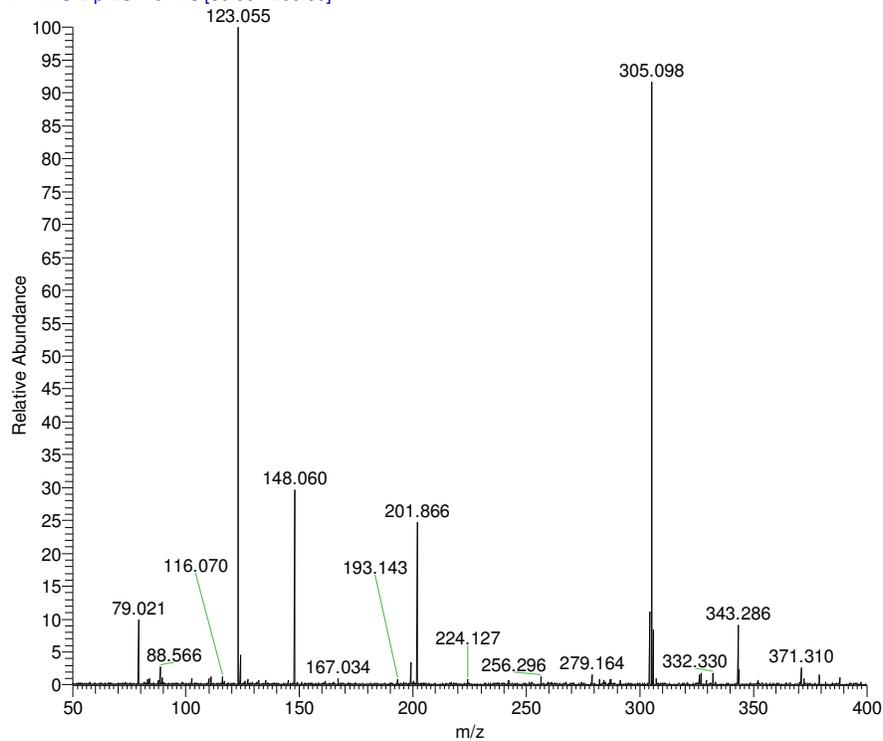
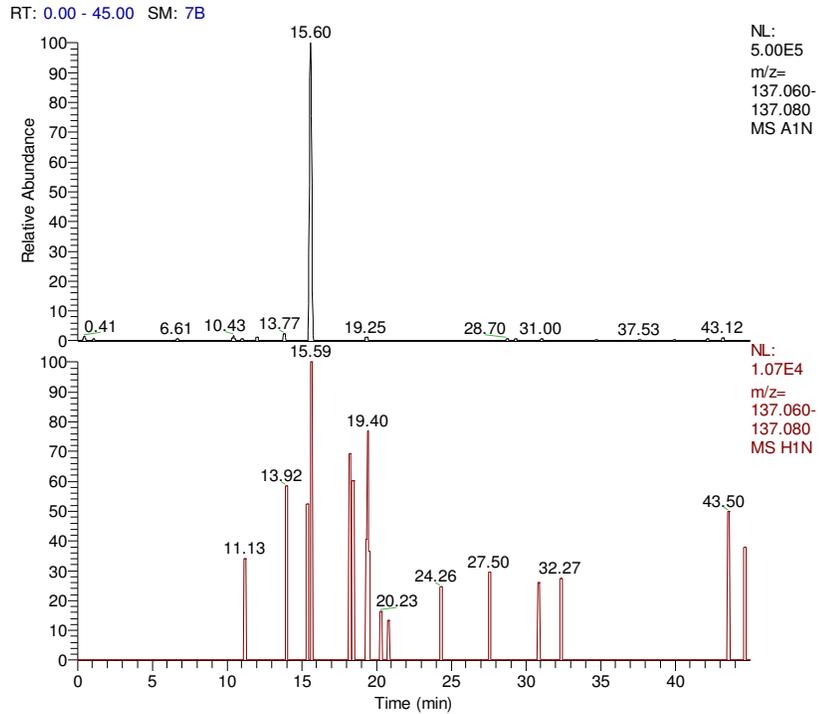
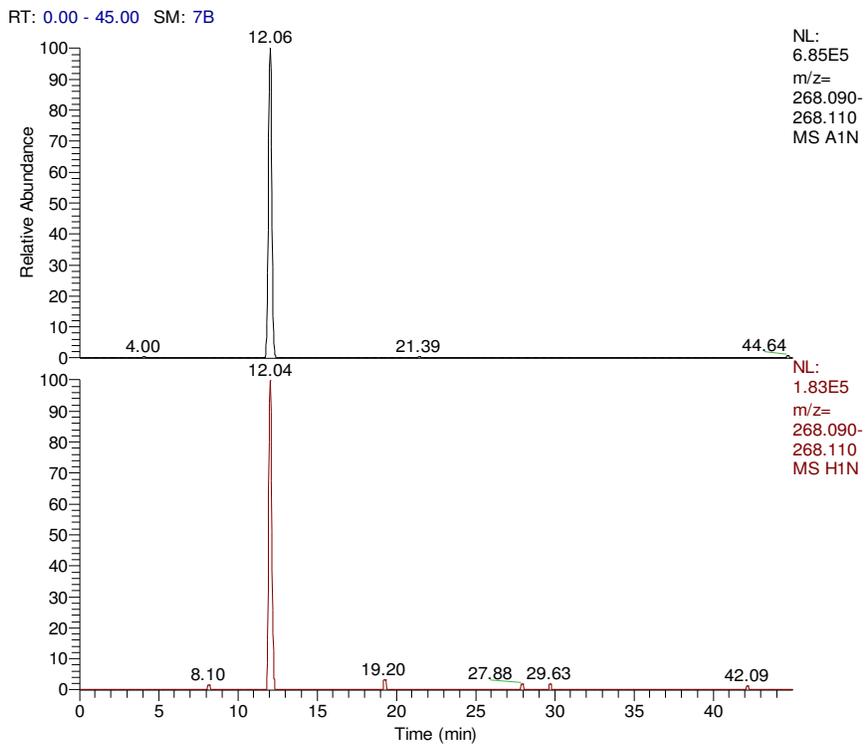
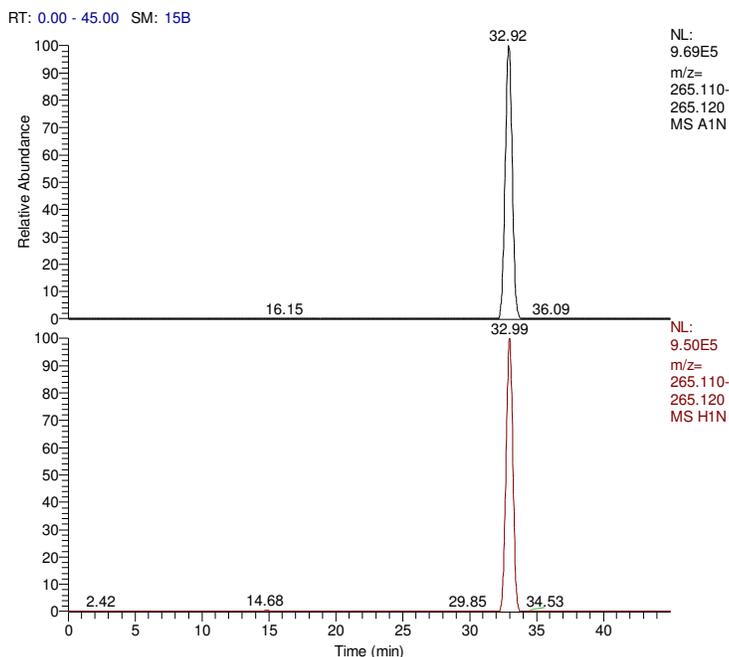


Figure 4.6 Elevated adenine levels in A375 cell (top) relative to HCT cells**Figure 4.7** Comparison of adenosine in A375 cells (top) compared with adenosine in HCT cells.

The differences in thiamine shown in table 4.1 are an artefact of the Sieve software which can occasionally misreport differences due to failure to align the extracted ion chromatograms properly. As can be seen from figure 4.8 this is the case with thiamine which is similar in both cell lines and in table 4.1 it is reported as a difference both at the top and the bottom of the table.

Figure 4.8 Comparison of thiamine between A375 cells (top) and HCT cells.



The pattern of relatively elevated metabolites in the HCT cells is more complex. The most notable group of metabolites which are elevated are the carnitine and the acyl carnitines including acetyl, propionyl, butyryl and valerylcarnitines. The acylcarnitine metabolites are associated with long chain fatty acid metabolism within the peroxisomes [153-155]. Longer chain fatty acids (>C18) can be metabolised initially in peroxisomes rather than in the mitochondria and once they are broken down into shorter chain lengths they are conjugated to carnitine so that they can be transported into the mitochondria for further metabolism. It is not clear if this is the role of these compounds in cell cultures since surprisingly there is little information about the content of long chain fatty acids in culture medium although it likely that

supplements such as calf serum will contain components such as phytanic acid. This might suggest that peroxisomal metabolism is more important in the HCT cells than the A375 cells. The HCT cells derive from gastrointestinal cancer cells and it might be expected that fatty acid breakdown would be given more emphasis in GI cells than in cells derived from neurological tissue. Certain types of cells such as liver and muscle cells do have very high rates of fatty acid oxidation. There is some evidence that peroxisome-proliferator activated receptor gamma (PPAR γ) is upregulated in colorectal cancers [156].

The other group of metabolites that is elevated in the HCT cells is comprised of the polyamine putrescine and the polyamine metabolites putrescine acetate and spermidine acetate. GI tumours have also been found to contain elevated levels of polyamines and their metabolites and it is well established that polyamines play an integral role in colorectal cancer [157-159]. Spermidine/spermine N1-acetyltransferase (SSAT), which down regulates polyamines, presumably by reducing their polycation character, may have anti-proliferative effects. It has been shown that acetylated polyamines are present in large amount in tumours [160] and it has been proposed that they represent an attempt by cancer cells to down regulate cell division in the face of over production of polyamines [158]. PPAR γ has been found to be involved in the regulation of SSAT since it heterodimerises with the retinoid X receptor and binds to DNA thus regulating the production of SSAT. This is somewhat in contradiction to the fact that PPAR γ is upregulated in GI tumours; however, the relationship may be more subtle than this. It has been found that where PPAR γ is overexpressed in GI tumours that the prognosis is better [161]. Figure 4.9 shows a comparison of the polyamine acetates in A 375 and HCT cells. As can be seen in figure 4.9, in the case of the acetylspermidine it may be one form of SSAT that is elevated since two monoacetate metabolites can be seen (three are

possible) and only one of these is clearly elevated in the HCT cells. In addition creatine, creatinine and creatinine phosphate levels are higher in HCT cells compared with A375 cells. Thus the two cell lines which might be thought to interchangeable in providing a model for hypoxia in solid tumours are quite different in their metabolic profiles.

Figure 4.9 Elevated levels of acetylated polyamines in HCT cells

RT: 0.00 - 45.00 SM: 15B

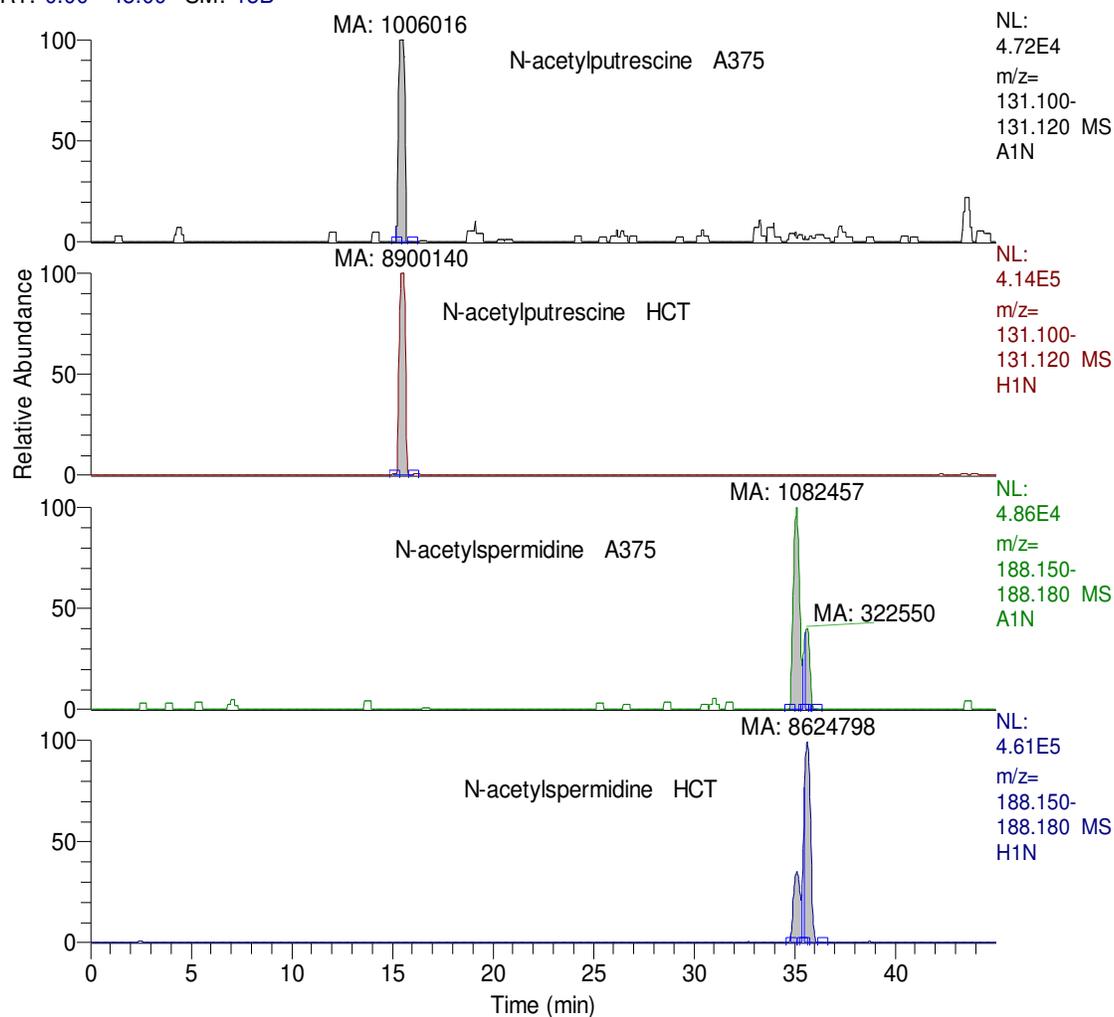


Table 4.3 shows the response of the A375 cells to chronic hypoxia for 24 hours. The significant changes are not extensive and some of them are not preserved between the two times of sampling. True to type the A375 cells respond with an

elevation of NAAG which is what might be expected in neurological tissue where the NAAG may be protective against reperfusion injury. GSH exhibits a clear effect at the two sampling times where it falls slightly. It is well established that GSH levels fall under hypoxic conditions probably due to its levels being reduced by the formation of reactive oxygen species (ROS). Increase in ROS may also account for the fall in levels of cysteine.

The effect of hypoxia on HCT cells is much more complex (table 4.4). The most marked effect of hypoxia on the HCT cells is that there is a large elevation in palmitoyl and linoyl carnitines (figure 4.9). The long chain acylcarnitines are produced in order to transport fatty acids into the mitochondria for oxidation and the elevation in their levels indicates impaired uptake of these compounds into the mitochondria. The levels of short chain acylcarnitines such as acetyl carnitine and propionyl carnitine also fall under hypoxic conditions. Since these conjugates generally result from metabolism of fatty acids by peroxisomes this indicates that peroxisomal metabolism is lower under hypoxic conditions. The levels of N-acetyl putrescine and N-acetyl spermidine in the HCT cells fall under hypoxic conditions. In the case of spermidine acetate it again appears that one of the acetylated forms of N-acetylspermidine is affected. This fall may reflect down regulation of the activity of PPARY. From the depressed levels of acylcarnitines it can be seen that peroxisome activity is lowered. This implies a lowered level of SSAT since this is regulated by PPARY. It has been proposed that SSAT can promote the degradation of HIF-1 α . SSAT over expression can stabilize the interaction of VHL and Elongin C in the absence of HIF-1 α thus promoting its ubiquitination [160, 162]. The remarkable complexity of the protein interactions required for the efficient hydroxylation, ubiquitination, and subsequent proteasomal degradation of HIF-1 α in oxygenated cells points to the critical importance of this process for cellular homeostasis and

provides a molecular basis for the extremely rapid hypoxic induction and post-hypoxic decay of HIF-1 α . Interestingly, SSAT has been shown to interact with the p65 subunit of NF κ -B

Table 4.3 A375 cells under hypoxic versus normoxic conditions (hyp/norm).

Compounds	MZ	Time	PValue	Ratio	PValue	Ratio
GSH	308.0912	14.509	0.000658	0.533	0.027640	0.79
N-acetyl aspartate	176.0554	7.663	0.000894	0.565	0.55	0.96
O-Acetylcarnitine	204.1231	12.866	0.007314	0.37	0.002464	0.42
2,3,4,5-Tetrahydrodipicolinate	172.0605	7.259	0.008622	2.576		
Glyceraldehyde phosphate	171.0053	16.116	0.01125	2.429		
Choline phosphate+	184.0734	21.758	0.01136	0.753	0.0353	0.773
NAAG	305.0981	8.112	0.01321	2.126	0.001952	1.72
L-Cysteine	122.027	14.961	0.01661	0.665	0.001937	0.58
Phosphocreatine	212.0432	15.29	0.01864	1.859	0.010560	1.54
S-Adenosyl-L-methionine+	399.1447	34.003	0.02056	0.667	0.58	0.89
Glutamic acid	148.0604	15.389	0.02384	0.819	0.040060	0.87
N-Acetyl-L-glutamate	190.0711	7.291	0.02928	2.316		
Carnitine	162.1124	15.071	0.03295	2.016	0.017650	1.23
*Glucose 6 phosphate	261.0370	16.5		1.69		
*Fructose 6-phosphate	261.0370	17.2		2.83		
*Phosphoenol pyruvate	168.9899	15.1				

* Only observed on one occasion

and function as a co-activator to stimulate NF- κ Bdependent transcriptional activation, although the precise molecular mechanisms were not determined. Thus, SSAT inhibits HIF-1 α and activates NF- κ B, two of the most important stress-induced transcription factors in metazoan species, by regulating protein stability and transactivation function, respectively. In addition to the acetate metabolite of putrescine another metabolite GABA is also depressed by hypoxic conditions.

Table 4.4 Metabolic changes hypoxic HCT cells versus normoxic HCT cells

(hyp/norm).

Compounds	MZ	Time	PValue	Rati	P value	
O-Acetylcarnitine	204.1231	12.9	0.0003	0.43	0.018020	0.61
L-Palmitoylcarnitine	400.3421	8.8	0.0029	8.58	0.19	2.27
N-acetylspermidine	188.1758	35.7	0.0031	0.71	0.018830	0.56
GABA	104.0706	15.3	0.0051	0.57	0.028050	0.77
Propionyl carnitine	218.1388	11.8	0.0053	0.34	0.000557	0.73
L-Asparagine	133.0608	16.9	0.0054	0.62	0.207	0.78
N-Acetylputrescine	131.1178	15.5	0.0072	0.32	0.004604	0.38
Pantothenate	220.1181	7.0	0.0078	0.55	0.199800	1.13
Glyceraldehyde phosphate	171.0053	16.0	0.0091	18.1		
GSH	308.0913	14.5	0.0117	0.71	0.000670	0.71
acetylcholine 1+	146.1176	14.4	0.0145	0.69		
2,3,4,5-	172.0604	7.3	0.0340	0.69		
Linolylcarnitine	426.3575	8.8	0.0360	25.3	0.17*	4.0
Glycine	76.03934	16.1	0.0415	0.68		
Glutamate 5-semialdehyde	132.0655	15.3	0.0433	0.84		
1H-Imidazole-4-ethanamine	112.0868	15.2	0.0496	0.95		
N-Acetyl-L-glutamate	190.071	7.3	0.0523	0.68	0.40	1.06
L-Histidine	156.0767	25.0	0.0569	0.60		

Figure 4.10 Comparison of long chain acylcarnitines under hypoxic and normoxic conditions.

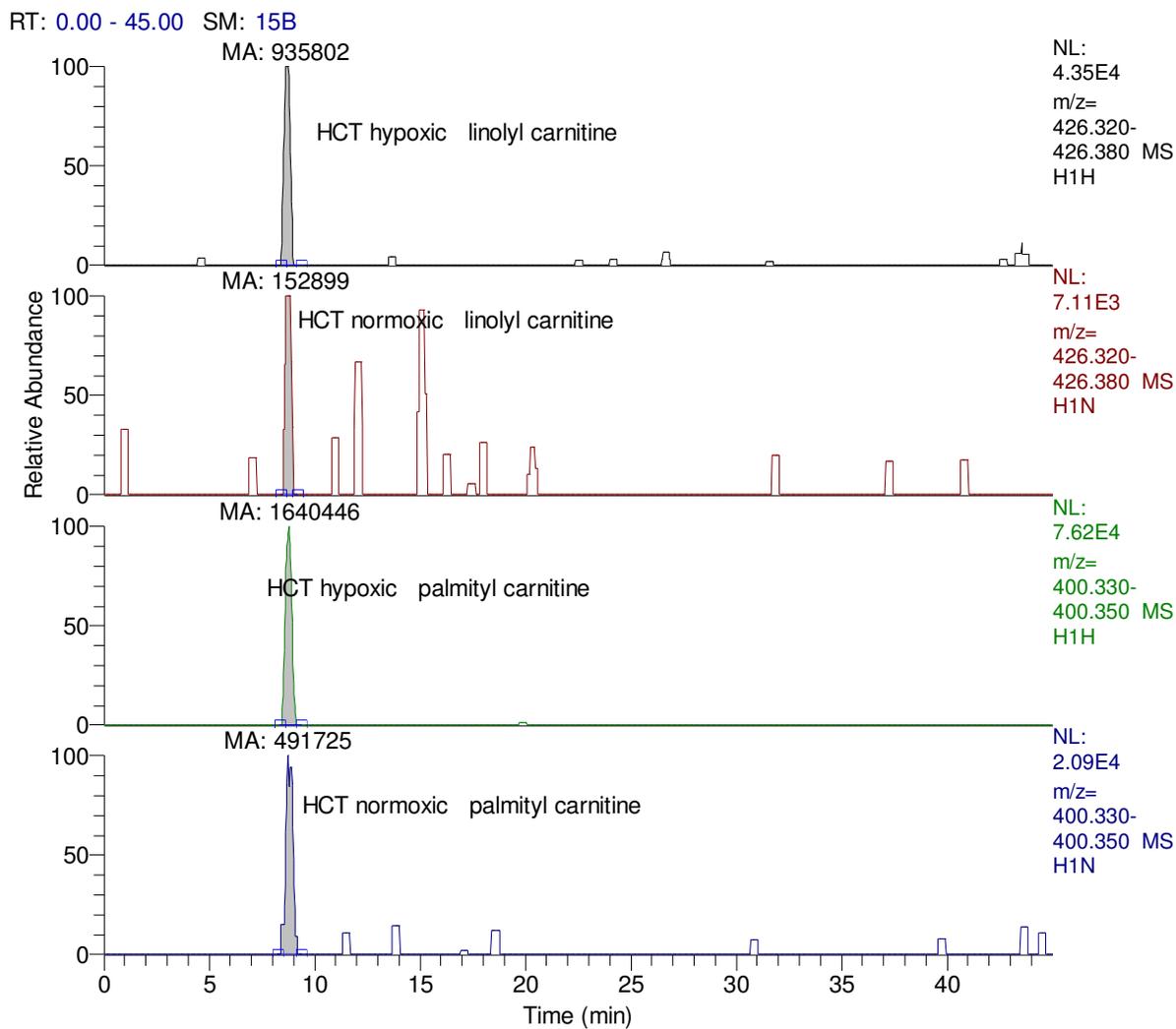


Figure 4.11 Decrease in spermidine acetate under hypoxic conditions showing a specific fall in one of the acetylated forms.

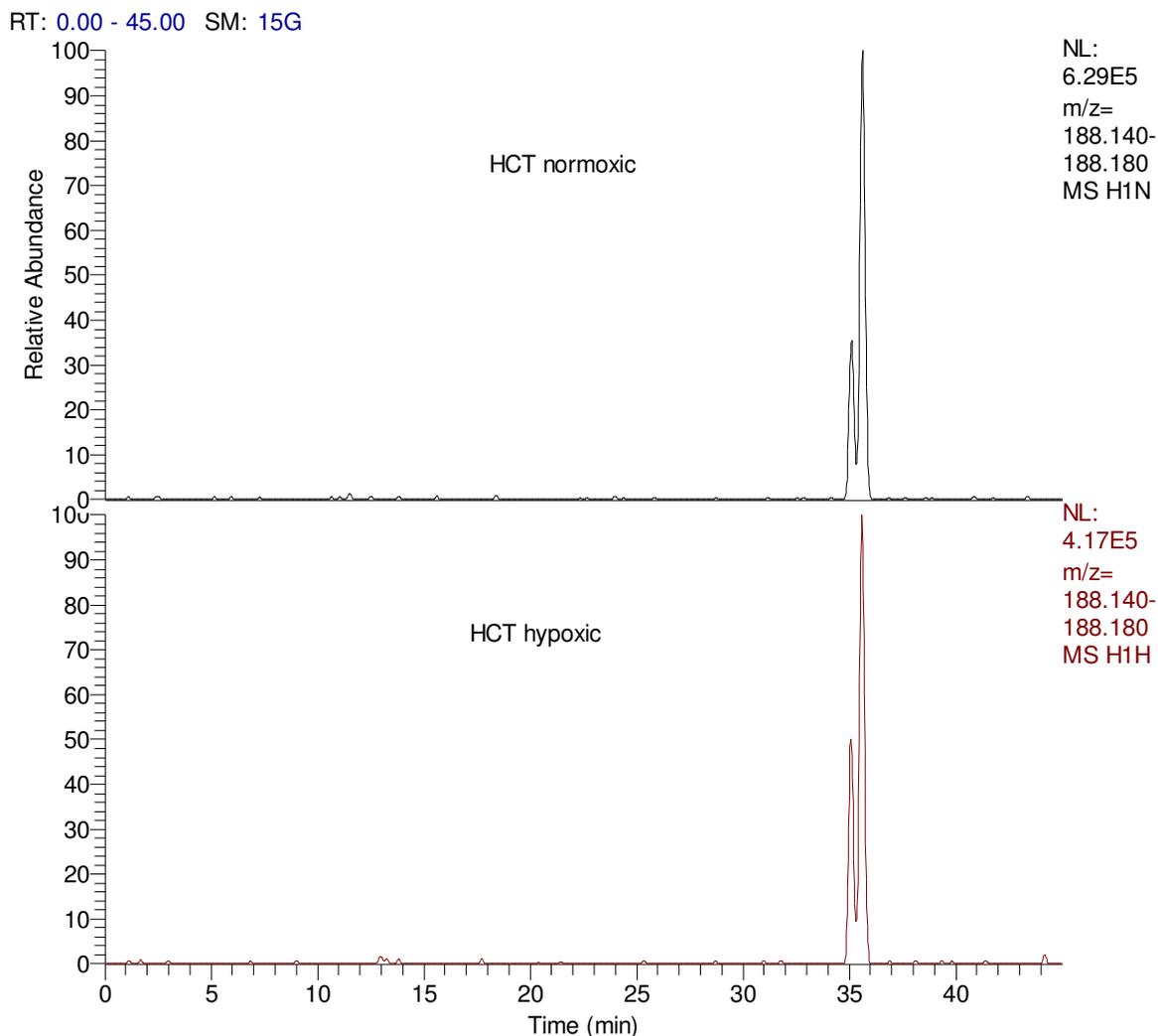


Table 4.5 shows a comparison of A375 and HCT cells under hypoxic conditions. Only the compounds which differ significantly are shown. As might be expected the A375 cells have high levels of NAAG and its precursor N-acetylaspartate relative to HCT cells. The relative elevation in thiamine in the A375 compared to HCT cells under hypoxia is genuine and may reflect a greater slowing down in the Krebs cycle in the A375 cells. The differences in table 4.5 largely re-enforce the observations made for each individual cell line when comparing normoxia and hypoxia.

Table 4.5 Metabolite levels in A375 vs HCT cells under hypoxic conditions

Compounds	MZ	Time	Ratio _Avs B	PValue_Av sB
NAAG	305.0982	8.1	0	0.000204
thiamine(+)	265.1119	32.9	0.12	0.004291
GPC	258.1102	18.1	0.19	0.001815
N-acetyl aspartate	176.0554	7.7	0.32	0.001462
adenosine	268.1042	12.1	0.36	0.003873
Urocanate	139.0503	13.3	0.86	0.004847
gamma-N- methylaminobutyrate	118.0862	16.5	1.61	0.034542
Choline phosphate+	184.0734	22.8	2.11	0.00026
L-Histidine	156.0768	13.9	1.72	0.009105
Putrescine	89.10731	35.9	1.91	0.0154
Taurine	126.0219	15.6	1.97	0.016635
Phosphocreatine	212.0432	15.3	2.07	0.002098
Glycine	76.03934	16.1	2.62	0.000137
Carnitine	162.1124	15.0	2.70	0.018301
(R)-Pantoate	149.0806	16.5	2.74	0.026035
Indolylmethylthiohydroxim ate	207.0587	21.5	2.76	0.002902
acetylcholine	146.1176	14.4	2.78	0.002193
GABA	104.0706	15.2	3.32	0.000149
valerylcarnitine	246.1701	10.5	3.56	0.009086

L-Asparagine	133.0608	16.8	3.72	4.41E-08
Propionyl carnitine	218.1388	11.8	3.87	0.002987
Butyrylcarnitine	232.1545	11.0	3.98	0.00462
O-Acetylcarnitine	204.1231	12.9	4.28	0.002143
5- Formiminotetrahydrofolate	459.1726	9.0	4.80	0.000105
Proline Betaine	144.1019	14.5	5.32	0.020025
N-Acetyl-L-glutamate	190.071	7.3	6.08	0.000121
2,3,4,5- Tetrahydrodipicolinate	172.0604	7.3	7.16	6.34E-05
N-acetylspermidine	188.1758	34.8	8.83	0.00088
L-Serine	106.0498	8.3	9.18	5.30E-05
N-Acetylputrescine	131.1178	15.5	12.23	0.005407
O-Acetyl-L-homoserine	162.0761	13.1	22.87	0.014088

The profiling of polar acids in LC-MS mode presents a problem since many anion exchange columns use strong acids or buffers to elute analytes which are incompatible with MS. The ZICHILIC column does not work well with acids since the sulfonic acid group in the phase, which is at the end of the ligand attached to the phase, repels acids so that they elute early and present as poorly shaped peaks. Aminopropyl phases would be an obvious choice but they do not produce satisfactory peaks either. GC-MS is a possible method for analysis of organic acids but it is not ideal for profiling at the level of acids present in the cell extracts since the ion current carried by the individual acids becomes fragmented under electron impact conditions and thus it is not suitable for global profiling although targeted

analysis can be carried out for individual acids in selected ion monitoring mode. Recently a new HILIC column has become available the Cogent diamond hydride column. This column is based on a reversed phase silica gel column in which the silanol groups have been converted to silicon hydride groups. Some features are quite attractive when coupling to MS.

Some time was spent in optimising this phase for the analysis of organic acids and it proved quite difficult to use. The optimised conditions used an isocratic method with high levels of ammonium formate in the mobile phase and a small injection volume. The effect of hypoxia on Krebs cycle acids in HCT cells is shown in table 4.6.

