University of Strathclyde

Strathclyde Institution of Pharamcy and Biomedical Sciences

The Combination of Hydrophilic Interaction Liquid Chromatography and High Resolution Mass Spectrometry (HILIC-HRMS). Optimisation and Application in Pharmaceutical Analysis.

by

Muhammad Anas Kamleh

A thesis presented in fulfilment of the requirements for the degree of

Doctor of Philosophy

October 2010

Declaration:

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Muhammad Anas Kamleh

15 October 2010

Acknowledgments:

I start this thesis with the name of Allah, the most beneficent, the most merciful. First and most important I thank Allah for giving me the health, power and will to proceed with the work in this thesis, to his majesty is attributed all the good and valid in this work and to my shortcomings attributed all the mistakes and flaws in it. Then I would like to thank my mother, Ikbal Kadoura for being my constant source of inspiration, charging me with her soft words urging me for a harder work. Her teachings were and still are the core of my ethics, the light illuminating my way and the motivation for me to always run the extra mile. Her wisdom and hard work have always brought me up when the life got me down. Her happiness with this work cannot even be a start to the pay back of all the efforts she invested in me, or the sacrifices she made for me. I would like as well to thank my sisters and brothers, Juhaina, Rasha, Yasser, Rula and Abdullah for all the support they provided since childhood, their efforts in teaching me number crunching started at the age of two and is paying off now.

I would like as well to deeply thank my supervisor Dr David G Watson for all the hard work and help he provided during the period of my PhD and for giving me the opportunity to participate in his awaited new book “Pharmaceutical Chemistry.” I am in debt to Damascus University for sponsoring my PhD and thankful to Professor Mhmd Ammar Al-khayat for approving the scholarship to Strathclyde University. I would like to thank my second supervisor Dr John J Parkinson for helping me with understanding advanced NMR techniques. A special thanks to Dr.Ruangelica Edrada-Ebel who saved no effort in helping all graduate students at BCDD with NMR experiments, and skilfully explained the principles of NMR and the interpretation of spectra.

I am appreciative to Professor Julian AT Dow for his help and suggestions about the *Drosophila* metabolome project and to Dr Yaha Hobani for sample extraction. I am indebted to professor Michel Barrett for all the fruitful discussions for the *Trypanosoma* metabolome project and to all members in his group who helped with sample extraction.

A sincere thanks to all the colleagues in BCDD and all the good friends and flatmates I was surrounded with during my life in Glasgow, especially Dr.Adel Daas and Mr.Hossam Reda. I am grateful as well to all my friends in Syria who were available during crisis times to ease my problems with their words and for helping my family during my absence, this includes all friends with a special thanks to Ms Lilas Jamaleden, Mr Ahmad Al-Kayal, and Ms Hanine Kayal.

I am deeply grateful for Professor Alastair Davidson for all the teaching he exerted , the advice he provided and for recommending me to the EDQM laboratory management for an internship which was an integral and essential experience which added significantly to my knowledge and for giving me the opportunity to attend ,as an observer, the meeting of group of experts. I am grateful for the wonderful reception and hospitality I received during my days in EDQM, Strasbourg, everyone in the laboratory was magnificent with a special salute to my supervisor Dr Stephan Aimling who was very generous fitting me into his extremely busy schedule and answering all my queries. I would like to thank the chromatographic

society and Dr Amjad Khan for the bursaries they provided allowing me to attend all the society`s events at no charge. I would like to thank Professor Ian D Wilson for recommending me to a great opportunity for an internship at the Imperial College of London, and would like to thank all my supervisors at IC who saved no effort in educating me and answering my questions, of them I am especially in debt to Dr Elizabeth J Want , Dr Timothy Ebbless and Dr Hector Keun. Last but never least I would like to give a big thanks to all my true friends in London, those who were not only a source of knowledge with all the fruitful discussions but as well were beside me all the time and were a second family to me especially to Ms Estitux Rico, Ms Konstantina Spagou and Mr Paul Benton.

The list is long and I am sure no matter how hard I try to be comprehensive my memory will never be good enough to encompass all the people who have done good to me and I am grateful to everyone of those, even if their names were not mentioned.

Dedication:

To the dad I never saw with my eyes, but has always been in my heart,

To *Saadi Kamleh*

1931-1980

To the memories of a sister never to be forgotten,

To Juhaina Saadi Kamleh

1966-2004

Abstract:

The development of analytical technology over the last few decades has opened the door to new and intriguing fields of research, allowing investigations at a level of quality not achieved before. Chromatography and mass spectrometry have undergone a revolution of advances in the last four decades with extensive developments in new stationary phases and new developments in ionisation and ion separation techniques. The work in this thesis investigated the application of two of these advances, hydrophilic interaction liquid chromatography (HILIC) and Fourier transformation mass spectrometry (FTMS) for the analysis of complex biological samples. This thesis focuses on the development of metabolic profiling methods which have applications in disease diagnosis, personalized healthcare and drug discovery. The complex nature of the type of samples is discussed and data simulation was used to show the importance of a separation step before detection, even with high resolution mass spectrometers. HILIC was compared to more traditional reversed phase stationary phases and found to be more suitable for the research questions in this thesis. Several commercially available chromatographic columns share the HILIC retention mechanism but differ in their chemistry. Four of these columns were compared using injections of 140 authentic standards and they were evaluated according to the retention range for the standards, peak width, isomer separation and background signal level generated in the mass spectrometer. The most "fit for purpose" column was then selected and further tested with different organic modifiers and mobile phase additives. The final method was tested for repeatability and reproducibility of retention time. The selected analytical method was then applied to the analysis of the metabolome of two biological systems i.e. *Drosophila melanogaster* and *Trypanosoma brucei* . This demonstrated the ability of this method to rapidly profile and extract biological information from the tested systems. The complex output of the analytical methods needed data processing tools, and for this purpose a computer programme was written and is described and tested in the work reported in this thesis. The in-house built code was compared with a commercially available and an academic software package and the pros and cons of each of these tools are discussed within this thesis.

List of tables:

List of figures:

List of schemes:

Abbreviations:

AA	amino acids
ACN	acetonitrile
ADP	Adenosine diphosphate
AmF	ammonium formate
AMP	Adenosine monophosphate
amu	atomic mass unit
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
CAD	corona aerosol discharge
CDT	Compound Danshen Tablets
CE	capillary electrophoresis
CLOUDS	classification of unknowns by density superposition
COMET	COnsortium for MEtabonomic Toxicology
COSY	correlation Spectroscopy
COW	Correlation optimised Warping
CRC	colorectal cancer
CS	Canton-S
CV	coefficient of variation
CYP450s	cytochrome P 450
DNA	Deoxyribonucleic acid
DOPA	dihydroxyphenylalanine
EAD	electrochemical array detection
ED	electrochemical detection
EIC	extracted ion chromatogram
ELSD	evaporative light scattering detection
EOC	endothelium ovarian cancer
EPA	eicosapentaenoic acid
ESI	electrospray ionisation
EtOH	ethanol
FA	formic acid
FAD	flavin adenine dinucleotide
FAIMS	field asymmetric waveform ion mobility spectrometer
FT-ICR	Fourier Transform Ion Cyclotron Resonance
FTIR	Fourier transformation infra red
FTMS	Fourier transofrmation mass spectrometry
FWD	forward searching
GABA	gamma amino butyric acid
GC	gas chromatography
GPC	glycerophosphocholine
GPE	glycerophosphoethanolamine
GUI	graphical User Interface
HILIC	Hydrophilic interaction liquid chromatography
HMDB	Human Metabolome Database
HPLC	high performance liquid chromatography
KEGG	Kyoto Encyclopedia of Genes and Genomes

LC	liquid chromatography
LOD	limit of detection
LOOCV	leave one out cross validation
LOQ	limit of quantification
MAK	metabolomics aid kit
MALDI	matrix assisted laser desorption ionisation
MeOH	methanol
MI	myocardial ischemia
MP	mobile phase
MS	Mass spectrometry
NAD	Nicotinamide adenine dinucleotide
NMR	Nuclear magnetic resonance
NP	normal phase
OC	Ovarian cancer
PCA	principal components analysis
PCVG	principal component variable grouping
PEG	polyethylene glycol
PFP	pentaflourophenyl
PK	pharmacokinetics
PLS	projection to latent structure
PPARs	peroxisome proliferator-activated receptors
ppm	part per million
QC	quality control
QTOFMS	quadrupole time of flight mass spectrometer
RESUME	reporting system for untargeted metabolomics
RNA	Ribonucleic acid
ROC	receiver operating characteristics
RP	reversed phase
RT	retention time
ry	rosy *(a Drosophila* genotype)
SHY	Statistical Heterospectroscopy
SMRS	standard metabolic reporting structure
SP	stationary phase
SPE	solid phase extraction
STOCSY	Statistical total correlation spectroscopy
SVM	Support vector machine
T2DM	type-2-diabetic mellitus
Tb	Trypanosoma brucei
TCA	tricarboxylic acid
TG	triglycerides
TIC	total ion chromatogram
TMAO	trimethylamine-N-oxide
TMS	trimethylsilyl
TOCSY	total correlation spectroscopy
TOF	time of flight
UPLC	Ultra high performance liquid chromatography
URL	uniform resource locator

UV	ultraviolet
VBA	Visual basic application
WT	wild type (*Drosophila*)

1.A Review of The Literature:

1.1. Introduction:

The word chromatography originates from two Greek syllables (***chroma*** meaning colour) and (***graphenie*** meaning to write). The name was assigned based on the first reported separation of compounds (late 19th century) by Mikhail Tswett who used a calcium carbonate column to separate plant pigments resulting in layers of coloured bands.Chromatographic separation is based on the relative preference of two phases to hold (retain) one compound over others. One of the two phases is called the stationary phase and the other is called the mobile phase. The latter could be a liquid, a gas or a mixture of liquid-gas. As mentioned above, the very first chromatographic column was made of calcium carbonate but since then chromatographic stationary phases have seen major advances, the most important of which happened in the last four decades and were associated with the introduction of the principles of high pressure (performance) liquid chromatography.Until recently, the two most common used types of stationary phases were normal phase (NP) and the reversed-phase (RP, non-polar groups bonded on silica gel or polymers). The mechanism of retention for the two types relied on the adsorption of polar groups in the analytes on polar groups of the stationary phase in the case of NP or the partitioning of the non-polar groups of the analytes into the lipophilic portions on the stationary phase in the case of RP.

NP chromatography is limited to non-polar solvents, and is not amenable to steep gradient systems because of slow equilibration problems. In addition using chromatographic systems with mobile phases such as hexane or ethyl acetate will generate an unhealthy environment (Waksmundzka-Hajnos et al., 2001).

Older types of silica gel used to support the reversed-phase suffered lack of stability and undesirable ion exchange mechanisms of retention due to residual silanol groups (Bij et al., 1981) often rendered more acidic by the presence of metal ions. The new generation silica gels are aimed at providing high purity materials with a better stability towards low and high pH values of the mobile phase, less effect from residual silanols and improved efficiencies (Michel et al., 2007).The new

chromatographic columns use higher purity silica (acid washed) and/or a protection of the silanol groups by end-capping (Sunseri et al., 2003, El Rassi, 1996) or very high purity gels produced by polymerisation of tetraethoxy silane.

Further advancement in stationary phase technology has included the intentional introduction of a secondary mode of action. The term “polar embedded groups” was coined for those stationary phases where a polar group was introduced in the linkage between the silica support and the reversed-phase lipophilic moiety. These stationary phases provide better retentivity for polar analytes and a means to prevent phase collapse during the use of mobile phases with high aqueous content (Layne, 2002).

1.2. Hydrophilic Interaction Liquid Chromatography (HILIC):

Modified versions of reversed-phases are not the only indicators of the recent developments in chromatography. New classes of stationary phases have been introduced. The term hydrophilic interaction liquid chromatography and the abbreviation HILIC were proposed by Alpert in 1990 (Alpert, 1990) describing the process where a hydrophilic stationary phase is combined with a hydrophobic mobile phase.The work by Alpert studied hydrophilic compounds including amino acids, carbohydrates and peptides with a comparison between a polysulfoethyl and a polyhydroxyethyl stationary phases being used to distinguish between retention due to electrostatic and retention due to HILIC mechanisms. Peptides being composed of a sequence of amino acids with different charge and hydrophilicity characteristics are good evaluation analytes as they can be used to test different parameters by manipulating their linear status, charge values, hydrophobic content, sites of interaction and any combination of these. Alpert identified HILIC as a variant of normal phase where analytes are retained by partitioning between the mobile phase and a thin layer of water enriched mobile phase immobilised at the surface of the stationary phase.

There is currently a selection of HILIC columns on the market (Schiesel et al., 2010), most of them have a secondary retention mechanism as well as exhibiting

the HILIC phenomenon (Ikegami et al., 2008). HILIC stationary phases can be classified according to the secondary mechanism of retention into:

- Neutral phases.
- Charged phases.
- Zwitterionic phases.

1.2.1. Neutral phases:

In these stationary phases, the secondary retention is caused by H-bonds or dipole-dipole interactions. These include diol, cyano bonded silica and amide bonded silica phases. The neutral stationary phases based on polyethylene glycole (PEG), hydroxyl or oxyethylene-hydroxy exhibit HILIC and RP retention behaviour in low and high aqueous mobile phase content, respectively. Retention of phenolic acids and flavones were studied on these phases and it was concluded that the best performance was obtained on the oxyethylene-hydroxy (Luna-HILIC) phase in terms of retention, selectivity and peak symmetry of the studied compounds (Jandera and Hájek, 2009). Luna-HILIC was again tested, along with other HIILC columns, for the separation of ascorbic acid from its degradation product dehydroscorbic acid with corona aerosol discharge (CAD) detection. The other tested columns were the zwitterionic phases (Obelisc R and Obelisc N). Luna-HILIC and Obelisc N were not successful for the target of the analysis on this occasion. Obelisc R gave satisfactory separation (Nováková et al., 2009). Luna-HILIC was tested for metabolic profiling purposes (profiling of fermentation broths of β-lactam antibiotics production) as well and found to be satisfactory but less efficient than the zwitterionic phase, ZIC-HILIC this time (Schiesel et al., 2010).

Neutral cyano columns were tested in HILIC mode and were found to show poor retention for peptides even under normal phase conditions (Yoshida and Okada, 1999).

Neutral amide stationary phases have been employed for different pharmaceutical applications. For example, TSK gel amide was used in HILIC mode with MS detection for the analysis of paralytic shellfish toxins. The shellfish poison

contains polar tetrahydropurine derivatives of different potencies requiring the separation and quantification of the individual components of the poison. An isocratic elution with 65% acetonitrile containing 2.0mM ammonium formate and 3.6mM formic acid was able to achieve the goal of separating the principal components of the poison. Even on this neutral column, residual silanols contributed significantly to the retention via an electrostatic mechanism. A retention time shift due to a matrix effect was reported (Dell'Aversano et al., 2005). TSK gel amide 80 was proposed as a suitable stationary phase for the analysis of polyols and carbohydrates. An example of this is presented in the work on phloem exudates from petioles of fully expanded *Cucurbita maxima* leaves where sugars, sugar amines and other important metabolites were retained and quantified using a gradient of acetonitrile and ammonium acetate buffer at pH 5.5 (Tolstikov and Fiehn, 2002). One of the two stationary phases on which the term HILIC was coined was a neutral amide column (Alpert, 1990).

1.2.2. Charged phases:

In these stationary phases electrostatic interactions dominate the retention. Charged phases can range from charged ion exchange resins to silica bonded phases (e.g. aminopropyl) or bare silica gel.

The bare silica gel is charged due to its silanol groups which carry a negative charge at a wide range of pH values. It was first employed in normal phase analysis but has been used in the HILIC mode as well. Glucoraphanine is a member of the glucosinolate family of compounds found in cruciferous vegetables. It is reported to have different health benefits including cancer chemopreventative activity. Glucoraphanine was quantified in rat and dog plasma using an LC-MS/MS method with a separation by a silica column (Luna) and a gradient between 200mM ammonium acetate buffer (A) and acetonitrile (B). The method was proposed for supporting clinical, pharmacological and pharmacokinetic studies (Cwik et al., 2010).

Aminopropyl columns are silica bonded phases with an aminopropyl group attached to the surface. They can be operated under NP or HILIC conditions. The use

of aminopropyl columns in analysing polar metabolites has been reported on several occasions (or many occasions if their long use for the analysis of sugars is considered), as in the example of the analysis of 2`-C-methyl-cytidine-triphosphate which is a RNA-dependent polymerase inhibitor with reported activity against hepatitis C virus. The quantification method used a gradient of ammonium acetate buffer (A) and acetonitrile (B) with an MS detection in the negative ion and multiple reaction monitoring mode (Pucci et al., 2009). Another example of the use of aminopropyl columns in the HILIC mode involved another pharmaceutical application for the analysis of a pharmaceutical excipient rather than an active ingredient. Dry inhalation products introduce the active substance as a mixture with the polar sugar lactose; hence, the deposition of lactose provides an indication of the deposition of the active ingredient. A method for analysing lactose with a very low limit of quantification was thus required and was suggested by Beilmann *et al* (2006) using a chromatographic system of an aminopropyl column with isocratic elution of acetonitrile (80%) and water (20%) and evaporative light scattering detection (ELSD) (Beilmann et al., 2006). Aminopropyl contains reactive amine groups and some analytes may bind covalently to the amine causing perturbation of the stationary phase chemistry (Ikegami et al., 2008).

1.2.3. Zwitterionic Phases:

Zwitterionic phases are permanently charged phases with a total charge value of zero, thus they exhibit electrostatic interactions that are weaker than those observed with charged phases bearing a single positive or negative charge. Zwitterionic phases provide a compromise between the two other types of phases minimising most of their flaws, especially the hygroscopicity for silica phases, the need for extreme pH values or high salt concentrations in the mobile phases for charged phases, and the poor retention and selectivity of neutral phases, making zwitterionic phases good candidates for reproducible LC-MS applications (Guo and Gaiki, 2005).

ZIC-HLIC columns (from Sequant-Merck) are zwitterionic columns with a stationary phase chemistry of a sulfobetaine bonded with its quaternary ammonium

moiety closest to the surface of silica gel or polymer beads. This leaves the negatively charged sulphonate on the surface and the positively charged ammonium embedded nearer to the support. ZIC-HILIC has been reported in different applications as an alternative solution for ion-pairing chromatography for polar and basic compounds. As can be seen from Table.1.1 the majority of these papers have used buffers in the mobile phases. These buffers were, in many cases, selected after multiple optimisation experiments which studied the column selectivity of the analytes (or a sub-group of analytes in case of metabolic profiling) on the column following the adjustment of parameters such as pH, ionic strength and organic modifiers. In these reports, the studied parameters were shown to affect the ZIC-HILIC selectivity.

One study was dedicated to probing the effect of mobile phase composition on ZIC-HILIC selectivity using glycerol, glycerol carbonate and urea as test analytes. The study investigated the mechanism of retention of each of the three HILIC column types, bare silica (charged) diol (neutral) and ZIC-HILIC (zwitterionic). The study showed the predominance of HILIC retention on these columns with effects of secondary mechanisms varying according to the analyte and the composition of mobile phase. An interesting conclusion of this study was about the effect of water layer depth on the mechanism of retention, showing that highest retention was achieved with the diol column where the bonded moiety is large leading to a thicker partitioning aqueous layer (Fourdinier et al., 2010).

Skerikova *et al* (2010) studied the effect of operation temperature on retention on a ZIC-HILIC monolith. The study used test analytes composed of 10 phenolic acids under RP and the HILIC modes, concluding that retention of these compounds was independent of temperature in HILIC mode. However, this was compound specific, as caffeic acid and protocatechuic acid were affected by temperature. The RP mode showed strong dependence on temperature for all compounds (Skeríková and Jandera, 2010). Additionally, glycoproteomics studies have shown the ZIC-HILIC is selective for different glycopeptide isomers (Takegawa et al., 2006a, Takegawa et al., 2006b). ZIC-pHILIC columns have the

same stationary phase chemistry of ZIC-HILIC with a polymer based support for an improved pH stability range (Pluskal et al., 2009).

ZIC-HILIC solid phase extraction (SPE) cartridges were used in bioanalytical work as well utilising the lack of retention of phospholipids of these cartridges in order to reduce matrix effects (Lindegardh et al., 2008; Lindegardh *et al*., 2005). These are shown as shaded cells in Table.1.1**.**

Although ZIC-HILIC has gained the most attention in the literature it is not the only commercially available zwitterionic phase. Obelisc columns are other variants of zwitterionic phases. The brand incorporates two models Obelisc N and Obelisc R. Obelisc N carries its positive charge on the surface and the negative charge closer to the silica support. Obelisc R is the reverse in charge orientation. Another fundamental difference between the two stationary phases lies in the linkage between the two charges which is hydrophilic silica in Obelisc N and a hydrophobic hydrocarbon chain in Obelisc R. The exact chemistry of the Obelisc columns is not explained explicitly by the manufacturer (Sielc, 2010a).