Not all the peak shapes for the acids are ideal. Figure 4.11 shows the extracted ion chromatogram for α -ketoglutaric acid extracted from the HCT cells which gives quite a good peak shape while figure 4.12 shows the peak obtained for lactic acid which is not quite as good. A citric acid standard gave a very poor peak shape. The requirement for high levels of ammonium formate in the mobile phase tended to make the instrument source dirty quite quickly.

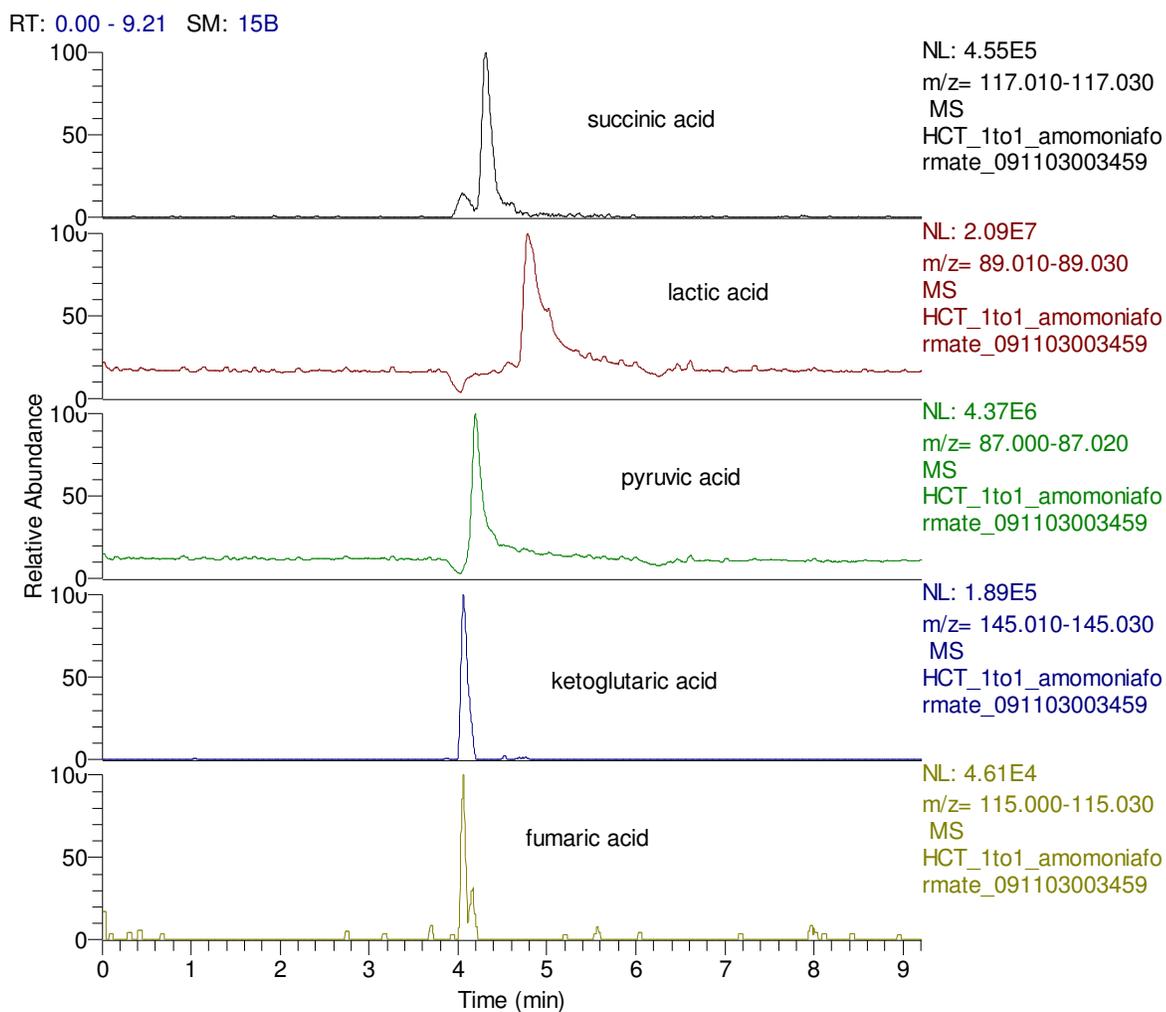
Table 4.6 The effect Krebs cycle acids in hypoxic vs normoxic HCT cells.

M/Z	Formula	Name	P value	Ratio
87.00884	C ₃ H ₄ O ₃	Pyruvate	0.0947	0.90
89.02449	C ₃ H ₆ O ₃	Lactate	0.116	1.15
115.0038	C ₄ H ₄ O ₄	Fumarate	0.00485	1.93
117.0194	C ₄ H ₆ O ₄	Succinate	0.13	0.81
133.0142	C ₄ H ₆ O ₅	Malate	ND	ND
130.9987	C ₄ H ₄ O ₅	Oxaloacetate	ND	ND
173.0092	C ₆ H ₆ O ₆	cis-Aconitate	ND	ND
145.0143	C ₅ H ₆ O ₅	α -Ketoglutarate	0.00366	0.62

The hydride column can be used over a range of mobile phase compositions from 100% aqueous to pure non-polar organic solvents. Thus, it can function in high water (reversed-phase), high organic with some water present (aqueous normal-phase, ANP) and pure organic (organic normal-phase). Since the reversed-phase and aqueous normal-phase modes are complementary on the hydride column, it is possible to rapidly switch between them due to the rapid equilibration of these separation materials. Also in some cases, both mechanisms operate simultaneously and thus retain both hydrophobic and hydrophilic compounds in a single isocratic run. The surface composition of the phase is a combination of the base silica hydride and the organic moiety attached to it. In the current case the metabolites investigated were organic acids. Since retention in the ANP mode is based on polarity it was anticipated that a mobile phase above the pKa of any acid studied would give enhanced retention. However, the width of peaks becomes greater with the increasing number of -COOH. For instance, the peak of pyruvic acid is shaper than cis-acontic acid. Citric acid has three carboxylic acid groups and even cannot form a good peak shape on the hydride. This column is very new and from direct correspondence with the inventor, Joseph Pesek, it was suggested that organic acids can be sensitive to the amount of sodium in the mobile phase. There is a competitive equilibrium between NH_4^+ and Na^+ during the cation ion exchange processes within the column. This sensitivity is magnified at the very low concentrations used for MS. The sodium often comes from the glass solvent bottles and from the solvents and additives themselves. Each solvent bottle can be different depending on its age, which solvents have been in it, and how long the mobile phase has been inside the bottle. Teflon or plastic bottles could be good alternatives. Thus there is still development work required with this column and currently it is quite difficult to use. In conclusion, using hypoxic cell cultures to

understand the adaption of solid tumour cells to hypoxia shows some promise, however, not all cell cultures are equivalent and it is important to select a cell-line which is likely to behave in a similar way to tumour cells in adapting to hypoxic conditions.

Figure 4.12 Analysis of acids on a Cogen Diamond Hydride column. Conditions as in section 4.2.2.2



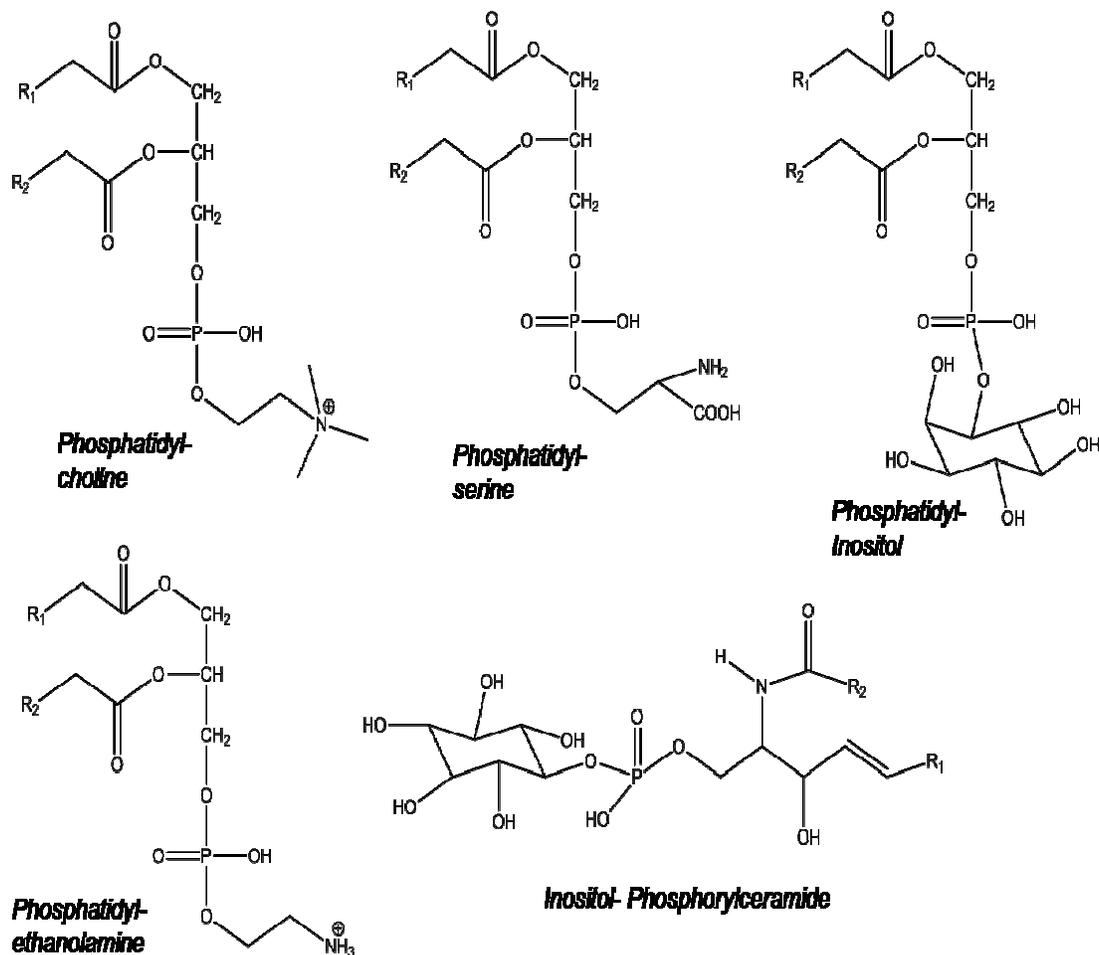
Chapter 5

LC-MS Profiling of Phospholipids in *Leishmania donovani* extracts by HILIC mode

5.1. Introduction

Biological lipids are a chemically diverse group of compounds (structures shown in Figure 5.1). They play important roles in living organisms and function diversely based on their chemistry. Fats and oils are used as the principal forms of energy storage and are converted to ATP, CO₂ and H₂O in order to supply an organism's nutrition. Phospholipids and sterols are the major building materials of biological membranes and act as barriers to the diffusion of ions, proteins etc. In recent years, evidence has emerged showing that lipid signalling is a vital part of cell signalling [163]. Several lipids have been identified as signalling molecules including sphingosine-1-phosphate, diacylglycerols (DAGs), Phosphatidylinositol-phosphates (PIPs), phosphatidyl serine (PS) and prostaglandins etc [164-166]. Lipid signalling is thought to be different from classic signalling pathways (like amine neurotransmitters) in that lipid messengers are biosynthesized on demand at the site of action [166]. Some studies have suggested that they are involved in the regulation of cell growth, apoptosis, calcium mobilisation and calcium-mediated activation of protein kinases etc. [167-169].

A series of lipid related studies on trypanosomes were conducted in order to understand the infection mechanisms and cell cycle progression of the parasites. In 2002, Lillico et al. controlled the expression of a dense surface coat of glycosylphosphatidylinositol (GPI)-anchored proteins in tsetse fly hosts which affected the survival of *Trypanosoma brucei*. They depleted *T. brucei* GPI8, the gene encoding the catalytic subunit of the GPI: protein transamidase, which resulted in the absence of GPI-anchored proteins from the cell membrane of the procyclic

Figure 5.1 The demonstration of structure of Phospholipids

form of trypanosomes. This treatment rendered the parasites unable to establish infections in the tsetse fly midgut and therefore demonstrated that GPI-anchored proteins play an essential role in parasite vector interactions. The bloodstream form trypanosomes showed a defined multinuclear, multikinetoplast and multiflagellar phenotype after exertion of GPI8 RNAi, showing that GPI-anchored proteins are also important for proper cell cycle progression [170, 171]. In Priotto *et al.* used methyl-beta-cyclodextrin or added an inhibitor of cholesterol synthesis, mevinnolin, to Hep2 cells and affected the *Trypanosoma cruzi* infection of the host cell. These treatments led to a leakage of membrane cholesterol and placental alkaline

phosphatase, which is anchored to the membrane by a GPI molecule, resulting in a reduction of the invasion process [172-174].

Recognition of the importance of lipids in cellular functions has led to progress in the technology of lipid analysis. The structure characterisation of lipids is difficult because they are composed of different polar head groups, fatty acid tails at the sn-1 (sn-2) sites and even fatty acid double bond positions. For several years, lipids were characterised by GC-MS. Since most lipids are non-volatile in nature, the carboxylic acid moiety of fatty acids is usually converted to a methyl- or ethyl-ester and polar groups such as carbonyls and hydroxyls can be derivatized to form methoxylamines and silyl ethers, respectively [175]. With the advent of fast atom bombardment mass spectrometry (FAB-MS), the analysis of intact lipids became possible and simplified the structure characterisation. However, the lack of reliability and reproducibility of FAB-MS and the difficulty of interfacing it with HPLC prevented it from being widely applied in the study of lipids. The softer ionisation techniques, ESI and MALDI revolutionised lipid analysis increasing detection capability, sensitivity and precision and also making structure elucidation more straightforward. Most phospholipids can be ionised in either in positive or negative ion mode. Negative ion mode is more frequently used because of the rich information obtainable when using collision-induced dissociation. Hsu *et al* used tandem mass spectrometry to analyse several classes of phospholipids and confirm their compositions by stable isotope labelling deuterated hydrogen in combination with a lithiated buffer [176]. MALDI is also used as an approach but seldom for structural information because lipids' affinity for sodium and potassium which results in adduct formation in this mode make spectral interpretations difficult [177].

HPLC-MS is now the most popular method in the study of lipids. Both reverse phase and normal phase have been used in the separations. The reverse phase

mode separated lipids based on carbon chain length and the number of double bonds, which means their lipophilicity [178, 179], longer chains will have longer retention times [180]. Each double bond is equivalent to approximately two methylene groups regarding retention properties [181], thus some lipids species will have very similar retention times due to this approximation, which leads to co-elution [182]. Sommer *et al* used normal phase to fractionate each lipid class and obtain a detailed the map of each subgroup followed by a reverse phase [183]. Ogiso *et al* developed a C-18 method to profile phospholipids especially phosphatidic acid and PS groups in cells [184]. They added phosphoric acid into and isopropanol/methanol/ammonia formate mobile phase as well as an injector-rinsing solvent in order to improve peak shapes. Peak broadening and tailing often occur in reverse phase analysis of lipids due to slow mass transfer for these large lipophilic molecules. Barroso *et al* [185] used a fused-core C-8 phase to analyse the phospholipids and lysolipids in human bronchial lavage fluid. Hankemeier *et al* [186] developed Barroso's work and validated the method. Both of them used temperature control at 60°C in order to improve the peak shapes; however, the high risk of lipid oxidation due to high temperature was not assessed.

Normal phase methods can separate the lipids into different classes. Thus it is straightforward for researchers to determine how much PC, PE, PS or PI are present in the unknown samples rather than to know what fatty acid tails they have exactly. The stationary phases used in normal phase mode include pure silica gel or chemically modified silicas such as diol- or cyanopropyl phases. Lesnefsky *et al* [187] used silica as stationary phase and eluted by using hexane/propanol/potassium acetate/ethanol/acetic acid in order to characterise the phospholipids and lysophospholipids. The run time of 120 minutes was very long and detection was carried out by UV at 260nm. Malabolta *et al* [188] used a silica gel

column coupled with ESI –MS in negative mode to separate phospholipids. They added NH_4OH into the $\text{CHCl}_3/\text{MeOH}$ mobile phase in order to adjust the pH, which can easily cause bleeding of the silica column and block MS probes. Pang *et al* [189] used a diol column in order to measure the phospholipids in diabetes nephropathy. The method had a run time of 45 mins and peak shapes were poor for PC. Xu *et al* [190] also used a diol column to analyse phospholipid species in human blood. The run time was 45 mins and the lipids were detected by negative ion mode MS. Chloroform/Hexane/Isopropanol are frequently used in the mobile phase in normal phase methods and considered to be not very friendly with some of the rubber and plastic parts in the LC instrument. Also it is time-consuming switching from normal phase eluents to reverse phase mode [191]. Thus there are disadvantages to separating lipids by the classical normal phase mode.

In the current study several phases were investigated in reverse normal phase mode for their ability to separate phospholipids in different classes.

5.2 Materials and Methods

5.2.1. Materials

All phospholipid standards were purchased from Sigma-Aldrich, UK. Ammonium acetate and ammonium formate were purchased from Sigma-Aldrich, UK. Formic acid HPLC grade was purchased from VWR (Poole, UK). Methanol HPLC grade, acetonitrile HPLC grade and Isopropanol HPLC grade from ThermoFisher, U.K. Water was obtained from an in-house Milli-Q water purification station.

5.2.2. Chromatography

The LC system consisted of an Agilent 1100 instrument. ACE C-4 and Silica gel columns (3 mm x 150 mm x 3 μm) were obtained from HiChrom Ltd., Reading, U.K. The Zic-Hilic column (4.6 mm x 150 mm x 5 μm) was also obtained from HiChrom along with a ZicHilic guard column (20 mm X 2.1 mm i.d. 5 μm)

5.2.3 Mobile Phase Composition

Elution of the C-4 column was carried out with A: 0.02M ammonia formate in water (pH=6.2). B: methanol. The flow rate was at 0.4 ml/min and gradient was 0 mins 80% of B, 40 mins 90% of B, 50 mins 90% of B with a 10 minute equilibration time at the end of the run.

Elution of the Zic-Hilic column was carried out with A: 0.02M ammonium formate in water (pH=6.2). B: 0.02M ammonium formate in acetonitrile. The flow rate was 0.4 ml/min and the gradient was 0 mins 90% of B, 30 mins 95% B, 40 mins 95% of B.

Elution of the silica gel column was carried with A: 20% IPA in 0.02M ammonium formate (v/v). B: 20% IPA in acetonitrile (v/v). The flow rate was 0.3 ml/min and gradient was: 0-5mins 90% of B, 9 mins 70% of B, 13 mins 65% B, 23 mins 60% B, 28-30 mins 55% B, 31-40 mins 90% B.

5.2.4 Mass spectrometry

The LC system described above was coupled to an LCQ DECA mass spectrometer (ThermoElectron Co. U.K.), equipped with an electrospray ionisation (ESI) source. The ion spray voltage was set to 4.5KV in the positive mode and negative mode. The capillary voltage and temperature was set to 35V and 260°C respectively. Tube lens was 25 V for +ve mode and -20 V for -ve mode respectively. The machine was tuned with freshly prepared 16:0/16:0 PC solution. Data acquisition was controlled with Xcalibur 2.0. All lipids were measured by a full scan from 100 -1000 amu. The run time for data acquisition was 30 mins. In order to improve fragmentation performance, the run time was divided into 8 segments. Each lipid class (PC, PE, PI, IPC) occupied one segment. PI and IPC segments used a CID of 59, PE a CID of 24 and PC a CID of 40. These four segments used a MS³

data dependent scan with an activation Q of 0.18. Direct infusion was also carried into the LCQ DECA using the same parameters as above.

5.2.5 Extraction of Leishmania Samples

The number of cells in the flask were counted and the flasks were quenched flasks in a dry ice/ethanol bath at 0°C while monitoring the temperature with a digital thermometer (taking no longer than 1 min). The flask was transferred immediately into ice and kept in ice during extraction. Take the amount of culture necessary to get 4×10^7 cells was taken and transferred into pre-chilled Eppendorf tubes (kept in ice). The samples were spun down at 13000 rpm, 0°C for 10 min, supernatant was removed and washed with 1 ml of PBS. The sample was spun down again (13000 rpm, 0°C, 10 min), the supernatant was removed and the pellet washed with 1ml of PBS. The washing step with PBS was repeated and the supernatant was then thoroughly removed. Then 200 μ l of cold (0°C) chloroform/methanol (50/50 v:v) was added to the cell pellet and the sample was extracted for 1 hr in a thermomixer at 4°C (1400rpm). The sample was spun down (13000 rpm, 0°C, 10 min) and the supernatant was immediately transferred into vials with (glass) inserts. The supernatant was stored at -70°C until analysis. During analysis the sample was stored at 4°C in the autosampler tray and 10 μ l of sample was injected in the LC-MS system.

5.3 Results

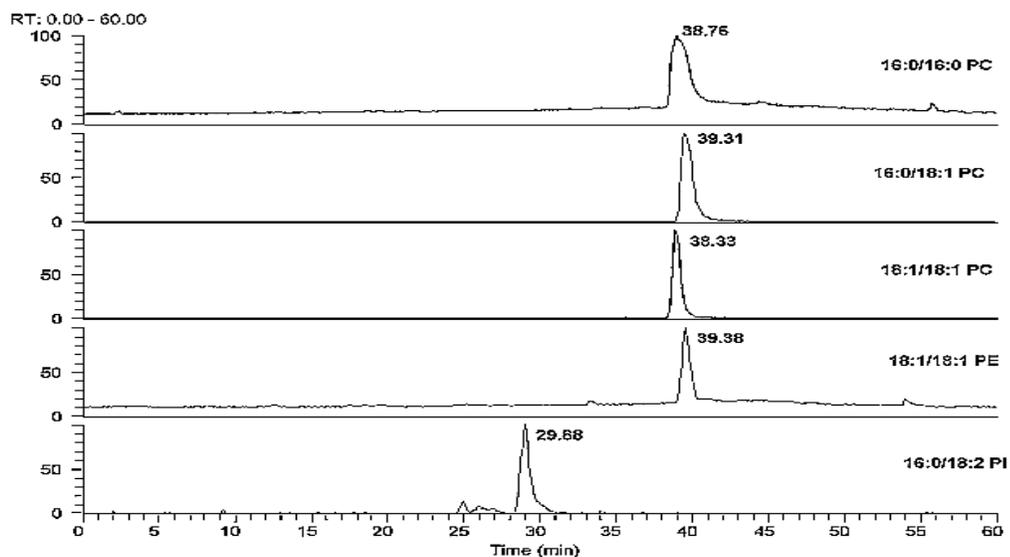
5.3.1. Characterisation of phospholipids by Infusion ESI-MS

All the standards for the four lipid classes were analysed in positive and negative mode by direct infusion. PC, PE, PS, PI can be detected in either ionisation mode. Among them, PC formed abundant positive ions $[M + H]^+$, while it formed very few $[M + \text{acetate}]^-$ or $[M + \text{formate}]^-$ ions in negative mode.

5.3.2 Analysis of Standards on the ACE C-4 Column

Initial experiments focused on optimizing the separation of the different phospholipid classes using an ACE C18 (150mm X 4.6 mm, 5 microns) column with a gradient between 0.02M ammonia formate and methanol. This method showed very broad peaks for phospholipids with extensive tailing probably due to the unfavourable mass transfer of such large lipophilic molecules between the mobile and stationary phases. A better chromatographic performance was obtained through the use of an ACE C4 column. Compared to the C-18, C-4 has a shorter acyl chain and thinner surface coating with 60% less of carbon load. Lipid molecules can diffuse through a short porous layer more rapidly thus providing faster mass transfer. Five phospholipids 16:0/16:0 PC, 16:0/18:1 PC, 18:1/18:1 PC, 18:1/18:1 PE and 16:0/18:2 PI were selected were tested as standards. Compared to the C18 column peak shapes were significantly improved (figure 5.1). Run time was reasonable for a 150 mm long column. The lipids 16:0/16:0 PC, 16:0/18:1 PC, 18:1/18:1 PC and 18:1/18:1 PE have very similar retention times and all elute at around 39 mins. The PI lipid eluted quite a bit earlier.

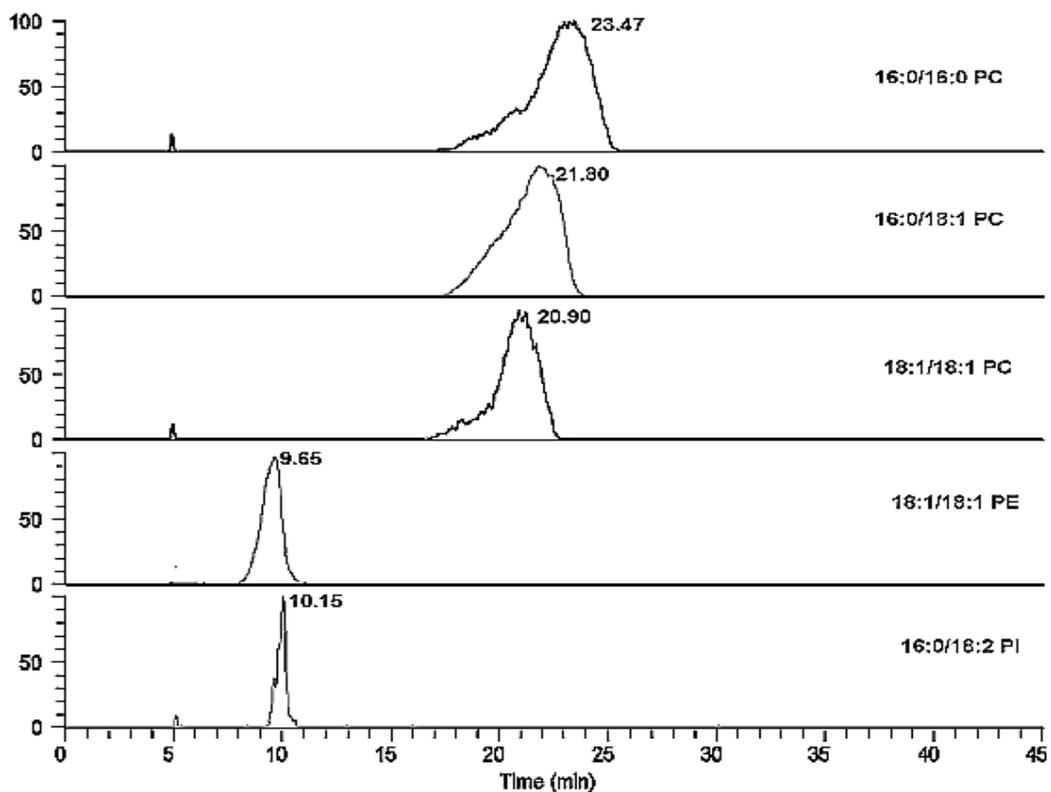
Figure 5.2 Standard phospholipids on an ACE C4 column. Mobile phase conditions as in section 5.2.2.



5.3.3 Analysis on a ZICHilic column

In principle, Hilic separates compounds by hydrophilic partitioning and electrostatic interactions. In this study, five phospholipids 16:0/16:0 PC, 16:0/18:1 PC, 18:1/18:1 PC, 18:1/18:1 PE, 16:0/18:2 PI were examined for their chromatographic behaviour. As seen in Figure 5.2, PE and PI have short retention times and good peak shapes. However, the PCs were retained on the HILIC column for a longer time of about 22 minutes. PC has a choline group with a quaternary amine and is therefore strongly retained by the anionic groups of the stationary phase. In LC-MS mode only ammonia is suitable as a counter cation to improve the elution characteristics and a more strongly basic counter ion might be more effective.

Figure 5.2 Analysis of phospholipid standards on a ZICHILIC column Conditions as in section 5.2.2.



5.3.4 Analysis of the Lipids on a Silica-gel Column

Silica gel has frequently been used in normal phase mode to separate lipids but as discussed in the introduction normal phase has the disadvantage that the solvents used, such as hexane, produce toxic vapours and can damage rubber and plastic parts of instruments. Five phospholipid standards 16:0/16:0 PC, 18:1/18:1 PE, 18:1/18:1 PS and 16:0/18:2 PI were analysed and as can be seen in figure 5.3 the peak shapes for all five lipids were good although the peak for the PS lipid was a little broad. The retention mechanism operating on the silica gel is open to some questions. In the case of the PC and PE lipids it is most likely the result of silanophilic interaction with the silanol groups in the phase which will be to some extent ionised at the pH of the mobile phase (*ca* 6.0 for NH₄OOCH solution). This would fit with PC lipids eluting later than PE lipids. However, there may be a purely HILIC type mechanism at work since the PI lipid peak, which is negatively charged, elutes at 5 minutes and at a flow rate of 0.3 ml/min. where the void time for the column is *ca* 2.5 min. At pH 6.0 the PS lipid is also predominantly negatively charged, since both the phosphate group and carboxylate group carry negative charges, yet it is retained by the column more strongly than the PE lipid.

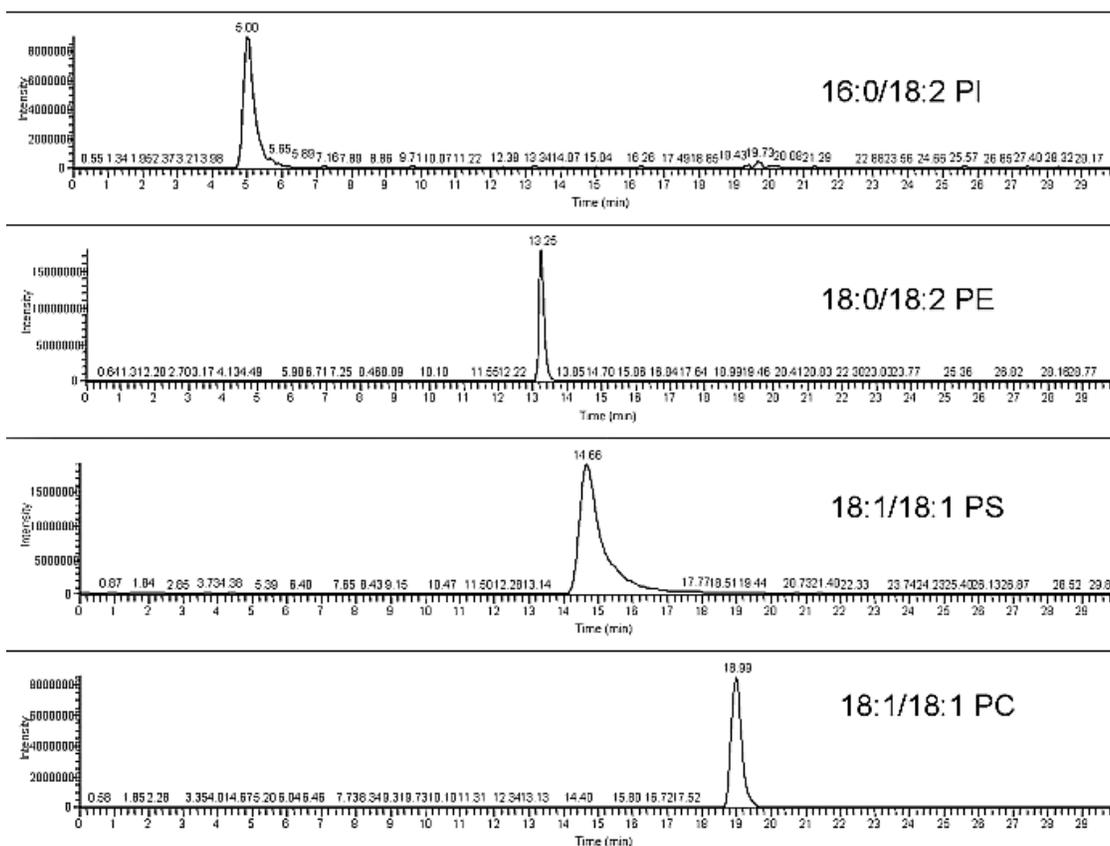
5.3.5 Analysis of *L.donovani* samples

L.donovani samples were then run on the silica gel column coupled with the LCQ DECA MS in negative ion mode using data dependent MS³ in order to determine the acyl groups in the PC lipids. The PI, PS and PE lipids gave diagnostic fragments in MS² mode. Figure 5.4 shows the profile obtained for the lipids in *L.donovani*, three lipid classes relating to the standards could be observed but there was no evidence of PS lipids. In addition a peak for inositol phosphate ceramides (IPC) could be observed.