These phases are relatively new with a very limited number of applications. Both stationary phases were tested and compared to the neutral Luna-HILIC for separating ascorbic acid from dehydroascorbic acid using a CAD detector. Obelisc R was the only stationary phase which could obtain satisfactory resolution for the two molecules operating with 75 mM ammonium acetate buffer (pH 4.2) : ACN [15:85] (Nováková et al., 2009). The columns were recently tested for metabolic profiling and found less successful than other HILIC columns with regards to retention, peak shape and isomer separation (Schiesel et al., 2010).

publication reference	application	analytes	matrix	mobile phase	elution mode & range	detector
(Idborg et al., 2005a)	metabolic profiling	-	rat urine	A= 5mM ammonium acetate (pH 4) B= acetonitrile and 0.025% formic acid	gradient 10% to 90% A	MS
(Kato et al., 2009)	metabolic profiling	Amino Acids	hydrolized protein/peptides	A=10 mM Acetic Acid, B= ACN	gradient 25-90% A	IDMS
(Schiesel et al., 2010)	metabolic profiling	hydrophilic metabolites	fermentation broth	A=20mM ammonium formate buffer, B= ACN	gradient 0 to 80% A	MS/MS
(Pluskal et al., 2009) †	Metabolic profiling	-	Fission yeast	A=acetonitrile, B= ammonium carbonate (pH=9.3)	Gradient 80% A to 20%A	FTMS
(Lakso et al., 2008)	metabolite assay	methylmalonic acid	plasma	A=100 mmol/L ammonium acetate buffer, pH 4.5, B= ACN (1:4)	isocratic	ESI-MS
(Nezirevic et al., 2007)	metabolites	4-amino-3-hydroxyphenylalanine (4-AHP) & 3-amino-4 - hydroxyphenylalanine (3-AHP)		acetonitrile:0.1 M ammonium acetate buffer, pH 4.5 (82:18, v/v)	isocratic	ECD
(Zheng et al., 2010)	lipidomics	lipid classes	*Leishmania* extracts	A= 20% isopropyl alcohol (IPA) in acetonitrile and B= 20% IPA in 0.02 M ammonium formate		MS
(Bengtsson et al., 2005)	drug & drug metabolite	Morphine & morphine metabolites	plasma	0.05% trifluoroacetic acid (TFA) for desalting and acetonitrile/5 mM ammonium acetate (70:30) for separation	isocratic	MS/MS
(Lindegårdh et al., 2005)*	drug & drug metabolite	Amoxicillin	plasma			UV
(Lindegårdh et al., 2007)	drug & drug metabolite	oseltamivir and oseltamivir carboxylate	plasma, saliva and urine	A=ammonium acetate buffer (pH 3.5; 10mM) , B=acetonitrile	gradient 15 to 40% of A	MS/MS

publication reference	application	analytes	matrix	mobile phase	elution mode & range	detector
(Lindegardh et al., 2008)*	drug & drug metabolite	peramivir	plasma			MS/MS
(Georgakakou et al., 2010)	drug & drug metabolite	perindopril & perindoprilat	plasma	10% 5.0 mM ammonium acetate water solution in a binary mixture of acetonitrile/methanol (60:40, v/v)	isocratic	MS/MS
(Martos et al., 2010)	drug & drug metabolite	aminoglycoside antibiotics	veal muscles	A=0.4% formic acid in water , B=ACN	gradient 85 to 95% of A	MS/MS
(Ndaw et al., 2010)	drug exposure evaluation	alpha-fluoro-beta-alanine	urine	A= 25mM ammonium formate and B= ACN	gradient 5 to 10% A	MS/MS
(Kopp et al., 2008)	carcinogenic compounds	acrylamide metabolites	urine	10% ammonium acetate buffer (20 mM, pH 6.9) and 90% acetonitrile	isocratic	MS/MS
(Heller and Nochetto, 2008)	toxic compounds	melamine, cyanuric acid and complex	animal feed	A=95% ACN in 0.1% formic Acid, B= 50% ACN in 20mM aqueous ammonium formate	gradient 100 to 25% A	MS/MS
(Al-Asmari et al., 2010)	Forensic toxicology	ethyl glucuronide and ethyl sulfate	urine	A= 5 mM ammonium acetate, B= ACN	gradient 10 to 50% A	MS/MS
(Nováková et al., 2008)	stability of drugs	Ascorbic Acid	solutions	ACN and 50 mM ammonium acetate buffer pH 6.8 (78:22 v/v)	isocratic	UV
(Matysová et al., 2009)	pharmaceutical analysis	dimethindene maleate	pharmaceutical topical preparations	acetonitrile and aqueous solution of acetic acid (25mM) and ammonium acetate (2.5mM) (87.5:12.5, v:v)	isocratic	UV

publication reference	application	analytes	matrix	mobile phase	elution mode & range	detector
(Risley and Pack, 2006)	pharmaceutical analysis	drug compounds and their counter ions	matrix free	A=75 mM ammonium acetate(pH=5) , B=ACN	gradient 100 to 0 % A	ELSD
(Takegawa et al., 2006b)	Glycoproteomics	isomeric N-Glycans and glycopeptides	matrix free protein	A=50% Acetonitrile, B=Acetonitrile, C=100 mM ammonium acetate buffer	gradient 36/59/5 to 64/31/5	MS / Flourescence
(Takegawa et al., 2006a)	Glycoproteomics	sialylated N-glycopeptides	matrix free protein	A=50% Acetonitrile, B=Acetonitrile, C=100 mM ammonium acetate buffer	gradient 36/59/5 to 64/31/5	MS / Flourescence
(Boersema et al., 2007)	proteomics	peptides	cellular nuclear lysate	A=ammonium buffer (pHrange 3 to 8), B= ACN	gradient	MS

Table.1.1 The use of ZIC-HILIC columns in literature.
*** Grey highlighted cells indicate the use of ZIC-HILIC SPE cartridges.**
† ZIC-pHILI C

1.3. Applications of HILIC in pharmaceutical analysis:

HILIC finds a role wherever chromatography is required for separating polar analytes. As can be seen from table.1.1 (which is specific for ZIC-HILIC) there is as wide an application profile for HILIC as that of RP. Nevertheless, we can summarise the three most important pharmaceutical applications of HILIC into the sections outlined below.

1.3.1. Drug and drug metabolite analysis:

HILIC analysis is important in the analysis of polar drugs or xenobiotics with reported bioactivity. Examples of this type of application are mentioned above as for the analysis of glucoraphanin by a neutral phase HILIC and the analysis of 2`-C-methyl-cytidine-triphosphate by charged phase HILIC chromatography. Other examples include the analysis of the polar anti-viral drugs. Zanamivir is a neuramidine inhibitor used to treat influenza. It was analyzed by HILIC-MS/MS using a silica column (Atlantis HILIC) with a mobile phase of ammonium acetate buffer (A) and acetonitrile (B) with a gradient elution (Baughman et al., 2007). Acyclovir is another antiviral drug widely used in tablets and in topical preparations for treating Herpes related diseases. It was analyzed in different biological matrices by using a silica column with an isocratic elution with a mobile phase composed of 80% acetonitrile and 20% ammonium formate buffer (Brown et al., 2002).

1.3.2. Quality control of pharmaceutical products:

Examples of the use of HILIC in analytical methods intended to be used in quality control laboratories include the use of ZIC-HILIC to analyse pharmaceutical positive and negative counterions in one single method using ELSD as a universal detector, and a collection of elution programs which were suitable for the analysis of 33 different counterions including inorganic ions (12 cations and 21 anions) (Risley and Pack, 2006). Ascorbic acid and its major oxidation product dehydroscorbic acid were resolved and analyzed by Obelisc R chromatography with CAD detection (Nováková et al., 2009). Quantitative analysis of eight water soluble vitamins was achieved using a diol neutral column (HILIC Inertsil) and a gradient elution with acetonitrile: ammonium acetate buffer (90:10) coupled with diode array detection

(Karatapanis et al., 2009). The experimental anti-cancer drug 5-fluoro-2'-deoxycytidine is a DNA methyltransferase inhibitor which is rapidly inactivated in the body. It is used as an injection in combination with tetrahydrouridine in order to block the metabolism of and enhance the pharmacokinetics of the anti-cancer agent. The stability and compatibility of this combination in injections and aqueous solutions was studied using a ZIC-HILIC column. The stability indicating method used acetonitrile: ammonium acetate buffer (85:15) in isocratic mode with UV detection and concluded that the combination was compatible although the anti cancer agent was unstable under acidic conditions (Guo et al., 2010).

1.3.3. Metabolic profiling:

The work of this thesis is focused on metabolic profiling of biological systems and thus it will be reviewed in greater detail.

1.3.3.1. Metabolism and metabolic profiling:

Metabolic profiling has been studied for a vast array of biological systems including unicellular systems such as bacteria and parasites (Jozefczuk et al., 2010, Szymanski et al., 2009, Breitling et al., 2006b, t'Kindt et al., 2010) , plants(t'Kindt et al., 2008a, t'Kindt et al., 2009, t'Kindt et al., 2008b) , animal models (Plumb et al., 2003, Plumb et al., 2005) , or human studies (Kenny et al., 2010, Kaddurah-Daouk et al., 2007, Steffens et al., 2010). Metabolic profiling in plants has been pursued for longer and more intensively than studies in other systems (Seger and Sturm, 2007, Ryan and Robards, 2006, Lay et al., 2006), but plants have unique metabolites which are only present in higher organisms as xenobiotics. It is easier to work with systems where the biological and environmental variables are easier to control, such as model animals or cell cultures. For this reason, in order to review a number of studies relevant to this thesis in reasonable depth we have drawn a, not altogether artificial, distinction between simple model systems and the metabolically distinct plant systems and the human systems, where controlling the variables involved in sample collection can be difficult.

Metabolomics, metabonomics and metabolic profiling, often used interchangeably, belong in the group of 'omics' technologies. In sequence genomics

identifies genes, transcriptomics indicates which genes are being converted into RNA, proteomics indicates whether or not the RNA is translated into protein and then further downstream what post-translation modifications are made to the protein and finally metabolomics/metabonomics/metabolic profiling indicates whether or not protein expression results in metabolic changes (Lay et al., 2006, Fiehn, 2002, Fiehn, 2001).

Metabonomics has been defined as:

'*quantitative measurement of time-related multiparametric metabolic responses of multicellular systems to pathophysiological stimuli or genetic modification'* (Brindle et al., 2002).

Metabolomics can be defined by quoting Fiehn: 'A comprehensive and quantitative analysis of all metabolites could help researchers understand such (*biological*) systems. Since such an analysis reveals the metabolome of the biological system under study, this approach should be called metabolomics.' (Fiehn, 2001).

Metabolic profiling is defined as the qualitative analysis of the range of metabolites amenable to detection by one or a set of analytical platform(s) in a specific biological system (García-Pérez et al., 2008). The terms are so often used interchangeably and "metabolic profiling" was selected to be used through this thesis.

Since the range of metabolites covered by metabolic profiling is so broad, certain specialist branches of metabolic profiling, such as lipidomics and glycomics, have also evolved (German et al., 2007, Watson, 2006, Willy Morelle, 2006).

1.3.3.2. Analytical techniques used in metabolic profiling:

1.3.3.2.1. Gas Chromatography Mass Spectrometry:

GC-MS is a very versatile technique for compounds which are volatile, or which can be rendered volatile. The applications and future developments of GC-MS in metabolomic studies have recently been reviewed (Pasikanti et al., 2008, Amador-Muñoz and Marriott, 2008, Adahchour et al., 2008). The important characteristics of GC-MS are as follows (Watson, 1999):

- It generally uses electron impact as a mode of ionisation, which promotes efficient ionisation of non-polar compounds (such as hydrocarbons and fatty acids) which are either not ionised or are poorly ionised under the commonly-used electrospray ionisation (ESI) conditions in LC-MS.
- Electron impact (EI) spectra are information-rich and very extensive libraries have been compiled, against which unknowns can be matched.
- GC-MS tends to have very high sensitivity since the solvent background, which is present in LC-MS, is absent; the mobile phase in GC-MS is generally the inert gas helium.
- A capillary gas chromatography column has typically about 10 times the resolving power of a typical HPLC column.

1.3.3.2.1.1. GC Injection Systems and Columns

Split/splitless injectors are most commonly used to load samples into the GC-MS system. Where there is no sensitivity issue, the split mode is used since it is easier to achieve focusing of the sample at the head of the column. Here, a large proportion (e.g. 95%) of the solvent cloud produced upon injection (which can be up to 0.5 ml) is vented while the rest of the sample is transferred to the column. In order to achieve rapid evaporation of the sample, the injection port is generally held at around 250ºC, causing flash evaporation of the sample. The starting column temperature is usually held at a low value e.g. 50ºC, to ensure the re-condensation and focusing of the more volatile components in the sample at the head of the column. The sample in the injector port is introduced into a glass or quartz liner; and since not all the sample is volatile, deposits may build up within liner, so it has to be cleaned or replaced regularly to ensure optimal chromatographic performance (Fiehn and Kind, 2006). Since the mobile phase in GC is inert, it is only the interactions of the analyte with the stationary phase that provide the selectivity. Compounds elute from the GC column in order of volatility with elution being assisted by a gradually increasing temperature.

GCxGC in combination with time of flight ion separation is becoming increasingly popular in metabolic profiling studies. In GCxGC two columns are used in series and

are connected via a thermo-modulator that condenses the eluent from column 1, moves upstream and then re-evaporates it so that it is transferred to column 2. The process may be carried out within the same GC oven or may use two separate GC ovens. Typically the process uses a relatively long first dimension column e.g. 30 m x 0.25 mm i.d. connected to a short second dimension column e.g. 1.5 m x 0.1 mm i.d. The modulation period is usually 3-4 seconds, corresponding to the time taken for the fractions condensed from the first column to pass through the second column. Usually the first column is a non-polar methylsilicone DB-1 type phase and the second short column is more polar. The technique produces very narrow peaks of 0.1 -0.6 seconds thus boosting sensitivity. The combination of GC with TOF separation is ideal because of the rapid data acquisition provided by this technique. The technique has been recently comprehensively reviewed (Amador-Muñoz and Marriott, 2008, Pasikanti et al., 2008, Adahchour et al., 2008)

1.3.3.2.1.2. Derivatisation

Some compounds are naturally volatile, but many molecules require derivatisation before analysis by GC-MS. The derivatisation process can itself contribute to the chromatographic peaks observed, since the derivatising agent is usually in great excess compared to the analytes. Fatty acids and Krebs cycle acids are often converted to their methyl esters by treatment with BF_3-methanol or 3% v/v acetyl chloride in methanol; this is a relatively 'clean' derivatisation process. A more general derivatisation method is trimethylsilylation which derivatises acids, alcohols and to some extent amines to form trimethylsilyl (TMS) ethers and esters. Trimethyl silylation, combined with methoximation (which produces stable derivatives with aldehydes and ketones), offers a useful general method suitable for the analysis of amino acids, sugars, polyols, Krebs cycle acids and fatty acids (Michaud et al., 2008). There are some disadvantages to TMS derivatisation: the reagent can contaminate the mass spectrometer, this is more of a problem with quadropole instruments than TOF instruments, and it has to be cleaned quite frequently; the reagents themselves produce peaks; and some groups such as amines, amides and indoles form unstable or multiple derivatives which complicate the analysis. There is a number of other derivatisation strategies to choose from (Watson, 1994, Halket et

al., 2005) and different approaches might be used to maximise information in a metabolic profiling experiment.

1.3.3.2.1.3. Ionisation Modes in GC-MS

EI is the technique most commonly used with gas chromatography. However, some complex molecules fragment extensively and if quantification of very low levels of a compound is required it may be an advantage to use soft-ionisation techniques such as positive or negative chemical ionisation. A very high level of sensitivity can be achieved with negative ion chemical ionisation (Stern et al., 1995) .

1.3.3.2.1.4. Ion Separation Methods

Single quadrupole GC-MS instruments are excellent for metabolite profiling of high abundance compounds such as sugars and fatty acids, and are also very effective for quantification of target metabolites when used in selected ion monitoring mode. Ion trap instruments offer the advantage that fragmentation can be selectively carried out on low level peaks to aid in identification. Ion traps work by trapping ions within an electrostatic field, so a particular ion of interest can be retained for fragmentation while ions of no interest are selectively ejected. Combination of GC with the high resolution ion separation (Mohler et al., 2008) afforded by time of flight separation potentially has the greatest power for metabolic profiling studies where minor peaks could be obtained with accurate mass measurement. It is thus possible to deconvolute minor peaks and their fragment ions from major overlapping peaks. Soft ionisation techniques in combination with GCTOF would also be very powerful and could produce simpler data than EI.

1.3.3.2.2. Liquid Chromatography Mass Spectrometry:

1.3.3.2.2.1. Ionization Methods

The most common ionization method used in conjunction with liquid chromatography is electrospray ionization (ESI) (Fenn et al., 1989). ESI can be operated in either positive or negative ion mode. While ESI is not a universal ionization technique, it gives coverage across a wide range of molecules. ESI is most sensitive to bases which give very strong signals in positive ion mode because they

are readily protonated. Compounds containing alcohol groups alone do not protonate particularly efficiently in positive ion ESI; thus for instance glucose and other sugars do not generally give strong signals in positive ESI mode, although complexation with metals can improve this (Qian et al., 2008). Sugars are better observed in negative ion mode where they form $[M-H]^-$ ions and also adducts, for instance with chloride and formic acid. Acids, apart from those that contain other protonatable groups such as amine groups, generally show up best in negative ion mode. Positive ion spectra have much better signal-to-noise performance than negative ion spectra (Keller et al., 2008).

In addition to molecular ions, many compounds form adducts with components present in the mobile phase. In positive ion mode, adducts are commonly formed with acetonitrile, methanol, ammonia, sodium, potassium and calcium. In the negative ion mode they form with formic acid, acetic acid and chloride. While adducts are usually quite minor species, a minor adduct formed from an abundant compound can give a substantial peak relative to minor components in a mixture (Keller et al., 2008). Adducts both increase the richness of the data which can be obtained from mass spectral profiles and add to the difficulty of interpreting data. In addition to adducts there are additional contributions particularly from carbon-13, sulphur-34 and nitrogen-15 isotope peaks. Thus care has to be taken in unambiguous assignment of a compound to a particular mass peak. Adduct clean-up can now be carried out within the instrument using high field asymmetric waveform ion mobility spectrometer (FAIMS) (Metz et al., 2008, Abu B. Kanu et al., 2008). FAIMS is an ion focusing technology where ions generated from the ESI (or other ion sources) will migrate in the asymmetric field according to their charge, size and molecular shape (Guevremont and Purves, 1999). Since its introduction in 1999, FAIMS has been developed for an optimal utilisation in background minimisation or isobaric resolution. A risk of running samples with a high background of abundant components in a matrix is that the abundant components will suppress the ionisation of more minor components in the mixture (Müller et al., 2002). The worst culprits for producing these types of effects are phospholipids, which are naturally present in most biological samples, and plasticizers (such as PEGs) which are common

environmental contaminants. Ion suppression in ESI can also be caused by electrolytes within the chromatographic mobile phase (Beaudry and Vachon, 2006).

1.3.3.2.2.2. Ion Separation Methods

1.3.3.2.2.2.1. Traps and Quadrupoles

The most basic LC-MS systems are based on single quadrupoles and the cost of these instruments has fallen greatly in recent years. Even a single quadrupole instrument is capable of delivering complex information, particularly if combined with good chromatography. The main drawbacks are their limited ability to produce fragmentation information on compounds; and in addition they are limited to a typical resolution of about 0.5 amu. Thus a standard method of expanding the capability of a single quadrupole instrument is to use a triple quadrupole or tandem MS instruments. Tandem MS systems are the standard configuration for determination of drugs and their metabolites in pharmacokinetic studies, and they deliver the most sensitive analyses in terms of detection limits. They are also widely used in proteomics studies where the output is readily integrated with database searching. Ion trap instruments do not deliver the same level of sensitivity as triple quadrupoles but are complementary in that they can carry out multiple fragmentation experiments in order to provide more detailed structural information on a compound (de Hoffmann and Stroobant, 2001, Hocart et al., 2010).

1.3.3.2.2.2.2. Time of Flight (TOF) Instruments

The TOF analyzer has been available since the 1970s, with many refinements to the technology over time. Initially TOF was mainly linked to pulsed ionization techniques, such as matrix assisted laser desorption ionization (MALDI). This technique provided a very sensitive method for determining proteins and protein digests. MALDITOFMS has good potential in metabolic profiling but has not yet been widely applied. Analysis in TOF instruments initially suffered from poor resolution due to variations in the kinetic energy of ions of the same mass leaving the source. The introduction of a device called a reflectron which enabled better focusing of ions and double reflectron instruments have increased mass resolution even further. Hybrid quadrupole TOF (QTOF) instruments have been developed over the

last 15 years, in which TOF separation can be coupled to a continuous flow of ions such as that generated from chromatographic systems interfaced to an ESI source. Thus QTOFMS is one of the methods of choice in metabolic profiling since it can deliver accurate mass measurement data. The technology has been gradually improved with regard to the ability of instruments to deliver a wide dynamic range. Linear dynamic ranges of up to 10^6 are claimed as well as a mass accuracy measurement of 3 ppm. There is a physical limitation to the resolving power of the instrument which is that greater resolution depends on a longer flight tube (Wren and Tchelitcheff, 2006, Chalmers and Gaskell, 2000). Table 1.2 summarises the specifications for QTOF instruments.

1.3.3.2.2.2.3. Fourier Transform Ion Cyclotron Resonance (FT-ICR)

FT-ICR offers the highest resolving power and accuracy of all MS instruments to date (McLuckey and Wells, 2001) . The high resolution is indispensible where overlapping ions can occur, such as when observing the content of a complex sample in a relatively narrow mass window, of which metabolic profiling is a conspicuous example (Breitling et al., 2006a). Although FT-ICR offers superb sensitivity, the ion-ion interaction in FT-ICR forms a serious challenge as it decreases the dynamic range of measurement (McLuckey and Wells, 2001). FT-ICR is based on measuring the frequency of oscillating ions. As frequency certainty can only be achieved by repetitive measurement the ultra high performance of FT-ICR is time dependent (McLuckey and Wells, 2001). That is why FT-ICR is tagged as a slow mass spectrometry, rendering hyphenation to a continuous high flow source of ions (like chromatography) difficult. Table 1.2 summarises the specifications for FT-ICR.