Figure 5.3 Separation of lipid standards on an ACE silica column. Conditions as in section 5.2.2

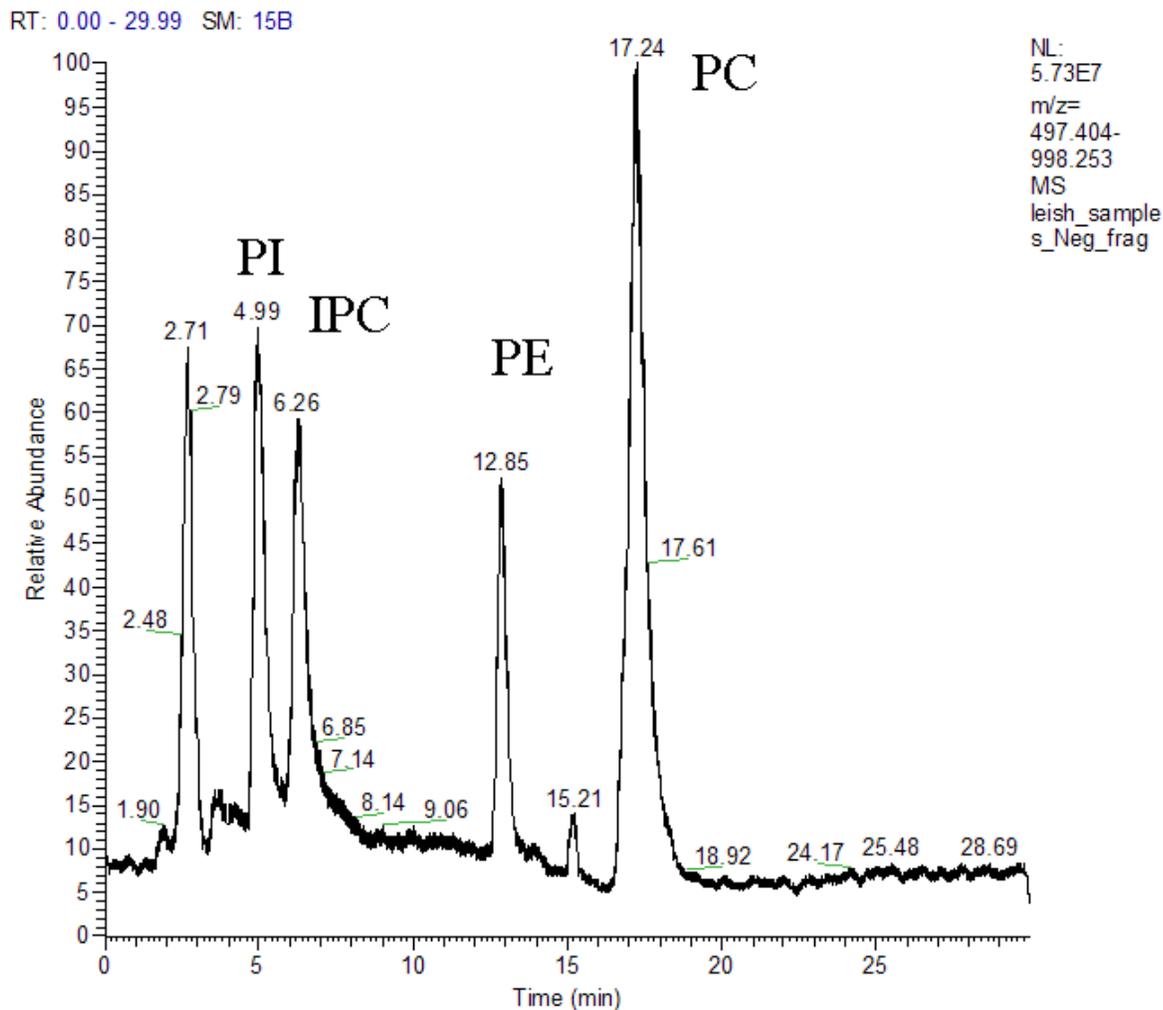
O:\20percentIPA_IN_Bothmix_gradient7

22/09/2009 01:07:52



Using the methodology 148 phospholipids could be identified in the sample (tables 5.1-5.4). The breakdown of classifications was: PC (38%), PI (22%), PE 23% and CPI (17%). Of these lipids 49% (the more intense peaks) could be classified with regard to the acyl substitution. Utilisation of targeted MS² and MS³ would be able to deliver a much larger % characterisation but would require a larger number of runs, the reported methodology delivers 49% characterisation of the acyl substitution in a single run in negative ion mode. The utility of MS³ in negative ion mode in characterising a PC lipid is illustrated in figure 5.5. The data dependant MS² results in the loss of 60 amu from the formic acid adduct of the molecular ion (m/z 830) of

Figure 5.4 TIC for Lipids extracted from *L.donovani* run in negative ion mode with data dependant MS². The sample was run on an ACE silica gel column, conditions as in section 5.2.2.



18:1, 18:1 PC due to loss of formic acid and a methyl group from the choline head group of the PC lipid. This results in an ion at m/z 770 which is picked up by the data dependent scan as the most intense ion in the MS² spectrum and fragmented further resulting in the formation of ions at m/z 606 and m/z 281 which result from the loss of $C_{17}H_{32}CO$ (ketene type fragment) from m/z 770 and from $C_{17}H_{33}COO^-$ respectively. In contrast, in positive ion mode the main fragment ion produced is at m/z 184 due to the phosphocholine head group, however, weak ions can be

observed which result from the loss of $C_{17}H_{32}CO$ and from the formation of $C_{17}H_{32}COH_2^+$. The PC peak also contains ions due to lysolipids which more or less co-elute.

Figure 5.5 MS^2 and MS^3 spectra of 18:1, 18:1 PC lipid at 40 V

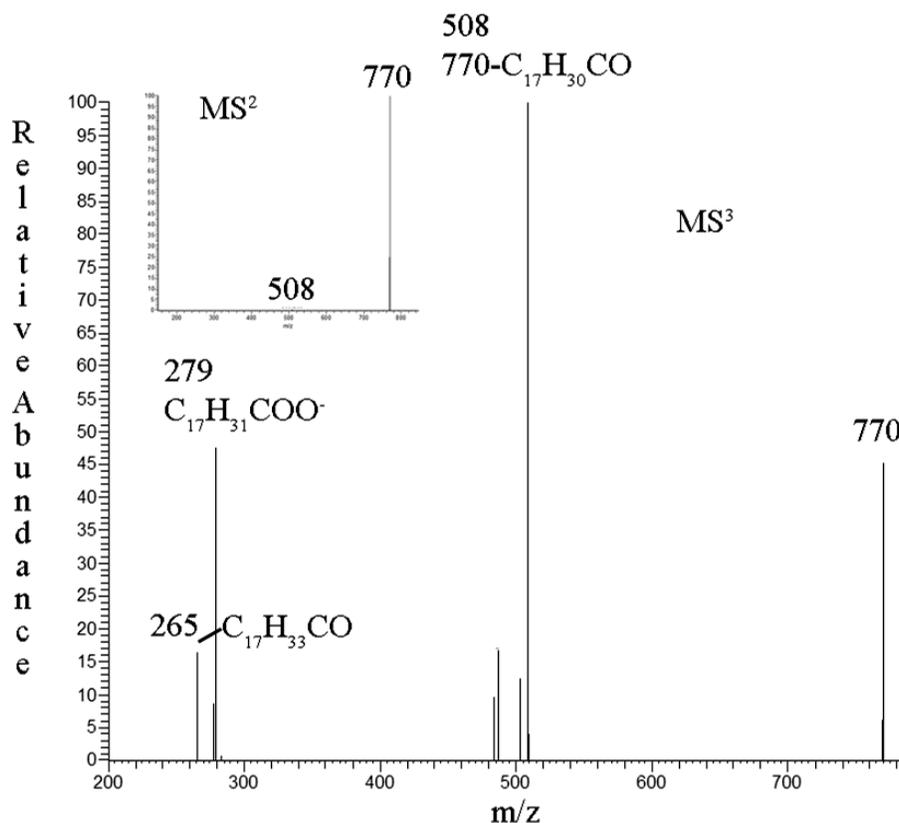


Figure 5.6 shows the MS^2 spectrum of 18:1/15:1 PE in negative ion mode. In this case the MS^2 spectrum is indicative of the acyl groups within the structure with prominent ions at m/z 478 due to loss of $C_{17}H_{32}CO$ from the molecular ion at m/z 698 and due to $C_{17}H_{33}COO^-$. The PE lipid fraction in leishmania is abundant in phospholipids containing a C15 chain.

PI which contains of a phosphatidylinositol nucleus readily yielded deprotonated anions in the negative ion mode. Figure 5.7 shows the MS^2 spectrum of 15:0/18:1 PI. The ions at m/z 553 and 577 arise from the losses of the linoleic acid (18:2) and palmitic acid (16:0), respectively. It can be observed that the m/z 553 ion is more

abundant than the m/z 577 ion, indicating that the 16:0 and 18:2 acyl substituents are located at *sn*-1 and *sn*-2 respectively. This is consistent with the notion that the ion reflecting the loss of the fatty acyl substituent at *sn*-2 should be more abundant than that reflecting the similar loss at *sn*-1 [192]. Ions at m/z 557 and m/z 539 are due to loss of $C_{17}H_{30}CO$ and $C_{17}H_{30}COOH$ respectively. Loss of inositol from the ion at m/z 539 ions gives rise to an ion at m/z 377. A small ion at m/z 241 is due to inositol phosphate. Such information rich spectra meant that many of the PI lipids in the leishmania sample could be completely assigned. Figure 5.8 shows the proposed fragmentation pathways for PI lipids.

Figure 5.6 the MS² spectrum of 18:1/15:1 PE in negative ion mode at 40V.

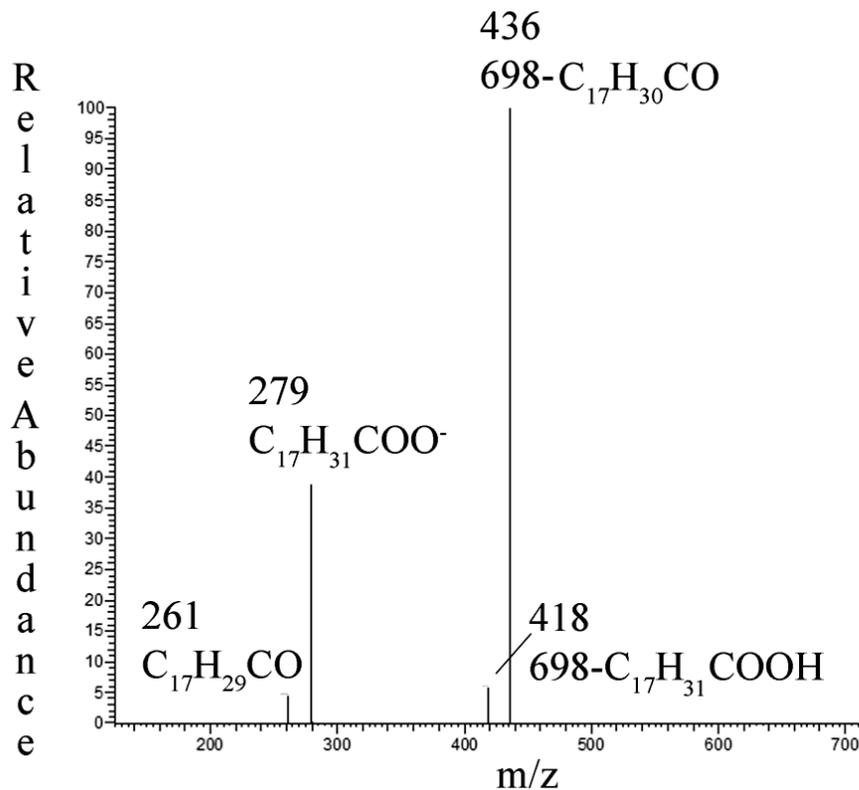
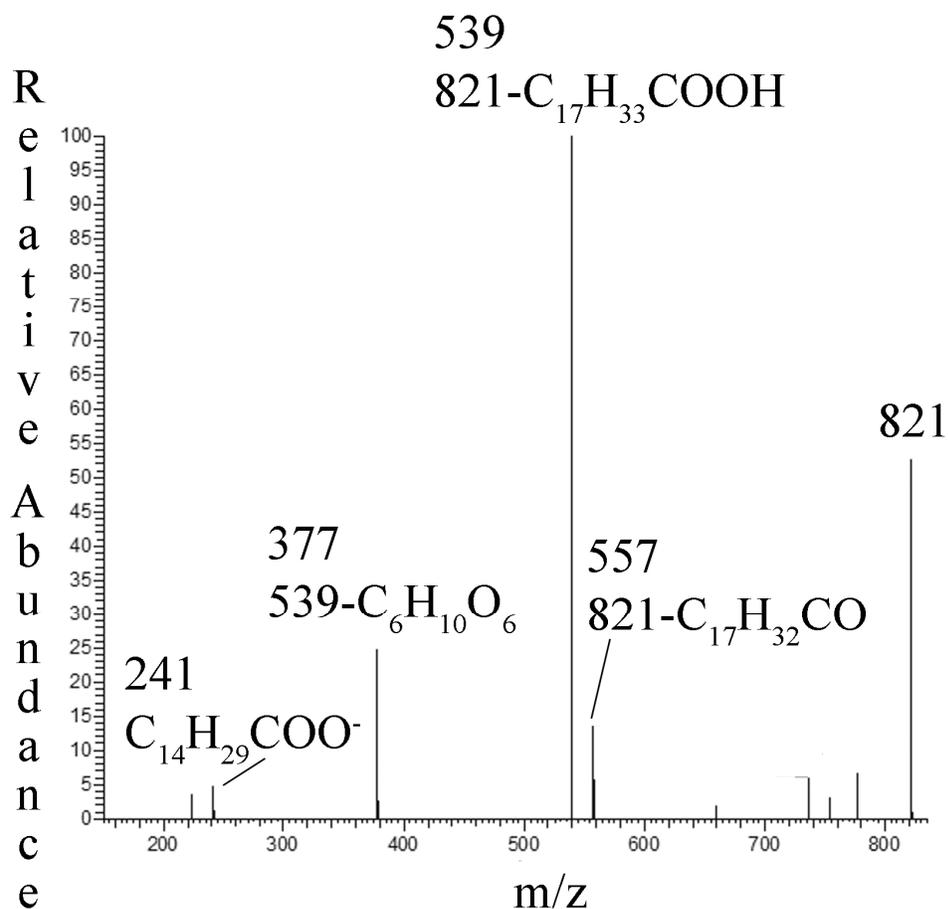


Figure 5.7 Fragmentation of 15:0/18:1 (palmitoyl linoleyl) PI in negative mode

The fourth major group of lipids in the leishmania extract were due to inositol phosphate ceramides (IPC). These compounds along with GPI lipids are important anchors for the surface glycoproteins of trypanosomatids. It is difficult to get an exact structural i.d. for these compounds since both the sphingosine portion of the molecule and the fatty acid acylating the amino group of the sphingosine base can vary. Figure 5.9 shows the MS² spectrum of the most abundant IPC compound in the extract which had a molecular ion at m/z 778 in negative ion mode and would correspond to palmityl sphingosine inositol phosphate.

The MS² spectrum of the IPC lipids all show an ion at m/z 241 which is due to the inositol phosphate portion of the molecule and in figure 5.10 an ion at m/z 616 is due

to neutral loss of inositol- H_2O from the molecular ion and the ion at m/z 598 results from loss of the complete inositol moiety. The modifying fatty acid is attached to sphingosine via a relatively strong amide bond thus there are no fragments indicative of the fatty acid substituents. The other abundant IPC in the sample has a molecular ion of m/z 806 which corresponds to the sphingosine core being modified with stearic acid.

Figure 5.8 Proposed fragmentation pathways for PI

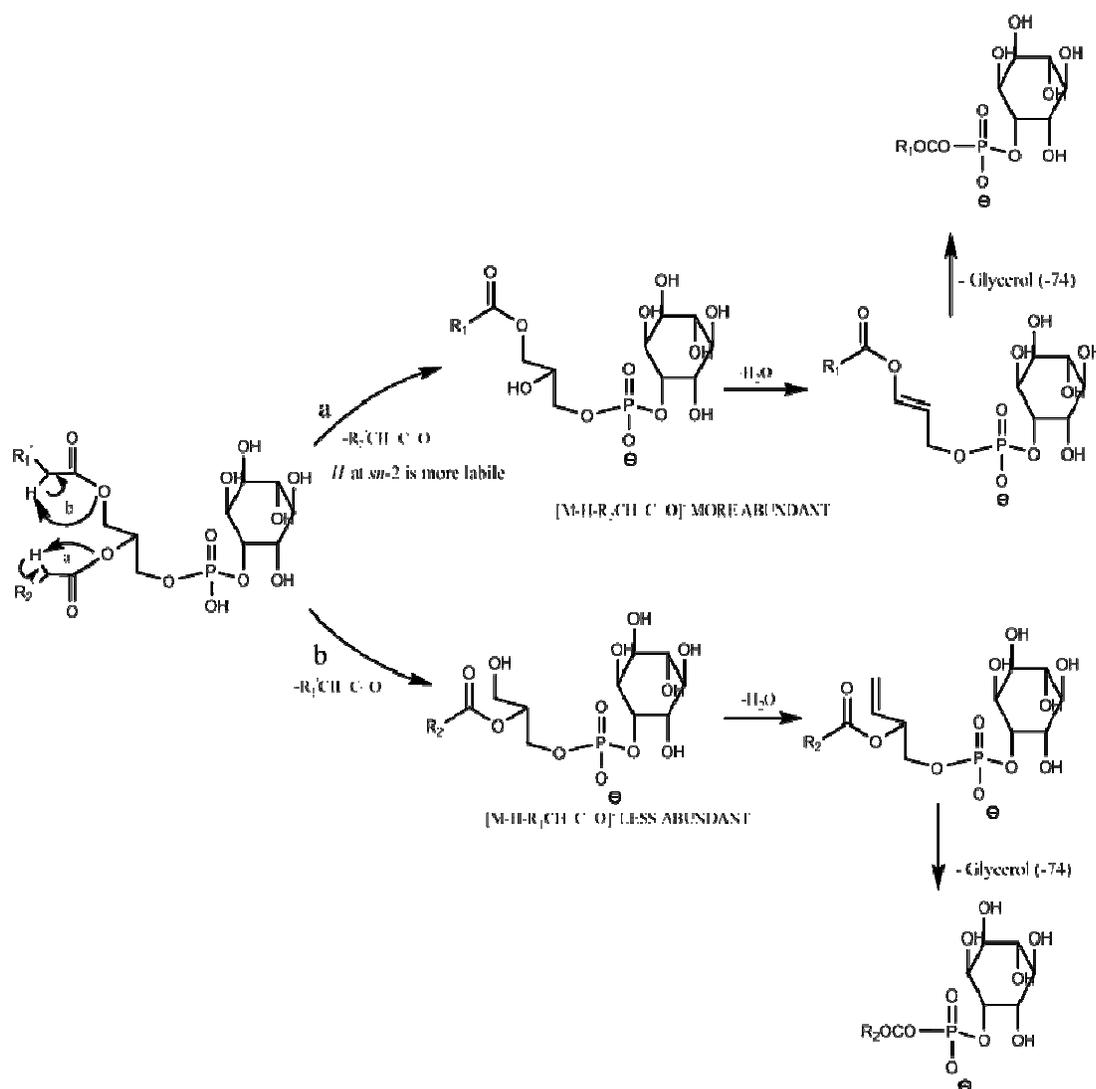
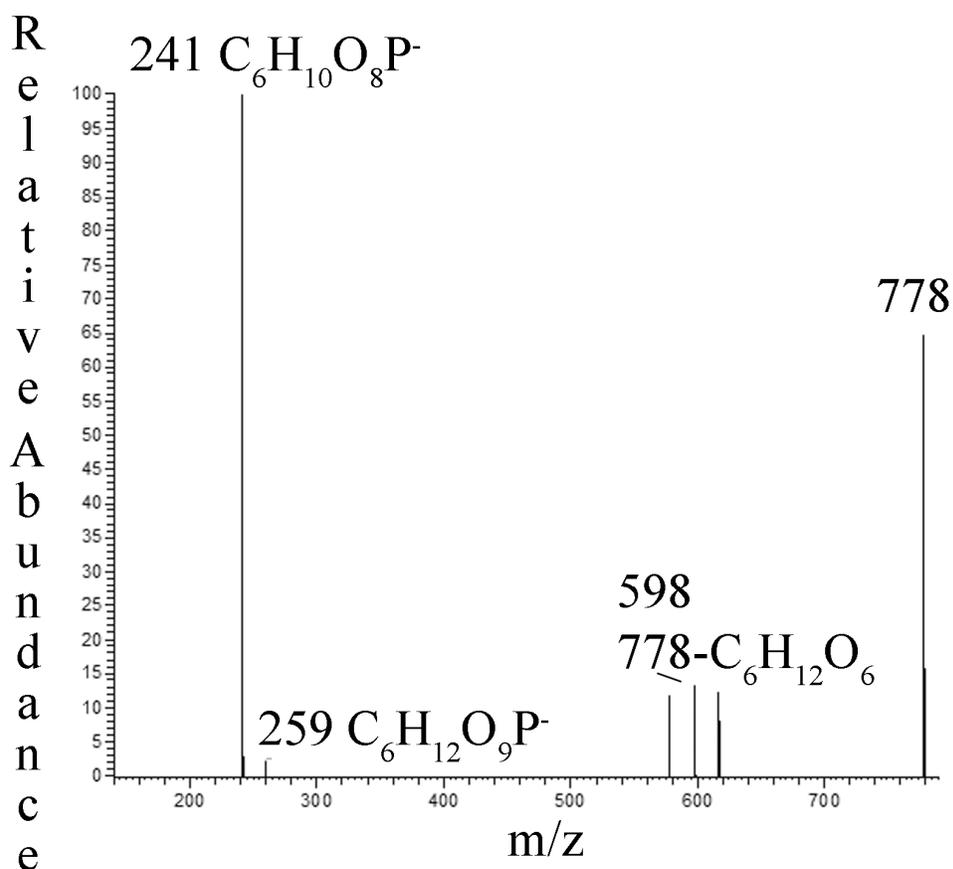


Figure 5.10 Fragmentation of the most abundant IPC lipid in leishmania.**Table 5.1** PC lipids extracts from *L.donovani* eluting in the peak at 17-19 min.

PC -ve	PC +ve	Comp.	Negative Ion	PC -ve	PC +ve	Comp.	Negative Ion
750	706	30:0		-	800	EL 38:2	
754	710	34:5		846	802	EL 38:1	
760	716	EL (32:2)		850	806	22:6, 16:0	480= 790- $C_{21}H_{30}CO$
762	718	EL (32:1)		852	808	22:5, 16:0 and 20:3,18:2	480 = 792- $C_{21}H_{32}CO$ 504=792- $C_{19}H_{32}CO$
770	726	32:4		854	810	38:4	
772	728	32:3		856	812	20:2,18:1	532=796- $C_{17}H_{32}CO$
774	730	32:2		858	814	20:2,18:0	508=798- $C_{19}H_{34}CO$
776	732	32:1		860	816	20:1,18:0	508=800- $C_{19}H_{36}CO$

778.32	734.32	16:0,16:0	m/z 480 = 718-	866	822	EL 40:5	
C ₁₅ H ₃₀ CO							
786	742	EL (34:3)		868.	824	EL 40:4	
788	744	EL (34:2)		870	826	EL 40:3	
790	746	EL (34:1)		872.	828	22:6,18:3	552 = 812- C ₁₇ H ₂₈ CO
792	748	33:0		874	830	22:6,18:2	552=814- C ₁₇ H ₃₀ CO
794	750	34:6		876	832	22:6,18:1	552=816- C ₁₇ H ₃₂ CO
796	752	34:5		878	834	22:6,18:0	552=818- C ₁₇ H ₃₄ CO
798	754	34:4		880	836	22:5,18:0	510=820- C ₂₁ H ₃₀ CO
800	756	34:3		882	838	40:4	
802	758	16:0,18:2	m/z 504 = 742-	-	840	40:3,	
C ₁₅ H ₃₀ CO							
804	760	16:0,18:1	m/z 506=744-	886	842	40:2,	
C ₁₅ H ₃₀ CO							
806	762	16:0,18:0	m/z 508=746-	-	844	40:1,	
C ₁₅ H ₃₀ CO							
814	770	EL 36.3		-	846	40:0,	
820	776	EL 36:0		896	852	EL 42:4	
822	778	36:6		898	854.	EL 42:3	
824	780	18:2,18:3	504=764- C ₁₇ H ₂₈ CO	916.34	872.34	42:1,	
826	782	18:2,18:2	504=766- C ₁₇ H ₃₀ CO	918.33	874.33	42:0,	
828	784	18:3,18:0	508=768- C ₁₇ H ₂₈ CO	LyPC	LyPCC		
830	786	18:2,18:0	508=770- C ₁₇ H ₃₀ CO	564.	520	18:2	279
832	788	18:1,18:0,	508=772- C ₁₇ H ₃₂ CO	566.16	524.27	18:1	281
834	790	18:0,18:0,	508=774- C ₁₇ H ₃₄ CO	568.17	526.28	18:2	283
840	796	EL 38:4		652	608	24:0	367

Table 5.2 Phosphoethanolamine lipids

PE -ve	PE+ve	Comp.	Negative ion	PE -ve	PE +ve	Comp.	Negative ion
696	698	15:2,18:2	279	738	740	36:4	
698	700	15:1,18:2	436=698- C ₁₇ H ₃₀ CO	740	742	36:3	
700	702	15:0,18:2	438=700- C ₁₇ H ₃₀ CO	742	744	17:1,19:1	464=742- C ₁₈ H ₃₄ CO
702	704	15:0,18:1	438= 702- C ₁₇ H ₃₂ CO	744	746	17:0,19:1	466=742- C ₁₈ H ₃₄ CO

710	712	34:4		-	748	36:0
712	714	16:1,18:2	450=712- C ₁₇ H ₃₀ CO	-	750	EL 38:6
714	716	16:0,18:2	452=714- C ₁₇ H ₃₀ CO	750	752	EL 38:5
714	716	15:1,19:1	436=714- C ₁₈ H ₃₄ CO	752	754	EL 38:4
716	718	34:1		756	758	EL 38:2
718	720	34:0		758	760	EL 38:1
726	728	17:1,18:2	464=726- C ₁₇ H ₃₀ CO	762	764	38:6,
728	730	16:1,19:1	436=714- C ₁₈ H ₃₄ CO	764	766	38:5,
728	730	17:1, 18:1	464=728- C ₁₇ H ₃₂ CO	766	768	38:4,
730	732	16:0,19:1	452=730- C ₁₈ H ₃₄ CO	768	770	38:3,
732	734	35:0		770	772	38:2,
-	736	36:6		772	774	38:1,
-	738	36:5		774	776	38:0,

Table 5.3 Phosphoinositol lipids

PI -ve	Comp.	Negative ion fragments	PI -	Comp.	Negative ion
815	15:0,18:4	539=815- C ₁₇ H ₂₇ COOH	859	36:3	
817	15:0,18:3	539= 817- C ₁₇ H ₂₉ COOH	861	18:0,18:2	581= 861 - C ₁₇ H ₂₉ COOH
819	15:0,18:2	539= 819- C ₁₇ H ₃₁ COOH	863	18:0,18:1	581= 863 - C ₁₇ H ₃₃ COOH
821	15:0,18:1	539= 821- C ₁₇ H ₃₃ COOH	865	18:0,18:0	581= 863 - C ₁₇ H ₃₃ COOH
823	15:0,18:0	539= 823- C ₁₇ H ₃₅ COOH	867	37:5	
831	16:0,18:3	553= 831- C ₁₇ H ₂₉ COOH	869	37:4	
833	16:0,18:2	553= 831- C ₁₇ H ₃₁ COOH	871	37:3	
835	16:0,18:1	553= 831- C ₁₇ H ₃₃ COOH	873	37:2	
837	16:0,18:0	553= 833- C ₁₇ H ₃₅ COOH	875	37:1,	
845	35:3		877	37:0	

847	17:0,18:2	567= 847- C ₁₇ H ₃₁ COOH	881	38:6
849	17:0,18:1	567= 849- C ₁₇ H ₃₃ COOH	883	38:5
851	17:0,18:0	567= 851- OH	885	38:4
853	36:6		887	38:3
855	36:5		889	38:2
857	36:4		907	40:7
-	36:5		909	40:6
857	36:4			

Table 5.4 Phosphoinositol ceramide

750	d 32:1	531=750-C ₁₇ H ₃₅ CO
768.84	t 32:0	241(IP-H ₂ O-H), 531=768-FA(256)+18
778.89	d16:1,18:0	512 = 778-C ₁₇ H ₃₅ CO
780.71	d16:0,18:0	223 (IP-2H ₂ O), 241.18 (IP-H-H ₂ O), 512=778-FA(FA284)+18, 616(778- inositol-H ₂ O)
792.67	t16:0,18:2	241.18(IP-H-H ₂ O), 259(IP-H), 531=796-(FA284)+18, 616.50(796- Inositol)
796.76	t16:0,18:0	241.18(IP-H-H ₂ O), 259(IP-H), 531=796-(FA284)+18, 616.50(796- Inositol)
802.81	d18:2,18:1	241(IP-H ₂ O-H),538=802-FA(282)+18
804.45	d18:2,18:0	241(IP-H ₂ O-H),538=806-FA(284)+18
806.57	d18:1,18:0	241(IP-H ₂ O-H),540=806-FA(284)+18
808.55	d18:0,18:0	241(IP-H ₂ O-H),542=808-FA(284)+18
820.72	d19:1,18:0	241(IP-H ₂ O-H),259(IP-H)(554=820- FA(284)+18
822.55	d19:0,18:0	241(IP-H ₂ O-H),259(IP-H)(556=822-

		FA(284)+18
830.56	d22:3,16:0	241(IP-H2O-H),592=830-FA(256)+18
832.71	38:2,	
834.51	d22:0,18:1	241(IP-H2O-H), 259 (IP-H), 571=834- FA(282)+18
836.63	d22:0,18:0	241(IP-H2O-H), 259 (IP-H), 570=836- FA(286)+18
842.52	41:4,	
846.91	d23:0,18:2	241(IP-H2O-H), 585=846-FA(280)+18
848.61	d23:0,18:1,	
850.34	d23:0,18:0	241(IP-H2O-H), 584=850-FA(284)+18
864.66	d24:0,18:0	241(IP-H2O-H), 599=864-FA(284)+18
874.08	43:2,	
878.50	43:0,	
906.65	46:2,	
914.33	46:3,	

5.4. Conclusions

Although it is possible to obtain structural information on lipids using direct infusion methods there are distinct advantages in carrying out a chromatographic separation of different lipid classes. These include avoidance of ion suppression effects and removing overlap between isobaric species. The current method gives a clear separation of five lipid classes using a commonly available column without the need for using environmentally and mass spectrometry unfriendly solvents. The main purpose of the work was to develop such a separation method but without any particular attempt to optimise structural elucidation of the different phospholipids it was possible observe 148 lipids and characterise many of these. There some

indication that the method also can be used for the analysis of triglycerides which elute around the void volume of the silica gel column. Use of a short C18 guard column in series with the silica gel column would move them slightly beyond the void volume of the column thus minimizing the risk of ion suppression. The next stage of this study will be to use the separation system in conjunction with the Orbitrap.