1.3.3.2.2.2.4. The Orbitrap Mass Spectrometer

In 2005 ThermoFinnegan introduced the Orbitrap, originally invented by Makarov (Makarov, 1999). The ion separation technology is based on trapping the ions injected into the trap between an outer barrel-like electrode and an inner spindle-like electrode. The ions exhibit angular, radial and axial oscillations at frequencies all of which are mass-dependent (Qizhi Hu et al., 2005). The mass/charge is measured by the image current generated by the axial oscillation which is completely independent

of energy and of the spatial spread of the ions (Makarov, 2000).This facilitates a fast accurate measurement of frequencies, which is well matched with the introduction of a continuous high flow ion flux such as the effluent of a chromatographic column (Hardman and Makarov, 2003). The limit of detection of the instrument is excellent (picomoles/ml) due to the ability to measure small changes in the propagated current image (Makarov et al., 2006). Other advantages of the Orbitrap include very high resolution, due to its ability to measure the image current with high accuracy,a high mass range, and a better dynamic range compared to Paul traps and FT-ICR (Qizhi Hu et al., 2005). The Orbitrap suffers to some extent from the limitations of the 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 of *ca* 10^5 can be achieved. Like the FT-ICR instrument, time is required to obtain high mass resolution, so scanning rates have to be reduced in order to get good resolution; but in practice a resolution of 30000 can be used with routine chromatographic techniques. However, very narrow chromatographic peaks such as those achieved by ultra high pressure liquid chromatography (UPLC) may require a reduction in resolution to permit frequent enough sampling of a peak across its width (Makarov and Scigelova, 2010).

Although it has been on the market only for a short time, the power of the Orbitrap has been extended with the addition of new technologies. Williams *et al* (2007) coupled the Orbitrap to a dual ion source which generated ions with both polarities (Williams et al., 2007). McAlister *et al* used the ion-ion interaction termed Electron Transfer Dissociation (ETD) to characterize peptides and proteins with the Orbitrap. The technique uses a chemical ionization source to cause

ionization/dissociation with a selectivity which is different from the widely used collision induced dissociation (CID). The fragments can be transferred into the Orbitrap to generate measurements with its characteristic high accuracies (McAlister et al., 2007). This all was introduced with minimal instrumental modification. Olsen *et al* (2005) described the introduction of background ions via direct injection into the C shaped, RF only storage trap, prior to injecting parent ions originating from ion source and/or fragments induced in the linear trap. This background ion serves as an internal mass calibrant to enhance the measurement accuracy. Sub ppm accuracy of parent ions, as well as fragments, was routinely achieved by using this technique (Olsen et al., 2005). The same group added later (2007) the HCD (High energy C-trap dissociation) approach, where masses are fragmented by immonium ions introduced into the C trap with the loss of the lower mass range, or by an additional dedicated octopole collision cell at the far end of the C-trap (Olsen et al., 2007). Combining both efforts, *de novo* peptide sequencing was achieved with very high accuracy mass measurement. All of these refinements are potentially beneficial to metabolic profiling.

Feature	FT-ICR	TOF	Orbitrap
mass resolving power	10^4-10^5	10^3-10^4	10^4-10^5
mass accuracy	1-5 ppm	5-50 ppm	1-5 ppm
mass range	$>10^4$	$>10^5$	$>10^4$
linear dynamic range	10^2-10^5	10^2-10^6	10^2-10^5
precision	0.3-5%	0.1-1%	
abundance sensitivity	10^2-10^5	up to10^6	
efficiency (transmission × duty cycle)	<1-95%	1-100%	
speed	0.001-10 Hz	10^1-10^4 Hz	10-100 Hz
compatibility with ionizer	Pulsed and continuous	pulsed and continuous	pulsed and continuous
cost	moderate to high	moderate to high	moderate to high
size/weight/utility requirements	lab instrument	benchtop	lab instrument/Benchtop
Hyphenability to Chromatography	Limited	YES	YES

Table.1.2 Characteristics of different high resolution mass analyzers.

1.3.3.2.2.3. Chromatographic Considerations

Reversed-phase chromatography is the most commonly used technique for analysis of metabolic profiling samples (Olsen et al., 2007). It is very useful for lipophilic compounds which elute in order of their lipophilicity. However, there is a disadvantage to reversed-phase chromatography with biological samples, since lipophilic components such as phospholipids and environmental contaminants can accumulate and elute in subsequent runs causing interference (Olsen et al., 2007) and ion suppression, unless a washout programme with high levels of organic solvent is included at the end of the run programme (McLuckey and Wells, 2001). In metabolic profiling, the resolving power of reversed-phase chromatography can be increased by using Ultra Performance Liquid Chromatography (UPLC) (Gika et al., 2008b, Gika et al., 2008d, Guy et al., 2008, Kim et al., 2010, Michopoulos et al., 2009). UPLC has the advantage that higher chromatographic efficiency can be produced: a typical 1.7μm column delivers about 2.5 times the efficiency of a 5 μm column and, because of the flat van Deemter plot obtained with low particle size, very high flow rates can be used without compromising the efficiency of the separation. While fast

separations can be obtained with small particle size columns, due to limitations of pressure, the highest number of theoretical plates (peak capacity) can be obtained by connecting several 5µm columns in series (Desmet, 2008). The disadvantage of this approach is that run times are relatively long. Many biomolecules are not well retained in reversed-phase chromatography, and thus polar amino acids such as glycine and alanine will elute in the void volume of most chromatographic columns - as will sugars. 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 inorganic salts present in the biological matrix also elute in the void (Rimmer et al., 2002). Hydrophilic interaction chromatography (HILIC) provides an alternative to reversed-phase chromatography (Pavel, 2008).

Furthermore, it seems like it is necessary to apply different separation techniques on the same sample, HPLC columns with various packings, or GCMS, LCMS, and CEC-MS (Ondarza et al., 2006a).

1.3.3.2.3. Nuclear Magnetic Resonance (NMR) :

Most of the initial clinical metabolic profiling work was carried out by using nuclear magnetic resonance (NMR) spectroscopy (Wilson and Nicholson, 1988, Nicholson et al., 1984, Nicholson et al., 1985, Bales et al., 1986, Anthony et al., 1994). NMR continues to advance through increases in the power of magnets and of the introduction of capillary and cryogenic probes in order to enhance both the sensitivity and resolution (Dunn et al., 2005, Wishart, 2008). A huge number of NMR based studies has been introduced into literature (Lenz et al., 2003, Griffin, 2003, Brindle et al., 2002, Brindle et al., 2003, Beckwith-Hall et al., 2002). It is considered a non destructive , reproducible method of analysis , where the global spectrum of chemical shifts of proton , carbon , nitrogen and phosphorus are observed (Robinson and Frame, 2005, Keun et al., 2002b).

As biological samples contain hundreds of compounds , each of which carries multiple resonating atoms , it is not unexpected to get complex spectra , these spectra are divided into small buckets of chemical shift regions , where these buckets are studied by chemometrics methods for inter-sample variability (Lindon et al., 2006).

Structure elucidation of a varied region in metabonomics NMR analysis is performed now with reference tables, which list the chemical shift regions for most abundant metabolites in the analysed biological matrix (Holmes et al., 1997) as in monitoring the sub-acute toxicity of aristolochic acid in rats (Zhang et al., 2006).

An NMR spectrum of a biological sample will be quite different depending on the sample matrix , with urine samples mainly containing small molecules giving rise to thousands of sharp peaks on top of each other , while plasma , semen and most of the other biological fluids are macro molecule rich and these will give a broad band with sharp peaks on top of it originating from small molecules (Lindon et al., 2007b).

Acquiring multiple types of spectra for the same samples enhances the structure elucidation and assignment of metabolites in metabolic profiling (Lindon et al., 2006) and in addition to the typical ^{1}H-NMR , the following techniques have been used in metabolic profiling:

Carr–Purcell–Meiboom–Gill (*CPMG*): where peaks of the macromolecules are attenuated , giving a fingerprint spectrum of only the most mobile small molecules in the sample (Claridge, 1999).

COSY and TOCSY: where a cancellation of the spin-spin coupling takes place in a progressive mode all over the spectrum. This leads to information about protons (or ^{13}C atoms) which are close to each other in space (Claridge, 1999).

One of the major drawbacks of NMR is the lack of sensitivity; however, improvements recently introduced to NMR systems could partially alleviate this problem. Higher magnetic power is one possibility and the cryogenic probe is another. In cryogenic probe, the detector coil and the preamplifier but not the samples are cooled to around 20K (around - 253°C). This reduces the thermal noise in the electronics , leading to an improvement of signal-to-noise ratio of about 5 times ,making shorter run times (1/25), and the detection of the low naturally abundant ^{13}C (~ 1.112%) with enhanced sensitivity feasible. However , even with this improvement in sensitivity, mass spectrometry retains a superioriority over NMR with regards of the number of detected metabolites (Keun et al., 2002a).

NMR is usually applied to samples in solutions, however, a newly introduced technique i.e. magic angle spinning (MAS-NMR) has enabled the analysis of solid samples. The technique is based on spinning the sample at a high frequency when it is positioned at an angle where $\cos^2(\Theta)=1/3$ with respect to the direction of the magnetic field. This leads to narrow lines and better resolution(Barton et al., 1999, Garrod et al., 1999).

1.3.3.2.4. Additional analytical platforms:

Other platforms used in metabolic profiling studies include Capillary electrophoresis coupled to MS, FTIR, and electrochemical array detection (EAD). FTIR suffers from a lack of sensitivity and EAD requires a library to compare the combination of retention time and redox properties of compounds in the samples analyzed. However , combining these methods with an MS data of the same sample, either separately (FTIR) or by directing some of the chromatographic effluent to an EAD would give an added value for metabolite identification (Lindon et al., 2006).

1.3.3.3. Sample preparation for metabolic profiling:

The sample preparation for metabolic profiling can be simple and straightforward for some matrices, such as urine where only dilution and centrifugation are needed (Want et al., 2010). However, sample pre-processing is necessary in some cases for samples where, for example, protein precipitation is required. Several protocols have been suggested for preparing plasma samples for metabolic profiling, sometimes with different conclusions. Alzweiri *et al* (2008) tested different solvents and procedures for the removal of salts, proteins and lipids from plasma samples reporting acetonitrile crash plates as the most suitable (Alzweiri et al., 2008). Want *et al* (2005) tested different solvent systems, solvent percentages and temperatures for protein precipitation in plasma and found the ratio of 3 parts cold MeOH to 1 part plasma to be the optimal method (Want et al., 2005). A similar, but more comprehensive, approach was introduced by Bruce *et al* (2009) for improved protein precipitation efficiency. In this report, the precipitated fractions were quantified by gel electrophoresis and two optimal solvent systems (methanol:ethanol 1:1) or (methanol:acetonitrile:ethanol 1:1:1) were proposed with a ratio of 4 parts solvent to 1 part plasma (Bruce et al., 2009). Reversed-phase SPE was compared to the

traditional organic solvent precipitation for removing proteins and it was found that SPE samples gave more reproducible results and needed less column conditioning time compared to precipitation methods. However, the number of features observed following SPE were generally fewer (Michopoulos et al., 2009). Turbulent flow chromatography has been very recently proposed for removing macro molecules from the plasma samples with no prior sample preparation. The method was able to deproteinate the samples introduced into the analytical columns, at the expense of losing some of the features which were shown to belong to the phospholipids class (Michopoulos et al., 2010).

Metabolic profiling of cells from cell cultures requires sample preparation steps which will quench metabolism quickly and keep the metabolites in the extraction solvent. Different cells have different characteristics and will require specific extraction methods. Examples of the optimisation of this process can be found in an investigation of different quenching and extraction techniques for yeast cells. The metabolic profile of the cells after extraction was investigated by LC-MS/MS and GC-MS. A standard method for rapid quenching of yeast cells while preserving the intracellular metabolome is to mix 1 ml of culture broth with 5 ml of methanol/water (40:60) at -40ºC. The cells are then pelleted by centrifugation and the pellet extracted with boiling ethanol. Different methods of extraction were tested using the ^{13}C-labelled metabolome as an internal standard. Prolonged contact of the cells with the quenching solution was found to cause slow leakage of metabolites from the cells. The effect of including buffers with the quenching solution was tested but they generally promoted leakage of representative metabolites. Better recovery of metabolites was obtained upon extraction with methanol/water (8:2) at -40ºC. The final conclusion was in order to avoid leakage, cultures should be harvested into pure methanol at -40ºC or below with a sample to quenching ratio of 1:5 or lower (Canelas et al., 2008).

Three extraction methods were investigated for their suitability for quenching and extracting metabolites from *E.coli* cultures. The extracts were converted to TMS-MO derivatives and then analysed by GC-TOFMS. The largest number of peaks was observed when the cells were quenched with either 60% aqueous methanol or 60%

aqueous methanol/tricine solution at -48ºC. Using hot ethanol as an extraction medium decreased the number of peaks observed. Extraction with KOH was better for recovery of organic acids while methanol/chloroform extraction was most suitable for lipids. Although no extraction method gave perfect recovery when compared against the other methods, 60% aqueous methanol at -48ºC appeared to give the most comprehensive coverage (Jackson et al., 2008)

In some cases a particular extraction method ,maybe be useful when looking for a specific range of the metabolome , such as in lipidomics studies (Kaddurah-Daouk et al., 2007), where the sample is liquid-liquid extracted with chloroform/ethanol mixtures. Lipid metabolism is a potential target for both resistance and treatment in the case of *Trypanosoma* (Sparagano, 1998) and is a target in the ketoconazole treatment of *Leishmania* (Mishra et al., 2007). Another example of selective extraction was used in the study of cis-diol derivatives, which are specifically extracted on a phenylboronic acid affinity column. This class of compounds is related to nucleotides and is modified during cancer pathology (Yang et al., 2005).

Combining data obtained from different sample preparation methods maybe of interest for obtaining a more comprehensive metabolic profile. A reversed-phase SPE method was used to fractionate urine samples into polar (non-retained) and non-polar (breakthrough) with the former introduced to a ZIC-HILIC and the later to a conventional reversed-phase column (Idborg et al., 2005b).

1.3.3.4. Statistical methods:

The complex data obtained from metabolic profiling makes the use of advanced statistical approach, such as multivariate analysis and pattern recognition, a key factor for the success of a metabolic profiling study (Robertson, 2005). The most commonly used multivariate analyses in metabolic profiling are described below.

1.3.3.4.1. Principal Components Analysis (PCA):

Of the different statistical approaches used for multivariate analysis, principal components analysis (PCA) seems to be the method of choice to observe the metabolic patterns of different groups (Robertson, 2005). PCA invovles arraying the data in a matrix, where rows represent different samples (observations) and columns

represent signals (variables.) The first principal component is the eigenvalue (general sense value) for the covariance matrix of the original matrix (Smith, 2002, Trygg et al., 2007, Umetrics, 2009). PCA can be assembled (figure.1.1) as the projection of the first column of the matrix (signal intensities of all observations for one variable) on a first axis, then the second column on a second axis, and so on for all variables. This will generate a hyperspace with a number of axes equal to the variables. The first PC is the axis which will maximize the variance among all variables, which is the one with the least square for all variables. The projection of the matrix rows (observation) on this axis will produce an individual value (score) for each sample (observation) expressing its contribution to the variability contained in this first PC, and its pattern of variation in the hyperspace determined by the whole matrix (Smith, 2002). Each signal (variable) will contribute to the principal component by a value which ranges from the absolute score value of its observation when the PC axis is parallel to the variable orientation (full variability), to the null value where the variable orientation is exactly orthogonal to the PC axis (Ewing, 1997, Robertson, 2005) Thus we can get the contribution of the original data variables (the loadings) to each PC by the equation :

$$\text{Loading value} = \text{Score value} \times \cos\Phi$$

where Φ is the angle between the variable orientation and the PC axis in the PCA plot.

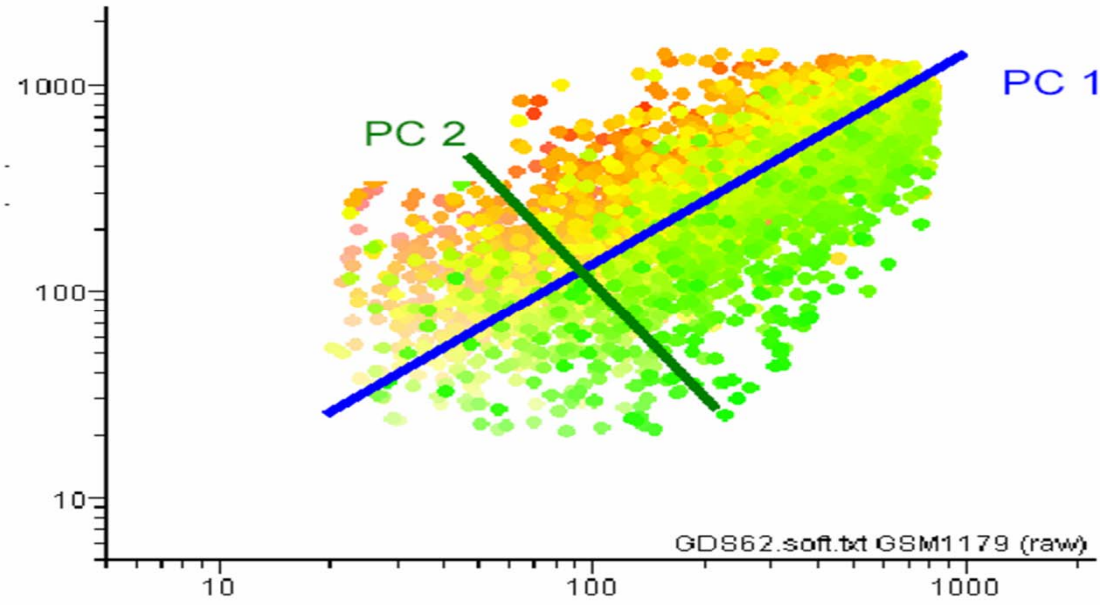


Figure.1.1 A representation of Principal Components Analysis.

The second PC is the axis which will maximize the variance among all variables after eliminating any variation expressed by the first PC, that is why it is represented in the hyperspace as an axis with an orthogonal direction to the first PC (no common variability will be included in both PCs). Similar to the first PC, a list of scores on the second PC is obtained, and combining these two scores, we can get a plane determined by the two axes with samples similar in variation trends clustered in the same region apart from other samples (Umetrics, 2009, Robertson, 2005).

Loadings of these two PCs happen in regions harmonised with the localisation of the observation in the scores plot, thus, if one group of observations is situated on the positive side of the scores plot, the positive side of the loadings plot will contain loadings (variables) highly expressed in this particular group compared to other groups (Brereton, 2002).

1.3.3.4.2. Partial Least Square (PLS):

The same PLS abbreviation may stand as well for the projection of latent structures, which is the main goal of PLS analysis. This statistical approach is used for relating two groups of data or two matrices by a linear multivariate method. Data are arrayed in the same way as in PCA, with values divided into two matrices, or in

some cases, one matrix with all data points, later divided into two matrices (Umetrics, 2009, Robertson, 2005).

For a set of two matrices X and Y, the first PLS component is a line in the X-space, and another line in the Y-space. These must:

- Approximate the data point projection of the variables.
- Provide a good correlation between the position of points along the lines in X and Y spaces.

In other words , the first PLS is the line which will maximize the variation within each matrix (X and Y) which is the PCA score, while at the same time maximizing the correlation between X and Y (Robertson, 2005).

1.3.3.4.3. Orthogonal PLS (OPLS) and OPLS Discrimination Analysis O-PLS-DA:

A further development in chemometrics is the O-PLS (Orthogonal Projection on Latent Structure) which is a factor analysis model where variation in matrix X (e.g. NMR spectra) and the matrix Y (the descriptive variable) is separated into three parts:

1. Variation common in X and Y.
2. specific variation for X (structured noise)
3. and the residual variation

It is an upgraded PLS method with improved interpretation because structured noise is modelled separately from the variations in X and Y which leads to a more straightforward and accurate model elements (loadings and regression coefficient) (Trygg et al., 2007).

1.3.3.4.4. Statistical Total Correlation Spectroscopy (STOCSY):

The STOCSY data analysis approach is based on correlating peaks obtained from NMR spectra with each other, which will generate a pseudo two-dimensional NMR spectrum for the correlation coefficients at each point with other points. When combined with statistical discrimination algorithms like O-PLS-DA, STOCSY was shown to be a powerful tool to infer the structure of the discriminating molecules in the spectrum.

Moreover, this approach can extract not only intra-molecular but intermolecular correlations as well, thus, molecules in the same pathway may correlate positively or negatively, and this can be inferred by this statistical model.

This approach was successfully applied, in combination with O-PLS-DA, to identify molecules responsible for the metabolic variation between urine samples of three different mice strains (C57BL/6Oxjr , BALB/cOxjr , and 129S6/SvEvOxjr) where different metabolites were shown to be highly variable (isovalerate , glutarate , dimethylglycerine and glycerate) across sample groups. These results were in agreement with previously applied more time consuming conventional methods (Cloarec et al., 2005b). This method has been applied to NMR data, but it is applicable to any type of data, such as MS, for inter-molecular correlations.

1.3.3.4.5. Statistical Heterospectroscopy (SHY) :

SHY is another highly promising approach which extracts the correlation between spectra from different analytical techniques. The importance of this methodology lies in getting the most out of and eliminating the drawbacks of each technique by integrating two types of spectra. This approach was applied to correlation of spectra from NMR and UPLC-MS analysis for the same urine samples of control and hydrazine-treated rats, and managed to correlate highly variable chemical shifts in the NMR data with corresponding masses in the mass spectrum (Crockford et al., 2006). This overcame the problems of peak overlapping in NMR analysis and the difficulty in obtaining molecular structure of unknowns rapidly and the variability in ionisation efficiency between structurally diverse analytes in MS analysis (Crockford et al., 2006). The method was able for the first time to assign spermine as a new biomarker for hepatic toxicity.

1.3.3.5. Other Visualization methods:

Among new trends in Metabonomics data interpretation , some attempts have been aimed at drawing a "metabonomics network" ,as in the work on the *Trypanosoma* metabolic network where mass to mass differences corresponding to different species of phospholipid molecules (Breitling et al., 2006b) or to the general metabolic network (Jourdan et al., 2010) are linked according to pre-defined sets of metabolic transformations, and the networks constructed from different samples are studied topographically Figure.1.2.

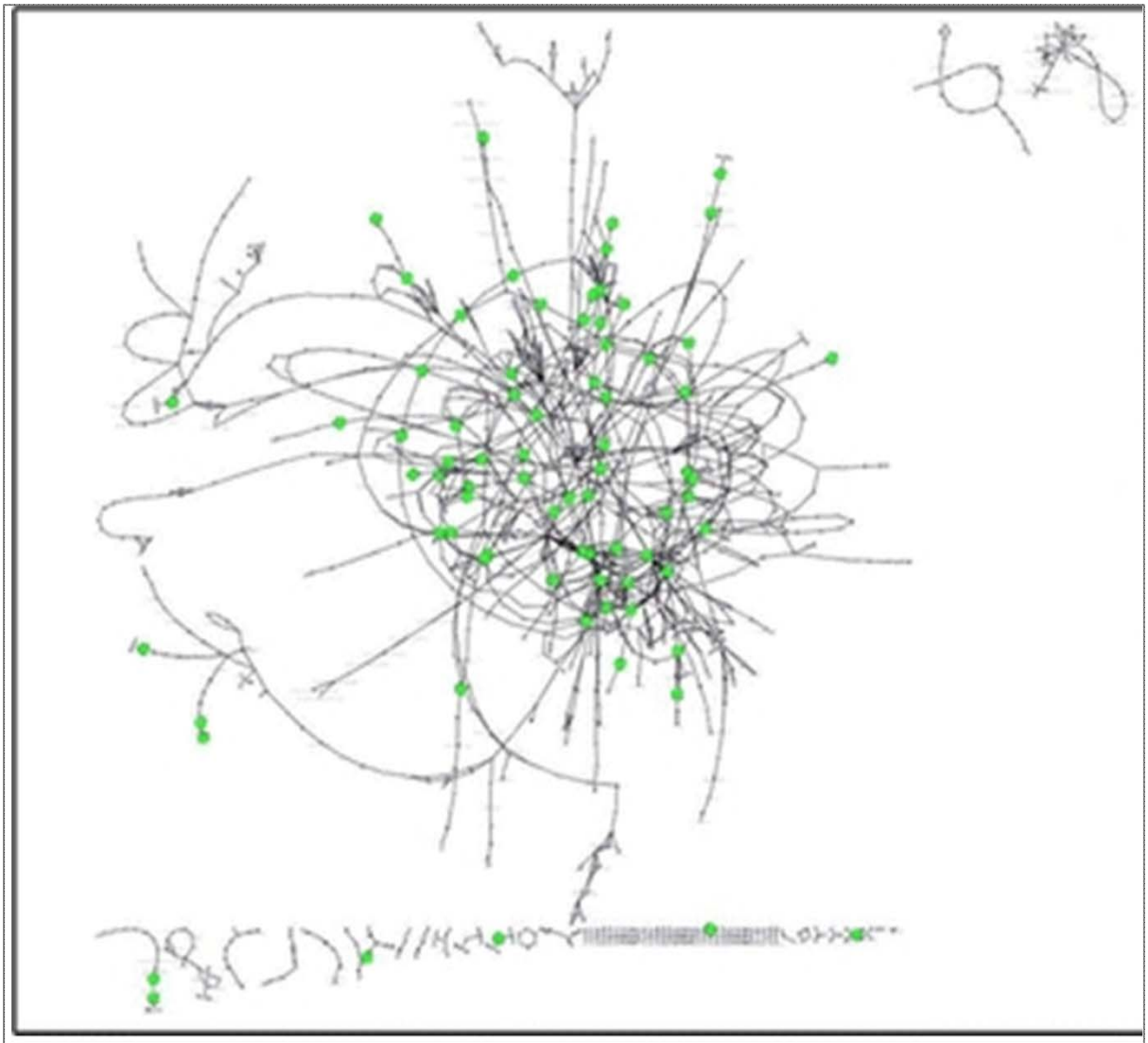


Figure.1.2 Metabolic network of *Trypanosoma brucei* with identified metabolites highlighted in green from (Jourdan et al, 2010)

1.3.3.6. Pharmaceutical Applications of Metabolic profiling:

Since the emergence of the discipline, metabolic profiling has gained an increasing importance, and applications are expanding from one day to another. The main applications are described in the sections below.

1.3.3.6.1. Phenotyping and physiological effects:

A normal sample from a biological system is metabolically different from a sample of a biological system which has been subjected to a toxic challenge. Metabolic profiling plays an indispensable role in both normal and toxicological biological profiling (Bollard et al., 2005). This phenotyping can be extended to detect many differences between groups such as male/female, estrus cycle in females, dietary and lifestyle status, age related differences, diurnal effects and interspecies differences and similarities (Plumb et al., 2003, Kim et al., 2010). Metabolic profiling can be applied to monitoring the homology of model animals whether normal or transgenic and their appropriateness for use in pharmaceutical trials, or even differentiation of unintended consequences of the process of the transfection from the intended results (German et al., 2003).

A significant merit of a metabolic profiling study is its ability to discriminate metabolic variations of multi-genome organisms, which cannot be achieved by genomics or transcriptomics. Mice of the same strain which had different gastrointestinal flora composition were shown to be metabolically dissimilar, and different as well from axenic animals (Robosky et al., 2005).

1.3.3.6.2. Preclinical drug candidate safety assessment:

Traditional screening methods, although reliable, have certain disadvantages. For example, histopathological examination is relatively slow and expensive, while clinical chemistry is highly selective for particular classes of analytes (Lindon et al., 2006).

The COMET “the consortium for metabonomics toxicology” is a consortium formed between six pharmaceutical industry companies and the Imperial College London, UK, aiming at generating and identifying a toxicity spectral database of urine and blood serum samples from rats and mice for pre-clinical toxicological

screening of candidate drugs. This ended up with developing the SMART (scaling multivariate trajectories) of these spectra (Ondarza et al., 2006a).

Additionally , a new classification method for identifying the toxicity class of a compound on basis of all available toxicity spectra has been developed and termed CLOUDS (classification of unknowns by density superposition), here, chemometrics methods are being used to model the time related and dose specific changes in the NMR spectra. If a sample from a specific treatment resides in the normal space, this may indicate a safety potential of a drug candidate and if the sample nests in the abnormal hyperspace, the technique can provide information about the organ or local region of toxicity e.g. liver or kidney (Ebbels et al., 2003). The project was continued with a second phase (COMET-II) aiming at elucidating the mechanisms of toxicity for toxins (Coen et al., 2007a, Coen et al., 2009).

1.3.3.6.3. Understanding disease pathology:

Transgenic mice expressing hepatitis B antigen were studied in order to better understand how these antigens can cause liver damage. Livers from transgenic mice and control mice were extracted and analysed by LC-MS. PCA and database searching revealed 12 biomarkers distinguishing the livers of the transgenic mice from the control livers. The metabolites observed related to effects on lipid metabolism and to increased oxidative stress (Yang et al., 2008).

A metabolomic approach was taken in the study of a mouse model of hepatic steatosis. One group of mice was fasted and the control group was allowed free access to food. LC-MS was used to determine variations in lipid composition within the liver between the two groups. After 24 h of fasting it was observed that two triglycerides (TGs) 44:2 and 48:3 were greatly elevated in the livers of the starved mice, in addition to an unusual TG 49:4 which appeared in the livers of starved mice and may be useful as a biomarker for the treatment of metabolic syndrome. Additionally, there was a large drop in a number of phosphatidyl choline lipids in the livers of starved mice particularly 38:6 and 34:3. Sphingolipids were unaffected (Fardet et al., 2008).

A combination of proton nuclear magnetic resonance spectroscopy and GC-MS was used to examine metabolic changes in the nematode *Caenorhabditis elegans* following the deletion of nuclear hormone receptor-49 (nhr-49). It has been proposed that this receptor is related to peroxisome proliferator-activated receptor (PPARs) in mammals. Thus *C. elegans* lacking the nuclear hormone receptor-49 was compared with an α-PPAR null mouse model. An aqueous extract was analyzed by NMR and an organic extract was analyzed by GC-MS after methylation with BF_3-methanol. It was observed that there were increases in lactate, alanine, leucine and decreases in lysine, glucose, malonate and aspartate in the mutant worms. In addition, α-linolenic acid, arachidonic acid, nonadecanoic acid, dihomo-γ-linolenic acid, heptadecenoic acid, γ-linolenic acid and docosanoic acid were elevated in the mutant worms. There were a number of common metabolites also elevated in the α-PPAR null mouse livers. A PLS model was built to compare the worm and mouse models. The overall conclusion was that nhr-49 had a role in regulating lipid synthesis, β-oxidation of fatty acids, glycolysis and gluconeogenesis in a similar manner to the metabolic regulation effected by the PPAR-α receptor in mouse (Atherton et al., 2008).

1.3.3.6.4. Understanding physiological and metabolic events:

The metabolic changes induced in the earthworm *Lumbricus rubellus* by exposure to varying levels of pyrene were studied by GC-MS. Extracts were analysed both by NMR and GC-MS. For GC-MS analysis the dried extracts were derivatised to form MO/TMS derivatives. PLS models indicated that increasing doses of pyrene were associated with decreasing levels of fatty acids and increasing levels of amino acids in the worms suggesting reduced food intake and increased breakdown of muscle proteins (Jones et al., 2008).

The metabolic changes which allow the larvae of the Antarctic midge *Belgica antarctica* to adapt to extreme conditions of heat and humidity were investigated by using GC-MS. The residue from after filtration was treated to produce a MO-TMS derivative. One hundred and forty one metabolites were observed which were mainly sugars, amino acids, short chain organic acids and fatty acids. The most significant metabolic changes were as follows:

- Heat shock caused elevation of ketoglutaric acid and putrescine and depression of glucose, glycerol and serine.
- Freezing caused elevation of alanine, aspartate, erythritol, glycerol, mannitol, succinate and urea and depression of glycine and serine.
- Dessication caused elevation of erythritol, GABA, glycerol, glycerophosphate, isocitrate, nonanoic acid, norvaline and succinate and depression of aspartate, serine and sorbitol.

The changes indicated that stress responses were tailored to the particular stress applied and serine which was common to all three stress conditions could be used a general marker of stress response(Michaud et al., 2008).

A similar study to the one described above was carried out on flesh fly larvae using GC-MS to observe metabolic changes during cold hardening, which enables flies to survive low temperatures. Sixty two out of 159 chromatographic peaks could be identified. It was found that in cold hardened flies glutamine, alanine, glycerol, glucose, sorbitol, pyruvate, cystathione and urea were elevated whereas β-alanine, ornithine, trehalose and mannose were depressed. In the diapausal pupae large elevations in alanine, glucose, pyruvate and glycerol were observed and marked falls in isocitrate, sorbitol, aspartate, phenylalanine, proline, tyrosine and fumarate were observed. In the cold hardened flies glycerol and glucose are produced as osmolyte protectants which had a knock on effect on the glycolytic pathway. In the diapausal pupae a similar picture was observed with even more marked increases in the osmolyte glycerol but with reduced activity in the Krebs cycle compared with cold hardened flies marked by large increases in pyruvate but marked falls in the Krebs cycle intermediates isocitrate and fumarate (Michaud and Denlinger, 2007).

Ammonia metabolism in *Aedes aegypti* is important since the female mosquito mainly uses the amino acids present in her blood meal as a carbon source for energy metabolism and lipid biosynthesis. The consequence of using amino acids in this way is that large quantities of ammonia are produced as a by product and the mosquito must have an efficient mechanism for ammonia detoxification. In order to investigate this process tandem mass spectrometry was used in conjunction with feeding of $^{15}NH_4Cl$, [5-^{15}N]-glutamine or [2,5-$^{15}N_2$]-glutamate to the mosquitoes.

$^{15}NH_4Cl$ was rapidly incorporated into glutamine following feeding and this was followed by an increase in ^{15}N-glutamate where the labeled ammonia fixed by incorporation into glutamine was transferred to ketoglutaric acid. The labeled glutamate was then converted to ^{15}N-proline (and to a lesser extent ^{15}N-labeled alanine) which persists for a long time after feeding thus acting as a storage molecule for nitrogen. The high activity of the glutamine synthase enzyme appears to be key to the mechanism for ammonia detoxification (Scaraffia et al., 2006).

Nitrogen metabolism was further studied in *Aedes aegypti* mosquitoes by feeding $^{15}NH_4Cl$, [5-^{15}N]-glutamine or [^{15}N]-proline. It was found that all these substrates resulted in the production of urea labeled at one position. Inhibitors of xanthine dehydrogenase and glutamine synthetase inhibited labeling of urea by $^{15}NH_4Cl$ indicating the urea synthesis proceeds via glutamine and uric acid biosynthesis. Uric acid in mosquitoes fed $^{15}NH_4Cl$ or [5-^{15}N]-glutamine became labeled at two positions. Fragmentation experiments indicated that one of the ^{15}N-atoms was at position 1 or 3 in uric acid and the other was probably at position 9. Feeding unlabeled allantoin and allantoic acid to the mosquitoes fed the labeled compounds at the same time diluted the ^{15}N-label in urea indicating that urea can originate from these intermediates. Thus it would appear that $^{15}NH_4Cl$ is incorporated into [5-^{15}N]-glutamine which is further metabolized via uric acid to urea. Formerly it was thought that urea in mosquitoes was produced from arginine via the action of arginase, however, the alternative pathway via uric acid also occurs and is probably an important mechanism for removing nitrogen derived from deamination of amino acids following the female mosquito's blood meal (Scaraffia et al., 2008).

Caenorhabditis elegans has 75 genes associated with cytochrome P450 (CYP450) enzymes. The functions of these enzymes are largely unknown. In order to determine whether or not some of the CYP450s were involved in metabolism of eicosapentaenoic acid (EPA), microsomes were isolated from the untreated worms and worms which had been treated with fenofibrate and were presented with [^{14}C]-EPA as a substrate. HPLC with radiochemical detection revealed that a number of epoxidated and hydroxylated products were formed and that these were higher in

worms treated with fenofibrate. When inhibitors of CYP450 were added to the microsomal incubations, metabolism of EPA was reduced. In order to narrow down the CYP450 involved in EPA metabolism RNA silencing was applied by feeding the worms *E.coli* strains producing double stranded RNAi corresponding to a range of P450 genes in pools which were capable of knocking down about 6 CYP450 genes at a time. Using this approach *cyp-29A3* or *cyp-33E2* could be proposed as being required for EPA metabolism. Tandem MS in the negative ion mode was used to determine EPA metabolites in whole worms. A full range of epoxidated and hydroxylated metabolites of EPA and arachidonic acid (AA) metabolites could be detected and quantified in the worms. Fenofibrate strongly induced EPA metabolism but not AA metabolism (Kulas et al., 2008).

An early example of metabolic profiling in insects is provided by the application of pyrolysis GC-MS to the profiling of melanin in black, grey and yellow strains of *Drosophila*. The melanin pigment breakdown products from the three strains were compared with synthetic DOPA melanin. Pyrolysis products included pyrroles, indoles, phenol, catechol, pyridine, toluene and benzene. The relative levels of these products differentiated the *Drosophila* melanins from synthetic melanin but did not clearly distinguish between the black and grey strains of *Drosophila.* The yellow strain of *Drosophila* yielded melanin with a high pheomelanin content which was readily distinguishable from the other melanins because it yielded sulphur compounds including: methanethiol, benzomethanothiol, 2-propyl-1,3-dithiolane and 3-hydroxybenzothiazine(Latocha et al., 2000).

1.3.3.6.5. Disease diagnosis and therapeutic effect:

Metabolic profiling, mainly using NMR have been studied for diagnosis purposes of many diseases , or for evaluating the disease state and severity(Lindon, 1999).

Metabolic profiling has been studied for extracting diagnostic biomarkers for many diseases of no satisfactory clinical or biochemical diagnostic test. However, cancer has been one of the most studied diseases and will be taken as an example in this thesis. Attempts to elucidate biomarkers for early (and correct) cancer diagnosis are essential for improved patient outcome, as early diagnosis can enable simpler and

more effective interventions (Galasso et al., 2010). Comparison of plasma LC-MS profiles from pancreatic cancer patients and healthy individuals revealed discrimination between the two groups, with altered metabolites due to pancreatic cancer including amino acids, bile acids and lipids (Shiro et al.). In a separate study, ^{1}H-MAS-NMR was used to analyze a relatively large number of tumour biopsies from breast cancer patients (160). Multivariate analysis of these biopsy data enabled the prediction of estrogen and progesterone receptor and axillary lymph node status, prognostic for breast cancer patient evaluation and intervention decisions, with >75% accuracy. These studies indicate the potential role of metabolic profiling as a tool for assessing cancer prognosis as well as for determining the type of treatment needed for individual breast cancer patients (Giskeodegard et al., 2009). Oesophageal cancer is another cancer with poor prognosis. Biopsies from oesophageal cancer patients were analyzed using GC-MS combined with both univariate and multivariate analyses, using PCA and receiver operating characteristics (ROC) curves. A set of 20 potential metabolic markers were revealed from univariate analysis and a good model to discriminate tumours from normal mucosa was obtained using multivariate analysis (Wu et al., 2009). Ovarian cancer (OC) has been the subject of multiple metabolic profiling investigations. ^{1}H NMR was used to discriminate sera from endothelium ovarian cancer (EOC) patients and healthy volunteers or benign cyst carriers. The multivariate model singled out a chemical shift at 2.77 (unidentified) (Odunsi et al., 2005), capable of predicting EOC with specificity and sensitivity of 100% and also of predicting the disease stage correctly. Another OC study followed the style of hypothesis-driven targeted LC-MS/MS metabolic profiling, focusing on only five metabolites, all lysophosphatidic acid derivatives. These metabolites were significantly up-regulated in OC patients ($P<0.001$) (Baker et al., 2002). An untargeted UPLC-MS metabolic profiling study of sera from OC patients and healthy volunteers, together with a detailed investigation of statistical algorithms used for feature selection and dimensionality reduction has been recently reported (Guan et al., 2009). Support vector machine (SVM) feature selection combined with a leave one out cross validation (LOOCV) model yielded a model with 13 identified biomarkers and a high prediction confidence (97%).

CE-MS was used to screen metabolites in healthy and infected tissues from colon and stomach cancer patients, leading to the identification of several cancer-specific metabolites. Levels of TCA intermediates e.g. glucose, lactate, citrate and succinate differentiated the organ-specific tumor (Hirayama et al., 2009). An efficient and high-throughput screening was established by Monleon *et al.* using NMR spectra of faecal water samples from colorectal cancer (CRC) patients or healthy volunteers followed by pattern recognition chemometrics, and was shown to be a valuable, cheap, and reproducible diagnostic tool for CRC (Daniel et al., 2009). While early cancer diagnosis is crucial for better patient outcome, metabolic profiling can also improve cancer knowledge and biomarker discovery, with numerous examples in the literature of the applicability of NMR or MS for these purposes (Sreekumar et al., 2009, Chen et al., 2009, Gowda, 2010, Fan et al., 2009). Validation of these results is required through large-scale studies, in some cases employing more specific biomarker assays. However, these metabolic profiling studies demonstrated the value of metabolic profiling as a tool for rapid hypothesis generation.

1.3.3.6.6. Drug efficacy and mechanism of action:

Metabolic profiling can also be used as a research tool for investigating drug and lifestyle effects on biological systems, whether the objective is to assess efficacy, potential additional therapeutic effects, unexpected side-effects or drug toxicity (Lindon et al., 2007a).

Herbal medicine use declined significantly with the introduction of modern, chemically synthesized drugs. However, there is currently a notable effort to improve understanding of both modes of action and the potential of herbal medicine. Compound Danshen Tablets (CDT), a plant mixture, is widely used as a traditional Chinese therapy to treat myocardial ischemic disease (MI) (Zhang, 2003). LC-MS metabolic profiling followed by supervised multivariate statistics was used to compare this treatment with vasodilating drugs in sham or MI rat models. The origin of the herbal protective effect was shown to occur through the reversal of the effects of MI on xanthine, hypoxanthine, inosine and allantoin, which are all involved in purine metabolism (Lv et al., 2010). In a separate study, GC-MS urinary profiles from a mild-stress rat model treated with different doses of

the traditional Chinese drug, Xiaoyaosan, or the tricyclic antidepressant amitriptyline, were compared with controls, together with behavioural scores of the animals. In terms of both metabolic and behavioural scores plots, high doses of Xiaoyaosan could reverse the induced stress, while amitriptyline showed similar efficacy to mid-dose Xiaoyaosan. Thirteen potential biomarkers were reported and annotated based on searching the Human Metabolome Database (HMDB), including the amino acids glycine and tyrosine, hippuric acid and long chain fatty acids (Dai et al., 2010). Hence, metabolic profiling was not only used to screen for active plant compounds, but could prove an existing hypothesis that a traditional herbal preparation has similar efficacy to a modern synthetic medication. Also it provided an understanding of the therapeutic mechanism of the herbal medicine and generated further hypotheses concerning the potential therapeutic profile for the medication.