Chapter 6

A Metabolomic Study of *Trichomonas vaginalis* and *Trichomonas foetus* Using High Resolution Mass Spectrometry and Hydrophilic Interaction Chromatography

6.1 Introduction

6.1.1 *T.vaginalis* Background

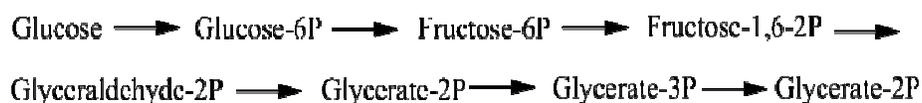
T.vaginalis is responsible for the sexual transmitted disease (STD) trichomoniasis [193]. The organism adheres to the epithelial cells in urogenital tract and this is responsible for its pathogenesis. Cysteine proteinase is responsible for the adherence of the organism to the epithelial cells. It manifests in women as vaginitis, increased vaginal secretion, lower abdominal pain and edema and in men it mainly causes proctitis and urethritis [194]. This disease is common worldwide and more found in the inner cities of the America. From current reports it has been noted that around 170 million people are infected with the disease per year. More commonly it is seen in students and young people. Trichomoniasis cause adverse effects on the pregnancy and increases the risk of HIV infection [195].

The currently used treatment for trichomoniasis is metronidazole (prodrug) in USA and in other parts of the world. After metronidazole enters into the cell through diffusion, it is activated through the hydrogenosome of the organism via reduction of the nitro group of the drug by an anaerobic pyruvate ferredoxin oxidoreductase enzyme into a nitro radical which is cytotoxic [196]. There is a chance of treatment failure due to low amounts of zinc, low bioavailability of the drug in vagina and bioinactivation of the drug. Resistance to this drug is currently at high levels. It has effects on both aerobic and anaerobic metabolism. In aerobic metabolism transcription of the ferredoxin gene may be reduced by mutational changes. In anaerobic metabolism level of pyruvate ferredoxin oxidoreductase and hydrogenase

may be down regulated, both of these changes result in inactivation of the drug [194, 195, 197, 198]. In addition, tinidazole can be used as an alternative and has been proven more effective requiring a lower dose in comparison to metronidazole. Other nitroimidazoles like ornidazole, secnidazole, carnidazole, flunidazole, nimorazole are also effective. *T. foetus* is similar to the *T.vaginalis* but has some morphological differences. *T.vaginalis* is the human pathogen whereas *T.foetus* infects cattle.

6.1.2 Metabolism of *T.vaginalis*

Energy and carbohydrate metabolism in trichomonas occurs in two parts. One is in the cytosol and another is in the hydrogenosome. Like other eukaryotics, respiratory metabolism differs in the trichomonas since it does not have mitochondria but has an organelle instead called hydrogenosome. Unlike the mitochondrion, the hydrogenosome does not have cristae or cytochromes and lacks DNA. Production of hydrogen is a characteristic of the hydrogenosome. Since there is incomplete oxidation of the glucose in trichomonas, carbohydrate metabolism in trichomonas is fermentative process in both aerobic and anaerobic conditions. Glycolysis happens in the classic way in the cytoplasm via conversion of glucose in the following manner.



Finally glycerate-2P is converted to pyruvate via formation of phosphoenolpyruvate. In this pathway several enzymes produce energy via substrate level phosphorylation. Pyruvate is formed in the cytoplasm via carbohydrate metabolism and can enter into the hydrogenosome. It is converted into the acetyl co-enzyme via decarboxylation by the hydrogenosomal enzyme pyruvate ferredoxin oxidoreductase (PFOD). Malate is also converted into pyruvate via decarboxylation

and dehydrogenation and in the mean while results in production of the acetyl coenzyme A. Coenzyme A is then released from acetyl Co-A to form the acetate. Coenzyme A is transferred to succinic acid and again it gets released from it via a succinate thiokinase enzyme which will cause the production of the ATP. Electrons released from the PFOD are transferred to ferredoxin which results in the formation of the hydrogen via coupling of the electrons to the two hydrogen ions by a hydrogenase enzyme and reaches to the end of the respiratory chain. Metronidazole enters into the hydrogenosome via non ATP required diffusion and then competes with the hydrogenase enzyme for the electrons from ferredoxin. This causes the reduction of the prodrug metronidazole and an active/toxic nitro radical form of the drug is produced.

Cholesterol, phosphatidylethanolamine, phosphatidylcholine and sphingomyelin are the main phospholipids found in the *T.vaginalis*. Formation of lipids in the membrane occurs by incorporation of glycerolphospholipids in almost all phospholipids. As there is deficiency of the many enzymes involved in the synthesis of essential and complex phospholipids. *T. vaginalis* is dependent on an exogenous supply. Synthesis of the glycolipids and glycoposphosphingolipids has been found to occur. Inositol phosphoceramide has been observed to be present in *T.vaginalis*. The exact role and metabolism of lipids in *T.vaginalis* remains to be elucidated.

Nucleotide metabolism efficiency is low in trichomonas. Purine salvage occurs by nucleoside phosphorylase and kinase and at the same time phosphoribosyl transferase and nucleoside kinase helps to recover pyrimidine. There are two carriers for nucleotide transport, one is believed to accommodate adenosine and pyrimidine nucleotides, and second has a high affinity for uridine and guanosine transport.

When carbohydrates are not available for the energy cycle, amino acids are used extensively as a substitution. Arginine, leucine, threonine and cysteine metabolism have all been found. Arginine dehydrolase is a major pathway for energy production. Putrescine is the end product of this pathway and at last excreted. Putrescine is produced from the ornithine via ornithine decarboxylase. Ornithine is produced from arginine by conversion of arginine to citrulline through arginine deaminase and further from citrulline by the action of the ornithine carbamoyltransferase which produces carbamoyl phosphate and results in ATP formation through removal of ammonia. Ornithine is also formed from proline via two steps, first is the production of pyrroline 5 carboxylate with the enzyme pyrroline 5 carboxylate reductase(6) and further action of lycine/ornithine aminotransferase(5) enzyme. Ornithine can also be directly produced from arginine and proline by the action of enzymes arginase (7) and ornithine cyclodeaminase(8) respectively. Main role of this alternative pathway is to synthesise putrescine independently of dihydrolase pathway which is essential for host parasite interaction.

In cysteine biosynthesis in *T.vaginalis* phosphoglycerate dehydrogenase converts phosphoglycerate to phosphohydroxy pyruvate. In the next step O-phosphoserine aminotransferase converts hydroxypyruvate to O-phosphoserine, and via the action of cysteine synthetase it gets converted to cysteine by combining with hydrogen sulphide produced from methionine via several steps. The first step is formation of S-adenosine methionine (SAM) via the action of an enzyme called methionine adenosine transferase. SAM sequentially is converted into S-adenosyl homocysteine through the action of S-adenosylmethionine methyl transferase. Finally hydrogen sulphide is produced by the action of methionine lyase (MGL). H₂S can also originate from 3-mercaptopyruvate via the enzymatic action of mercaptopyruvate sulphur transferase (MST). This hydrogen sulphide combines with the o-

phosphoserine via the action of cysteine synthase (CS) to form cysteine. Both CS and MGL which are involved in cysteine production in *T.vaginalis* are not present in humans so they could be a good target for drug action.

Spermine, spermidine and putrescine have effects on cell proliferation and differentiation. Degradation, synthesis, uptake and excretion of cationic polyamines are regulated by the polyamine metabolic pathway [199]. Since they are positively charged at normal biological pH, they bind to high molecular weight compounds like phospholipids, nucleic acids and proteins and stabilise these molecules. Formation of the higher polyamines occurs by the transfer of the aminopropyl group from decarboxylated S-adenosylmethionine to putrescine or spermidine forming methylthioadenosine and spermidine or spermine respectively. Thus putrescine (2 moles) obtained from the arginine pathway is excreted with the uptake of the spermine.

The aim of the current study was to carry out metabolomic profiling of *T. vaginalis* and to compare its metabolism with that of *T.foetus*, which is relatively unknown, by using high resolution mass spectroscopy and hydrophilic interaction chromatography. Also to find out differences in the level of metabolites and metabolic pathway patterns between *T.vaginalis* and *T.foetus*.

Figure 6.1 Energy productions in *T.vaginalis* via metabolism of arginine

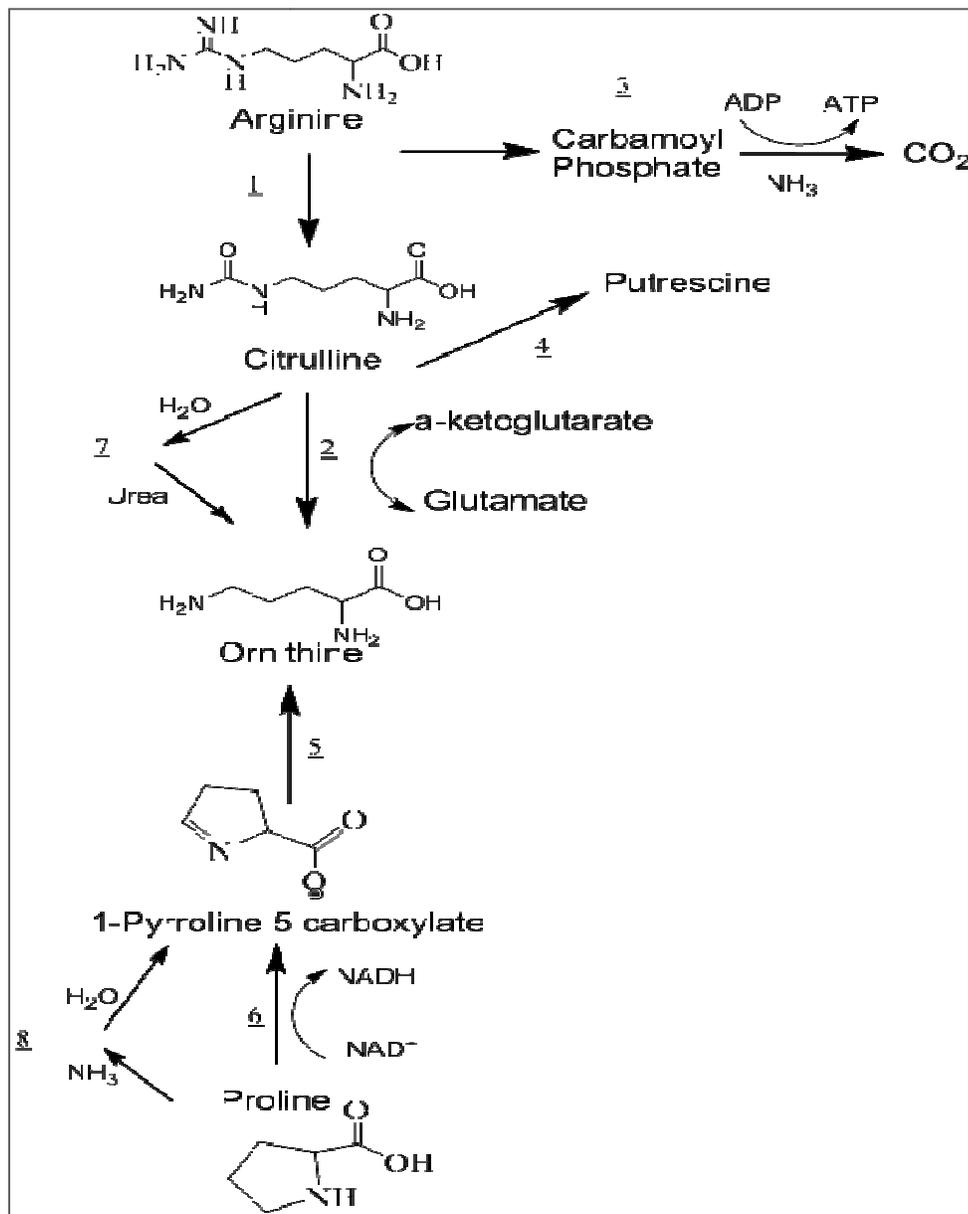
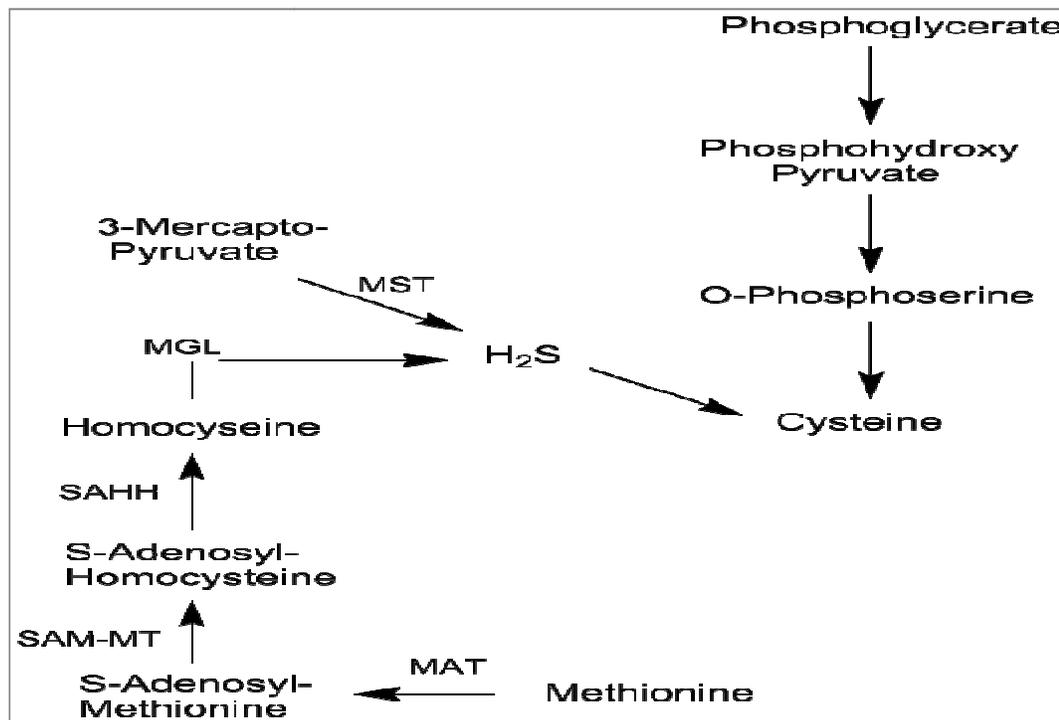


Figure 6.2 Biosynthesis of cysteine in *T.vaginalis*



6.2 Materials and Methods

6.2.1 Chemicals and reagents

Chemicals were obtained from the following sources. Phosphate buffered saline, pH 7.4 (PBS) was obtained from (GIBCO, UK). Extraction buffer which was comprised of 80% ethanol (Sigma-Aldrich, Dorset, UK) and 20 mM HEPES (GIBCO, UK). Acetonitrile, methanol and water of HPLC grade (from Fisher Scientific, Leicestershire, UK). Formic acid (98%) was purchased from VWR International Ltd. (Lutterworth, UK). The Orbitrap calibration solution kit was obtained from Thermo Finnigan, Hemel Hempstead U.K.. Other standards such as amino acids were also obtained from Sigma Aldrich.

6.2.2 Sample Preparation

Cells were grown in modified Diamond's Medium containing 10% (v/v) heat-inactivated Horse Serum (Invitrogen) in sealed culture flasks with little gas phase.

Modified Diamond's Medium

Trypticase.....20g (BD Biosciences)

Yeast Extract.....10g (Sigma)

Maltose.....5g (Sigma)

Ascorbic acid.....1g (Sigma)

KCl.....1g

KHCO₃.....1g

KH₂PO₄1g

K₂HPO₄0.5g

FeSO₄.7H₂O.....0.1g

Made up to 1 litre. pH 6.3 - 6.4 with 1 M HCl. Sterilised by autoclaving.

Culturing and extraction conditions were as follows:

Inoculate 90 ml cultured with 5 x 10⁵ cells / ml of the parasite.

Grew culture for 16 hrs at 37°C (to stationary phase).

Harvested cells by centrifugation at 1000 rpm (750 x g) for 15 minutes at 4°C.

Removed supernatant and resuspended cells in 12 ml cold PBS.

Stored cells at 4°C and counted cells using a haemocytometer.

Centrifuged for at 1000 rpm for 15 minutes at 4°C.

Removed supernatant.

Resuspended cells in hot HEPES/Ethanol (80°C) (HEPES/Ethanol = 80% Ethanol in 20 mM HEPES pH 8.0 PBS = Phosphate buffered saline).

Heated for 3 minutes at 80°C.

Centrifuged at 13000 rpm (13,000 x g) for 10 minutes at 4°C.

Transferred supernatant to a fresh tube.

Centrifuged at 13000 rpm for 10 minutes at 4°C.

Stored supernatants at -80°C until ready for HPLC/MS

6.2.3 Instrumentation

LC-MS was carried out by using a Surveyor MS pump plus and an Autosampler plus . The HPLC was interfaced to LTQ Orbitrap MS system. The system was controlled by Xcalibur software, version 2.0 (Thermo Electron Corporation). The LC-MS system was run in binary gradient mode in the +ve ion ESI mode. , The needle voltage was 4.00 kV, and the sheath gas and auxiliary gases were set at 50 and 20 respectively (arbitrary units), the heated capillary temperature was 200°C. The HPLC was operated in gradient mode. Solvent A was 0.1 %v/v formic acid/water and solvent B was 0.1 %v/v formic acid/acetonitrile (0.1% FA/ACN). The flow rate was 0.3 ml/min. A ZIC-HILIC column of (150 × 4.6 mm with 5 µm particle size , HiChrom, Reading U.K.) was used for all analyses. The gradient system was started at 80% B /20% A and programmed to reach 60% A /40% B after 8 minutes. The ratio was maintained at 60% A: 40% B for a further 12 min. Subsequently, the ratio decreased to 40% A within 6 min, and then that ratio was kept for 4 min. The ratio was then returned to A:B 20:80 in 1 min and maintained at this ratio for the rest of the run time (36 min).

6.2.4 Data Processing

Finally, two packages were used to evaluate the results: SIMCA-P software version 11.0 (copyright © Umetrics AB) and SIEVE version 1.1 (Thermo Fisher Scientific Inc., San Jose, USA).

6.3 Results and Discussion

6.3.1 Normalisation of Data

Normalization is an important step to used remove the variation in the level of metabolites detected. This variation may occur due to inhomogeneity of the sample, variations in sample preparation, instrumental effects or it may also be due to phenotype variation in between two groups in a metabolomics study. Data from

manual integration and Picker Shell software results showed differences in the level of many metabolites in between samples of the two strains of trichomonas. Table 6.1 shows such effects on the levels detected for the amino acids.

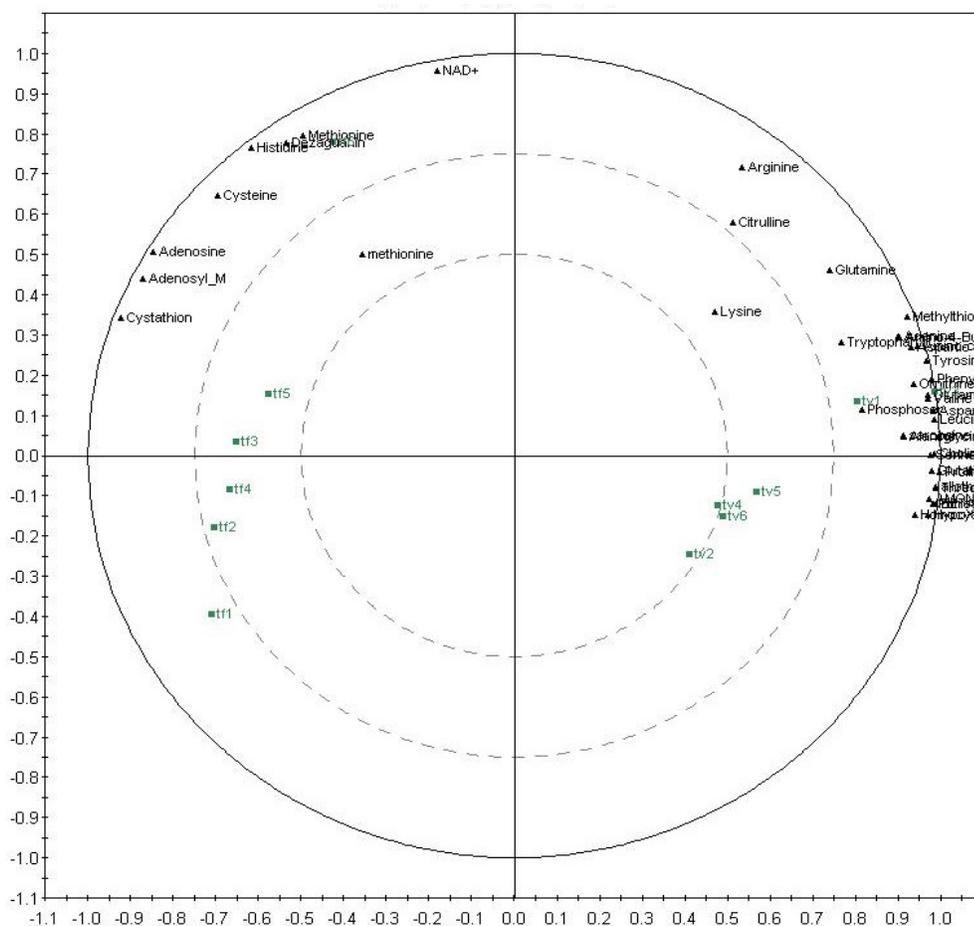
Table 6.1 shows means of the peak intensities of amino acids in different samples of *T.vaginalis* and *T.foetus*. Around 70% of total amino acids show the ratio of their peak intensities in *T.vaginalis* over *T.foetus* around 2 except threonine, cystathionine, homocysteine, lycine, cysteine and methionine. The overall the total intensities for aminoacids in *T.vaginalis* is almost double than that in *T. foetus*. The mean of ratio of all the amino acids is also 2.1. This clearly indicates that relative levels of metabolites is effected by some factor as their ratio is consistent for many amino acids. This is probably because of the size difference of the cells of *T. vaginalis* and *T. foetus*. The size of *T.vaginalis* cells is much bigger than those of *T.foetus*. Thus this could lead to a false interpretation of the data. In order to remove this error from the results normalization was done. In order to achieve this, the relative intensity for each the amino acids and other compounds were calculated by dividing peak intensity of a particular metabolite with total of peak intensities of all the metabolites. Figure 3.2 shows the PCA plot from the SIMCA- P™ Software (version 11.0 by

Table 6.1 Amino acids peak intensities without normalization of the data from manual integration of peaks (n of replicates= 3)

NAME	MEAN T.V	R.S.D T.V(%)	MEAN T.F	R.S.D T.F (%)	RATIO T.V/T.F
Valine	270645.2	8.31	131477.2	11.84	2.1
Threonine	11823.6	17.75	1626.2	12.10	7.3
Cystathionine	0		159.5	23.01	0
Homocysteine	2976.2	20.48	357.6	16.36	8.3

Proline	84968.6	14.63	30965.1	15.61	2.7
Leucine	829106.3	13.38	384294.8	13.84	2.2
Asparagine	973.5	15.25	333.7	30.18	2.9
Glycine	5284.4	15.71	2299.2	11.86	2.3
Serine	3821.9	19.62	1342.7	7.19	2.8
Phenylalanine	187985.7	9.90	94826.8	24.47	2.0
Glutamate	75837.1	15.95	39611.8	17.74	1.9
Tyrosine	17284	12.31	9103.6	29.34	1.9
Aspartate	5674.7	14.53	2071.6	67.12	2.7
Tryptophan	15224.7	13.81	7061.2	46.64	2.2
Lysine	1164.9	19.87	3012.4	30.77	0.4
Cysteine	71.3	159.61	347.7	29.94	0.2
O- Phosphoserine	1068.6	35.27	419	36.32	2.6
Histidine	353.9	24.24	940.2	43.65	0.4
Alanine	88411.2	30.44	51754.8	11.98	1.7
Methionine	11325	14.82	16127.4	23.96	0.7
Glutamine	2197.2	16.44	1429	41.95	1.5
Arginine	792.5	21.07	672.5	29.92	1.2
TOTAL	1616990.4	10.96	780233.6	13.72	2.1

Figure 6.3 PCA manual integration of method 2 following manual integration without normalisation



between *T.vaginalis* and *T.foetus*. So unnormalized data would give the wrong interpretation of the results. The two groups are better separated in the PCA than before normalization. Hence, normalization of data was proved to be necessary for drawing the correct conclusions from the results.

6.3.2 Comparison of Extraction Methods

Two methods of extract preparation were used and data obtained from each method was processed separately. In method 1, the cell concentration was 10^8 cells in 1ml 80% ethanol with no buffer added in the samples whereas in the second method it was 2.5×10^7 cells in 1 ml of 80% ethanol with 20mM of HEPES as buffer with a pH 7.5. All the data was manually processed and in order to determine the RSD for the metabolites and thus find out which extraction method gave the lowest

RSD. An RSD of $\pm 15\%$ was set as a target. Tables 6.2-6.5 summarise the amino acids found to have RSD values $< 15\%$ by the extraction method.

All the figures in the above tables are for normalized data. In table 6.2 (method 1), the number of amino acids with RSD below ± 15 is three for *T.foetus*, the corresponding number for *T.vaginalis* as shown in table 6.3 was 13. The results from extraction method 2 shown in tables 6.4 and 6.5, indicate that the numbers of amino acids having an RSD below are 9 and 11 for *T.foetus* and *T.vaginalis* respectively. Although the cells extracted in method 2 are half the number extracted in method 1, method 2 shows more amino acids having an RSD below 15 over both trichomonas species. From the Picker Shell result for method 1, the total number of ions detected was 2573 where as the same figure for method 2 was 2844. In method 2 347 ions corresponding to compounds were identified and 291 ions corresponding to compounds were identified in method 1. Thus the number of compounds identified was greater for method 2. There were some metabolites which were not detected with method 1, such as glycine and some phosphatidyl choline lipids and glycerophosphoethanolamine lipids which are two main phospholipids found in trichomonas but they were detected by using method 2 for sample preparation. Differences in the number of metabolites detected by the two methods may be due to differences in the degradation of some metabolites in the absence of buffer. Thus method 2 proved to be a better extraction method than method 1 for metabolite capture.

6.3.3 Statistical Analysis

Multivariate analysis was performed using principle components analysis (PCA) of normalized datasets obtained from the Pickershell results from extraction method 2. Sieve results and manual integration from both methods 1 and 2. PCA transformed the large set of related variables into a new, small set of independent

variables which showed that the two strains were easily separated in two groups. Graph 1 shows the score plot and loading plot for normalized Pickershell results from method 2. Graphs 2 and 3 describe the score plot and loading plot for the normalized data from Sieve analysis and manual integration. In the loading plot metabolites which are present on the edge of circle indicate high levels in that particular group compare to the other group. Metabolites away from the edge show no significant difference in their levels between the two groups. PCA for these data was carried out for further conformation of the metabolites which shown significant differences in their levels between *T.vaginalis* and *T.foetus*. Metabolites were grouped on according to their p-values. The score plot (figure 6.6) shows the degree of variation inbetween the two groups.

Figure 6.4 PCA plot obtained following manual integration of the data and normalisation.

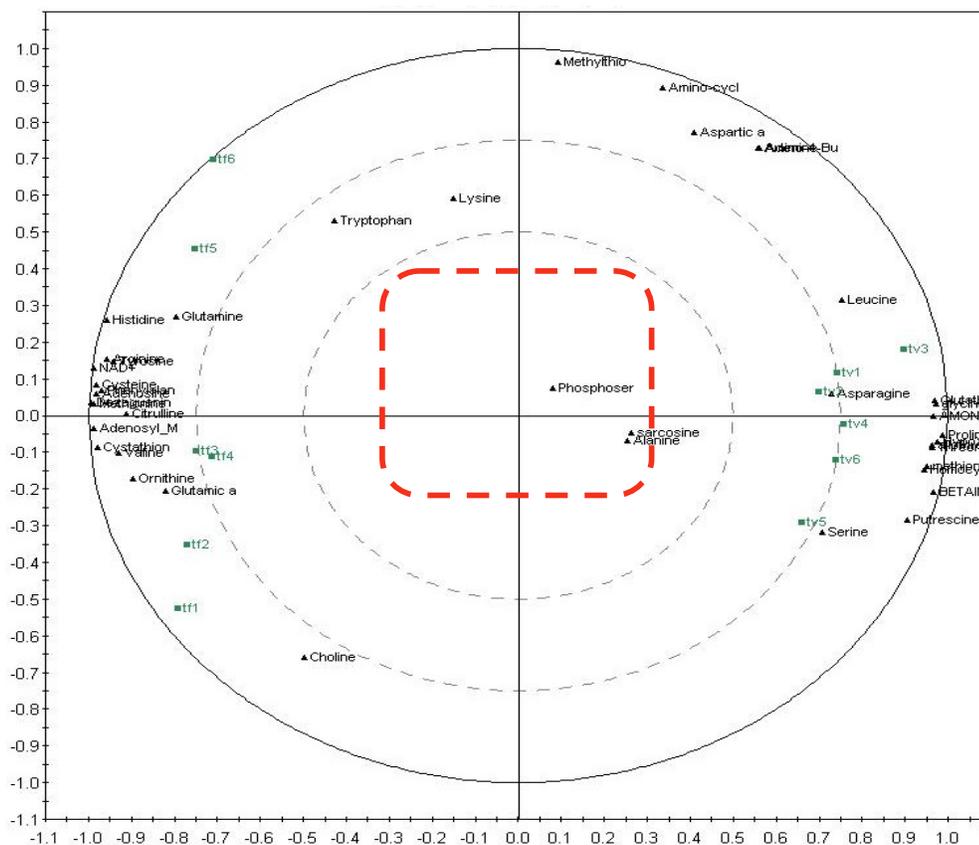


Table 6.2 *T. foetus*-amino acids with RSD by method 1

No	Compound	RSD (%)
1	Tyrosine	3.5
2	Phenylalanine	9.0
3	glutamine	10.4

Table 6.3 *T. vaginalis*- Amino acids with RSD for method 1

No	Compound	RSD (%)
1	Glutamate	4.5
2	Alanine	6.2
3	Tyrosine	6.9
4	Glutamine	6.9
5	Phenylalanine	9.4
6	Valine	10.2
7	Homocysteine	10.3
8	Tryptophan	10.5
9	Methionine	13.0
10	Histidine	14.0
11	Leucine	14.0
12	Threonine	14.3
13	Asparagine	14.4

Table 6.4 *T. foetus* amino acids with RSD for method 2.

No	Compound	RSD
1	Valine	7.0
2	Serine	7.1
3	Threonine	8.2
4	Leucine	8.8
5	Glycine	9.9
6	Alanine	10.9
7	Homocysteine	13.7
8	Glutamate	14.0
9	Proline	15.0

Table 6.5 *T. vaginalis* amino acids with R.S.D. for method 2.

No	Compound	RSD
1	Leucine	1.2
2	Phenylalanine	2.8
3	Valine	4.0
4	Tyrosine	4.5
5	Proline	6.4
6	Asparagine	9.3
7	Histidine	10.8
8	Threonine	12.2
9	Ornithine	12.9
10	Methionine	13.8
11	Serine	14.4

Figure 6.5 PCA score plot of Pickershell results comparing *T.vaginalis* and *T.foetus* using extraction method 2.