Metabolic profiling for drug efficacy/safety assessment has recently been applied to clinical research. One of the few reported clinical studies assessed the utility of metabolic profiling for evaluating the impact of genetic and environmental influences on the pharmacokinetics (PK) of Tacrolimus, an immunosuppressant drug. LC-MS profiles of pre-dose urine from 29 healthy volunteers were compared with the post-dose PK of Tacrolimus up to 72 hours. PLS for the pre-dose (Block-X) and PK results (Block-Y) were initially used as a feature selection filter, and this was followed by a second PLS for selected pre-dose ions (Block-X`) and PK (Block-Y). The second model could predict the correct PK of the individuals with sensitivity of 79.4% showing the ability of metabolic profiling to reliably predict the correct PK, thus paving the way for this technique to be considered for dose selection, especially for drugs with low therapeutic index (Phapale et al., 2010).

Serum metabolic profiles of type-2-diabetic patients (T2DM), either treatment-naive or treated with metformin for 3 months were generated using ^{1}H NMR spectroscopy and UPLC-MS. These approaches proved complementary in capturing metabolite differences between the two groups, ranging from observations as predictable as lower glucose levels in the metformin group, to extending beyond this to detect differences further away from the site of metabolic intervation, such as lower lysophosphatidylcholine, trimethylamine-N-oxide (TMAO), phenylalanine and

tryptophan. These findings confirmed changes previously elucidated through laborious non-metabolic profiling approaches (e.g. clinical chemistry tests) and provided new discoveries, thus opening the scope for further investigations and understanding of diabetes (Huo et al., 2009). Thiazolidinediones is another anti-diabetic drug class which have undergone ^{1}H-NMR spectroscopy metabolic profiling investigations (Martijn van et al., 2007). Plasma and urine metabolic profiles from 16 T2DM patients were compared to 16 healthy volunteers taking rosiglitazone or placebo 4 mg twice daily for 6 weeks. Metabolic changes observed between the drug-treated and placebo groups centred around hippurate and aromatic amino acids. The urine and plasma metabolites were able to discriminate patients according to disease state, showing metabolic profiling to be a promising tool for the multiparametric monitoring of disease progression, which currently depends solely on tracking blood glucose levels (Van Doorn et al., 2007).

1.3.3.6.7. Pharmacometabonomics and Personalized healthcare:

Pharmacometabonomics, first coined by (Clayton et al., 2006) arose from the inadequacy of pharmacogenomics to address specific environmental influences which may play an essential role in drug metabolism and efficacy. The observation of variation in inter-animal response was initially observed during toxicity studies of galactosamine (galN) and allyl alcohol, driving the same group to design an experiment specifically to evaluate metabolic profiling as a tool for drug response prediction (Clayton et al., 2006). A single toxic dose of acetaminophen was administered to male Sprague-Dawley rats, urine collected pre- (-48 to -24h) and post-dose (0 to 24h) and blood collected post-dose (24h). Samples were analyzed by ^{1}H-NMR spectroscopy and the results were compared with clinical chemistry and histopathological data. The authors showed the ability of metabolic profiling to predict the response of an individual to a specific drug treatment. Although the rats were all of the same genetic type, the pre-dose metabolic profiles were not the same, leading to variation in the toxicological profiles following the administration of the hepatotoxic dose of acetaminophen. Most interesting was the ability of metabolic profiling to predict the responders and non-responders from urinary *pre-dose* profiles. The examination of post-dose spectra identified the ratio of paracetamol glucuronide (G) to paracetamol (P) to have the strongest association with drug

toxicity. Pre-dose spectra were matched against this G/P ratio and the most intensely correlated regions (both positively and negatively) were isolated for further analysis and metabolite identification. The spectral regions of interest corresponded to taurine, betaine and trimethyl-N-oxide (TMAO). The latter is a gut microbial product, indicating the gut bacteria may play a role in the extent of liver damage (Clayton et al., 2006).

These researchers performed the same study on humans, where after acetaminophen administration at a typical therapeutic dose (1 g), different metabolic profiles were observed in 6 hours post-dose urine samples from 99 healthy male volunteers (Clayton et al., 2009). A disadvantage of human clinical experiments is the lack of a controlled environment, producing less informative multivariate statistics. Here, univariate approaches and visual observation of spectra were employed to obtain the most important variables from the ^{1}H-NMR spectra. Gut microbial metabolism again played a central role in the observed patterns, however, this time variation in p-cresol producing bacteria was reported. p-Cresol is a gut microbial product which appears to compete with acetaminophen for sulphate conjugation, leading to different metabolic profiles (Clayton et al., 2009).

This was followed by a hepatotoxicity study involving 71 male and female volunteers given a daily dose of acetaminophen (4 g) for seven days. Urine was collected for 14 days (pre-dose and post-dose) and analyzed by H^{1}-NMR, PCA and O-PLS. Separation between responders and non responders was investigated and response evaluated through serum levels of alanine aminotransferase (ALT), with a 2.0 fold increase from baseline to be considered as a strong responder, and below 1.5 as non-responder. Samples at time points associated with higher ALT levels, i.e. towards the end of dosing, showed a better separation model compared to pre-ALT elevation time points, however, both models were statistically significant. Pre-dose models did not give sufficient predictive information. Loadings separating the responders/non responders included exogenous (acetaminophen and metabolites) and endogenous compounds (e.g. TMAO, betaine, alanine), with the latter playing a bigger part in discriminating the groups (Winnike et al., 2010). The first published MS-based pharmacometabonomics study investigated the response of Wistar rats to streptozotocin and Sprague-Dawley rats to high energy diet, two common

approaches for inducing diabetes in animal models. Urine samples were analyzed using GC-MS and multivariate statistics (PCA) (Li et al., 2007). This experimental design was intended to generate hypotheses concerning individual differences in the potential to develop metabolic syndrome or diabetes. PCA analysis showed some outlying pre-dose Wistar rat urine profiles. These animals were metabolically "different" and, with the inclusion of clinical chemistry data, were considered to be diabetes-resistant. Sprague-Dawley rats showed high variation in weight gain in response to the high energy diet. Using PCA, the pre-dose urine samples clustered according to the individual animal's weight gain. In both diabetes induction approaches, the source of variation was due to gut microbial related metabolites such as hippurate, *p*-cresol and 4-ethyl-phenol, indicating a potential role of the gut microbiota in diabetes progression.

These studies emphasize the importance of metabolic profiling and gut microbial variation when developing new drug candidates, especially in the Phase-I of new drug applications. Moreover, there is a high need to replacing the current empirical based treatment selection towards a knowledge based approach. Metabolic profiling, through its predictive power, may present a valuable tool on the way to a systematic personalised selection of drug treatment.

1.3.4. Integrated Metabonomics (metabolic profiling):

Integrated metabonomics is defined as:

"Obtaining multiple NMR spectroscopic (or indeed other types of analysis) data sets from various biofluid samples and tissues of the same animals collected at different time points." This approach is intended to provide indications of the organ responsible for metabolic variation as well as the time course and mechanism of pathology. Metabolic profiling can be integrated with other complex data arising from genomic, transcriptomics, and proteomics analysis. This serves as directive integrative methodology combining metabolic profiling with the other more expensive omics technologies (Nicholson et al., 2002).

In order to harmonise metamolic profiling work all over the globe a standard metabolic reporting structure (SMRS) group has been formed in the purpose of generating a unified reporting system. This system should contain, origin of

biological sample , analysis of materials from that sample, and chemometrics and the statistical approach used (van der Werf et al., 2007).

2. Selection of The Chromatographic Technique:

2.1. Hydrophilic interaction liquid chromatography:

The last two decades have witnessed huge advances in the field of analytical chemistry. Liquid chromatography (LC) was one of the fields which has seen an improvement in technology ranging from better performance of machines, to the many fold improvement in separation power which has lately produced the combination of sub 2 μm columns and ultra high performance liquid chromatography (UPLC). While efficiency has been a major interest for chromatographers, selectivity has never been out of the sight. This is reflected by the numerous stationary phases introduced lately. Among these introductions hydrophilic interaction liquid chromatography (HILIC) has seen a remarkable renaissance as a solution for efficiently retaining compounds with high polarity or those with wetting difficulties. The principle of HILIC was first described in 1952, but the term "HILIC" was coined only in 1990 by Alpert who proposed the water layer mechanism of retention.

In spite of all the improvements in Reversed-phase (RP) chromatography, which has included embedding polar groups, better end capping in order to prevent silanol related interactions and end capping with polar bonded phases, RP is still refractory to retaining polar compounds such as amino acids (AA), nucleotides and many drug substances and this has made room for using HILIC for these purposes.

Metabolic profiling is a new branch of the "omics" aiming at the quantitative measurement of *all* small molecules in a given biological system (Fiehn, 2002). Mass Spectrometry (MS) has been employed ,on many occasions, for this purpose owing to its universality and sensitivity (Lindon et al., 2006). MS based metabolic profiling has been used either with direct infusion (van der Werf et al., 2007, Heng-Keang et al., 2007, McDougall et al., 2008, Erve et al., 2009) or with a prior separation. Pre-detection separation of metabolites has been achieved using gas chromatography (GC), capillary electrophoresis (CE) and liquid chromatography (LC). The three methods of separation are not identical, and give different profiles of metabolites with pros and cons specific to each method (see chapter 1 for further details.)

A comparison between the three separation methods has been made, using a mixture of 91 authentic standard of metabolites and a general suggestion of preferences was in the order of LC followed by GC and finally CE (Buscher et al., 2009). Another cross-platform comparison included GC, UPLC and HPLC with different LC stationary phases (SPs) where the conclusion with regard to which platform was preferable was not as obvious as the conclusion mentioned above, but rather gave an indication that different platforms would give complementary information(Kind et al., 2007).

The literature is rich with examples comparing the efficiency/suitability of different LC stationary phases (SP) for polar pharmaceuticals and metabolic profiling analysis. For example, t`kindt *et al* used 25 metabolites with *logD* values ranging from -7.85 to +7.08 to test extended polar reversed-phases (RPLC) which were : Atlantis dC18, Intersil ODS-3 , Zorbax XBD and Altima HP C18, concluding that the first and second were most suitable for plant metabolic profiling analysis with regards of reproducibility, selectivity and analyte retention. However, even these columns failed to properly retain polar compounds even when operating under RP conditions with high % aqueous (t'Kindt et al., 2007). It was even observed that polar embedded phases retained polar compounds less than conventional RP, reflecting previous reports on the same issue (Majors and Przybyciel, 2002).The preferred column Atlantis dC18 operated in RP mode was further challenged against a HILIC TSKGel Amide-80 column by the same group. The evaluation criteria included RT repeatability, peak area and shape and matrix effect. The test sample was an *Arabdopsis thaliea* extract crude or spiked with 22 metabolites and found lower ion suppression effects but less features and comparable peak shapes for the RP column (t'Kindt et al., 2008b).

Another RP vs HILIC comparison was made where a RP C18 column was compared to a Zwitterionic-HILIC (ZIC-HILIC) column for the analysis of urine samples from healthy volunteers. This was achieved through comparing PLS-DA models constructed from the volunteers` urine data run on both columns and detected in positive and negative modes, making a total of four data sets (RP+, RP-, ZIC+ and ZIC-). The models were constructed with regards to: gender, diurnal variation and

age. The most important discriminating features in a data set did not match any variable in any of other sets, although HILIC based datasets gave better models than those built from RP data (Cubbon et al., 2007).

Zucker rat urine samples were analyzed by RP (1.7 μm Acquity C18 BEH) or HILIC (Acquity BEH HILIC). A few standards of metabolites (21 metabolites) were used to optimise the chromatographic conditions but no comprehensive peak shape, peak area or reproducibility comparisons were shown in this report. The judgment of performance was made according to the number of reproducible features and the goodness of principal component analysis (PCA) model obtained by each method using two animal strains (lean Vs fa/fa obsese rats) at different ages to construct the model. The number of features for RP was higher than those of HILIC using optimised parameters for the peak picking software MarkerLynx (3284 Vs 2098, respectively.) The report concluded that HILIC was complementary to RP as different biomarkers were responsible for separating lean from obese rat samples analyzed according to the method of analysis (Gika et al., 2008c) similar to the findings by Cubbon *et al* (2007).

Another human urine comparison was reported by Mohamed *et al* (2009) who used a urine matrix spiked with increasing concentrations of 38 authentic standards of metabolites to compare a RP Acclaim Polar Advantage column with an Acquity BEH HILIC column. Performance was assessed by multiple criteria including PCA models and the ability to tightly cluster each concentration group, followed by a principal component variable grouping (PCVG) for grouping features according to PCA behaviour. Other assessment criteria involved were coefficient of variation (CV) for repeat injections and determined a general stability for the system over 24 hours where the CV was less than 20% for the spiked metabolites in both RP and HILIC modes, while the overall CV was less for HILIC than for RP. The final outcome was not meant to go into further comparison but the authors opted to combine the two sets and export the holistic data to a metabolite database search software for a test run for the biomarker discovery ability of the analytical systems (Mohamed et al., 2009).

Preinerstorfer *et al* (2009) compared a different type of RP column, a pentaflourophenyl (Luna-PFP) column with ZIC-HILIC. The matrix used was a cell culture from beta-lactam producing fungi. The experiment included testing of a large number of authentic standards of metabolites (258) for which an MS/MS method was optimised. Precision, accuracy, LOQ, and linearity were tested for the compounds at three levels of concentration (low, med and high). The authors indicated that some compounds were not adequately eluted from either of the two stationary phases and thus recommended metabolites such as propionic, butyric, 2-methylbutyric, 2,2-dimethylbutyric acids should be analyzed by GCMS. According to this record, HILIC performed better for most metabolites in terms of precision, recovery and linearity with the best recoveries and precisions obtained at the med and low and at med and high concentrations, respectively. On the other hand, metabolites amenable to detection with more than one method agreed well in terms of validation parameters between the two methods (Preinerstorfer et al., 2009).

An extensive study was made on *E.coli* extracts in order to assess the ability of different stationary phases to efficiently elute a total of 164 metabolites corresponding to 20% of the genome based estimate of the *E.coli* metabolome. The study included testing of 18 different chromatographic methods encompassing a range of stationary phases and elution modes which included: extended polarity RP in RP mode, graphitized carbon in RP mode, pure bare silica (NP) and TSK Gel Amide 80 in HILIC mode and Luna-NH_2 and Luna-CN HILIC columns operated under both RP and HILIC modes. An initial stability/limit of detection (LOD) study which excluded 20 metabolites was followed by a direct infusion with automatic MS/MS parameters optimisation for 144 metabolites. The remaining compounds were run under the 18 chromatographic conditions and the methods were ranked using a scoring system which took into consideration the following parameters: sensitivity, peak shape, peak symmetry and retention. The ranking indicated a preference for the aminopropyle column operated under HILIC conditions and further investigations led to concluding this method was best for a fast and comprehensive analysis of 141 metabolites (two metabolites were dropped for non linearity and one for low reproducibility) (Bajad et al., 2006).

Buscher *et al* (2009) tested the suitability of different separation platforms (GC, CE and LC) for the analysis of 91 metabolites. The study excluded isomer separations thus further reducing the target number of metabolite identifications to 75. The LC methods they used were either ion pairing with tributylamine followed by RP separation or separation on a HILIC Luna-NH_2 column. The authors concluded that no single method could analyze all the metabolites involved in the study, even though only a portion of the metabolome was studied, and only by employing six methods (2*LC , 2*GC and 2* CE) could comprehensive analysis of the selected metabolites be carried out (Buscher et al., 2009).

Kind *et al* (2007) provided another example of using a cross platform approach for the metabolic profiling of urine, aimed at identifying biomarkers for kidney carcinoma. One GCMS and two LCMS methods were employed. The LC methods were a UPLC RP method based on an Acquity C18 BEH column and a method based on a HPLC HILIC Luna-NH_2 column. A comparison of the different freely available data processing packages (XCMS and MZmine) was also reported and a projection to latent structure (PLS) statistical model was built for the 50 most important features in each of the methods in order to separate healthy from kidney cancer patients. In this study, RP gave a higher number of features in MZmine but a much lower number than HILIC when using XCMS (Kind et al., 2007).

Non metabolic profiling comparisons of stationary phases have been reported as well. Several studies have opted to select specific chemical groups of compounds as a selectivity challenge for stationary phases. Purines, pyrimidines and organic acids were run on different stationary phases including bare silica, polyhydroxyethyl aspartamide, aminopropyl and TSK gel amide 80 in order to observe the ability of the columns to separate the members of each group. Although the chemical classes here look like metabolites, the study included some polar drug molecules (e.g. 5-fluorouracil, acyclovir) retention of which provides a challenge for RP columns. The different stationary phases behaved differently towards the different classes, and the selectivity could be manipulated via changing the mobile phase components as well. For example, the order of resolving power for pyrimidines was polyhydroxethyl aspatamide < bare silica < TSK gel amide < aminopropyl but for small molecule

amides, although selectivity was similar, bare silica performed better than Luna-NH_2. The authors dug deeper and tested different brands of aminopropyl and silica columns (YMC-NH_2, Nucleosil-NH_2 and Zobrax-NH_2 for aminopropyl and YMC-SIL, Nucleosil-silica and Zobrax-SIL for bare silica) finding comparable results in retention and selectivity for the aminopropyl columns and similar selectivity but not retention and peak shape for silica. The study concluded by providing a real life example of using HILIC for impurity profling of pharmaceutical products (Olsen, 2001).

Analysis of natural products is another example of the potential for HILIC. Seven stationary phases were tested. One carbohydrate based, one aminopropyl, two cyano columns, a TSK gel amide 80 phase, a polymeric carbamoyl based phase (Polyhydroxyethylene aspartamide) and a cyclodextrin based phase (CycloBond III) were challenged with the separation of six hydrophilic natural products. The only two common features of the test compounds were their hydrophilicity and natural origin. The six compounds were run on all seven stationary phases along with a panoptic study of the effect of mobile phase additives (type, concentration and pH) and according to the authors, for this very specific application, the TSK gel amide 80 was deemed best among these column (Strege, 1998).

Another dimension of stationary phase comparison could be found by comparing in-house synthesised stationary phases with commercially available phases. An example was introduced in the work of Lammerhofer *et al* (2008) who tested four in-house made mixed mode, reversed-phase and weak anion exchange RP-WAX, phases with eleven commercially available columns encompassing interaction mechanisms of NP, RP, RP-WAX, XPRP, carbamoyl based and zwitterionic. These 15 stationary phases were tested with four compounds in order to assess their ability to retain by the different mechanisms. PCA was used on the retention data in order to group the phases where all the mixed mode RP-WAX clustered in the same region and away from columns like ZIC-HILIC and TSKGel Amide 80 (Lämmerhofer et al., 2008).

In this chapter, the importance of the pre-MS separation step will be discussed with the aid of in-silico simulations and experimental work is presented in

order to shed light on several LC stationary phases screened with the objective of maximizing the information obtained from a metabolic profiling experiment, using extracts of different genotypes of *Drosophila melanogaster* as a test model.

2.2. Materials and methods:

2.2.1. Chemicals and authentic Standards:

Authentic Standards were obtained from Sigma-Aldrich(Poole, UK), Fluka Biochemika (Poole, UK) Alfa Aesar (Heysham, UK),BDH Chemicals LTD (Dorset UK). The sources of the authentic standards are shown in Table 2.1 along with the concentrations of the stock solutions used. The amounts indicated in the table (in mg) were dissolved in 1.8 ml of a 50:50 mixture of Water:Methanol and further diluted in order to get the reported concentrations. Water was of HPLC grade obtained from a Millipore Q system (Millipore Watford, UK) and MeOH and acetonitrile were of HPLC grade and obtained from VWR International Ltd. (Lutterworth, UK). AnalaR grade formic acid (98%) was obtained from BDH-Merck (Dorset, UK).

Exact mass values were calculated online from Scientific Instrument Services SISWEB.(Services, 2010)

logP of all compounds was imported from Pubchem (http://pubchem.ncbi.nlm.nih.gov/).