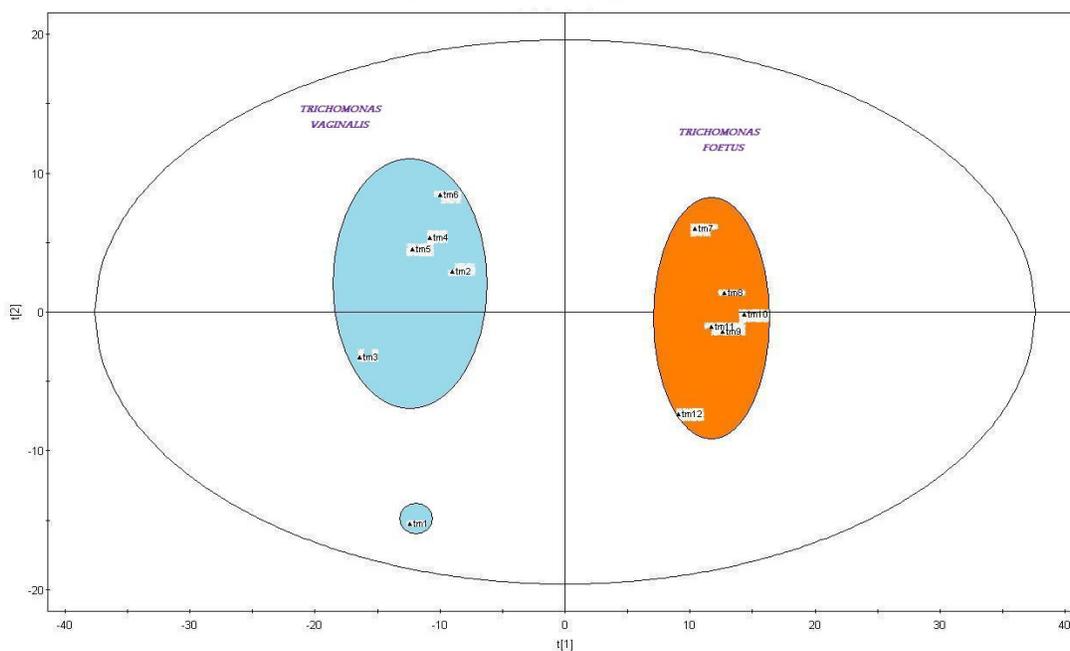


Figure 6.6 Loading plot of Pickershell results comparing *T.vaginalis* and *T.foetus* using extraction method 2.

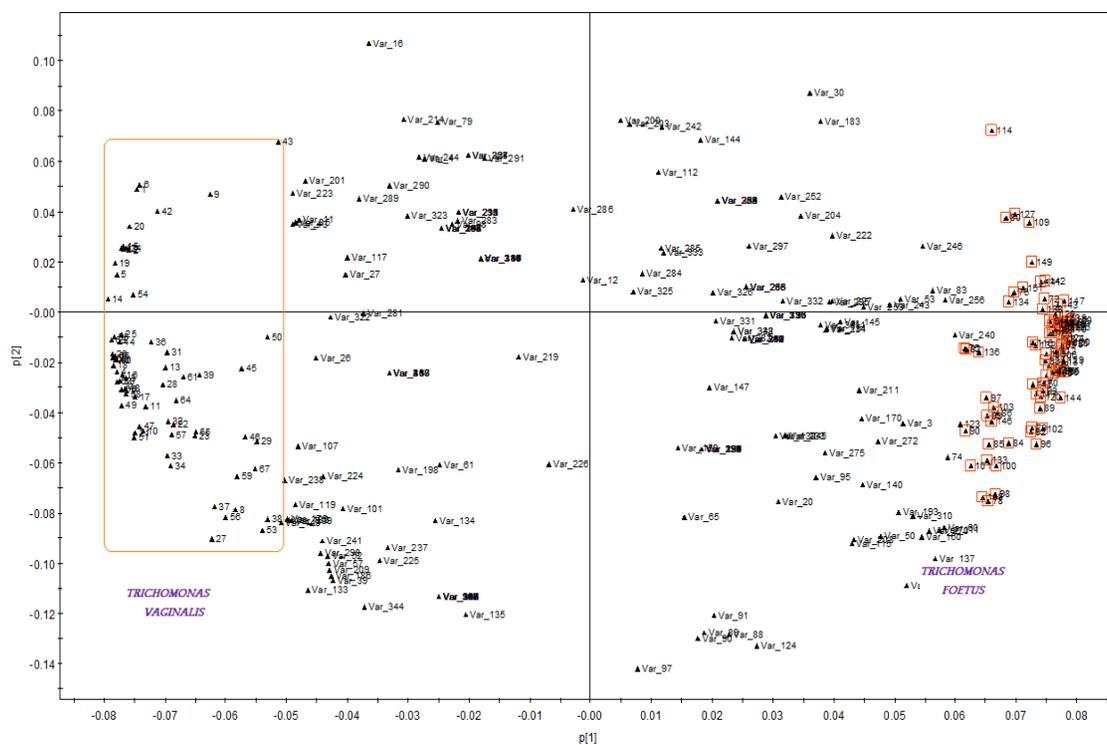


Table 6.6 Metabolites exhibiting major differences between *T. vaginalis* and *T. foetus* (TV/TF). Comparison of 6 against 6.

Compound	MW	RT	P value	Ratio
O-Acetyl-L-homoserine	162.076	11.48	0.0000514	0.1
amiohexadecanediol	274.2739	7.76	0.0000348	0.1
Sphingosine	300.2897	7.23	0.0000019	0.1
GPC	258.1099	17.60	0.0003634	0.1
Sphinganine	302.3053	7.70	0.0000030	0.2
C17 sphinganine	288.2896	7.74	0.0000147	0.2
S-Adenosyl-L-methionine+	399.1443	30.78	0.0000162	0.2
L-Cysteine	122.0269	13.40	0.0024330	0.4
phosphoethanolamine	216.0632	15.27	0.0106400	0.4
L-Histidine	156.0767	23.42	0.0170900	0.4
Val Val	217.1548	8.71	0.0000653	0.4
L-Lysine	147.1127	26.15	0.0037870	0.4
Glycine	76.03918	15.17	0.0000077	3.1
Adenine	136.0617	11.63	0.0063830	3.1
L-Serine	106.0498	15.41	0.0004301	3.1
gamma-N-methylaminobutyrate	118.0862	13.45	0.0000043	3.2
Xanthine	153.0406	8.25	0.0001204	3.2
1-Stearoylglycerophosphocholine ,	524.371	7.58	0.0001275	4.5
Putrescine	89.10725	33.49	0.0000601	4.6
butanedione	87.04401	13.45	0.0695800	4.7
L-Proline	116.0705	13.17	0.0000313	4.7
GABA	104.0705	13.45	0.0070040	5.0
Aminoacetone	74.05995	14.16	0.0002190	5.6
Niacin/Nicotinate	124.0392	7.23	0.0004318	7.0
L-Threonine	120.0655	14.15	0.0000804	7.5
oxoproline	130.0498	7.19	0.0000153	8.3
Homocysteine	136.0425	12.15	0.0000968	8.5
methylpropanitrile	70.06507	13.17	0.0000614	10.2
methylproline	130.0862	12.53	0.0000140	13.1
Pro Pro Phe	360.1919	9.25	0.0008745	14.5
Piperidine	84.08076	12.53	0.0001890	24.3
Hypoxanthine	137.0456	8.65	0.0000180	77.2
Acetylarginine	217.1297	13.11	0.0000136	109.8
5-Acetamidopentanoate	160.0969	12.01	0.0001622	117.4
Glyceryl phosphate	173.0208	16.24	0.0000265	148.1
1-Aminopropan-2-ol	76.07565	14.03	0.0001014	161.9

6.3.4 Metabolic Pathway Analysis

Large numbers of metabolites were found to have significant differences having high levels either in *T.vaginalis* or in *T.foetus*. The metabolites are discussed below on the basis of metabolic pathways (Kegg pathways) to which they belong. Table 6.6 shows the most varying metabolites between the two organisms.

6.3.4.1 Methionine metabolism

One of the major significant differences was found with sulfur containing amino acids like methionine, cysteine and homocysteine. The first observation for variation in a pathway was noticed in the methionine metabolic pathway. From the methionine metabolism pathway methionine, S-adenosyl methionine, homocysteine, cystathionine, cysteine, amino cyclopropane carboxylate (ACC) and S-methylthioadenosine (SAM) were identified with low p-values indicating significant differences with these compounds being higher in *T.foetus*. Sulfur containing amino acids serve various functions. Cysteine has an important role in stability and catalytic activity of proteins and is incorporated into glutathione, which has a role as an antioxidant. Methionine is important in cell proliferation. Figure 6.7 shows the metabolism of methionine in trichomonas and other organisms. As shown in figure 6.7, methionine metabolism initiates with production of SAM from methionine. From which it produces adenosylhomocysteine and homocysteine. Cysteine can be formed from homocysteine by two possible routes one is by the action of cysteine synthase and other is via a trans-sulfuration reaction via the formation of cystathionine. The enzyme responsible for the trans-sulfuration is absent in *T.vaginalis* [200-203] and this is supported by the absence of cystathione (figure 6.8). The gene for all the other enzymes in the metabolism of methionine have been found and proved [204-206]. Cysteine can be produced from homocysteine by the

action of cystathionine β -synthase and cystathionine γ -lyase. In *T. vaginalis* cysteine is formed from homocysteine by the action of cysteine synthase with the help of H_2S produced via the conversion of homocysteine. H_2S is produced from homocysteine and reacts with O-phosphoserine via the action of methionine γ -lyase. Elevated levels of phosphoglycerol a precursor of O-phosphoserine are observed in *T. vaginalis* (figure 6.9). Methionine may be regenerated from adenosyl methionine via formation of amino cyclopropane carboxylate (ACC) and methyl thio adenosine (MTA).

From the Pickershell results, Sieve results and manual integration results, the methionine ratio of 0.29 for *T. foetus* vs *T. vaginalis* shows a highly significant difference with a high level in *T. foetus* along with adenosyl methionine, methyl thioadenosine and cysteine. Homocysteine was present in a high level in *T. vaginalis*. Regeneration of methionine from adenosyl methionine did not show any difference as the p-values for MTA and ACC were not significant and their ratios for the two organisms were almost 1:1. The methionine and S-adenosyl methionine pool was greater in *T. foetus* and homocysteine was present at levels 3.5 fold higher in *T. vaginalis* compared with *T. foetus*. The ratio of cysteine in *T. vaginalis* over *T. foetus* was very low with a very low p-value indicating that cysteine is present in high levels in *T. foetus*. It is difficult to completely explain these results. The absence of cystathione in *T. vaginalis* (figures 6.8 and 6.9) is expected since it does not utilise it in the biosynthesis of cysteine and the gene required for encoding of protein for the enzymes responsible for production of cysteine by transulfuration are not present in *T. vaginalis* [4-7]. Otherwise it is not obvious why homocysteine should be higher in *T. vaginalis* since both organisms use it as a biosynthetic precursor of cysteine and the same is true for the methionine and its metabolites.

Figure 6.7: Methionine Metabolism (Cysteine biosynthesis). Enzyme involved 1: Methionine adenosyl transferase; 2: S-Adenosyl methionine dependent methyl transferase 3: S-Adenosyl homocysteine hydrolase 4: methionine gamma lyase 5: 1-amino cyclo propane carboxylate synthase 6: Methionine regeneration pathway (enzyme not found) MS: Methionine synthase; CBS: Cystathionine β -synthase CGL: Cystathionine γ -lyase

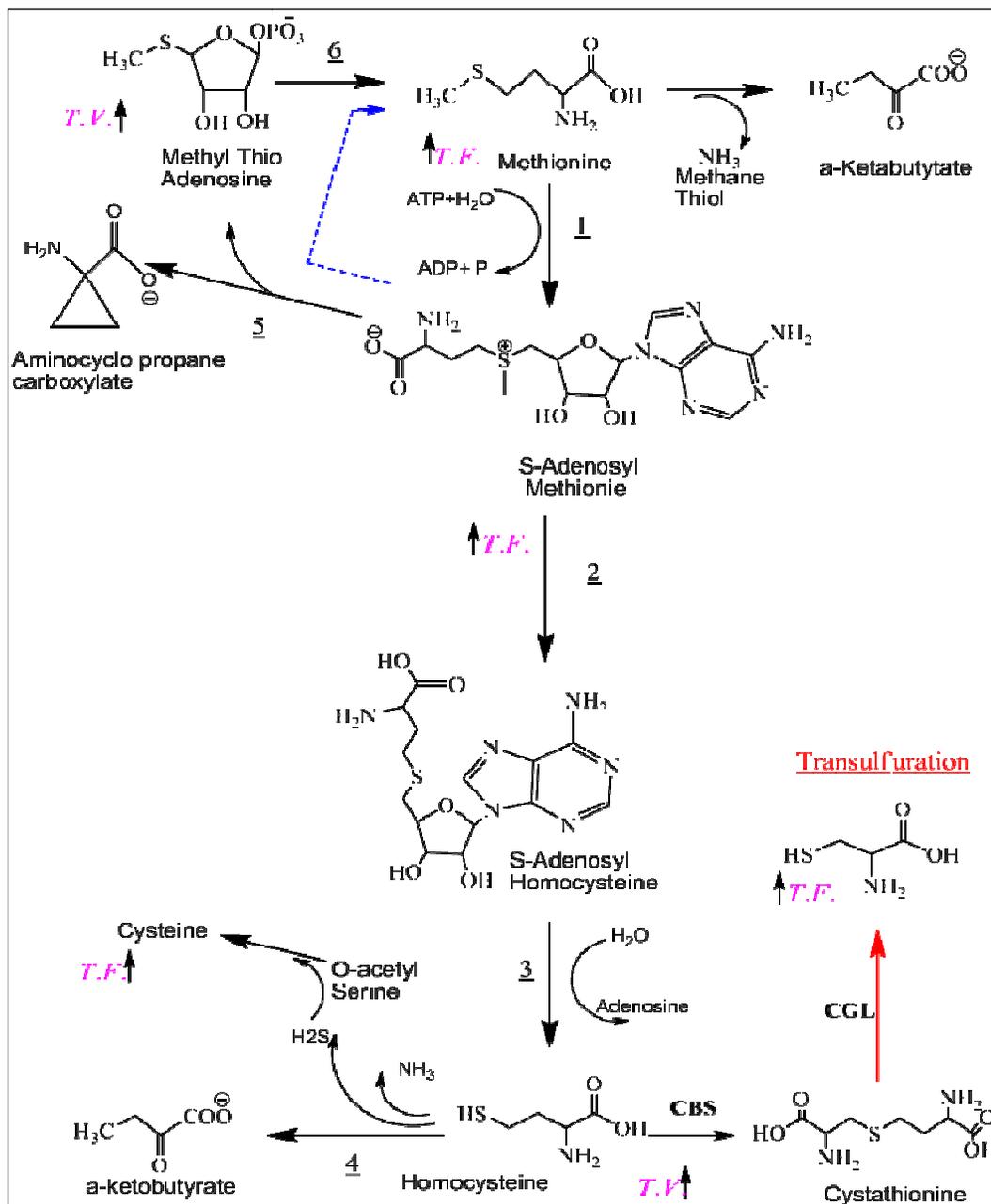
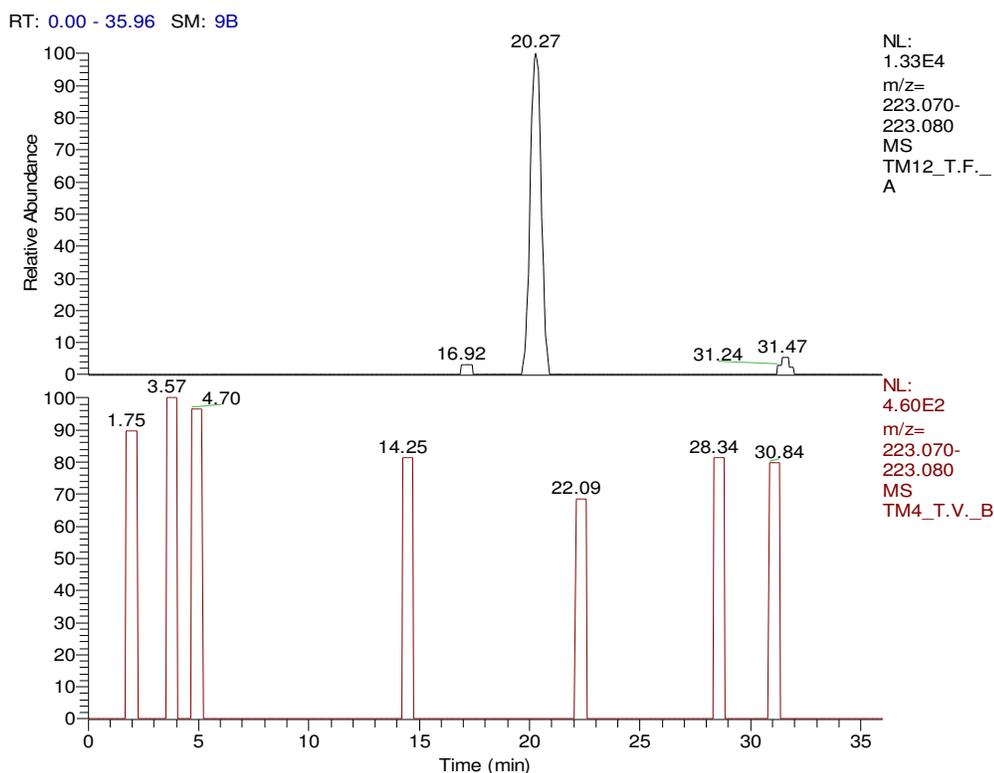


Figure 6.8 Extracted ion chromatogram for cystathione indicating the absence of cystathionine in *T.vaginalis* (lower trace) and its presence in *T.foetus*.

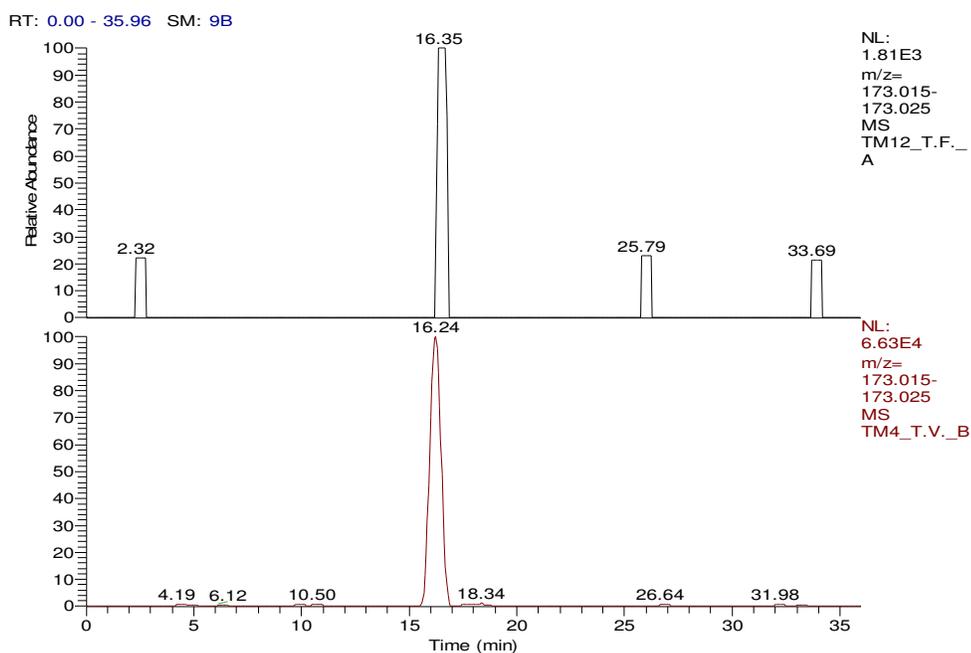


6.3.4.2 Proline, Arginine and Ornithine Metabolism

The arginine dihydrolase pathway is a major pathway for the production of energy in trichomonas species which, in the absence of mitochondria, cannot fully utilise glucose. The end products of arginine metabolism are one mole of ATP and one mole of putrescine (figure 6.10). Arginine, citrulline and ornithine are present in relatively high amounts in *T.foetus* when compared with *T.vaginalis*, however, putrescine showed its highest levels in *T.vaginalis*. Again it is difficult to read much biological significance into this except that the presence of putrescine at quite high levels in *T.foetus* suggests that arginine metabolism is also important in this organism. Levels of proline are high in *T.vaginalis* and proline is supposed to be another energy source in trichomonas species and its utilisation in *T.vaginalis* might

be supported by the presence of oxoproline in high amounts in this organism. In order to form this metabolite two protons have to be lost from proline which would result in the production of NADH.

Figure 6.9 Extracted ion chromatogram showing elevated levels of phosphoglycerol in *T.vaginalis* (lower trace).



6.3.4.3 Sphingolipid Metabolism:

Sphingolipids and other phospholipids play an important role in cell-cell adhesion, differentiation, cellular growth and in other cell functions. There were significant differences in the sphingolipid pathway between *T.vaginalis* and *T.foetus*. Serine, dehydrosphinganine and dihydrosphingosine (sphinganine) showed significant differences. L-serine was present in high levels in *T.vaginalis* whereas dehydrosphingosine and sphinganine appeared in very high levels in *T.foetus* when compared with *T.vaginalis*. The enzyme responsible for conversion of serine to dehydrosphinganine is serine palmitoyl amino transferase. Different levels with

significant p-values indicate production of sphinganine in *T.vaginalis* occurs to a lesser extent than in *T.Foetus*.

Apart from sphingolipids other lipids which are of importance are glycerophospholipids like phosphotidyl choline and phosphotidyl ethanolamine which are more apparent in *T.vaginalis* than *T.foetus*. Phosphotidylcholine and hexadecyl glycerophosphocholine acetate was present in high amounts in *T.Vaginalis* whereas glycerophosphoethanolamine and glycerophosphorylcholine were significantly higher in *T.foetus*.

.6.3.4.4 Glycine, Serine and Threonine Metabolism

Figure 6.12 shows pathway of glycine, serine and threonine metabolism with some metabolites marked with green which were identified in both samples except for cystathionine which was only detected in *T.foetus*. From the results of manual integration, the levels of serine, glycine and threonine were high in *T.vaginalis*. Whereas no difference was found in the levels of phosphoserine. Glycine is also produced from choline via production of betaine aldehyde, betaine, dimethyl glycine and sarcosine which were all identified in the samples and all were significantly higher in *T.vaginalis*. In addition in the same pathway amino acetone and aminopropanol were also found to be greatly elevated in *T.vaginalis*. (figure 6.13).

As shown in figure 6.13 aminopropanol enters into porphyrin biosynthesis although there is no evidence for the occurrence of porphyrin in *T.vaginalis*.

Apart from the above biosynthetic pathways, other metabolites also varied in their levels between the two organisms.

6.3.4.5 Variations in nucleotide metabolism

Xanthine and hypoxanthine were greatly elevated in *T.vaginalis* compared with *T.foetus*. These compounds are presumably scavenged from the growth medium (hypoxanthine levels are high in calf serum) and it is not clear why one organism accumulates these compounds to a greater extent than the other.

6.4 Conclusion

Metabolic profiling of *T.vaginalis* and *T.foetus* identified a number of metabolic differences between the two organisms in addition to the expected differences in the pathway leading to cysteine biosynthesis. It is difficult to explain the origin of most of these differences although it is perhaps not unexpected that two closely related but different micro-organisms would have differences in their metabolome. It might be possible to link these differences to differences in the evolution of the two protozoa. The observations in this short study underline the possibility of looking at evolutionary biology in metabolic terms. *T.vaginalis* and *T.foetus* and trypanosomatids which include leishmania and trypanosomes which are major disease producing organisms as well as the more distantly related plasmodium species which cause malaria. The ability to understand and exploit metabolic differences between disease bearing microbes and their hosts could lead to the rational development of more effective anti-microbial agents.

Figure 6.10 Arginine dihydrolase pathway Enzymes involved in pathway. 1: Arginine deiminase, 2: Ornithine carbamoyl transferase, 3: Carbamate kinase, 4: Ornithine decarboxylase, 5: Lysine/ornithine amino transferase, 6: Pyrroline 5-carboxylate reductase, 7: Arginase, 8: Ornithine cyclodeaminase

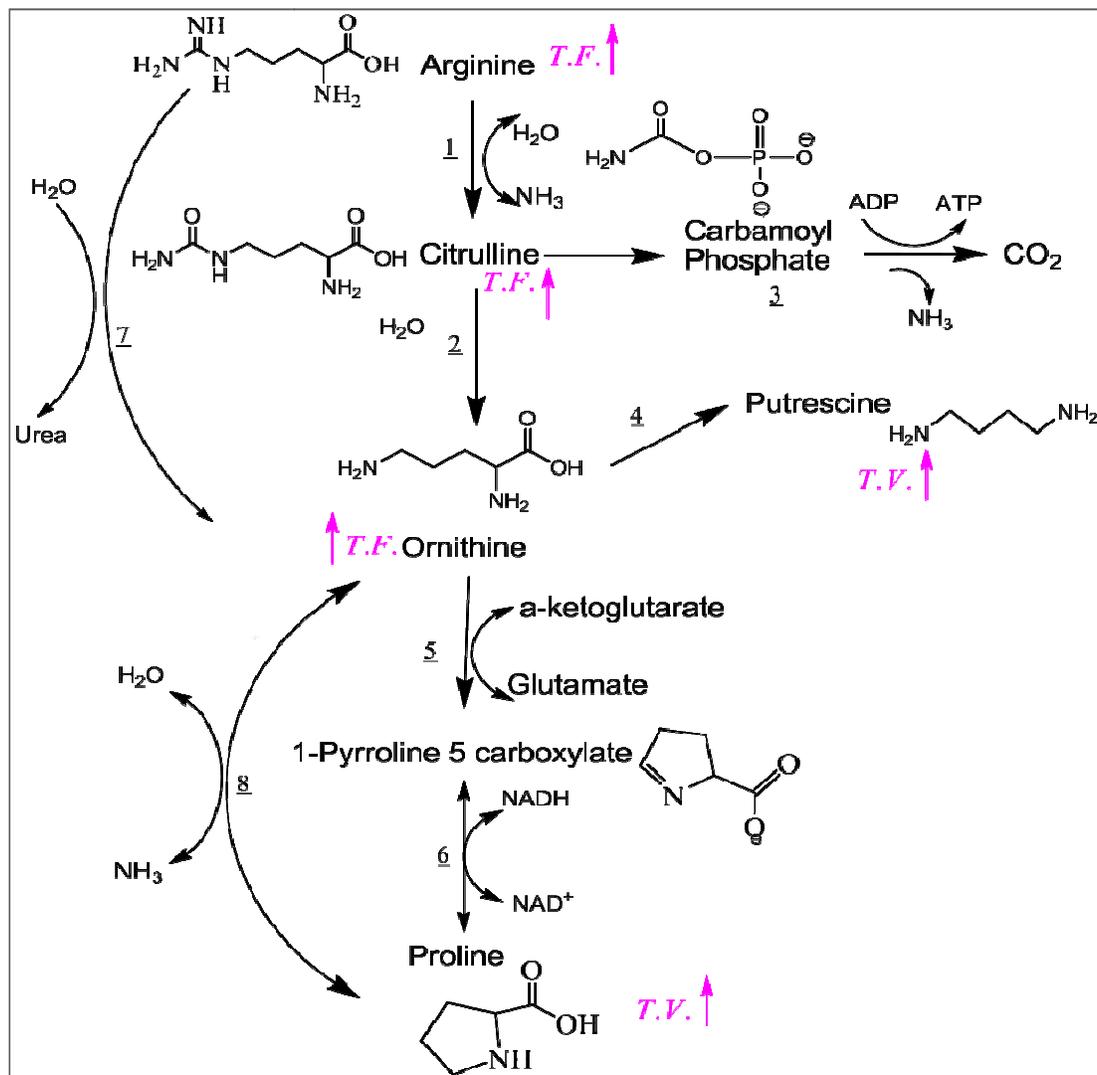


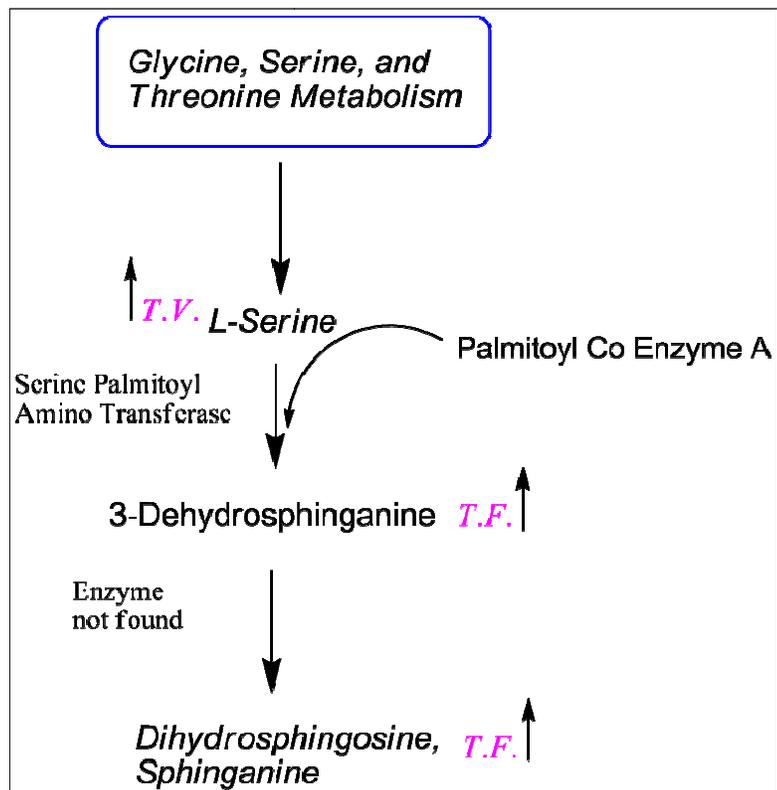
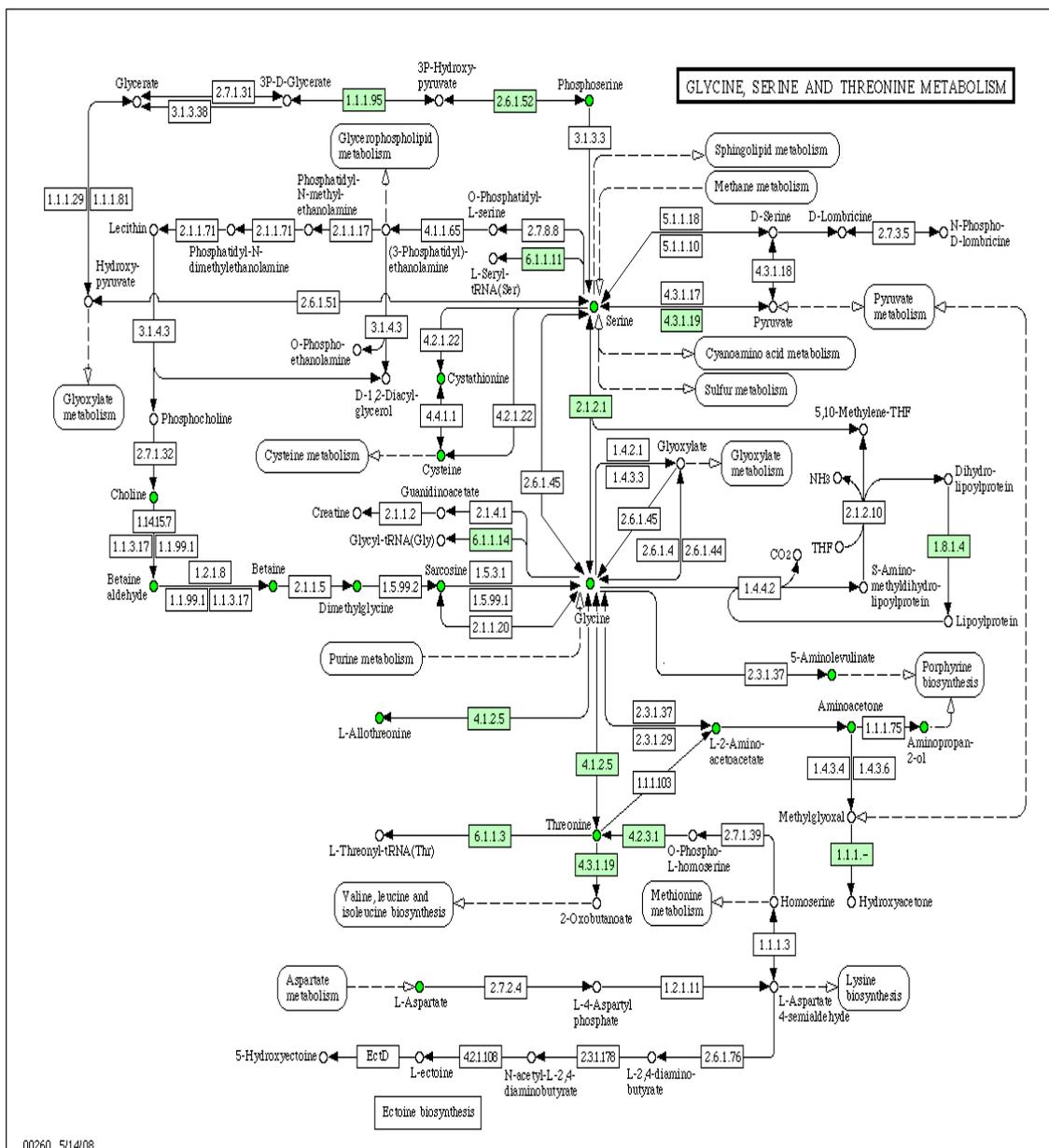
Figure 6.11 Sphingolipid Metabolism

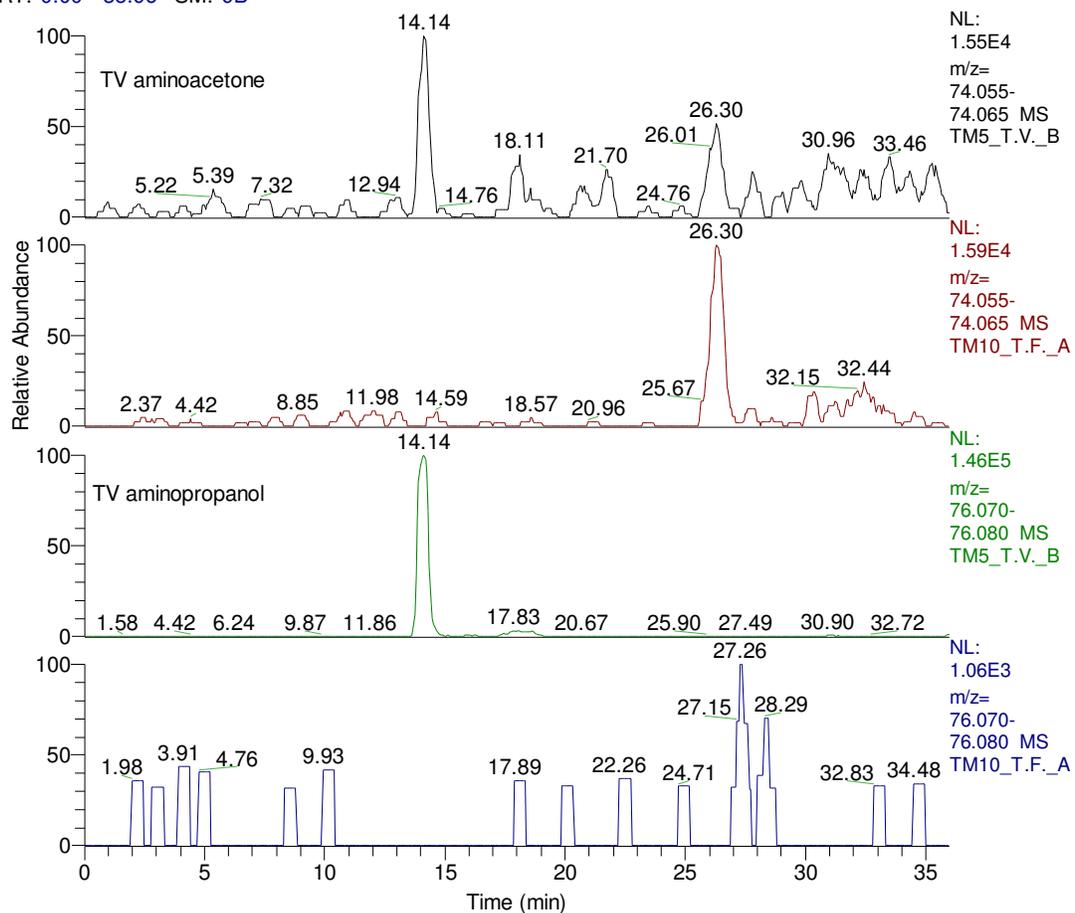
Figure 6.12 Glycine, serine and threonine pathway



00260 5/14/08

Figure 6.13 Production of aminoacetone and aminopropanol in *T.vaginalis*

RT: 0.00 - 35.96 SM: 9B



Reference:

- [1] Chace D H, Kalas T A, Taylor E W, Use of Tandem Mass Spectrometry for Multianalyte Screening of Dried Blood Specimens From Newborns. *Clinical Chemistry* 2003, 49, 1797-1817.
- [2] Frank R, Hargreaves R, Clinical Biomarkers in Drug Discovery and Development. *Nature Reviews Drug Discovery* 2003, 2, 566-580.
- [3] Wang T J, Gona P, Larson M G, Tofler G H, *et al.*, Multiple Biomarkers for The Prediction of First Major Cardiovascular Events and Death. *New England Journal of Medicine* 2006, 355, 2631-2639.
- [4] Xu Y, Shen Z Z, Wiper D. W, M. Z. Morton, *et al.*, Lysophosphatidic Acid As A Potential Biomarker For Ovarian And Other Gynecologic Cancers. *Jama-Journal of the American Medical Association* 1998, 280, 719-723.
- [5] Zytovicz T H, Fitzgerald E F, Marsden D Larson, C A Shih, *et al.*, Tandem Mass Spectrometric Analysis For Amino, Organic, and Fatty Acid Disorders in Newborn Dried Blood Spots: A Two-Year Summary From The New England Newborn Screening Program. *Clinical Chemistry* 2001, 47, 1945-1955.
- [6] Bar, N. A., Ott, K.-H., Roongta, V., Mueller, L., Metabolomic Analysis Using Optimized NMR and Statistical Methods. *Analytical Biochemistry* 2006, 355, 62-70.
- [7] Cheng, L. L., Pohl, U., John, C. L., Jeremy, K. N., Elaine, H., *The Handbook of Metabonomics and Metabolomics*, Elsevier Science B.V., Amsterdam 2007, pp. 345-374.
- [8] Clarke, C. J., HASELDEN, J. N., Metabolic Profiling as a Tool for Understanding Mechanisms of Toxicity. *Toxicologic Pathology* 2008, 36, 140-147.
- [9] Coen, M., Holmes, E., Lindon, J. C., Nicholson, J. K., NMR-Based Metabolic Profiling and Metabonomic Approaches to Problems in Molecular Toxicology. *Chem. Res. Toxicol* 2008, 21, 9-27.
- [10] Ebbels, T. M., Cavill, R., Bioinformatic methods in NMR-based metabolic profiling. *Progress in Nuclear Magnetic Resonance Spectroscopy* 2009, 55, 361-374.
- [11] Grivet, J.-P., Delortb, A.-M., Portaisc, J.-C., NMR and microbiology: from physiology to metabolomics. *Biochimie* 2003, 85, 823-840.
- [12] Tjeerdema, R. S., Application of NMR-based techniques in aquatic toxicology: Brief examples. *Marine Pollution Bulletin* 2008, 57, 275-279.
- [13] Wei, L., Liao, P., Wu, H., Li, X., *et al.*, Toxicological effects of cinnabar in rats by NMR-based metabolic profiling of urine and serum. *Toxicology and Applied Pharmacology* 2008, 227, 417-429.
- [14] Wishart, D. S., Quantitative metabolomics using NMR. *TrAC Trends in Analytical Chemistry* 2008, 27, 228-237.
- [15] Denkert C, Budczies J, Kind T, Weichert W, *et al.*, GC-TOF based metabolic profiling reveals different metabolite patterns in invasive ovarian carcinomas and ovarian borderline tumors. *Cancer Res* 2006, 66, 10795-10804.
- [16] Fiehn O, Kopka J, Trethewey RN, Willmitzer L, Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal. Chem* 2000, 72, 3573-3580.
- [17] Oliver Fiehn, Metabolomics - the link between genotype and phenotype. *Plant Mol. Biol.* 2002, 48, 155-171
- [18] Roessner U, Luedemann A, B. D., Fiehn O, Linke T, *et al.*, Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* 2001, 13, 11-29.

- [19] Dunn, W. B., Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Phys. Biol.* 2008, *5*, 1-24.
- [20] Georgios Theodoridis, Helen G. Gika, Wilson, I. D., LC-MS-based methodology for global metabolite profiling in metabonomics/metabolomics *TrAC Trends in Analytical Chemistry* 2008, *27*, 251-260.
- [21] Muhammed, A., K., J.Dow, Dave.G.Watson, Applications of mass spectrometry in metabolomic studies of animal model and invertebrate systems. *Brief Funct Genomic Proteomic.* 2009, *8*, 28-48.
- [22] W.B. Dunn, M. Brown, S.A. Worton, I.P. Crocker, *et al.*, Changes in the Metabolic Footprint of Placental Explant-Conditioned Culture Medium Identifies Metabolic Disturbances Related to Hypoxia and Pre-Eclampsia. *Placenta* 2009, *30*, 974-980.
- [23] Vicki H Wysocki, Katheryn A Resing, Qingfen Zhang, Guilong Cheng, Mass spectrometry of peptides and proteins. *Methods* 2005, *35*, 211-222.
- [24] Qingyu Song, Wei Xu, Scott A Smith, Liang Gao, *et al.*, Ion trap mass analysis at high pressure: an experimental characterization. *Journal of Mass Spectrometry* 2009, *10*, 1-9.
- [25] Karen R Jonscher, John R Yates, The Whys and Wherefores of Quadrupole Ion Trap Mass Spectrometry.
<http://www.abrf.org/ABRFNews/1996/September1996/sep96iontrap.html> 2006.
- [26] Bruker, Bruker Daltonics Introduces MaXis™, A Revolutionary New Ultra-High Resolution TOF Mass Spectrometer.
<http://www.medicalnewstoday.com/articles/109607.php> 2006.
- [27] Bruker, Fast and Highly Accurate QTOF ETD MS/MS for Top-Down Protein Identification. *HUPO 2009* http://www.bdal.de/uploads/media/HUPO2009_ETD-maXis.pdf 2007.
- [28] Alan G Marshall, C L Hendrickson, G S Jackson, Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Reviews* 1998, *17*, 1-35.
- [29] Coyne, K., What's in an Oil Drop?
<http://www.magnet.fsu.edu/education/tutorials/magnetacademy/fticr/> 2007.
- [30] Kaiser, N. K., Skulason, G. E., Weisbrod, C. R., Bruce, J. E., A Novel Fourier Transform Ion Cyclotron Resonance Mass Spectrometer with Improved Ion Trapping and Detection Capabilities. *Journal of the American Society for Mass Spectrometry* 2009, *20*, 755-762.
- [31] Bruker, Apex-ultra hybrid Qq-FTMS. <http://www.bdal.de/products/qh-ftms/apex-ultra.html>.
- [32] Bruker, Solarix Achieve Record Breaking Results with a Unique 18 Tesla (18T) Magnet System for FTMS.
http://www.bdal.de/fileadmin/PDF/literature2009/brochures/solarix_18T.pdf 2009.
- [33] Makarov A, Mass Spectrometer. *U.S. Patent Ed* 1999, *5*.
- [34] Hu Q Z, Noll R J, Li H Y, Makarov A, *et al.*, The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry*, 2005, *40*, 430-443.
- [35] Makarov A, Denisov E, Lange O, Horning S, Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. *Journal of the American Society for Mass Spectrometry*, 2006, *17*, 977-982.
- [36] Thermo, LTQ Orbitrap XL.
<http://www.thermo.com/com/cda/product/detail/1,1055,10131071,00.html> 2009.
- [37] Alexander Makarov, Eduard Denisov, Alexander Kholomeev, Wilko Balschun, *et al.*, Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. *Analytical Chemistry* 2006, *78*, 2113-2120.

- [38] Thermo, LTQ Orbitrap Discovery.
<http://www.thermo.com/com/cda/product/detail/1,1055,10131074,00.html>.
- [39] Thermo, Thermo Scientific Exactive.
<http://www.thermo.com/com/cda/product/detail/1,,10142061,00.html?ca=exactive>.
- [40] Thermo, Thermo Scientific LTQ Orbitrap Velos.
http://www.thermo.com/eThermo/CMA/PDFs/Product/productPDF_51627.pdf 2009.
- [41] Dunn W B, Broadhurst D, Brown M, Baker P N, *et al.*, Metabolic profiling of serum using Ultra Performance Liquid Chromatography and the LTQ-Orbitrap mass spectrometry system. *Journal of Chromatography B-analytical Technologies in the Biomedical and Life Sciences*, 2008, 871, 288-298.
- [42] Gika H G, Theodoridis G, Extnance J, Edge A M, Wilson I D, High temperature-ultra performance liquid chromatography-mass spectrometry for the metabonomic analysis of Zucker rat urine. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences*, 2008, 871, 279-287.
- [43] Plumb R S, Granger J H, Stumpf C L, Johnson K A, *et al.*, A rapid screening approach to metabonomics using UPLC and oa-TOF mass spectrometry: application to age, gender and diurnal variation in normal/Zucker obese rats and black, white and nude mice. *Analyst*, 2005, 844-849.
- [44] Zelena E, Dunn W B, Broadhurst D, Francis-McIntyre S, *et al.*, Development of a Robust and Repeatable UPLC-MS Method for the Long-Term Metabolomic Study of Human Serum. *Analytical Chemistry*, 2009, 81, 1357-1364.
- [45] Desmet G, Comparison techniques for HPLC column performance. *LC-GC Europe* 2008, 21, 310-320.
- [46] Gika H G, Theodoridis G A, Wilson I D, Hydrophilic interaction and reversed-phase ultra-performance liquid chromatography TOF-MS for metabonomic analysis of Zucker rat urine. *Journal of Separation Science* 2008, 31, 1598-1608.
- [47] Pesek J J, Matyska M T, Fischer S M, Sana T R, Analysis of hydrophilic metabolites by high-performance liquid chromatography-mass spectrometry using a silica hydride-based stationary phase. *Journal of Chromatography A* 2008, 1204, 48-55.
- [48] Pesek J J, Matyska M T, Loo J A, Fischer S M, R, S. T., Analysis of hydrophilic metabolites in physiological fluids by HPLC-MS using a silica hydride-based stationary phase. *Journal of Separation Science* 2009, 32, 2200-2208.
- [49] Weisenberg S A, Butterfield T R, Fischer S M, Rhee K Y, Suitability of silica hydride stationary phase, aqueous normal phase chromatography for untargeted metabolomic profiling of *Enterococcus faecium* and *Staphylococcus aureus*. *Journal of Separation Science* 2009, 32, 2262-2265.
- [50] Lapainis T, Rubakhin S S, Sweedler J V, Capillary Electrophoresis with Electrospray Ionization Mass Spectrometric Detection for Single-Cell Metabolomics. *Analytical Chemistry*, 2009, 81, 5858-5864.
- [51] Baidoo E E K, Benket P I, Neuss C, Pelzing M, *et al.*, Capillary electrophoresis-Fourier transform ion cyclotron resonance mass Spectrometry for the identification of cationic metabolites via a pH-mediated stacking-transient isotachopheric method. *Analytical Chemistry*, 2008, 80, 3112-3122.
- [52] Umetrics, SIMCA-P. <http://www.umetrics.com/>.
- [53] Hiroyuki Yamamoto, Hideki Yamaji, Yuichiro Abe, Kazuo Harada, *et al.*, Dimensionality reduction for metabolome data using PCA, PLS, OPLS, and RFDA with differential penalties to latent variables. *Chemometrics and Intelligent Laboratory Systems* 2009, 98, 136-142.
- [54] P Ciosek, Z Brzozka, W W oblewski, E Martinelli, *et al.*, Direct and two-stage data analysis procedures based on PCA, PLS-DA and ANN for ISE-based electronic tongue-Effect of supervised feature extraction. *Talanta* 2005, 67, 590-596.

- [55] Tapp, H., Kate Kemsley, OPLS: an ideal tool for interpreting PLS regression models? *CAC2008* 2008.
- [56] Tominaga, Y., Comparative study of class data analysis with PCA-LDA, SIMCA, PLS, ANNs, and k-NN. *Chemometrics and Intelligent Laboratory Systems* 1999, *49*, 105-115.
- [57] Preti George, Metabolomics comes of age. *The Scientist* 2005, *19*, 8-18.
- [58] Horning EC, Metabolic Profiles: Gas-Phase Methods for Analysis of Metabolites. *Clinical Chemistry* 1971, *17*, 802-809.
- [59] Mamer OA, Crawhall JC, Tjoa SS, The Identification of Urinary Organic Acids by Coupled Gas Chromatography Mass Spectrometry. *Clin. Chim. Acta* 1971, *32*, 171-184.
- [60] Robinson, Pauling, Quantitative Analysis of Urine Vapor and Breath by Gas-Liquid Partition Chromatography. *Proceedings of the National Academy of Sciences* 1971.
- [61] Oliver S G, Winson M K, Kell D B, Baganz F, Systematic functional analysis of the yeast genome. *Trends Biotechnol* 1998, *16*, 373-378.
- [62] J K NICHOLSON, J C LINDON, E HOLMES, Metabonomics: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *xenobiotica* 1999, *29*, 1181- 1189.
- [63] John C Lindon, Elaine Holmes, Jeremy K Nicholson, Metabonomics Techniques and Applications to Pharmaceutical Research & Development. *Journal Pharmaceutical Research* 2006, *23*, 1075-1088.
- [64] Brindle J T, Holmes E, Tranter G, Nicholson J K, Rapid and non-invasive diagnosis of the presence and severity of coronary heart disease using ¹H NMR - based metabonomics. *Nature Medicine* 2002, *8*, 1439-1444.
- [65] Goodacre R, Vaidyanathan S, Dunn W B, Harrigan G G, Kell D B, Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol* 2004, *22*, 245-252.
- [66] Kell D B, Metabolomics and systems biology: making sense of the soup. *Curr. Opin. Microbiol* 2004, *7*, 296-307.
- [67] Lindon J C, Holmes E, Bollard M E, Stanley E G, Nicholson J K, Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. *Biomarkers* 2004, *9*, 1-31.
- [68] Nicholson J K, Holmes E, Lindon J C, Wilson I D, The challenges of modeling mammalian biocomplexity. *Nature Biotechnol* 2004, *22*, 1268-1274.
- [69] van der Greef, T., J Stroobant R, van der Heijden R, The role of analytical sciences in medical systems biology. *Curr. Opin. Chem. Biol* 2004, *8*, 559-565.
- [70] Fiehn O, Kloska S, Altmann T, Integrated studies on plant biology using multiparallel techniques. *Current Opinion in Biotechnology* 2001, *12*, 82-86.
- [71] Jenkins H, Hardy N, Beckmann M, Draper J, *et al.*, A proposed framework for the description of plant metabolomics experiments and their results. *Nat Biotechnol* 2004, *22*, 1601-1606.
- [72] Sang Hee Lee, Han Min Woo, Byung Hwa Jung, Jeongae Lee, *et al.*, Metabolomic Approach To Evaluate the Toxicological Effects of Nonylphenol with Rat Urine. *Analytical Chemistry* 2007, *79*.
- [73] Salek RM, Maguire ML, Bentley E, Rubtsov DV, *et al.*, A Metabolomic Comparison of Urinary Changes in Type 2 Diabetes in Mouse, Rat and Man. *Physiol Genomics* 2006, *29*, 99-108.
- [74] Silas G Villas-Bôat, Albert Koulman, Geoffrey A Lane, Analytical methods from the perspective of method standardization. *Book Metabolomics* 2007, *18*, 11-52.

- [75] Stella C, Beckwith-Hall B, Cloarec O, Holmes E, *et al.*, Susceptibility of human metabolic phenotypes to dietary modulation. *J Proteome Res* 2006, *5*, 2780-2788.
- [76] Mlawule R Mashego, Karl Rumbold, Marjan De Mey, Erick Vandamme, *et al.*, Microbial metabolomics: past, present and future methodologies. *Journal Biotechnology Letters* 2007, *29*, 1-16
- [77] R A Triebkorn, I F Henderson, A P Martin, Detection of iron in tissues from slugs (*deroceras reticulatum müller*) after ingestion of iron chelates, by means of energy-filtering transmission electron microscopy (EFTEM). *Pesticide Science* 1999, *55*, 55 - 61.
- [78] Henderson, I. M., Hendy, M. D., Penny, D., Influenza viruses, comets and the science of evolutionary trees. *Journal of Theoretical Biology* 1989, *140*, 289-303.
- [79] G S Puritch, D S Almond, R M Matson, W M Mason, Ingestible mollusc poisons. *US Patent* 1995, *5437870*.
- [80] Langan, A. M., Shaw, E. M., Responses of the earthworm *Lumbricus terrestris* (L.) to iron phosphate and metaldehyde slug pellet formulations. *Applied Soil Ecology* 2006, *34*, 184-189.
- [81] Appel, M. J., Kuper, C. F., Woutersen, R. A., Disposition, accumulation and toxicity of iron fed as iron (II) sulfate or as sodium iron EDTA in rats. *Food and Chemical Toxicology* 2001, *39*, 261-269.
- [82] Le Zhu, R P Glahn, Chi Kong Yeung, Dennis D Miller, Iron Uptake by Caco-2 Cells from NaFeEDTA and FeSO₄: Effects of Ascorbic Acid, pH, and a Fe(II) Chelating Agent. *Journal of Agricultural and Food Chemistry* 2006, *54*, 7924-7928.
- [83] Le Zhu, Dennis D Miller, Tissue Iron Distribution and Urinary Mineral Excretion Vary Depending on the Form of Iron (FeSO₄ or NaFeEDTA) and the Route of Administration (Oral or Subcutaneous) in Rats Given High Doses of Iron. *Journal of Agricultural and Food Chemistry* 2007, *55*, 8793-8799.
- [84] Watson, D., Logarithms of stability constants of miscellaneous complexones. *Complexometry* 1969, 228-230.
- [85] A T Hubbard, Encyclopaedia of Surface and Colloid Science, . *CRC Press* 2002.
- [86] G Owens, V K Ferguson, M J McLaughlin, I Singleton, *et al.*, Determination of NTA and EDTA and Speciation of Their Metal Complexes in Aqueous Solution by Capillary Electrophoresis. *Environ. Sci. Technol.* 2002, *34*, 885-891.
- [87] Blatny, P., Kvasnicka, F., Kenndler, E., Trace determination of iron in water at the [µg/l] level by on-line coupling of capillary isotachopheresis and capillary zone electrophoresis with UV detection of the EDTA-Fe(III) complex. *Journal of Chromatography A* 1997, *757*, 297-302.
- [88] Sillanpaa, M., Kokkonen, R., Sihvonen, M.-L., Determination of EDTA and DTPA as their Fe(III) complexes in pulp and paper mill process and waste waters by liquid chromatography. *Analytica Chimica Acta* 1995, *303*, 187-192.
- [89] Lucena J J, Barak P, Hernández-Apaolaza L, Isocratic ion-pair high-performance liquid chromatographic method for the determination of various iron(III) chelates. *Journal of Chromatography A* 1996, *727*, 253-264.
- [90] Chen, Z., Megharaj, M., Naidu, R., Confirmation of iron complex formation using electrospray ionization mass spectrometry (ESI-MS) and sample stacking for analysis of iron polycarboxylate speciation by capillary electrophoresis. *Microchemical Journal* 2007, *86*, 94-101.
- [91] Quintana, J. B., Reemtsma, T., Rapid and sensitive determination of ethylenediaminetetraacetic acid and diethylenetriaminepentaacetic acid in water samples by ion-pair reversed-phase liquid chromatography-electrospray tandem mass spectrometry. *Journal of Chromatography A* 2007, *1145*, 110-117.
- [92] Álvarez-Fernández Ana, Orera, I., Javier, A., Anunciación, A., Determination of Synthetic Ferric Chelates Used as Fertilizers by Liquid Chromatography-

- Electrospray/Mass Spectrometry in Agricultural Matrices. *Journal of the American Society for Mass Spectrometry* 2007, *18*, 37-47.
- [93] Yeh, C.-F., Jiang, S.-J., Speciation of V, Cr and Fe by capillary electrophoresis-bandpass reaction cell inductively coupled plasma mass spectrometry. *Journal of Chromatography A* 2004, *1029*, 255-261.
- [94] Xuan, Y., Scheuermann, E. B., Meda, A. R., Hayen, H., *et al.*, Separation and identification of phytosiderophores and their metal complexes in plants by zwitterionic hydrophilic interaction liquid chromatography coupled to electrospray ionization mass spectrometry. *Journal of Chromatography A* 2006, *1136*, 73-81.
- [95] John A Neal, Norman J Rose, Stereospecific ligands and their complexes. I. A cobalt(III) complex of ethylenediaminedisuccinic acid. *Inorganic Chemistry* 1968, *7*, 2405-2412.
- [96] P C Vandevivere, H Saveyn, W Verstraete, T C J Feijtel, D R Schowanek, Biodegradation of Metal-[S,S]-EDDS Complexes. *Environmental Science & Technology* 2001, *35*, 1765-1770.
- [97] Tandy, S., Ammann, A., Schulin, R., Nowack, B., Biodegradation and speciation of residual SS-ethylenediaminedisuccinic acid (EDDS) in soil solution left after soil washing. *Environmental Pollution* 2006, *142*, 191-199.
- [98] George S Puritch, Enhancement of metal molluscicides by ethylene diamine disuccinic acid (EDDS). <http://www.freepatentsonline.com/EP1052899B1.html> 2002.
- [99] W Pongsuwan, E Fukusaki, Takeshi Bamba, Tsutomu Yonetani, *et al.*, Prediction of Japanese Green Tea Ranking by Gas Chromatography/Mass Spectrometry-Based Hydrophilic Metabolite Fingerprinting. *Journal of Agricultural and Food Chemistry* 2007, *55*, 231-236.
- [100] Hsieh, M.-M., Chen, S.-M., Determination of amino acids in tea leaves and beverages using capillary electrophoresis with light-emitting diode-induced fluorescence detection. *Talanta* 2007, *73*, 326-331.
- [101] Zhao, J., Chen, Q., Huang, X., Fang, C. H., Qualitative identification of tea categories by near infrared spectroscopy and support vector machine. *Journal of Pharmaceutical and Biomedical Analysis* 2006, *41*, 1198-1204.
- [102] Lillo, M. P. y., Sanderson, P. N., Wantling, E. L., Pudney, P. D. A., *et al.*, *Developments in Food Science*, Elsevier 2006, pp. 537-540.
- [103] Daszykowski, M., Vander Heyden, Y., Walczak, B., Robust partial least squares model for prediction of green tea antioxidant capacity from chromatograms. *Journal of Chromatography A* 2007, *1176*, 12-18.
- [104] Kumar, S. D., Narayan, G., Hassarajani, S., Determination of anionic minerals in black and kombucha tea using ion chromatography. *Food Chemistry* 2008, *111*, 784-788.
- [105] Wang, D., Lu, J., Miao, A., Xie, Z., Yang, D., HPLC-DAD-ESI-MS/MS analysis of polyphenols and purine alkaloids in leaves of 22 tea cultivars in China. *Journal of Food Composition and Analysis* 2008, *21*, 361-369.
- [106] X B Yang, Yan Zhao, Ying Yang, Yun Ruan, Isolation and Characterization of Immunostimulatory Polysaccharide from an Herb Tea, *Gynostemma pentaphyllum* Makino. *Journal of Agricultural and Food Chemistry* 2008, *56*, 6905-6909.
- [107] Xiang, L., Xing, D., Lei, F., Wang, W., *et al.*, Effects of Season, Variety, and Processing Method on Ellagic Acid Content in Pomegranate Leaves. *Tsinghua Science & Technology* 2008, *13*, 460-465.
- [108] Daszykowski M, W Wu, A W Nicholls, R J Ball, *et al.*, Identifying potential biomarkers in LC-MS data. *Journal of Chemometrics* 2007, *21*, 292-302.
- [109] Nielsen, N.-P. V., Carstensen, J. M., Smedsgaard, J., Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *Journal of Chromatography A* 1998, *805*, 17-35.

- [110] Katajamaa, M., Oresic, M., Data processing for mass spectrometry-based metabolomics. *Journal of Chromatography A* 2007, *1158*, 318-328.
- [111] C A Smith, E J Want, G O'Maille, R Abagyan, G Siuzdak, XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Analytical Chemistry* 2006, *78* 779-787.
- [112] Johnson, K. J., Wright, B. W., Jarman, K. H., Synovec, R. E., High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analysis. *Journal of Chromatography A* 2003, *996*, 141-155.
- [113] Minhø Chae, R J Shmookler, J J Thaden, An iterative block-shifting approach to retention time alignment that preserves the shape and area of gas chromatography-mass spectrometry peaks. *BMC Bioinformatics* 2008, *9*, 15-26.
- [114] T Yann, Wilson, J., Machine Learning. <http://hunch.net/> 2005.
- [115] Kam HT, Random decision forest. *Proc. of the 3rd int'l conf. on document analysis and recognition* 1995, 278-282.
- [116] Kim H T, The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1998, *20*, 832-844.
- [117] Kim, H.-C., Pang, S., Je, H.-M., Kim, D., Yang Bang, S., Constructing support vector machine ensemble. *Pattern Recognition* 2003, *36*, 2757-2767.
- [118] Vapnik V N, , Support Vector Machines. *The Nature of Statistical Learning Theory 2nd EDI* 2000.
- [119] Vapnik V N, Chervonenkis A Y, On the uniform convergence of relative frequencies of events to their probabilities. . *Theory of Probability and its Applications* 1971, *16*, 264-281.
- [120] Yali Amit, Donald Geman, Shape quantization and recognition with randomized trees, . *Neural Computation*, 1997, *9*, 1545-1588.
- [121] Kim K H, Kim S S, Kim S J, A Superiority of nonlinear mapping in decoding multiple single-unit neuronal spike trains: a simulation study. *J. Neurosci. Methods* 2005, *150*, 202-211.
- [122] Kim H, Park H, Protein secondary structure prediction based on an improved support vector machines approach *Protein Eng.* 2003, *16*, 553-560.
- [123] I Guyon, J Weston, S Barnhill, V Vapnik, Gene Selection for Cancer Classification using Support Vector Machines. *Journal Machine Learning* 2002, *46*, 389-422.
- [124] Di Wu, Xiaojing Chen, Yong He, Application of Image Texture for Discrimination of Tea Categories Using Multi-spectral Imaging Technique and Support Vector Machine. *2007 International Conference on Computational Intelligence and Security Workshops* 2007, 291-294.
- [125] D Wu, X J Chen, Y He, Random Forest. *International Conference on Computational Intelligence and Security Workshops* 2007.
- [126] Leo Breiman, Random Forests. *Journal Machine Learning* 2001, *45*, 5-32.
- [127] Leo Breiman, Adele Cutler, Random Forests. http://www.stat.berkeley.edu/~breiman/RandomForests/cc_manual.htm 2002.
- [128] Leo Breiman, Jh Friedman, Ra Olshen, Cj Stone, Classification and Regression Trees. *Wadsworth International Group*, 1984.
- [129] Grimm, R., Behrens, T., Märker, M., Elsenbeer, H., Soil organic carbon concentrations and stocks on Barro Colorado Island -- Digital soil mapping using Random Forests analysis. *Geoderma* 2008, *146*, 102-113.
- [130] Perdiguero-Alonso, D., Montero, F. E., Kostadinova, A., Raga, J. A., Barrett, J., Random forests, a novel approach for discrimination of fish populations using parasites as biological tags. *International Journal for Parasitology* 2008, *38*, 1425-1434.