No	Chemical	logP	MW (amu)	quant mg	Conc µg/ml	vendor
1	Cis-Aconitic Acid	-1	174.0164	0.944	0.262	Sigma
2	Acetylcholine Chloride	0.2	146.1181	4.1	1.139	Aldrich
3	Adenine	-0.1	135.0545	0.812	0.226	Sigma
4	Adenosine Tri-Phosphate	-5.7	506.9958	0.538	0.149	Sigma
5	Alanylglycine	-3.7	146.0691	0.878	0.244	Aldrich
6	gamma-Aminobutyric Acid	-3.2	103.0633	0.806	0.224	Sigma
7	Arabinose	-2.5	150.0528	0.658	0.183	Aldrich
8	Arginine	-4.2	174.1117	1.332	0.370	BDH Chemicals
9	Ascorbic Acid	-1.8	176.0321	4.29	1.192	BDH Chemicals
10	Biopterin	-2.4	237.0862	0.392	0.109	Fluka BiochemiKa
11	Caffeic Acid	1.2	180.0423	0.966	0.268	Fluka BiochemiKa
12	Cinnamic Acid	2.1	148.0524	0.578	0.161	Aldrich
13	(p-)Coumaric Acid	1.5	164.0473	1.054	0.293	Sigma
14	Creatinine	-1.8	113.0589	1.792	0.498	Aldrich
15	Cysteine	-2.5	121.0198	0.564	0.157	BDH Chemicals
16	Cystine	-6.3	240.0239	1.082	0.301	Fluka BiochemiKa
17	Deoxyglucose	-2.9	164.0685	1.146	0.318	Sigma-Aldrich
18	Dihydrobiopterin	-2.2	239.1018	0.348	0.097	Fluka BiochemiKa
19	(1,7)Dimethylxanthine	-0.2	180.0647	0.856	0.238	Sigma
20	Galactose	-2.6	180.0634	0.612	0.170	Sigma-Aldrich
21	D(+) Galacturonic Acid	-2.3	194.0427	1.204	0.334	Fluka BiochemiKa
22	alpha D (+) Glucose	-2.9	180.0634	2.6	0.722	Sigma
23	(D) Glucose-6-Phosphate	-4.4	260.0297	1.062	0.295	Sigma
24	D(+) Glucosamine	-2.8	179.0794	2.924	0.812	Sigma
25	Glutamic Acid	-3.7	147.0532	1.188	0.330	Aldrich
26	Glutathione (reduced)	-4.5	307.0838	0.494	0.137	Sigma
27	Glutathione (Oxidised)	-9.1	612.152	0.6	0.167	Sigma
28	Glycine (D5)	-3.2	77.04458	3.818	1.061	Aldrich
29	Guanine	-1	151.0494	1.804	0.501	Fluka BiochemiKa
30	Fructose	-3.2	180.0634	4.352	1.209	Sigma-Aldrich
31	L(-)Fucose	-2.1	164.0685	1.166	0.324	Sigma
32	DOPA (Dioxyphenylalanine)	-2.7	197.0688	0.796	0.221	Sigma-Aldrich
33	Epinephrine Bi-tartarate	-1.4	183.0895	1.012	0.281	Sigma-Aldrich
34	Histamine	-0.7	111.0796	2.14	0.594	Sigma-Aldrich
35	L(-)Histidine	-3.2	155.0695	0.844	0.234	Sigma
36	(5) HydroxyDopamine	NA	169.0739	0.296	0.082	Sigma-Aldrich
37	alpha-HydroxyIsobutyric Acid	-0.4	104.0473	1.61	0.447	Sigma
38	DL(Alpha)Hydroxyisovaleric Acid	0.5	118.063	0.848	0.236	Sigma

No	Chemical	logP	MW (amu)	quant mg	Conc µg/ml	vendor
39	Hydroxy(β)methyl(β)glutaric Acid	-1.2	162.0528	0.808	0.224	Sigma-Aldrich
40	(3)HydroxyTyramine	-1	153.079	0.838	0.233	Sigma-Aldrich
41	Hypoxanthine	-1.1	136.0385	0.658	0.183	Fluka BiochemiKa
42	Inosine	-2.1	268.0808	0.906	0.252	Sigma
43	Isocitric Acid (trisodium salt)	-1.8	192.027	1.478	0.411	Sigma
44	(DL) IsoLeucine	-1.7	131.0946	0.96	0.267	Sigma
45	(Alpha)Ketoglutaric Acid (Disodium)	-0.9	146.0215	0.688	0.191	Sigma
46	Kynurenic Acid	1.3	189.0426	0.882	0.245	Fluka BiochemiKa
47	(DL)Kynurenine	-2.2	208.0848	0.416	0.116	Fluka BiochemiKa
48	(D) Lactic Acid (sodium salt)	-0.7	90.0317	1.126	0.313	Fluka BiochemiKa
49	(L) Lysine	-3	146.1055	0.898	0.249	Aldrich
50	Maleic Acid	-0.3	116.011	1.882	0.523	Fluka BiochemiKa
51	Maltopentaose	-11.5	828.2747	0.802	0.223	Sigma
52	Maltotetraose	-9	666.2219	0.472	0.131	Sigma
53	Mannose	-2.6	180.0634	0.624	0.173	Sigma-Aldrich
54	Methionine	-1.9	149.0511	0.824	0.229	Aldrich
55	Methyl(α)carboxyl(4)phenylglycine	-2	209.0688	0.044	0.012	Sigma
56	Methyl(1)histamine	-0.6	125.0953	0.458	0.127	Sigma
57	(L) Methyl(3)histidine	-3.3	169.0851	0.492	0.137	Sigma
58	Methyl(6)uracil	-0.8	126.0429	0.552	0.153	Sigma
59	Methyl(1)xanthine	-0.3	166.0491	0.488	0.136	Aldrich
60	N-AcetylHistamine	-0.4	153.0902	1.302	0.362	Aldrich
61	**Nicotinamide Adenine Dinucleotide (NAD)**	-6	663.1091	1.936	0.538	Sigma
62	(B) Nicotinamide adenine dinucleotide Phosphate (NADP)	-7	744.0833	1.24	0.344	Sigma
63	(DL) NorAdrenaline (HCl)	-1.2	169.0739	1.464	0.407	Fluka BiochemiKa
64	Octopamine	-1.1	153.079	0.636	0.177	Sigma-Aldrich
65	Oxaloacetic Acid	-0.6	132.0059	1.128	0.313	Sigma-Aldrich
66	Oxalic Acid	-0.3	89.99531	4.236	1.177	BDH Chemicals
67	Phenylalanine	-1.5	165.079	1.062	0.295	Aldrich
68	(L-Alpha) Phosphatidylethanolamines		271.0457	1.402	0.389	Sigma
69	Pyruvic Acid	-0.3	88.01605	9.222	2.562	Sigma-Aldrich
70	Rhamnose	-2.4	164.0685	2.8	0.778	Sigma-Aldrich
71	Riboflavin	-1.5	376.1383	0.906	0.252	Sigma-Aldrich
72	Serine	-3.1	105.0426	2.672	0.742	Sigma-Aldrich
73	Serotonin (HCl)	0.2	176.095	0.678	0.188	Sigma-Aldrich
74	Spermidine (Tri-HCl)	-1	145.1579	0.948	0.263	Sigma
75	Stachyose Hydrate	-8	666.2219	2.116	0.588	Sigma
76	Succinic Acid	-0.6	118.0266	2.166	0.602	ANALAR

No	Chemical	logP	MW (amu)	quant mg	Conc µg/ml	vendor
77	Sucrose	-3.7	342.1162	3.402	0.945	Sigma-Aldrich
78	Taurocholate Hydrate (sodium)	2.2	515.2917	0.814	0.226	Fluka BiochemiKa
79	Thiamine (HCl)	1	265.1123	0.376	0.104	Sigma
80	Tryptophan	-1.1	204.0899	0.74	0.206	Sigma-Aldrich
81	Tyrosine	-2.3	181.0739	2.41	0.669	Sigma-Aldrich
82	Uric Acid	-1.9	168.0283	0.794	0.221	Fluka BiochemiKa
83	(DL) Valine	-2.3	117.079	1.346	0.374	Alfa-Aesar
84	Xanthine	-0.7	152.0334	0.988	0.274	Sigma-Aldrich
85	(D(+)) Xylose	-2.5	150.0528	4.368	1.213	Sigma
86	Alanine	-3	89.04768	2.944	0.818	Sigma-Aldrich
87	Asparagine	-3.4	132.0535	1.736	0.482	Sigma-Aldrich
88	Aspartic Acid	-2.8	133.0375	1.574	0.437	Aldrich
89	Betaine	0.5	117.079	1.554	0.432	Sigma-Aldrich
90	Carbamylcholine (Chloride)	1	147.1134	2.56	0.711	Sigma-Aldrich
91	Choline (base)	-0.4	104.1075	4.5	1.250	Sigma-Aldrich
92	Creatine	-1.2	131.0695	0.714	0.198	Sigma-Aldrich
94	Dimethyl(3,3)glutaric Acid	0.4	160.0736	2.978	0.827	Sigma-Aldrich
95	Dimethyl(N,N)glycine (HCl)	-2.9	103.0633	2.46	0.683	Sigma-Aldrich
96	Ferulic Acid	1.5	194.0579	1.142	0.317	Sigma-Aldrich
97	Fumaric Acid	-0.3	116.011	2.9	0.806	Sigma-Aldrich
98	Glycylproline	-3.1	172.0848	1.46	0.406	Sigma-Aldrich
99	Glutaric Acid	-0.1	132.0423	4.494	1.248	Sigma-Aldrich
100	Glycine Methyl Ester (HCl)	-0.8	89.04768	2.182	0.606	Sigma-Aldrich
101	Heptadecanoic Acid	6.9	270.2559	1.048	0.291	Sigma-Aldrich
102	Hydroxy(6)nicotinic Acid	-0.5	139.0269	1.194	0.332	Sigma-Aldrich
103	Inositol (myo)	-3.7	180.0634	2.254	0.626	Sigma-Aldrich
104	Lactose	-4.7	342.1162	1.246	0.346	Sigma-Aldrich
105	Leucine	-1.5	131.0946	1.056	0.293	Sigma-Aldrich
106	Levulinic Acid	-0.5	116.0473	9.8	2.722	Sigma-Aldrich
107	Malonic Acid	-0.8	104.011	3.43	0.953	Sigma-Aldrich
108	Methyl(N)thioUrea	-0.7	90.02517	4.244	1.179	Sigma-Aldrich
109	Methylurea	-1.4	74.04801	5.192	1.442	Sigma-Aldrich
110	Nicotinamide	-0.4	122.048	1.44	0.400	Sigma-Aldrich
111	Palmitic Acid	6.4	256.2402	3.454	0.959	Sigma-Aldrich
112	Phenyl(Alpha)glycine	-1.7	151.0633	0.55	0.153	Aldrich
113	Phosphorylcholine Chloride	-1.5	183.066	1.86	0.517	Sigma-Aldrich
114	Pyridoxine	-0.8	169.0739	0.918	0.255	Sigma-Aldrich
115	Ribose	-2	150.0528	1.368	0.380	Sigma-Aldrich
116	Spermine	-1.1	202.2157	10	2.778	Sigma-Aldrich
117	Tartaric Acid	-1.9	150.0164	1.908	0.530	Sigma-Aldrich
118	Thymine	-0.6	126.0429	0.84	0.233	Sigma-Aldrich

No	Chemical	logP	MW (amu)	quant mg	Conc µg/ml	vendor
119	Trehalose (Dihydrate)	-4.2	342.1162	2.542	0.706	Sigma-Aldrich
120	Uracil	-1.1	112.0273	1.22	0.339	Sigma-Aldrich
121	Urea	-1.4	60.03236	4.606	1.279	Sigma-Aldrich
122	Cytidine(5)Diphosphoethanolamine	-7.2	446.0604	0.786	0.218	Sigma
123	Cytidine(5)diphosphocholine	-3.9	489.1152	0.222	0.062	Sigma
124	Sphingomyelin	13.4	730.5989	0.466	0.129	Sigma
125	Glycerophosphocholine	-2.3	258.1107	1	0.278	Sigma
126	S-Adenosylmethionine	-2.8	398.1372	0.144	0.040	Sigma
127	O-acetyl-N-carnitine (HCl)	0.4	203.1158	1.002	0.278	Sigma
128	(L) Citrulline	-4.3	175.0957	0.438	0.122	Sigma
129	Spermine	-1.1	202.2157	0.252	0.070	Sigma-Aldrich
130	Spermidine (Tri-HCl)	-1	145.1579	0.252	0.070	Sigma
131	(+-) Carnitine (HCl)	-0.2	161.1052	1.45	0.403	Sigma
132	HomoCysteine	-3.4	135.0354	0.724	0.201	Sigma
133	(D) Pantothenic acid (Calcium Salt)	-1.1	219.1107	0.612	0.170	Fluka BiochemiKa
134	Thymidine	-1.2	242.0903	0.53	0.147	Fluka BiochemiKa
135	Guanidine	-1.3	59.04835	1.504	0.418	Sigma
136	(DL) Ornithine	-4.4	132.0899	1.24	0.344	Fluka BiochemiKa
137	UDP-Glucuronate	-6.4	580.0343	1.762	0.489	Sigma-Aldrich
138	Proline	-2.5	115.0633	1.432	0.398	Sigma-Aldrich
139	Adenosine	-1.1	267.0968	1.446	0.402	Sigma
140	Taurine	-4.1	125.0147	5.746	1.596	Fluka BiochemiKa
141	(L) Glutamine	-3.1	146.0691	1.134	0.315	Fluka BiochemiKa
142	Guanosine	-2.7	283.0917	1.078	0.299	Fluka BiochemiKa
143	Putrescine (Diaminobutane) (2*HCl)	-0.9	88.10005	0.852	0.237	Fluka BiochemiKa

Table.2.1. Test solution concentration, source, logP and exact molecular weight of the authentic standards used in chapter 2 experiments. Grey highlighted compounds have limited solubility in the solvent used. logP values were extracted from XlogP3 pubchem®.

2.2.2. *Drosophila melanogaster* extraction:

Drosophila melanogaster were kept in vials of standard medium at 25°C and 55% relative humidity (r.h.) on a 12:12 hour photoperiod. Stocks used were Canton S, Oregon wild-type flies, and ry 506 homozygotes, and Maroon-like (Bloomington Stock Center). Seven day old adult flies were anaesthetized by chilling on ice and then the whole flies were ground in 50% methanol:50% water. Ten flies were found to provide ample experimental material. Five males and five females were used for

each extraction apart from extractions were males and females were compared. Following extraction, samples were stored at -20ºC until the day of analysis. Immediately prior to analysis samples were brought to room temperature, vortexed and then centrifuged for 5 min at 7500 rpm. Canton S and Oregon wild type samples were combined in a separate 5 ml glass vial (Kinesis, Cambridgeshire, UK) to obtain a stock of wild type whole flies extracts which will be referred to as pooled_1. Maroon-like and ry 506 extracts were combined in a separate 5 ml glass vials in order to obtain a stock of xanthine-oxidase deficient whole flies extracts which will be referred to as pooled_2. 180 µl portions of both stocks were withdrawn and transferred to chromatographic vials.

2.2.3. LC-MS conditions:

LC–MS data were acquired using a LTQ Orbitrap instrument (Thermo Fisher Scientific, Hemel Hempstead, UK) set at 30000 resolution with scan time of 200 ms. Sample analysis was carried out under positive and negative ion mode. The mass scanning range was m/z 50–1200 and the capillary temperature was 200ºC, sheath and auxiliary gas was set to flow at 30 and 10 arbitrary units, respectively. The spray needle voltage was set to + 4 kV for positive mode and – 4 kV for negative mode ionization.

The LC–MS system was run in binary gradient mode. Unless otherwise mentioned, the 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. The column temperature was set to 25ºC.

Chromatographic columns used in this experiment were:

A ZIC-HILIC (Sequant) column 5 um* 150 * 4.6 mm.

Obilesc_N (Sielc) column 5 um* 250 * 4.6 mm.

Obilesc_R (Sielc) column 5 um* 250 * 4.6 mm.

Luna-HILIC column (Phenomenex) 5 um* 150 * 4.6 mm.

ACE C18 column (ACE) 5 um* 150 * 4.6 mm.

Phenyl column (Phenomenex) 5 um* 150 * 4.6 mm

All columns were purchased via (HiChrom, Reading, UK).

The gradient used was as follows: 80% B (0 min) – 40% B (12 min) – 40% B at 22 min – 80% B at 30 min. The injection volume was 25 µl in a full loop mode.

Mass measurement was externally calibrated just before commencing the experiment.

LC-MS retention time data were acquired using an LCQ-Deca Mass spectrometer (Thermo Fisher Scientific, Hemel Hempstead, UK) operated in positive or negative ion mode using conditions similar to those used for Orbitrap. All mass spectrometric data (Orbitrap and LCQ) were acquired under the control of Xcalibur 2.0.6 in centroid data collection mode.

2.2.4. Data processing:

2.2.4.1. Data pre-processing:

For the RP Vs HILIC experiment, Sieve®1.2 was used to process the data.
For the retention behaviour experiments: retention time, peak width, area and height were manually exported from Xcalibur 2.0.6 QualBrowser to Microsoft Excel 2007 spreadsheets. For the Luna-HILIC, ZIC-HILIC comparison a full peak list was exported from Xcalibur 2.0.6 QualBrowser to Microsoft Excel 2007 spreadsheets. Filtration used mass peak widths of 2 ppm for positive ion mode and 5 ppm for negative ion mode. Data alignment and comparison were performed using in-house built Microsoft Excel Visual Basic for Application (VBA) macros.

2.2.4.2. Multivariate analysis:

Multivariate analysis was performed using the commercially available software SIMCA-P+11.5.0.0, Umetrics AB, Sweden.

2.3. Results and discussion:

2.3.1. The need for pre-detection chromatographic separation in *Drosophila* metabolic profiling:

Metabolic profiling experiments necessitate the detection of a very high number of compounds in the shortest time, therefore direct infusion seems like the optimal method. Nevertheless, ion suppression and lack of resolution of isomers are known problems associated with direct infusion. The conversion of exact masses into molecular formulae have been discussed before, and conclusions were drawn with regard to the inability of current instrumentation (accuracy as low as 1 ppm) to

provide one single unique formula for a feature when current available databases were used(Kind and Fiehn, 2006). The work here, will explore this same issue, but from a different angle focused on the *Drosophila* metabolome. As the queried databases in Kind`s work were large databases containing formulae of human endogenous metabolites as well as xenobiotics, the investigation here was aimed at assessing if these conclusions are specific to large databases or can be generalised to simpler challenges with smaller number of metabolites and with less interferences from exogenous compounds.

For this purpose, an *in-silico* technique was used for plotting mass interference points on the x-axes of mass distribution. The *Drosophila* metabolome (hypothetical metabolome, as it is based on available but not experimental literature information) and the *Drosophila* pathways were exported from [Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000) Release 49.0, January 1, 2009] into Microsoft Excel spreadsheets and the monoisotopic exact masses were calculated using an in-house built VBA macro along with the ^{13}C and ^{34}S isotopes and Sodium (Na) adducts of each metabolite.

The generated list represents a very simple challenge to the required resolution as only metabolites with pathway involvement ,only one isotopic (most

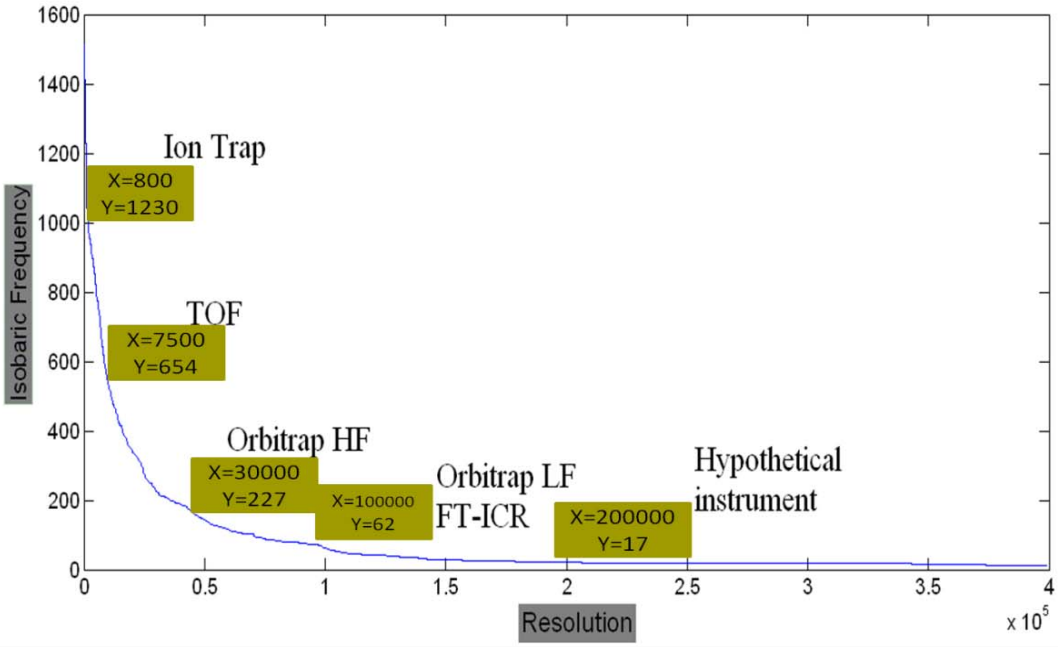


Figure.2.1 :Required resolution power to achieve a direct infusion resolution of isobaric pairs in *Drosophila* Metabolome.

abundant isotope) in most cases and one adduct type were included. This is an over simplification of the real life experiments, as other isotope peaks (^{15}[N] 2*^{13}[C] and ^{18}[O]) and other adducts (K^+, $NH4^+$, Cl^- , HCOOH...etc) can be encountered.

Figure.2.1 is a plot of the number of isobaric pairs (y axes) against the resolution of the instrument. It shows that improvement in resolution gives, as expected, less isobaric pairs. However, even with such a simple challenge, the best resolving power did not travel beyond "reducing" the frequency of occurrence of isobaric pairs, but never removing it, in fact , to reduce the number of isobaric pairs to zero, a resolution of 6,006,100 (sixty times more than the current best) is required.