- [131] G Tomasi, F van den Berg, C Andersson, Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics* 2004, 18, 231-241.
- [132] Pravdova, V., Walczak, B., Massart, D. L., A comparison of two algorithms for warping of analytical signals. *Analytica Chimica Acta* 2002, 456, 77-92.
- [133] G Tomasi, Thomas Skov, Frans van den Berg, Dynamic Time Warping (DTW) and Correlation Optimized Warping (COW). http://www.models.kvl.dk/source/DTW_COW/index.asp 2006.
- [134] Aizerman M A, Brave E M, Rozonoer L I, Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 1964, 25, 821-837.
- [135] Burges C J C, A Tutorial on Support Vector Machines for Pattern Recognition. *Journal of Data Mining and Knowledge Discovery* 1998, 2, 121-167.
- [136] Meyer, D., Leisch, F., Hornik, K., The support vector machine under test. *Neurocomputing* 2003, 55, 169-186.
- [137] Leo Breiman, Wald Lecture: Looking Inside The Black Box. <ftp://ftp.stat.berkeley.edu/pub/users/breiman/> 2001.
- [138] Angulo, C., Parra, X., Català, A., K-SVCR. A support vector machine for multi-class classification. *Neurocomputing* 2003, 55, 57-77.
- [139] Leo Breiman, Bagging predictors. *Journal Machine Learning* 1996, 24, 123-140.
- [140] Van Vliet E, Morath S, Eskes C, A novel in vitro metabolomics approach for neurotoxicity testing, proof of principle for methyl mercury chloride and caffeine. *Neurotoxicology* 2008, 29, 1-12.
- [141] Rademakers S E, Span P N, Kaanders J, Sweep F, *et al.*, Molecular aspects of tumour hypoxia. *Mol. Oncol* 2008, 2, 41-53.
- [142] Semenza GL, Hypoxia response elements in the aldolase A, enolase 1, and lactate dehydrogenase A gene promoters contain essential binding sites for hypoxia-inducible factor 1. *J. Biol. Chem* 1996, 271, 32529-32537.
- [143] Kim J W, Tchernyshyov I, Semenza GL, Dang CV, HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia. *Cell Metab* 2006, 3, 177-185.
- [144] Papandreou I, Cairns R A, Fontana L, Lim A L, Denko N C, HIF-1 mediates adaptation to hypoxia by actively downregulating mitochondrial oxygen consumption. *Cell Metab* 2006, 3, 187-197.
- [145] J Martin Brown, The Hypoxic Cell A Target for Selective Cancer Therapy. *Cancer Research* 1999, 5863-5870.
- [146] Kamleh M A, Hobani Y Dow, J A T, Zheng L, Watson D G, Towards a platform for the metabolomic profiling of different strains of *Drosophila melanogaster* using liquid chromatography-Fourier transform mass spectrometry. *Febs Journal* 2009, 276, 6798-6809.
- [147] Zhou J, Neale J H, Pomper M G, Kozikowski A P, NAAG peptidase inhibitors and their potential for diagnosis and therapy. *Nature Reviews Drug Discovery* 2005, 4, 1015-1026.
- [148] Baslow M H, NAAG peptidase as a therapeutic target: Potential for regulating the link between glucose metabolism and cognition. *Drug News & Perspectives* 2006, 19, 145-150.
- [149] Van Hemelrijck, Hachimi-Idrissi S, Sarre S, Ebinger G, Michotte Y, Neuroprotective effect of N-acetyl-aspartyl-glutamate in combination with mild hypothermia in the endothelin-1 rat model of focal cerebral ischaemia. *Journal of Neurochemistry* 2005, 95, 1287-1297.

- [150] Cristalli G, Lambertucci C, Marucci G, Volpini R, Dal Ben D, Adenosine receptor and its modulators: Overview on a druggable GPCR and on structure-activity relationship analysis and binding requirements of agonists and antagonists. *Current Pharmaceutical Design* 2008, *14*, 1525-1552.
- [151] Neale J H, Olszewski R T, Gehl L M, Wroblewska B, Bzdega T, The neurotransmitter N-acetylaspartylglutamate in models of pain, ALS, diabetic neuropathy, CNS injury and schizophrenia. *Trends in Pharmacological Sciences* 2005, *26*, 477-484.
- [152] Takahashi R N, Pamplona F A, Prediger R D S, Adenosine receptor antagonists for cognitive dysfunction: a review of animal studies. *Frontiers in Bioscience* 2008, *13*, 2614-2632.
- [153] Grimaldi P A, Peroxisome Proliferator-Activated Receptors as sensors of fatty acids and derivatives *Cell. Mol. Life Sci*, 2007, *64*, 2459 - 2464.
- [154] Steiber A, Kerner J, Hoppel C L, Carnitine: a nutritional, biosynthetic, and functional perspective. *Mol. Aspects Med.* 2004, *25*, 455-473.
- [155] Westina M A K, Hunt M C, Alexson S H, Short- and medium-chain carnitine acyltransferases and acyl-CoA thioesterases in mouse provide complementary systems for transport of β -oxidation products out of peroxisomes. *Cell. Mol. Life Sci.* 2008, *65*, 982 - 990.
- [156] Kersten S, Desvergne B, Wahli W, Roles of PPARs in health and disease. *Nature*. 2000, *405*, 421-424.
- [157] Babbar N, Gerner E W, Polyamines as modifiers of genetic risk factors in human intestinal cancers. *Biochemical Society Transactions* 2003, *31*, 388-392.
- [158] Linsalata M, R Giannini, Notarnicola, M., A. Cavallini, Peroxisome proliferator-activated receptor gamma and spermidine/spermine N-1-acetyltransferase gene expressions are significantly correlated in human colorectal cancer. *Bmc Cancer* 2006, *6*, 191-198.
- [159] Milovic V, Turchanowa L, Polyamines and colon cancer. *Biochemical Society Transactions* 2003, *31*, 381-383.
- [160] Baek J H, Liu Y V, McDonald K R, Wesley J B, *et al.*, Spermidine/spermine N-1-acetyltransferase-1 binds to hypoxia-inducible factor-1 alpha (HIF-1 alpha) and RACK1 and promotes ubiquitination and degradation of HIF-1 alpha. *J. Biol. Chem.* 2007, *282*, 33358-33366.
- [161] Ogino, S., Shima, K., Baba, Y., Nosho, K., *et al.*, Colorectal Cancer Expression of Peroxisome Proliferator-Activated Receptor [gamma] (PPARG, PPARgamma) Is Associated With Good Prognosis. *Gastroenterology* 2009, *136*, 1242-1250.
- [162] Pegg A E, Spermidine/spermine-N-1-acetyltransferase: a key metabolic regulator. *Am. J. Physiol.-Endocrinol. Metab* 2008, *294*, 995-1010.
- [163] Wang XM, Lipid signalling. *Current Opinions in Plant Biology* 2004, *7*, 329-336.
- [164] Dayanjan S, Substrate specificity of human ceramide kinase. *J. of Lipid Research* 2005, *46*, 2706-2716.
- [165] Eyster K M, The membrane and lipids as integral participants in signal transduction: lipid signal transduction for the non-lipid biochemist. *Adv Physiol Educ* 2007, *31*, 5-16.
- [166] Xu R, Golgi alkaline ceramidase regulates cell proliferation and survival by controlling levels of sphingosine and S1P. *FASEB J.* 2006, *20*, 1813-1825.
- [167] Klein C, Mechanism of nuclear calcium signalling by inositol 1,4,5-trisphosphate produced in the nucleus, nuclear located protein kinase C and cyclic AMP-dependent protein kinase. *Adv Physiol Educ* 2008, *25*, 7-18.

- [168] Maceyka M, Sphingosine kinase, sphingosine-1-phosphate, and apoptosis. *Biochimica et Biophysica Acta* 2002, 1585, 193- 201.
- [169] Okada T, Kajimoto T, Sphingosine kinase/sphingosine 1-phosphate signalling in central nervous system. *Cellular Signalling* 2009, 21, 7-13.
- [170] Ferguson M.J, The GPI biosynthetic pathway as a therapeutic target for African sleeping sickness. *Biochimica et Biophysica Acta* 1999, 1455, 327-340.
- [171] Lillico S, Essential Roles for GPI-anchored Proteins in African Trypanosomes Revealed Using Mutants Deficient in GPI8, . *Molecular Biology of the Cell* 2003, 14, 1182-1194.
- [172] Duncan M.J, Bacterial Penetration of Bladder Epithelium through Lipid Rafts. *J. Biological chemistry* 2004, 279, 18944-18951.
- [173] Goot G van der, Raft membrane domains: from a liquid-ordered membrane phase to a site of pathogen attack,. *IMMUNOLOGY* 2001, 13, 89-97.
- [174] Priotto S, Trypanosoma cruzi: Participation of cholesterol and placental alkaline phosphatase in the host cell invasion, . *Experimental Parasitology* 2009, 122, 70-73.
- [175] Watson A D, Lipidomics: A Global Approach to Lipid Analysis in Biological Systems. *J Lip Res* 2006, 15, 334-355.
- [176] Hsu F F, Electrospray Ionization/Tandem Quadrupole Mass Spectrometric Studies on Phosphatidylcholines: The Fragmentation Processes. *J Am Soc Mass Spectrom* 2003, 14, 352-363.
- [177] M Pulfer, R C Murphy, Mass spectrometry of phospholipids,. *Mass Spectrometry Reviews* 2003, 22, 332-364.
- [178] Perona J S, Ruiz-Gutierrez V, Simultaneous determination of molecular species of monoacylglycerols, diacylglycerols and triacylglycerols in human very-low-density lipoproteins by reversed-phase liquid chromatography. *Analytical Technologies in the Biomedical and Life Sciences* 2003, 785, 89-99.
- [179] Ruiz-Gutierrez V, Perez-Camino MC, Update on solid-phase extraction for the analysis of lipid classes and related compounds *Journal of Chromatography A* 2000, 885, 321-341.
- [180] Pacetti D, Malavolta M, Bocci F, Boselli E, Frega NG, Highperformance liquid chromatography/electrospray ionization iontrap tandem mass spectrometric analysis and quantification of phosphatidylcholine molecular species in the serum of cystic fibrosis subjects supplemented with docosahexaenoic acid. *Rapid Communications in Mass Spectrometry* 2004, 18, 2395-2400.
- [181] Andrikopoulos NK, Chromatographic and Spectroscopic Methods in the Analysis of Triacylglycerol Species and Regiospecific Isomers of Oils and Fats. *Critical Reviews in Food Science and Nutrition* 2002, 42, 473-505.
- [182] Brianna L Peterson, Brian S Cummings, A review of chromatographic methods for the assessment of phospholipids in biological samples. *Biomedical Chromatography* 2006, 20, 227-243.
- [183] Ulf Sommer, Haya Herscovitz, Francine K Welty, Catherine E Costello, LC-MS-based method for the qualitative and quantitative analysis of complex lipid mixtures. *Journal of Lipid Research* 2006, 47, 804-814.
- [184] Ogiso H, T Suzuki, R Taguchi, Development of a reverse-phase liquid chromatography electrospray ionization mass spectrometry method for lipidomics, improving detection of phosphatidic acid and phosphatidylserine. *Analytical Biochemistry* 2008, 375, 124-131.
- [185] Barroso B, Bischoff R, LC-MS analysis of phospholipids and lysophospholipids in human bronchoalveolar lavage fluid. *J. of Chrom B* 2005, 814, 21-28.

- [186] Hu C X, J van Dommelen, R van der Heijden, G Spijksma, *et al.*, RPLC-Ion-Trap-FTMS Method for Lipid Profiling of Plasma: Method Validation and Application to p53 Mutant Mouse Model. *Journal of Proteome Research* 2008, 7, 4982-4991.
- [187] Lesnefsky, E. J., Stoll, M. S. K., Minkler, P. E., Hoppel, C. L., Separation and Quantitation of Phospholipids and Lysophospholipids by High-Performance Liquid Chromatography. *Analytical Biochemistry* 2000, 285, 246-254.
- [188] Malavolta, M., Bocci, F., Boselli, E., Frega, N. G., Normal phase liquid chromatography-electrospray ionization tandem mass spectrometry analysis of phospholipid molecular species in blood mononuclear cells: application to cystic fibrosis. *Journal of Chromatography B* 2004, 810, 173-186.
- [189] Pang, L.-Q., Liang, Q.-L., Wang, Y.-M., Ping, L., Luo, G.-A., Simultaneous determination and quantification of seven major phospholipid classes in human blood using normal-phase liquid chromatography coupled with electrospray mass spectrometry and the application in diabetes nephropathy. *Journal of Chromatography B* 2008, 869, 118-125.
- [190] Wang, C., Xie, S., Yang, J., Yang, Q., Xu, G., Structural identification of human blood phospholipids using liquid chromatography/quadrupole-linear ion trap mass spectrometry. *Analytica Chimica Acta* 2004, 525, 1-10.
- [191] Paul D Rainville, Chris L Stumpf, John P Shockcor, Robert S Plumb, Jeremy K Nicholson, Novel Application of Reversed-Phase UPLC-oeTOF-MS for Lipid Analysis in Complex Biological Mixtures: A New Tool for Lipidomics. *Journal of Proteome Research* 2007, 6 552-558.
- [192] F F Hsu, Turk J, Kai Zhang, Beverley, S. M., Characterization of Inositol Phosphorylceramides from *Leishmania major* by Tandem Mass Spectrometry with Electrospray Ionization. *J Am Soc Mass Spectrom.* 2007, 18, 1591-1604.
- [193] Petrin, D., Delgaty, K., Bhatt, R., Garber, G., Clinical and Microbiological Aspects of *Trichomonas vaginalis*. *Clinical Microbiology Reviews* 1998, 11, 300-317.
- [194] Dunne R L, Dunn L A, Upcroft P, O'Donoghue P J, Upcroft J A, Drug resistance in the sexually transmitted protozoan *Trichomonas vaginalis*. *Cell Research* 2003, 13, 239-249.
- [195] PETER UPCROFT, JACQUELINE A UPCROFT, Drug Targets and Mechanisms of Resistance in the Anaerobic Protozoa. *CLINICAL MICROBIOLOGY REVIEWS* 2001, 14, 150-164.
- [196] SL, C., KL, D., SF, H.-M., DP, P., GE, G., Treatment of infections caused by metronidazole-resistant *Trichomonas vaginalis*. *Clin. Microbiol. Rev* 2004, 17, 783-793.
- [197] Austin D J, M Kakehashi, R M Anderson, The transmission dynamics of antibiotic-resistant bacteria: the relationship between resistance in commensal organisms and antibiotic consumption. *Proc. R. Soc Lond* 1997, 264, 1629-1638.
- [198] P Ayala, J Samuelson, D Wirth, E Orozco, *Entamoeba histolytica*: physiology of multidrug resistance. *Exp. Parasitol* 1990, 71, 169-175.
- [199] Yarlett, N., artinez, M. P., Goldberg, B., Kramer, D. L., W, C., Dependence of *Trichomonas vaginalis* upon polyamine backconversion. *Microbiology* 2000, 146, 2715-2722.
- [200] Amanda E McKie, Thomas Edlind, John Walker, Jeremy C Mottram, Coombs, G. H., The Primitive Protozoon *Trichomonas vaginalis* Contains Two Methionine γ -Lyase Genes That Encode Members of the γ -Family of Pyridoxal 5'-Phosphate-dependent Enzymes. *The Journal of Biological Chemistry* 1998, 273, 5549-5556.
- [201] Müller, S., Liebau, E., Walter, R. D., Krauth-Siegel, R. L., Thiol-based redox metabolism of protozoan parasites. *Trends in Parasitology* 2003, 19, 320-328.

- [202] Walker, J., Barrett, J., Thong, K.-W., The identification of a variant form of cystathionine [beta]-synthase in nematodes. *Experimental Parasitology* 1992, 75, 415-424.
- [203] Westrop, G. D., Goodall, G., Mottram, J. C., Coombs, G. H., Cysteine biosynthesis in *Trichomonas vaginalis* involves cysteine synthase utilizing O-phosphoserine. *J Biol Chem* 2006, 281, 25062-25075.
- [204] Dominy, J. E., Hwang, J., Stipanuk, M. H., Overexpression of cysteine dioxygenase reduces intracellular cysteine and glutathione pools in HepG2/C3A cells. *Am J Physiol Endocrinol Metab* 2007, 293, 62-69.
- [205] Thong, K.-W., Coombs, G. H., The effects of inhibitors of sulphur-containing amino acid metabolism on the growth of *Trichomonas vaginalis* in vitro. *Journal of Antimicrobial Chemotherapy* 1987, 19, 429-437.
- [206] Westrop, G. D., Georg, I., Coombs, G. H., The mercaptopyruvate sulfurtransferase of *Trichomonas vaginalis* links cysteine catabolism to the production of thioredoxin persulfide. *The Journal of Biological Chemistry* 2009, 284, 33485-33494.

Appendix A

Table 1.2 Metabolomics data-processing Software

Software	Type
SIEVES (Thermo)	LC-MS, LC-MS/MS data
Mzmine, Mzmine2 (http://mzmine.sourceforge.net/)	LC-MS, GC-MS, MS data (mzdata, netCDF XCalibur format)
MetAlign (RIKILT)	LC-MS , GC-MS dataBinBase
Xcms (Metlin Scripps)	LC-MS, GC-MS, MS,MS/MS data (netCDF format)
MassLynx (Waters)	LC-MS (netCDF)
MsInspect (Proteomics Fred Hutchinson Cancer Center)	MS Spectral
Progenesis PG600 (Nonlinear)	MALDI and SELDI MS
XAlign (Bindley Bio Ctr, Purdue Univ.)	LC-MS
Malign (Matlab Bioinformatics Toolbox)	MS spectral
Genedata Expressionist (genedata.com)	LC-MS, MS spectral
MS Align (Connitect Univ.)	High Res MS spectral
MarkerView (ABI/Sciex)	LC-MS, MALDI, peak picking, alignment and statistics
MathDAMP (Keio Uni.)	GC-MS, LC-MS, CE-MS
CPM MatLab toolbox (Toronto Uni.)	LC-MS, GC-MS
MS-Xelerator (MSMETRIX)	LC/MS
ApLCMS - Adaptive processing	LC/MS data (R package)
COMSPARI (Biomechanic)	GC/MS LC/MS (netCDF)

HiRes (Columbia Univ. Medical Center)	NMR data
BluFuse (BlueGnome)	for MS and NMR data
SpecAlign University of Oxford	Alignment of SELDI, MALDI, NMR, RAMAN
ChenomX Profiler (Chenomx)	NMR binning, alignment (+ Database search)
Automics	NMR alignment and statistics
KnowItAll Metabolomics (BioRad)	NMR alignment, bucketing and binning (+Database search)

Appendix 1 MATLAB codes for COW alignment

```

Function [Warping,XWarped,Diagnos] = cow(T,X,Seg,Slack,Options)
% function [Warping,XWarped,Diagnos] = cow(T,X,Seg,Slack,Options);
% Correlation Optimized Warping function with linear interpolation
%
% in: T (1 x nt) target vector
%     X (mP x nP) matrix with data for mP row vectors of length nP to be
warping/corrected
%     Seg (1 x 1) segment length; number of segments N = floor(nP/m)
%     or (2 x N+1) matrix with segment (pre-determined) boundary-points
%         first row = index in "xt", must start with 1 and end with "nt"
%         second row = index in "xP", must start with 1 and end with "nP"
%     Slack (1 x 1) 'slack' - maximum range or degree of warping in segment length
"m"
%     Options (1 x 5) 1 : triggers plot and progress-text (note: only last row/object in
"xP" is plotted)
%         2 : correlation power (minimum 1th power, maximum is 4th power)
%         3 : force equal segment lengths in "xt" and "xP" instead of filling up
"xt" with N boundary-points
%         (notice that different number of boundaries in "xt" and "xP" will
generate an error)
%         4 : fix maximum correction to + or - options(4) points from the
diagonal
%         5 : save in "diagnos" the table with the optimal values of loss
function and predecessor (memory
consuming for large problems - on how to read the tables are in
the m-file
%         default [0 1 0 0 0] (no plot; power 1; no forced equal segment lengths; no
band constraints; no Table in "diagnos")
%
% out: Warping (mP x N x 2) interpolation segment starting points (in "nP"
units) after warping (first slab) and before warping (second slab)
%     (difference of the two = alignment by repositioning segment
boundaries; useful for comparing correction in different/new
objects/samples)
%     XWarped (mP x nt) corrected vectors (from "xP" warped to mach "xt")
%     Diagnos (struct) warping diagnostics: options, segment, slack,
%         index in target ("xt", "warping" is shift compared to this) and sample ("xP"),
search range in "xP", computation time
%     (note: diagnostics are only saved for one - the last - signal in "xP")
%
%
Comments: The codes were the development version by Leon and Blair on the
basis of core programming by Giorgio Tomasi / Frans van den Berg from
Copenhagen Univsity.

```

```
%% Check Input values
```

```
if nargin < 4
    help cow;
```

```

    return;
end
if nargin < 5
    Options = [0 1 0 0 0];
end
if length(Options) < 5
    Options_def = [0 1 0 0 0];
    Options(length(Options) + 1:5) = Options_def(length(Options) + 1:5);
end
if (Options(2) < 1) || (Options(2) > 4)
    error('ERROR: "Options(2)" (correlation power) must be in the range 1:4');
end
if any(isnan(T)) || any(isnan(X(:)))
    error('ERROR: function "cow" can not handle missing values');
end

%% Initialise
[nX,pX] = size(X);    % nX    : number of signals that are to be aligned
                    % pX    : number of data points in each signal
pT    = size(T,2);    % pT    : number of data points in the target
XWarped = zeros(nX,pT); % XWarped: initialise matrix of warped signals
Time    = zeros(1,1); % Time  : processing time
%% Initialise segments
Seg      = round(Seg); % Only integers are currently allowed as segment
boundaries
Pred_Bound = length(Seg) > 1; % True if segment boundaries are predefined
if Pred_Bound

    if not(isequal(Seg(:,1),ones(2,1)) & isequal(Seg(:,end),[pT,pX]))
        error('End points must be equal to 1 and to the length of the pattern/target'); end

    LenSeg = diff(Seg,1,2); % LenSeg(1,:): Length of the segments in the - 1
    if not(all(LenSeg >= 2))
        error('Segments must contain at least two points'); end

    nSeg = size(LenSeg,2); % nSeg: number of segments
else

    if Seg > min(pX,pT)
        error('Segment length is larger than length of the signal'); end

    if Options(3) % Segments in the signals can have different length from those in the
target
        nSeg      = floor((pT - 1)/Seg);
        LenSeg(1,1:nSeg) = floor((pT - 1)/nSeg);
        LenSeg(2,1:nSeg) = floor((pX - 1)/nSeg);
        fprintf('\n Segment length adjusted to best cover the remainders')
    else

        nSeg      = floor((pT - 1) / (Seg - 1));
        LenSeg(1:2,1:nSeg) = Seg - 1;
    end
end

```

```

    if floor((pX - 1) / (Seg - 1)) ~= nSeg
        error('For fixed segment lengths the target and the signal do not have the
same number of segments'); end

    end
    temp = rem(pT - 1,LenSeg(1,1)); % The remainders are attached to the last
segment in the target and in the reference
    if temp > 0
        LenSeg(1,nSeg) = LenSeg(1,nSeg) + temp;
        if Options(1)
            fprintf('\n Segments: %i points x %i segments + %i (target)',LenSeg(1,1) +
1,nSeg - 1,LenSeg(1,end) + 1); end
        else
            if Options(1)
                fprintf('\n Segments: %i points x %i segments (target)',LenSeg(2,1) + 1,nSeg);
            end
        end
    end
    temp = rem(pX - 1,LenSeg(2,1));
    if temp > 0
        LenSeg(2,nSeg) = LenSeg(2,nSeg) + temp;
        if Options(1)
            fprintf('\n          %i points x %i segments + %i (signals)\n',LenSeg(2,1) +
1,nSeg - 1,LenSeg(2,end) + 1); end
        else
            if Options(1)
                fprintf('\n          %i points x %i segments (signals)\n',LenSeg(2,1) + 1,nSeg);
            end
        end
    end
end
end

end
if any(LenSeg(:) <= Slack + 2) % Two points are the minimum required for linear
interpolation
    error('The slack cannot be larger than the length of the segments'); end

bT = cumsum([1,LenSeg(1,:)]);
bP = cumsum([1,LenSeg(2,:)]);
Warping = zeros(nX,nSeg + 1);

%% Check slack
if length(Slack) > 1 % Different slacks for the segment boundaries will be
implemented
    if size(Slack,2) <= nSeg
        error('The number of slack parameters is not equal to the number of optimised
segments');
    end
    fprintf('\n Multiple slacks have not been implemented yet')
    return
end
Slacks_vec = -Slack:Slack; % All possible slacks for a segment
boundary
%% Set feasible points for boundaries
Bounds = ones(2,nSeg + 1);

```

```

% Slope Constraints
offs      = (Slack * [-1,1]) * (0:nSeg);
Bounds_a  = bP(ones(2,1),1:nSeg + 1) + offs;
Bounds_b  = bP(ones(2,1),1:nSeg + 1) + offs(:,nSeg + 1:-1:1);
Bounds(1,:) = max(Bounds_a(1,:),Bounds_b(1,:));
Bounds(2,:) = min(Bounds_a(2,:),Bounds_b(2,:));
% Band Constraints
if Options(4)

    if abs(pT - pX) > Options(4)
        error('The band is too narrow and proper correction is not possible');end

    Bounds(1,:) = max(Bounds(1,:),max(0,pX/pT * bT - Options(4)));
    Bounds(2,:) = min(Bounds(2,:),min(pX,pX/pT * bT + Options(4)));
    if any(diff(Bounds < 0))
        error('The band is incompatible with the fixed boundaries'); end

end

%% Calculate first derivatives for interpolation
Xdifff = diff(X,1,2);

%% Calculate coefficients and indexes for interpolation
Int_Coeff = cell(nSeg,1);
Int_Index = Int_Coeff;
if ~Pred_Bound
    [A,B] = Intercoeff(LenSeg(1,1) + 1,LenSeg(2,1) + Slacks_vec + 1,Slacks_vec);
    [Int_Coeff{1:nSeg - 1}] = deal(A);
    [Int_Index{1:nSeg - 1}] = deal(B);
    [Int_Coeff{nSeg},Int_Index{nSeg}] = Intercoeff(LenSeg(1,nSeg) + 1,LenSeg(2,nSeg) + Slacks_vec + 1,Slacks_vec);
else
    for i_seg = 1:nSeg
        [Int_Coeff{i_seg},Int_Index{i_seg}] = Intercoeff(LenSeg(1,i_seg) + 1,LenSeg(2,i_seg) + Slacks_vec + 1,Slacks_vec);
    end
end

%% Dynamic Programming Section
Table_Index = cumsum([0,diff(Bounds) + 1]); % Indexes for the first node
(boundary point) of each segment in Table
Table = zeros(3,Table_Index(nSeg + 2),nX); % Table: each column refer to a
node
% (1,i) position of the boundary point in the
signal
% (2,i) optimal
% value of the loss function up to node (i)
% (3,i) pointer to optimal preceding node (in
Table)
Table(2,2:end,1:nX) = -Inf; % All loss function values apart from
node (1) are set to -Inf

```

```

for i_seg = 1:nSeg + 1                                % Initialise Table
    v = (Bounds(1,i_seg):Bounds(2,i_seg));
    Table(1,Table_Index(i_seg) + 1:Table_Index(i_seg + 1),:) = v(:,ones(nX,1));
end
warning('off','MATLAB:divideByZero')                % To avoid warning if division for zero
                                                    occurs

tic
% Forward phase
for i_seg = 1:nSeg                                    % Loop over segments

    a = Slacks_vec + LenSeg(2,i_seg);                % a,b,c: auxiliary values that
    depend only on segment number and not node
    b = Table_Index(i_seg) + 1 - Bounds(1,i_seg);
    c = LenSeg(1,i_seg) + 1;
    Count = 1;                                        % Counter for local table for segment
i_seg
    Node_Z = Table_Index(i_seg + 2);                % Last node for segment
i_seg
    Node_A = Table_Index(i_seg + 1) + 1;            % First node for segment
i_seg
    Bound_k_Table = zeros(2,Node_Z - Node_A + 1,nX); % Initialise local table
    for boundary

        Int_Index_Seg = Int_Index{i_seg}' - (LenSeg(2,i_seg) + 1); % Indexes for
        interpolation of segment i_seg
        Int_Coeff_Seg = Int_Coeff{i_seg}';          % Coefficients for interpolation
        of segment i_seg

        TSeg = T(bT(i_seg):bT(i_seg + 1));          % Segment i_seg of target T
        TSeg_centred = TSeg - sum(TSeg)/size(TSeg,2); % Centred TSeg (for
        correlation coefficients)
        Norm_TSeg_cen = norm(TSeg_centred);          % (n - 1) * standard
        deviation of TSeg

        for i_node = Node_A:Node_Z                    % Loop over nodes (i.e.
        possible boundary positions) for segment i_seg

            Prec_Nodes = Table(1,i_node) - a;          % Possible
            predecessors given the allowed segment lengths
            Allowed_Arcs = Prec_Nodes >= Bounds(1,i_seg) & Prec_Nodes <=
            Bounds(2,i_seg); % Arcs allowed by local and global constraints
            Nodes_TablePointer = b + Prec_Nodes(Allowed_Arcs); %
            Pointer to predecessors in Table
            N_AA = sum(Allowed_Arcs);                 % Number of
            allowed arcs
            if N_AA % Sometimes boundaries are ineffective and few nodes are allowed that
            cannot be reached
                % It has to be further investigated

                Index_Node = Table(1,i_node) + Int_Index_Seg(:,Allowed_Arcs);
                % Interpolation signal indexes for all the allowed arcs for node i_node