Isobaric compounds are not the only problem in feature annotation in metabolic profiling. Isomeric compounds provide another challenge for the methodology used, as these cannot be separated regardless of the resolution of the mass spectrometer and definitely postulate a pre-detection separation. To evaluate the size of the isomeric problems, the isomers of the *Drosophila* metabolome were plotted on the x-axes of m/z.

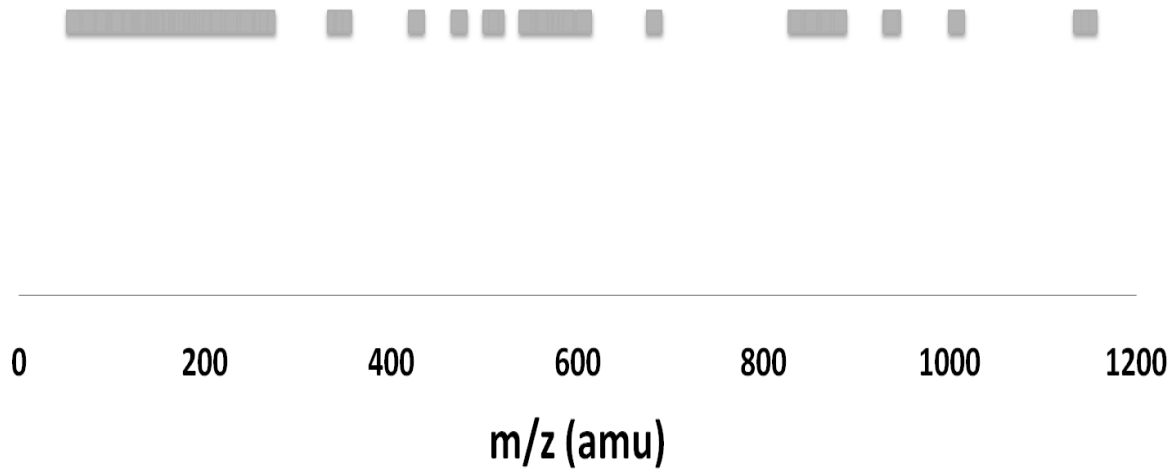


Figure.2.2 a representation of the isomeric frequency in the *Drosophila* metabolome. Each square represents an isomeric pair.

As direct infusion will not distinguish isomeric or isobaric pairs, and the number of these for Orbitrap, the instrument to be used in all metabolic profiling experiments included in this work, is significant (103 isomeric pairs in the mass range 50-1200 amu), chromatographic separation is shown to be a must for metabolic profiling purposes, even when the expected number of detectable metabolites small. Ion suppression phenomenon adds to the importance of the last conclusion as well.

2.3.2. Effect of chromatography type on the outcome of a metabolic profiling experiment:

In the last paragraph, a pre-detection chromatographic separation was deemed inevitable for a valid and successful metabolic profiling analysis of *Drosophila*. The analytical market is currently rich with a vast array of chromatographic techniques, each may give a different answer for exactly the same samples, as can be seen in Figure.2.3 where three samples of *Drosophila* wild type (pooled_1) and three samples of *Drosophila* deficient with xanthine-oxidase (pooled_2) were run on different columns (two reversed-phase a C18 and a Phenyl) and two HILIC phases (a ZIC-HILIC and a Luna-HILIC) operated in a HILIC gradient elution mode.

The variation of chromatogram shapes is further explained by the principal component analysis (PCA) models of the m/z features obtained from each chromatographic set (Figure.2.4).

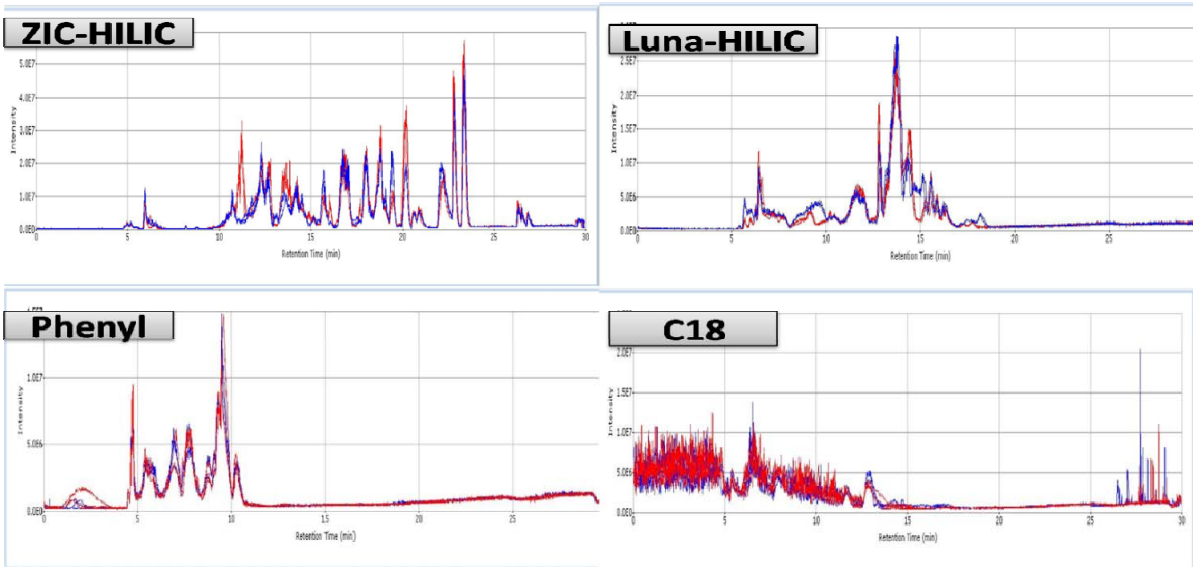


Figure.2.3 Representative chromatograms of *Drosophila* samples run through four stationary phases. Two HILIC columns (ZIC-HILIC and Luna-HILIC) and two reversed phase (Phenyl and C18).

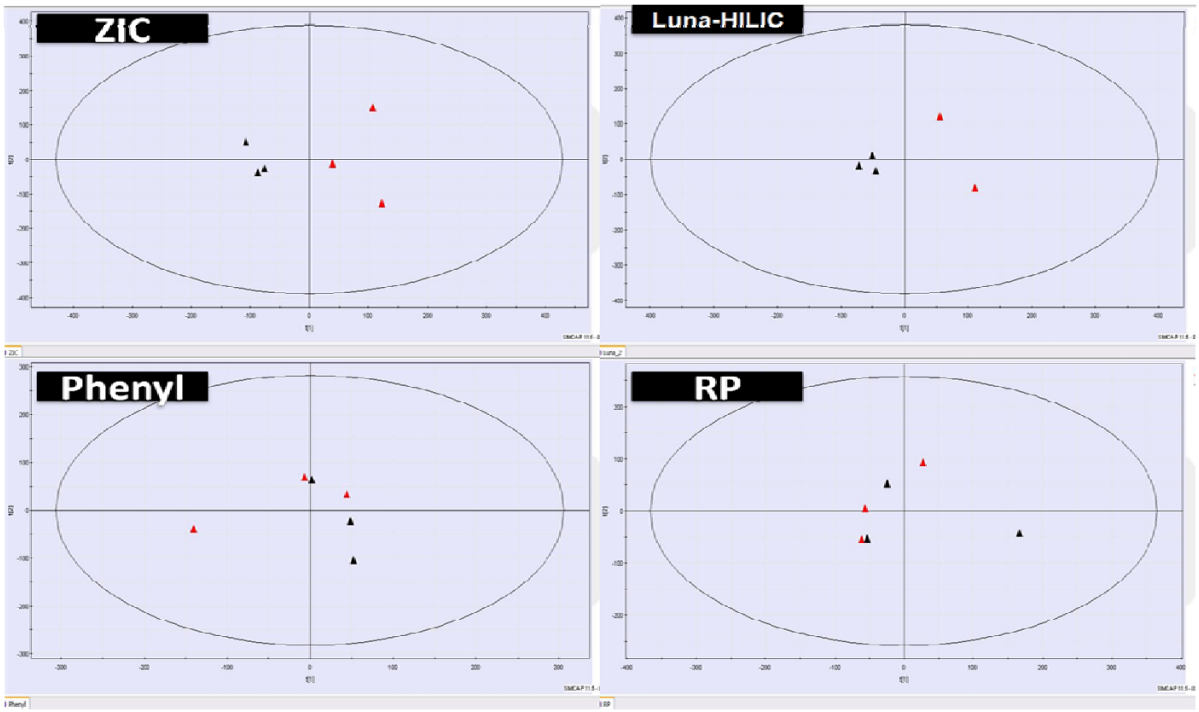


Figure.2.4 Principal component analysis of Drosophila samples. Red triangles are of the ry (n=3) type and black triangles from wild type (n=3). In the Luna-HILIC experiment (n=2) for the ry due to a missed injection.

Disregarding the time domain, as different chromatography will result in different retention; we can observe the complementarities (in feature detection) of these techniques one to another in Table.2.2 and the corresponding Venn diagram Figure.2.5. Neither HILIC nor RP could account for more than 70% of the total feature count obtained by both techniques. This shows the amount of information potentially lost when relying on a single chromatographic platform, as is the practice in many metabolic profiling studies (Bruce et al., 2009, Guy et al., 2008, Chang et al., 2009, Eric et al., 2007). The feature here represent a unique m/z which reproducibly occur in all repeats per column (n=6) with a tolerance of 2 ppm for positive and 5 ppm for negative modes.

I should mention here that a comparision between the two chromatography platforms (HILIC vs RP) is not the subject of this investigation, but I am rather using this test as a show case that neither platform can give a complete picture of the content of the same sample. A fair comparision of platforms should include data from optimised conditions on both platforms, while here; the extraction method and the used gradient are more suitable for HILIC than RP analyses.

column	Total number of detected features	Specific to the column	Common with one other column	Common with two other columns	Common in all four columns
ZIC-HILIC	786	392	125	92	177
Luna-HILIC	547	179	121	70	177
RP-C18	540	200	76	87	177
RP-phenyl	509	146	96	90	177
	Specific to HILIC	Specific to RP	common in HILIC & RP	total	
HILIC vs. RP	651	389	376	1416	

Table.2.2 Positive ion mode feature distribution across the tested stationary phases. A HILIC vs. RP comparison. Specific to the column means features encountered in that column but not any other column. Specific to HILIC (or RP) means specific to C18, Phenyl or found in both columns but not in any HILIC column.

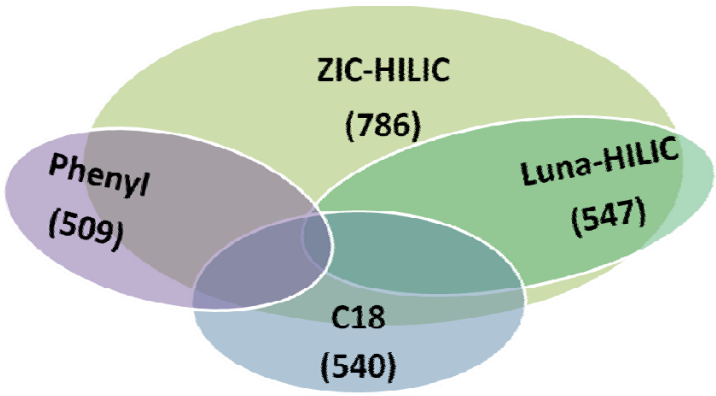


Figure.2.5 A venn diagram showing the distribution of features detected using the four different stationary phases.

In the following paragraphs, the focus will be on HILIC separation, taking on board a few of the HILIC columns available in the market and minimally or not tested, at the date of this work, for metabolic profiling.

2.3.3. Comparing the performance of different HILIC columns:

2.3.3.1. Retention behaviour:

In metabolic profiling, as in all types of analysis, there is a desire to accomplish the whole analysis in the shortest amount of time. On the other hand, ion suppression phenomenon in ESI-MS requires a good separation of the components of very complex samples e.g. containing a few thousands metabolites per sample for a human serum sample (Wishart et al., 2009, Eric et al., 2007) which reduces speed of analysis if good separation is required. Thus, a compromise is required in this case, and a good method should separate most components in a convenient total run time. This implies that the chromatographic system should have a good retention factor for all types of analytes included in the analysis i.e. all metabolite classes.

Since the rediscovery of the HILIC concept by Alpert, many HILIC stationary phases have emerged each of which has its unique features and selectivity. Suitability testing of all HILIC phases in the market would be of interest, but falls beyond the scope of one thesis, especially when chromatography is not the major component of the work, thus only four HILIC columns were selected for investigation. However, the general method in this work could provide a basis for testing other columns.

Obelisc N is a liquid cell technology bonded silica stationary phase. The bonded phase is composed of a silica skeleton separating a latent negative charge (closer to the silica support) and a superficial positive charge. The analytes will interact with the silanol groups on the silica support and silica backbone (HILIC interaction) and with the latent and superficial charges (weak ion-exchange interactions).

Obelisc R is another Sielc liquid cell technology stationary phase with mixed mode properties. It is a silica bonded phase, with the bond composed of a reversed-phase C18 skeleton separating a latent positive charge and superficial negative charge. The analytes will have a hydrophobic London force interaction with the alkyl ligand and a weak ion exchange interaction with the latent and superficial charges

which will retain a water layer leading to further hydrophilic interaction.(Sielc, 2010b)

Luna-HILIC is a diglycol type bonded silica stationary phase where the silanol groups are capped with ethylene glycol to yield a polar non-charged stationary phase. The mechanisms of retention here could be hydrogen bonds with the hydroxyl /silanol moieties and the HILIC proposed interactions for the retained water layer (Chirita et al., 2010).

ZIC-HILIC is a Zwitterion sulfobetaine silica (or polymer) bonded phase with a short hydrophobic (C3) skeleton separating a latent positive charge and a superficial negative charge. The total charge of the stationary phase is zero. Analytes will have a HILIC interaction at the water layer as well as a weak ion exchange with the charges (for charged analytes) (Nesterenko et al., 2009) The structure of the phases are shown in Figure 2.6.

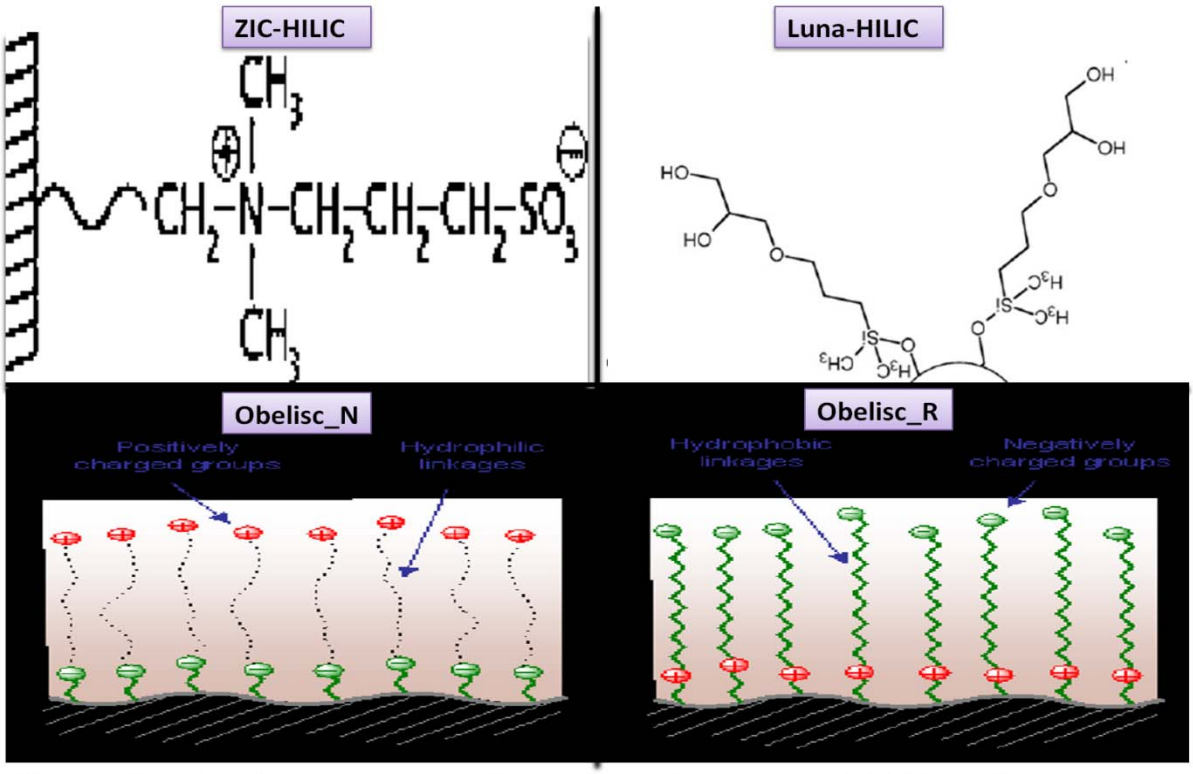


Figure.2.6 Stationary phase Chemistry of the tested HILIC columns.

Authentic standards were run through each of the four columns in positive and negative polarities, and retention times, peak borders, areas and heights were manually exported to spreadsheets. The LC-MS parameters were chosen to suit a metabolic profiling experiment as there is a growing evidence the composition of the stationary phase affects the outcome of metabolic profiling experiments more than the composition of the mobile phase (Legido-Quigley et al., 2010). Gradient systems give better chromatographic efficiency, hence, peak capacity which is a crucial requirement for the suitability of a chromatographic system for metabolic profiling analysis. As all stationary phases here share the HILIC mechanism of retention, it was convenient to make a general HILIC mode gradient elution swinging between 80% and 40% of organic modifier (acetonitrile). One concentration was chosen for each reference standard in the proximity of 0.2 µg/ml (0.1-1.2 µg/ml), a concentration which is good enough for medium response compounds but still not high enough to saturate the detector even for strong response compounds. Retention factors of all compounds are summarised in the colour coded Table.2.3. A blue background indicates detection, at the concentration shown in Table.2.1, only in positive mode, a yellow background reflects detection only in negative mode, a green background indicates detection in both polarities and a transparent cell indicates no detection. V_o was calculated from column volume ($Vm = 0.5 * L * d_c^2$) where L is the column length (cm) and d_c is the column internal diameter (cm). A fixed volume was added to the calculated volume to compensate for the extra column volume. Vm was calculated to be 1.5 ml (t_0=5 minutes for the method used) for ZIC-HILIC and Luna-HILIC and 2.5 ml (t_0=8 minutes for the method used) for Obelisc N and Obelisc R.