```

```

    Coeff_b = Int_Coeff_Seg(:,Allowed_Arcs);
% Interpolation coefficients for all the allowed arcs for node i_node
    Coeff_b = Coeff_b(:);
    Coeff_b = Coeff_b(ones(nX,1),:);
    Xi_Seg = X(:,Index_Node);
    Xi_diff = Xdiff(:,Index_Node);
    Xi_Seg = reshape((Xi_Seg + Coeff_b .* Xi_diff)',c,N_AA * nX);
% Interpolate for all allowed predecessors
    Xi_Seg_mean = sum(Xi_Seg)/size(Xi_Seg,1);
% Means of the interpolated segments
    Norm_Xi_Seg_cen = sqrt(sum(Xi_Seg.^2) - size(Xi_Seg,1) *
Xi_Seg_mean.^2); % Fast method for calculating the covariance of T and X (no
centering of X is needed)
    CCs_Node = (TSeg_centred * Xi_Seg)./(Norm_TSeg_cen *
Norm_Xi_Seg_cen); % Correlation coefficients relative to all possible
predecessors
    CCs_Node(~isfinite(CCs_Node)) = 0; %
If standard deviation is zero, update is not chosen
    CCs_Node = reshape(CCs_Node,N_AA,nX);
    if Options(2) == 1
        Cost_Fun = reshape(Table(2,Nodes_TablePointer,:),N_AA,nX) +
CCs_Node; % Optimal value of loss function from all predecessors
    else
        Cost_Fun = reshape(Table(2,Nodes_TablePointer,:),N_AA,nX) +
CCs_Node.^Options(2);
    end
    [ind,pos] = max(Cost_Fun,[],1); %
Optimal value of loss function from all predecessors
    Bound_k_Table(1,Count,:) = ind;
    Bound_k_Table(2,Count,:) = Nodes_TablePointer(pos);
% Pointer to optimal predecessor
    Count = Count + 1;

end

end % i_node
    Table(2:3,Node_A:Node_Z,:) = Bound_k_Table; % Update general table (it
turned out to be faster than using Table directly in the loop over nodes

end % i_seg
Time = toc;

for i_sam = 1:nX % Loop over samples/signals
    % Backward phase
    Pointer = size(Table,2); % Backtrace optimal boundaries using
the pointers in Table
    Warping(i_sam,nSeg + 1) = pX;
    for i_bound = nSeg:-1:1
        Pointer = Table(3,Pointer,i_sam);
        Warping(i_sam,i_bound) = Table(1,Pointer,i_sam);
    end
% if Options(1) % Some output if requested

```

```

%   fprintf('\n Sample %i: %g sec',i_sam,Time(i_sam));
%   end

end
Warping(:,2) = bT(ones(nX,1),:);
warning('on','MATLAB:divideByZero')
% fprintf('\n')

%% Output
if nargout > 1 % Reconstruct aligned signals

    for i_seg = 1:nSeg

        indT = bT(i_seg):bT(i_seg + 1);
        lenT = bT(i_seg + 1) - bT(i_seg);
        for i_sam = 1:nX
            indX = Warping(i_sam,i_seg):Warping(i_sam,i_seg + 1);
            lenX = Warping(i_sam,i_seg + 1) - Warping(i_sam,i_seg);
            % NB the right handside expression must be transposed to fit MATLAB
            % version 6.5
            XWarped(i_sam,indT) = interp1q(indX' - Warping(i_sam,i_seg) +
            1,X(i_sam,indX)',(0:lenT)/lenT * lenX + 1)';
            end

        end

    end

end

%% Plot
if Options(1)

    figure
    minmaxaxis = [1 max([pT pX]) min([T X(nX,:)]) max([T X(nX,:)])];
    subplot(2,1,1);
    plot(1:pT,T,'b',bT,T(bT),'b',1:pX,X(nX,:),'g',bP,X(nX,bP),'g');
    hold on
    for a = 2:length(Warping(nX,:,1))

        plot([bT(a) Warping(nX,a,1)],[T(Warping(nX,a,2)) T(Warping(nX,a,2))],'r');
        if (Warping(nX,a,2) > Warping(nX,a,1))
            plot(Warping(nX,a,2),T(Warping(nX,a,2)),'>r');
        else
            plot(Warping(nX,a,2),T(Warping(nX,a,2)),'<r');
        end

    end

    end
    hold off
    axis(minmaxaxis)
    grid
    title(['COW reference = blue, Sample ' num2str(nX) '('/ ' num2str(nX) ') = green,
    Segment-boundary movement = red']);
    subplot(2,1,2);

```

```

plot(1:pT,T,'b',1:pT,XWarped(nX,:),'g');
grid;
axis(minmaxaxis);
title('Warped sample')

```

```
end
```

```
%% Function to calculate coefficients for interpolation
```

```
function [Coeff,Index] = InterpCoeff(n,nprime,offs)
```

```
p = length(nprime);
```

```
q = n - 1;
```

```
Coeff = zeros(p,n);
```

```
Index = zeros(p,n);
```

```
for i_p = 1:p
```

```
    pp          = 1:nprime(i_p);
```

```
    p           = (0:q) * (nprime(i_p) - 1)/q + 1;
```

```
    [ignore,k]  = histc(p,pp);
```

```
    k(p < 1)    = 1;
```

```
    k(p >= nprime(i_p)) = nprime(i_p) - 1;
```

```
    Coeff(i_p,:) = (p - pp(k));
```

```
    Index(i_p,:) = k - offs(i_p);
```

```
end
```

Appendix 2 MS-SQL script for Picker Shell

```
@ Title : Mass Chrom Picker
@ Version: 2.0
@ Author : Leon
@ Date : 4/1/2008
@ Desc : -
@ History: V1.0, V1.1, V1.2, V1.3.
```

declare

```
@f_param_A decimal(18,8),
@s_path varchar(100),
@s_flms varchar(8000),
@i_flnum int,
@i_flnumA int,
@i_flnumB int,
@f_Peak_Index float,
```

```
@s_dpath varchar(100),
@s_dfile varchar(100),
@t_dttm1 datetime,
@t_dttm2 datetime,
@t_dttm3 datetime;
```

/*=====

=====

Define the parameters and Variable

=====

=====*/

```
set @f_param_A = %P_param_A;
set @f_param_B = %P_param_B;
set @f_param_C = %P_param_C;
set @f_param_D = %P_param_D;
set @f_Peak_Index = %P_Peak_Index;
set @s_path = '%P_path';
set @s_flms = '%P_flms';
set @i_flnum = %P_flnum;
set @i_flnumA = %P_flnumA;
set @i_flnumB = %P_flnumB;
set @s_dpath = '%P_dpath';
set @s_dfile = '%P_dfile';
```

```
set @f_param_A = @f_param_A;
set @f_param_B = @f_param_B/1000000;
set @f_param_C = @f_param_C/2;
set @f_param_D = @f_param_D/1000000;
```



```

set @s_dtb = @s_dtb + 'drop table ##definethb;';
set @_s_sql = replace(@_s_sqlt, '@FileName@', @s_dfile);
exec(@_s_sql);

```

```

/*=====
=====

```

Step 2.1 Prepare

```

=====
=====*/
set @_s_sqlt = 'if object_id("tempdb.##@FileName@_step_1") is not null drop
table ##@FileName@_step_1;' +
               'if object_id("tempdb.##@FileName@_bnd") is not null
drop table ##@FileName@_bnd;' +
               'Declare @f_int_sum decimal(38,8),' +
               '        @f_param_A decimal(18,8),' +
               '        @f_param_B decimal(18,8),' +
               '        @f_param_C decimal(18,8);' +
               'set @f_param_A = ' + str(@f_param_A, 20, 8) + ';' +
               'set @f_param_B = ' + str(@f_param_B, 20, 8) + ';' +
               'set @f_param_C = ' + str(@f_param_C, 20, 8) + ';' +
               'select      @f_int_sum      =      sum(Intensity)      from
##@FileName@;' + --sum Intensity
               'select [ID], M_Z, Intensity, (Intensity/@f_int_sum*100) as
Normalisation, [time]' +
               'INTO ##@FileName@_step_1 ' +
               'FROM ##@FileName@ WHERE @f_param_A<=Intensity;' +
               'select ' +
               --M_Z Çø¼ä
               '[ID], ((1+@f_param_B) * M_Z) as mz_mx, ' +
               '((1-@f_param_B) * M_Z) as mz_mn,' +
               --Ê±¼äÇø¼ä
               '[time] + @f_param_C as tm_mx, ' +
               '[time] - @f_param_C as tm_mn, ' +
               'M_Z, Intensity, Normalisation, [time]' +
               'into ##@FileName@_bnd ' +
               'from ##@FileName@_step_1 order by M_Z asc, [time]
asc;';
set @i =0;
set @_i_sp = 0;
set @_i_ep = 0;
while @i<@i_flnum
begin
    set @_i_sp = @_i_ep + 1;
    set @_i_ep = charindex('|', @s_flnms, @_i_sp);

```

```

set @_s_fn = substring(@s_fnms, @_i_sp, @_i_ep - @_i_sp);
--í¼ÖÉ¾ÿÀÛÊ±±íµÁÓí¾ä£¬ÖÚ'!Áí½äÊø°óO»ÆðÉ¾ÿÀÛÊ±±í
set @s_dtb = @s_dtb + 'drop table ##' + @_s_fn + '_step_1';
set @s_dtb = @s_dtb + 'drop table ##' + @_s_fn + '_bnd';

set @_s_sql = replace(@_s_sqlt, '@FileName@', @_s_fn);
EXEC(@_s_sql);
set @i = @i + 1;
end

/*=====
=====

Step 2.2 Create table

=====
=====*/
if object_id('tempdb..##alltb') is not null drop table ##alltb;
if object_id('tempdb..##allbnd') is not null drop table ##allbnd;

set @s_dtb = @s_dtb + 'drop table ##alltb;drop table ##allbnd;';--
ìí¼ÓÉ¾ÿÀÛÊ±±íÓí¾ä

CREATE TABLE ##alltb (
    [M_Z] decimal(18,8) NULL ,
    [Intensity] decimal(18,8) NULL ,
    [Normalisation] decimal(18,8) NULL ,
    [time] decimal(18,8) NULL ,
    [time_max] decimal(18,8) NULL ,
    [time_min] decimal(18,8) NULL ,
    [peak_idx] decimal(18,8) NULL ,
    [fname] varchar(200) NULL
)

--ìí¼ÓÖ÷¼ü
exec('ALTER TABLE ##alltb add [ID] bigint PRIMARY KEY identity(1,1);')

/*=====
=====

Step 2.3 Process Single File

=====
=====*/
set @_s_sqlt = '
declare @id    bigint,
        @m_z   decimal(18,8),
        @mz_mx decimal(18,8),
        @mz_mn decimal(18,8),
        @tm_mx decimal(18,8),

```

```

@tm_mn decimal(18,8),
@time decimal(18,8),
@inty decimal(18,8),
@norm decimal(18,8),

@id2 bigint,
@m_z2 decimal(18,8),
@mz_mx2 decimal(18,8),
@mz_mn2 decimal(18,8),
@tm_mx2 decimal(18,8),
@tm_mn2 decimal(18,8),
@time2 decimal(18,8),
@inty2 decimal(18,8),
@norm2 decimal(18,8),

@_i_cnt int,
@_f_total decimal(18,8),
@_f_intmax decimal(18,8),
@_f_peak_idx decimal(18,8),
@_f_intenhalfsum decimal(18,8),
@_f_intensity decimal(18,8),
@_f_normalisation decimal(18,8),
@_f_timemax decimal(18,8),
@_f_timemin decimal(18,8),
@s_intensity varchar(8000),
@i int,
@_i_sp int,
@_i_ep int;

```

```

declare cur_bnd cursor SCROLL for select [id], mz_mx, mz_mn, M_Z, tm_mx,
tm_mn, [time], Intensity, Normalisation from ##@FileName@_bnd order by M_Z
asc, [Time] asc for READ ONLY;
open cur_bnd;
fetch next from cur_bnd INTO @id, @mz_mx, @mz_mn, @m_z, @tm_mx,
@tm_mn, @time, @inty, @norm;
if @@FETCH_STATUS = 0
begin
    set @_i_cnt = 1;
    set @_f_total = @m_z;
    set @_f_intensity = @inty;
    set @_f_normalisation = @norm;
    set @_f_timemax = @time;
    set @_f_timemin = @time;
    set @_f_intmax = @inty;
    set @s_intensity = CAST(@inty AS VARCHAR) + "|";
    fetch next from cur_bnd INTO @id2, @mz_mx2, @mz_mn2, @m_z2,
@tm_mx2, @tm_mn2, @time2, @inty2, @norm2;
    while @@FETCH_STATUS = 0
    begin
        if (@m_z2 between @mz_mn and @mz_mx) and (@time2 between
@tm_mn and @tm_mx)
        begin

```

```

set @_i_cnt = @_i_cnt + 1;
set @_f_total = @_f_total + @m_z2;
set @_f_intensity = @_f_intensity + @inty2;
set @_f_normalisation = @_f_normalisation + @norm2;
if @time2 > @_f_timemax
    set @_f_timemax = @time2;
if @time2 < @_f_timemin
    set @_f_timemin = @time2;
if @inty2 > @_f_intmax
    set @_f_intmax = @inty2;
set @_s_intensity = @_s_intensity + CAST(@inty2 AS
VARCHAR) + "|";
end else begin
    if @_i_cnt > 2
        begin
            set @i = 0;
            set @_i_sp = 0;
            set @_i_ep = 0;
            set @_f_intenhalfsum = 0;
            while @i < @_i_cnt
                begin
                    set @_i_sp = @_i_ep + 1;
                    set @_i_ep = charindex("|", @_s_intensity,
@_i_sp);
                    set @_f_peak_idx =
CAST(substring(@_s_intensity, @_i_sp, @_i_ep - @_i_sp) AS DECIMAL(18,8));
                    set @_f_peak_idx = @_f_peak_idx -
@_f_intmax / 2;
                    if @_f_peak_idx > 0
                        set @_f_intenhalfsum =
@_f_intenhalfsum + @_f_peak_idx;
                    set @i = @i + 1;
                end
                set @_f_peak_idx = @_f_intenhalfsum /
@_f_intensity;
            end else begin
                set @_f_peak_idx = 0;
            end;
            insert into ##alltb
([M_Z],Intensity,Normalisation,[time],[time_max],[time_min],[peak_idx],fname)
values(@_f_total/@_i_cnt,@_f_intensity,@_f_normalisation,(@_f_timemax+@_f_ti
memin)/2,@_f_timemax,@_f_timemin,@_f_peak_idx,"@FileName@");
            set @_i_cnt = 1;
            set @_f_total = @m_z2;
            set @_f_intensity = @inty2;
            set @_f_normalisation = @norm2;
            set @_f_timemax = @time2;
            set @_f_timemin = @time2;
            set @_f_intmax = @inty2;
            set @_s_intensity = CAST(@inty2 AS VARCHAR) + "|";
        end;
    set @id = @id2;

```

```

        set @mz_mx = @mz_mx2;
        set @mz_mn = @mz_mn2;
        set @m_z = @m_z2;
        set @tm_mx = @tm_mx2;
        set @tm_mn = @tm_mn2;
        set @time = @time2;
        set @inty = @inty2;
        set @norm = @norm2;
        fetch next from cur_bnd INTO @id2, @mz_mx2, @mz_mn2,
@m_z2, @tm_mx2, @tm_mn2, @time2, @inty2, @norm2;
    end;
    if @_i_cnt > 2
    begin
        set @i = 0;
        set @_i_sp = 0;
        set @_i_ep = 0;
        set @_f_intenhalfsum = 0;
        while @i < @_i_cnt
        begin
            set @_i_sp = @_i_ep + 1;
            set @_i_ep = charindex('|', @s_intensity, @_i_sp);
            set @_f_peak_idx = CAST(substring(@s_intensity, @_i_sp,
@m_i_ep - @_i_sp) AS DECIMAL(18,8));
            set @_f_peak_idx = @_f_peak_idx - @_f_intmax / 2;
            if @_f_peak_idx > 0
                set @_f_intenhalfsum = @_f_intenhalfsum +
@m_f_peak_idx;
            set @i = @i + 1;
        end
        set @_f_peak_idx = @_f_intenhalfsum / @_f_intensity;
    end else begin
        set @_f_peak_idx = 0;
    end;
    insert                                into                                ##alltb
([M_Z], Intensity, Normalisation, [time], [time_max], [time_min], [peak_idx], filename)
values(@_f_total/@_i_cnt, @_f_intensity, @_f_normalisation, (@_f_timemax + @_f_ti
memin)/2, @_f_timemax, @_f_timemin, @_f_peak_idx, "@FileName@");
end;

CLOSE cur_bnd;
DEALLOCATE cur_bnd;

set @i = 0;
set @_i_sp = 0;
set @_i_ep = 0;
while @i < @i_flnum
begin
    set @_i_sp = @_i_ep + 1;
    set @_i_ep = charindex('|', @s_flnms, @_i_sp);
    set @_s_fn = substring(@s_flnms, @_i_sp, @_i_ep - @_i_sp);

    set @_s_sql = replace(@_s_sqlt, '@FileName@', @_s_fn);

```

```

EXEC(@_s_sql);
set @i = @i + 1;
end

/*=====
=====

Step 3.1 Prepare

=====
=====*/
set @_s_sql = replace('select [ID], ((1+@f_param_B) * M_Z) as mz_mx, ((1-
@f_param_B) * M_Z) as mz_mn,[time] + @f_param_C as tm_mx,
[time] - @f_param_C as tm_mn,
M_Z, Intensity, Normalisation, [time], [time_max], [time_min], [peak_idx],
fname
into ##allbnd
from ##alltb order by M_Z asc,[time] asc;', '@f_param_B', str(@f_param_B,
20, 8));
set @_s_sql = replace(@_s_sql, '@f_param_C', str(@f_param_C, 20, 8));
exec(@_s_sql);

/*=====
=====

Step 3.2 Create a new table

=====
=====*/
if object_id('tempdb..##alltbR') is not null drop table ##alltbR;

set @s_dtb = @s_dtb + 'drop table ##alltbR;';--Ìí¼ÓÉ¾¾ýÁÙÊ±±íÓĩ¾ä

CREATE TABLE ##alltbR (
[M_Z] float NULL ,
[time_max] float NULL ,
[time_min] float NULL ,
[Peak_Index_Max] float NULL ,
[Peak_Index_Min] float NULL ,
[Formula] varchar(1300) NULL,
[Name] varchar(6600) NULL,
[Ratio] float NULL,
[ID] [bigint] NOT NULL,
PRIMARY KEY CLUSTERED
(
[ID] ASC
)
) ON [PRIMARY]

--Ìí¼Ó²çÐÐÁÐ
set @_s_sqlt = 'ALTER TABLE ##alltbR add @FileName@_Intensity decimal(18,8)
NULL, @FileName@_Normalisation decimal(18,8) NULL;';

```

```

-----
set @i = 0;
set @_i_sp = 0;
set @_i_ep = 0;
while @i < @i_flnum
begin
    set @_i_sp = @_i_ep + 1;
    set @_i_ep = charindex('|', @s_flnms, @_i_sp);
    set @_s_fn = substring(@s_flnms, @_i_sp, @_i_ep - @_i_sp);

    set @_s_sql = replace(@_s_sqlt, '@FileName@', @_s_fn);
    EXEC(@_s_sql);
    set @i = @i + 1;
end

```

```

/*=====
=====

```

Step 3.3 Process

```

=====
=====*/

```

```

declare @id  bigint,
        @m_z  DECIMAL(18,8),
        @mz_mx DECIMAL(18,8),
        @mz_mn DECIMAL(18,8),
        @tm_mx DECIMAL(18,8),
        @tm_mn DECIMAL(18,8),
        @time DECIMAL(18,8),
        @time_max DECIMAL(18,8),
        @time_min DECIMAL(18,8),
        @peak_idx DECIMAL(18,8),
        @inty DECIMAL(18,8),
        @norm DECIMAL(18,8),
        @fnm varchar(200),

        @id2  bigint,
        @m_z2 DECIMAL(18,8),
        @mz_mx2 DECIMAL(18,8),
        @mz_mn2 DECIMAL(18,8),
        @tm_mx2 DECIMAL(18,8),
        @tm_mn2 DECIMAL(18,8),
        @time2 DECIMAL(18,8),
        @time_max2 DECIMAL(18,8),
        @time_min2 DECIMAL(18,8),
        @peak_idx2 DECIMAL(18,8),
        @inty2 DECIMAL(18,8),
        @norm2 DECIMAL(18,8),
        @fnm2 varchar(200),

```

```

        @_s_sql2 varchar(1300),
        @_s_rf varchar(6600),
        @_i_cnt int,
        @_f_total DECIMAL(18,8)

declare @_s varchar(2000);
declare @_sA varchar(2000);
declare @_sB varchar(2000);

declare cur_bnd cursor SCROLL for select [id], mz_mx, mz_mn, M_Z, tm_mx,
tm_mn, [time], time_max, time_min, peak_idx, Intensity, Normalisation, flname from
##allbnd order by M_Z asc, [Time] asc for READ ONLY;
open cur_bnd;
fetch next from cur_bnd INTO @id, @mz_mx, @mz_mn, @m_z, @tm_mx,
@tm_mn, @time, @time_max, @time_min, @peak_idx, @inty, @norm, @fnm;
set @_i_cnt = 1;
set @_f_total = @m_z;

if @@FETCH_STATUS = 0
begin
        set          @_s          =          'insert          into          ##alltbR
([ID],[M_Z],time_max,time_min,Peak_Index_Max,Peak_Index_Min,Formula,[Name],
@fnm_Intensity,@fnm_Normalisation)
values(@id,@m_z,@time_max,@time_min,@peak_idx,@peak_idx,NULL,NULL,@i
nty,@norm);'
        set @_s = replace(@_s, '@fnm', @fnm);
        set @_s = replace(@_s, '@id', @id);
        set @_s = replace(@_s, '@m_z', @m_z);
        set @_s = replace(@_s, '@time_max', @time_max);
        set @_s = replace(@_s, '@time_min', @time_min);
        set @_s = replace(@_s, '@peak_idx', @peak_idx);
        set @_s = replace(@_s, '@inty', @inty);
        set @_s = replace(@_s, '@norm', @norm);
        exec(@_s)
        fetch next from cur_bnd INTO @id2, @mz_mx2, @mz_mn2, @m_z2,
@tm_mx2, @tm_mn2, @time2, @time_max2, @time_min2, @peak_idx2, @inty2,
@norm2, @fnm2;
        while @@FETCH_STATUS = 0
        begin
                if (@m_z2 <= @mz_mx) and (@m_z2 >= @mz_mn) and (@time2 <=
@tm_mx) and (@time2 >= @tm_mn)
                begin
                        set @_i_cnt = @_i_cnt + 1;
                        set @_f_total = @_f_total + @m_z2;
                        set          @_s          =          'update          ##alltbR          set
@fnm2_Intensity=ISNULL(@fnm2_Intensity,0)+@inty2,@fnm2_Normalisation=ISN
ULL(@fnm2_Normalisation,0)+@norm2,time_max          =          case          when
@time_max2>time_max then @time_max2 else time_max end,time_min = case
when          @time_min2<time_min          then          @time_min2          else          time_min
end,Peak_Index_Max          =          case          when          @peak_idx2>Peak_Index_Max          then
@peak_idx2          else          Peak_Index_Max          end,Peak_Index_Min          =          case          when

```

```

@peak_idx2<Peak_Index_Min then @peak_idx2 else Peak_Index_Min end where
[ID]=@id;

        set @_s = replace(@_s, '@fnm2', @fnm2);
        set @_s = replace(@_s, '@inty2', @inty2);
        set @_s = replace(@_s, '@norm2', @norm2);
        set @_s = replace(@_s, '@time_max2', @time_max2);
        set @_s = replace(@_s, '@time_min2', @time_min2);
        set @_s = replace(@_s, '@peak_idx2', @peak_idx2);
        set @_s = replace(@_s, '@id', @id);
        exec(@_s)
end else begin
        set @m_z = @_f_total / @i_cnt;
        set @_s = 'update ##alltbR set [M_Z]=@m_z where
[ID]=@id;

        set @_s = replace(@_s, '@m_z', @m_z);
        set @_s = replace(@_s, '@id', @id);
        exec(@_s)

        set @_s = 'insert into ##alltbR
([ID],[M_Z],time_max,time_min,Peak_Index_Max,Peak_Index_Min,Formula,[Name],
@fnm_Intensity,@fnm_Normalisation)
values(@id,@m_z,@time_max,@time_min,@peak_idx,@peak_idx,NULL,NULL,@i
nty,@norm);'

        set @_s = replace(@_s, '@fnm', @fnm2);
        set @_s = replace(@_s, '@id', @id2);
        set @_s = replace(@_s, '@m_z', @m_z2);
        set @_s = replace(@_s, '@time_max', @time_max2);
        set @_s = replace(@_s, '@time_min', @time_min2);
        set @_s = replace(@_s, '@peak_idx', @peak_idx2);
        set @_s = replace(@_s, '@inty', @inty2);
        set @_s = replace(@_s, '@norm', @norm2);
        exec(@_s)

        set @i_cnt = 1;
        set @_f_total = @m_z2;

        set @id = @id2;
end;
set @mz_mx = @mz_mx2;
set @mz_mn = @mz_mn2;
set @m_z = @m_z2;
set @tm_mx = @tm_mx2;
set @tm_mn = @tm_mn2;
set @time = @time2;
set @time_max = @time_max2;
set @time_min = @time_min2;
set @peak_idx = @peak_idx2;
set @inty = @inty2;
set @norm = @norm2;
set @fnm = @fnm2;

```

```

        fetch next from cur_bnd INTO @id2, @mz_mx2, @mz_mn2,
@m_z2, @tm_mx2, @tm_mn2, @time2, @time_max2, @time_min2, @peak_idx2,
@inty2, @norm2, @fnm2;
        end;

```

```

        set @m_z = @f_total / @i_cnt;
        set @_s = 'update ##alltbR set [M_Z]=@m_z where [ID]=@id;';
        set @_s = replace(@_s, '@m_z', @m_z);
        set @_s = replace(@_s, '@id', @id);
        exec(@_s)

```

```
end;
```

```

CLOSE cur_bnd;
DEALLOCATE cur_bnd;

```

```
-- Delete record(time_max eq time_min)
```

```
delete from ##alltbR where time_max=time_min
```

```
delete from ##alltbR where Peak_Index_Max <= @f_Peak_Index
```

```

/*=====
=====

```

Step 3.4 Calculate Ratio

```

=====
=====*/

```

```

set @_sA = "";
set @_sB = "";

```

```

set @i = 0;
set @i_sp = 0;
set @i_ep = 0;
while @i < @i_fnum
begin

```

```

    set @i_sp = @i_ep + 1;
    set @i_ep = charindex('|', @s_flnms, @i_sp);
    set @_s_fn = substring(@s_flnms, @i_sp, @i_ep - @i_sp);
    if @i < @i_fnumA
    begin
        -- Group A
        if @_sA = ""
        begin
            set @_sA = @_s_fn + '_Intensity';
        end else begin
            set @_sA = @_sA + ' ' + @_s_fn + '_Intensity';
        end
    end else begin
        -- Group B
        if @_sB = ""

```

```

begin
    set @_sB = @_s_fn + '_Intensity';
end else begin
    set @_sB = @_sB + ' ' + @_s_fn + '_Intensity';
end
end
end
set @i = @i + 1;
end
set @_s = 'update ##alltbR set Ratio = case when (' + @_sB + '=0 or (' + @_sB + ')
is null then 0 else ((' + @_sA + ') / ' + CAST(@i_flnumA AS varchar(3)) + ') / ((' +
@_sB + ') / ' + CAST(@i_flnumA AS varchar(3)) + ') end;';
exec(@_s);

```

```

/*=====
=====

```

Step 4 Match Name

```

=====*/
declare @_f_rule decimal(18,8),
        @_s_formula varchar(200),
        @_s_name varchar(200),

        @_i_ID integer,
        @_f_MZ decimal(18,8),
        @_f_mxMZ decimal(18,8),
        @_f_mnMZ decimal(18,8),
        @_f_mxRL decimal(18,8),
        @_f_mnRL decimal(18,8),
        @_i_dcount integer,
        @_i_mcount integer;

select @_f_mxRL = Max(M_ZRule), @_f_mnRL = Min(M_ZRule) from ##definnetb;

declare cur_dtb cursor SCROLL for select [id],M_Z from ##alltbR where M_Z
between @_f_mnRL*(1-@f_param_D) and @_f_mxRL*(1+@f_param_D);
open cur_dtb;
fetch next from cur_dtb INTO @_i_ID, @_f_MZ;
while @@FETCH_STATUS = 0
begin
    declare cur_dfn cursor SCROLL for select M_ZRule, Formula, Name from
##definnetb where @_f_MZ between M_ZRule*(1-@f_param_D) and
M_ZRule*(1+@f_param_D) Order By ABS(@_f_MZ - M_ZRule) asc for READ
ONLY;
    open cur_dfn;
    fetch next from cur_dfn into @_f_rule, @_s_formula, @_s_name;
    while @@FETCH_STATUS = 0
    begin

```

```
        update ##alltbR set [Name]=case when [Name] is null then
@s_s_name else [Name] + '|' + @_s_name end,Formula=case when Formula is null
then @_s_formula else Formula + '|' + @_s_formula end where [ID]=@_i_ID;
        fetch next from cur_dfn into @_f_rule, @_s_formula, @_s_name;
    end
    CLOSE cur_dfn;
    DEALLOCATE cur_dfn;

    fetch next from cur_dtb INTO @_i_ID, @_f_MZ;
end;
CLOSE cur_dtb;
DEALLOCATE cur_dtb;
```

Appendix 3 Macro codes for Sieves Extactor V.10.0

```

Sub hit()
Sheets("metabolites").Select

Dim RTW, MW, mzdif, RTdiff As Double
Dim counter, anticounter, D, C As Long

metano = ActiveSheet.UsedRange.Rows.Count

Dim MZarray(2000) As Double
'maximum metabolome DT below 2000
Dim RTarray(2000) As Double
Dim FMarray(2000) As String
Dim CMarray(2000) As String
Dim PTarray(2000) As String
Dim PCarray(2000) As String

counter = 0
anticounter = 0

RTW = cells(1, 12) 'parametres
of ppm and RTwidth
MW = cells(3, 12)

For i = 2 To metano ' metabolites below
1000(set)Transfer DT to array
FMarray(i - 1) = cells(i, 3).Value
CMarray(i - 1) = cells(i, 4).Value
MZarray(i - 1) = cells(i, 2).Value
RTarray(i - 1) = cells(i, 7).Value Array will read RT info.
PTarray(i - 1) = cells(i, 5).Value
PCarray(i - 1) = cells(i, 6).Value

Next i

Sheets("sieves").Select

sievesrowno = Sheets("sieves").UsedRange.Rows.Count
sievescolno = Sheets("sieves").UsedRange.Columns.Count

Range(cells(1, 1), cells(sievesrowno, (sievescolno - 6))).Copy 'Copy without
PCA

Sheets("hit").Select

Range(cells(1, 1), cells(sievesrowno, (sievescolno - 6))).PasteSpecial

```

```

Range(Columns(1), Columns(sievescolno)).Select
'Sorting by M/Z
Selection.Sort Key1:=Range("b1"), Order1:=xlAscending, Header:=xlYes,
OrderCustom:=1, MatchCase:=_
False, Orientation:=xlTopToBottom, DataOption1:=xlSortNormal, DataOption2
:=xlSortNormal

```

```

Columns("B:m").Select
Selection.Insert Shift:=xlToRight 'Two Column Insertion for
compounds and formula
Selection.Insert Shift:=xlToRight
Selection.Insert Shift:=xlToRight
Selection.Insert Shift:=xlToRight

```

```

hitrowno = Sheets("hit").UsedRange.Rows.Count
hitcolno = Sheets("hit").UsedRange.Columns.Count

```

```

cells(1, 2) = "Formula"
cells(1, 3) = "Compounds"
cells(1, 4) = "Pathway"
cells(1, 5) = "Pathway code"

```

```

For k = 2 To hitrowno
For j = 2 To metano

```

```

mzdiff = Abs(cells(k, 6) - MZarray(j - 1)) * 1000000 / MZarray(j - 1)
'RTdiff = Abs(Cells(k, 7) - RTarray(j - 1))

```

```

If mzdiff <= MW Then '(And RTdiff <= RTW) is suspended until RT
on

```

```

cells(k, 2).Value = FMarray(j - 1)
cells(k, 3).Value = Carray(j - 1)
cells(k, 4).Value = Parray(j - 1)
cells(k, 5).Value = Parray(j - 1)

```

```

cells(k, 1).Interior.ColorIndex = 4

```

```

counter = counter + 1

```



```
Sheets("metabolites").Select  
cells(14, 12) = matchingno
```

```
End Sub
```