		Retention factor							
No	Chemical	ZIC-HILIC		Luna-HILIC		Obelisc_R‡		Obelisc_N	
		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
1	Cis-Aconitic Acid		2.51		1.32		0.29		1.35
2	Acetylcholine Chloride	4.23		1.04		0.78			
3	Adenine	4.26	4.35		2.29			33.66	34.53
4	Adenosine Tri-Phosphate								
5	Alanylglycine	5.65	5.75		2.79	0.82			
6	gamma-Aminobutyric Acid	5.07		1.89		0.80		28.19	
7	Arabinose						0.11		
8	Arginine	6.98	7.07	3.52	3.59	1.27			
9	Ascorbic Acid*						0.53		4.988
10	Biopterin	3.77	3.70	2.38	2.65	1.39		4.38	4.454
11	Caffeic Acid	1.82	1.05		1.21		1.83		1.55
12	Cinnamic Acid								
13	(p-)Coumaric Acid	0.98	0.91		1.16		2.59		1.266
14	Creatinine	5.11	5.30	1.38		0.81		35.24	
15	Cysteine								
16	Cystine								
17	Deoxyglucose						0.16		26.41
18	Dihydrobiopterin	4.06	3.97	3.09	3.07				78.67
19	(1,7)Dimethylxanthine	1.07		1.40		1.18		1.56	
20	Galactose		4.43		2.77		0.13		
21	D(+) Galacturonic Acid		4.80		2.74		1.86		5.232
22	alpha D (+) Glucose		4.43		2.71		0.14		2.458
23	(D) Glucose-6-Phosphate	5.81	5.60		4.69				1.288
24	D(+) Glucosamine	6.46				0.52		29.31	
25	Glutamic Acid	5.18	5.23	2.89	3.01	1.24		4.12	4.328
26	Glutathione (Oxidised)								
27	Glutathione (reduced)	5.96	5.94		4.43				
28	Glycine (D5)	5.58		2.97		1.00			
29	Guanine	4.02	3.98	2.84				3.90	
30	Fructose		4.04		2.46		0.14		2.322
31	L(-)Fucose		3.17						1.954
32	DOPA (Dioxyphenylalanine)	5.33	5.36	2.76	3.01	0.83		10.40	
33	Epinephrine BiTartarate	5.49	5.70	1.80					
34	Histamine	7.82		3.02					
35	L(-)Histidine	6.83	6.88	3.53	3.59				71.39
36	(5) HydroxyDopamine								
37	alpha-HydroxyIsobutyric Acid		1.16		1.21				2.336
38	DL(α)Hydroxyisovaleric Acid		0.94		1.22				2.394
39	Hydroxy(β)methyl(β)glutaric Acid		1.08		1.23				4.1

		Retention factor							
No	Chemical	ZIC-HILIC		Luna-HILIC		Obelisc_R‡		Obelisc_N	
		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
40	(3)HydroxyTyramine	5.24		1.82		0.62			
41	Hypoxanthine	3.05	2.92	2.19	2.18	1.18		3.08	3.13
42	Inosine	3.42	3.35	2.46	2.46			2.41	2.43
43	Isocitric Acid (trisodium salt)		4.60		1.47				3.78
44	(DL) IsoLeucine	4.13		2.32		0.85		11.32	
45	(α)Ketoglutaric Acid (Disodium)		3.59		1.91				53
46	Kynurenic Acid	2.44	2.32	2.58	2.67				
47	(DL)Kynurenine	3.87	3.90	2.44	2.61	0.67		16.43	
48	(D) Lactic Acid (sodium salt)				1.25				2.738
49	(L) Lysine	6.98		3.64					
50	Maleic Acid		1.06		2.17				27.76
51	Maltopentaose	5.65							
52	Maltotetraose								
53	Mannose		4.06		2.65		0.12		2.382
54	Methionine	4.28	4.13	2.57	2.71	0.73		10.49	
55	Methyl(α)carboxyl(4)phenylglycine	4.33		2.78					
56	Methyl(1)histamine			2.52					
57	(L) Methyl(3)histidine	6.91	6.97	3.46	3.57				
58	Methyl(6)uracil	1.91						1.48	
59	Methyl(1)xanthine	1.92	1.12	1.73	1.37			1.84	1.74
60	N-AcetylHistamine	4.96	5.10	1.76	1.95	0.87		11.11	16.18
61	Nicotinamide Adenine Dinucleotide (NAD)	5.78	5.50	4.94	4.91			10.88	10.82
62	(B) Nicotinamide adenine dinucleotide Phosphate (NADP)								
63	(DL) NorAdrenaline (HCl)		6.26	2.23	2.55				
64	Octopamine	5.09		1.86					
65	Oxaloacetic Acid								
66	Oxalic Acid								2.746
67	Phenylalanine	3.80	3.79	2.58	2.56	0.80		10.72	10.78
69	Pyruvic Acid		1.60		1.36				23.27†
70	Rhamnose		3.17		2.04		0.12		3.788
71	Riboflavin	2.67		2.19				2.04	
72	Serine	5.57	5.60	3.57	3.64	0.87		9.85	9.996
73	Serotonin (HCl)	4.91		1.80		0.48			
74	Spermidine (Tri-HCl)								
75	Stachyose Hydrate		5.61		4.31				
76	Succinic Acid		1.09		1.23				2.108
77	Sucrose	4.86	4.79	3.62	3.50		0.06		

		Retention factor							
No	Chemical	ZIC-HILIC		Luna-HILIC		Obelisc_R‡		Obelisc_N	
		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
78	Taurocholate Hydrate (sodium)	2.54	2.38	2.96	2.93				
79	Thiamine (HCl)	7.60		1.27		2.72			
80	Tryptophan	4.06	4.07	2.39	2.64			11.24	
81	Tyrosine	4.59	4.64	2.69	2.86			4.51	4.738
82	Uric Acid		3.89		2.61				2.552
83	(DL) Valine	4.53		2.53		0.84			
84	Xanthine	2.91	2.81	3.55					2.036
85	(D(+)) Xylose		3.99		2.36		0.13		2.106
86	Alanine	5.33		2.76		0.84		11.00	
87	Asparagine	5.56	5.54	3.44	3.68	0.96		9.69	9.882
88	Aspartic Acid	5.36	5.36	3.44	3.16	2.69			9.626
89	Betaine	4.99		2.63		1.51			
90	Carbamylcholine (Chloride)	5.05		1.16		0.54		47.22	
91	Choline (base)	5.33		1.38		0.47		46.99	
92	Creatine	5.12	5.21	2.25	2.65	0.93		17.40	
94	Dimethyl(3,3)glutaric Acid	0.99	0.92		1.19				3.68
95	Dimethyl(N,N)glycine (HCl)	5.10		2.69		1.55			
96	Ferulic Acid	0.94	0.88		1.09				1.294
97	Fumaric Acid		1.05		1.21				8.426
98	Glycylproline					0.91			
99	Glutaric Acid		1.05		1.21		1.93		1.494
100	Glycine Methyl Ester (HCl)	4.98				0.46		34.32	
101	Heptadecanoic Acid								
102	Hydroxy(6)nicotinic Acid	2.01	1.15	1.70	1.38			2.69	2.832
103	Inositol (myo)		5.31				0.07		2.838
104	Lactose	5.18		3.81			0.05	2.83	
105	Leucine	3.97		2.33		0.91		11.06	
106	Levulinic Acid		0.94		1.21		0.92		1.296
107	Malonic Acid								11.08
108	Methyl(N)thioUrea	1.82		1.52				1.35	
109	Methylurea	2.40		1.67		1.27		1.98	
110	Nicotinamide	2.16		1.62		0.90		9.73	
111	Palmitic Acid								
112	Phenyl(Alpha)glycine	4.19	4.07	2.78	2.79	1.50			
113	Phosphorylcholine Chloride	6.96						4.36	
114	Pyridoxine	4.43	4.62	1.38	1.59	0.50		32.62	
115	Ribose		3.34				0.16		2.24
116	Spermine								
117	Tartaric Acid		5.55		2.02				

No	Chemical	Retention factor							
		ZIC-HILIC		Luna-HILIC		Obelisc_R‡		Obelisc_N	
		Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
118	Thymine	1.82	1.16		1.32				3.724
119	Trehalose (Dihydrate)						0.05		2.89
120	Uracil	2.16	1.65	1.58	1.36				3.78
121	Urea					1.32			
124	Sphingomyelin								
125	Glycerophosphocholine	5.97		3.84		1.84		3.81	
126	S-Adenosylmethionine	7.38		3.74					
127	O-acetyl-N-carnitine (HCl)	4.09		1.07		0.78		33.39	
128	(L) Citrulline	5.65	5.65		3.78	1.48		10.88	10.94
129	Spermine								
130	Spermidine (Tri-HCl)	8.47							
131	(+-) Carnitine (HCl)	5.06		1.38		0.84		16.72	
132	HomoCysteine	6.03	4.66	2.71					
133	(D) Pantothenic acid (Calcium Salt)	1.79	1.06	1.40	1.24	1.68		1.82	1.848
134	Thymidine	1.92	1.11		1.51				1.65
135	Guanidine								
136	(DL) Ornithine	7.00		3.69		1.87			
137	UDP-Glucuronate								
138	Proline	4.93		2.75		1.70		5.36	
139	Adenosine	3.63	3.64	2.35	2.31	1.05		16.41	
140	Taurine	5.14	5.07	3.15	3.14	1.21		2.59	2.61
141	(L) Glutamine	5.45	5.45	3.57	3.66	1.25		9.90	10.15
142	Guanosine	4.02	3.95	2.84	2.66	1.26		3.89	3.94

Table.2.3 Retention factors for standard metabolites on ZIC-HILIC, Luna-HILIC, Obelisc R and Obelisc N columns.

detected only in Negative	detected only in Positive	detected in both polarities	NOT detected

* stability rather than detection of used ascorbic acid maybe the cause of its absence.
† the peak is very weak and distorted so this maybe an outlier.
‡ Obelisc R negative mode RT data were obtained from a higher concentration run on LCQ deca.

Clear selectivity differences were noticed across the four HILIC columns. For example, sugars were well retained by the ZIC-HILIC and Luna-HILIC columns, with the order of pentoses ➔ hexoses then the disaccharide lactose (same order of the negative LogP for the groups). On Obelisc_R, all sugars eluted in the void volume (k`<0.2) and in the case of Obelisc_N in a region of the chromatogram later than the void but still early compared to the elution of other components. To make an

easier visual judgment figure 2.7 shows ranking, rather than retention factor, of the order of elution for the 45 compounds, commonly observed in all columns, and the sugars were pulled aside and coloured by subgroup as indicated by the legend of Figure.2.7.A.

Another example of selectivity is for organic acids, where six organic acids commonly detected in the four stationary phases have been moved to one side to facilitate observing their elution order. Acids eluted early from all but the Obelisc R stationary phases with the k` values correlated to negative *logP* values for Obelisc R (r=0.9) and inversely correlated to negative *logP* values for Luna-HILIC (r =-0.85). There was no correlation between *logP* and k` for Obelisc N and ZIC-HILIC (r=-0.1 and r=-0.3, respectively).Figure.2.7.B

The different selectivity information can be utilised when selecting the right column for pharmaceutical applications, for example, developing a pharmaceutical related substances test for polar compounds may use an Obelisc_R column for LC analysis when organic acids are among the related substances.

However, neither a mechanistic nor an extensive retention behaviour study is the subject of this work, and the rest of this chapter will focus only on the suitability and stationary phase selection for *Drosophila* metabolome analysis.

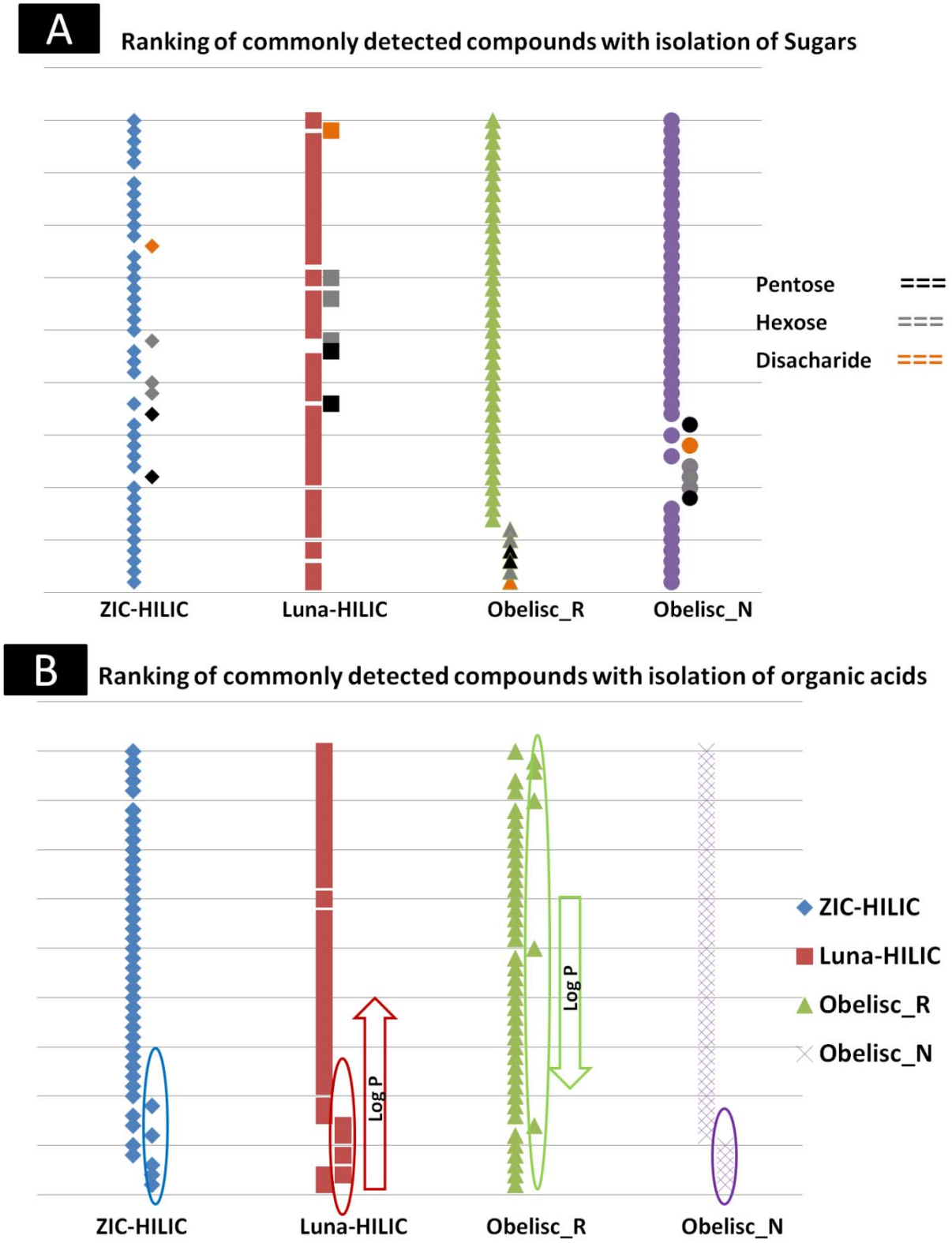


Figure.2.7 the rank of the retention factors of commonly detected metabolites in the four HILIC stationary phases. (A) Focusing on sugars, (B) focusing on organic acids. The plots were derived from data in table.2.2.

2.3.3.2. Peak parameters:

While a resolution value of 1.5 between chromatographic peaks is generally accepted in LC-UV for pharmaceutical purposes, no resolution in the time domain is required when the MS detector is used, provided ion suppression (matrix effect) or isobaric compounds are guaranteed not to affect the outcome. However, untargeted metabolic profiling enforces different criteria due to the complex sample content. The range of metabolites, chemical classes, polarity, concentrations and relative responses are all unknown. The general approach in MS bioanalysis is to follow the Bonfiglio ion suppression test (Bonfiglio et al., 1999), but again aiming at the qualitative analysis of all components of one sample may reduce the suitability of this test for metabolic profiling purposes. Instead, here a set of mathematical modules is proposed, based on concatenating data from individual injections of a representative subset of the metabolome for use in comparing the preference of one stationary phase over another.

On the other hand, the selectivity of the chromatography should provide sufficient resolution between components of the metabolic profiling samples. Thus, judging the suitability of chromatography for a metabolic profiling experiment can be viewed in terms of width, height and resolution.

Method comparison in the literature has employed the number of reproducible features for each stationary phase when analyzing the same samples. Although this approach makes a very good initial screen, it does not take into consideration the number of background features, the number of lipid derived features and the number of small molecular weight metabolites (< 500 amu).

Here, the suitability of stationary phases for *Drosophila* metabolic profiling has been compared in terms of peak width, peak height, resolving power, dynamic range of retention and background components using the authentic standards for 126 (16 metabolites were excluded for undetectability reasons) metabolites.

2.3.3.2.1. Peak widths and heights :

Good chromatographic peak shape will reduce the time range within each chromatogram occupied by a peak (peak capacity and peak efficiency), hence, reducing the potential for co-eluting isobaric peaks and ion suppression. The average peak width (at base) of standards run on the four columns is compared in Figure.2.8. The little number over each bar is the number of compounds which were included in the calculation for each column, as all the metabolites were listed in table.2.3 but not all of these metabolites were detected on each column.

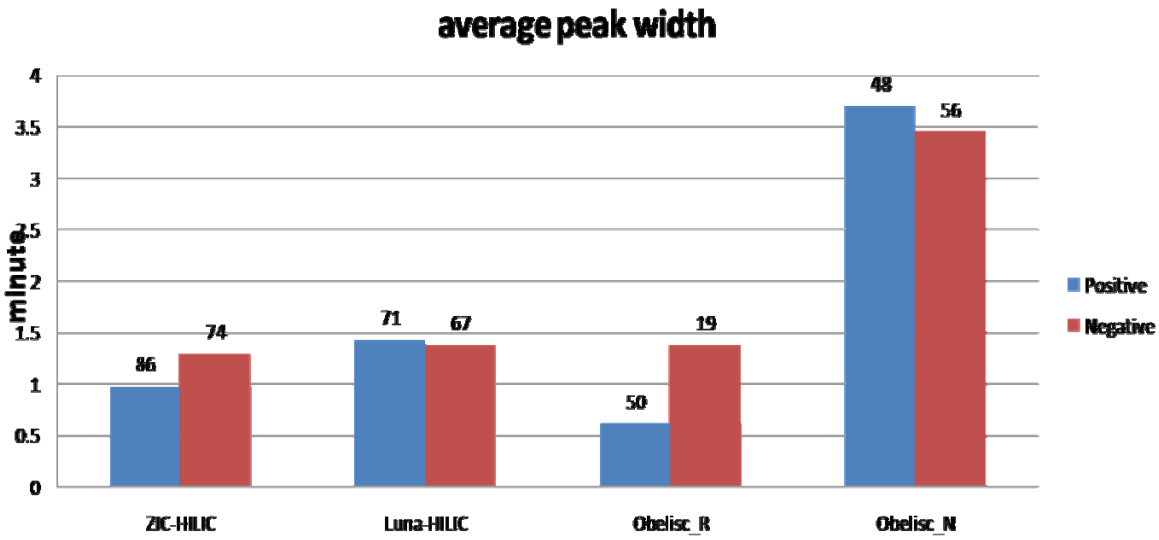


Figure.2.8 Average peak width of the detected metabolites for each stationary phase. The label indicates the number of detected metabolite at each polarity.

Obelisc R apparently gives the best combination of peak widths in negative and positive modes. However, the peak width in this work cannot be viewed separately from the peak height concept, as when the peak height is so low, the MS will detect only the apex of the peak, because of the trap cut off for low intensity ions, giving an apparent excellent peak shape, far better than its reality (Figure.2.9). This phenomenon does not happen in UV detection as the UV dynamic range (order of E3) is very low compared to that of MS (order of E5).

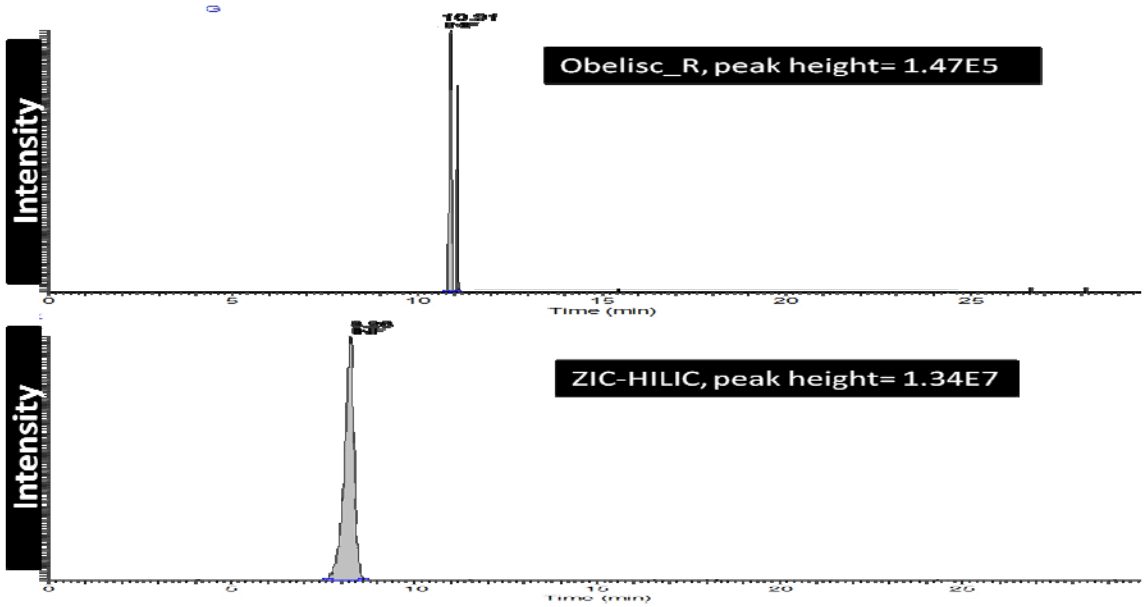


Figure.2.9 Effect of peak height on peak width in mass spectrometry.

ZIC-HILIC gave the best peak shape among the four columns, especially when considering the average peak height for a subset of the analysed compounds Figure.2.10.

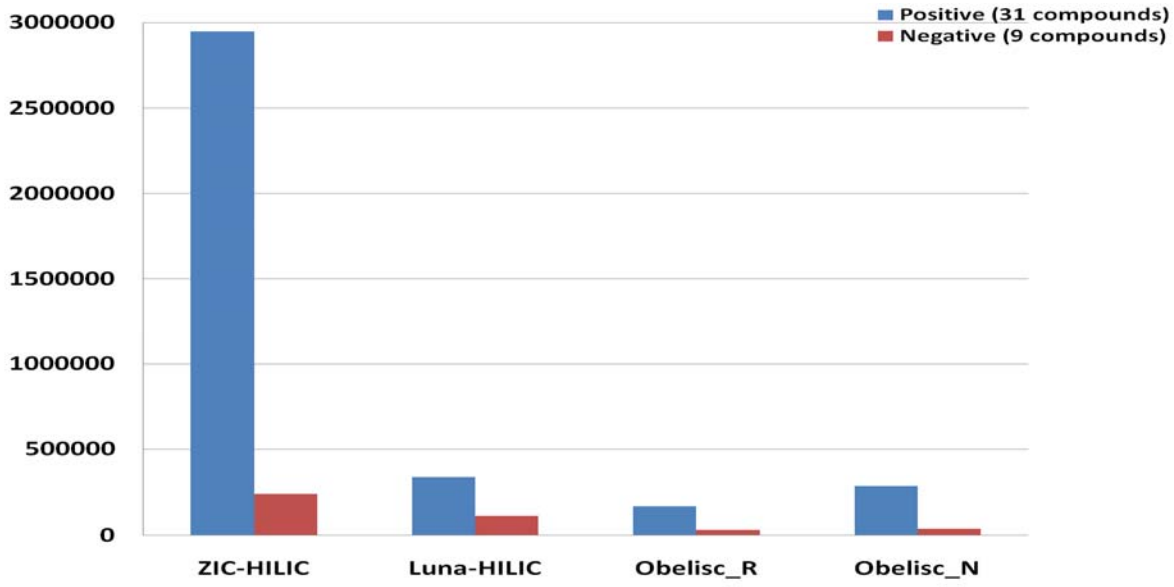


Figure.2.10 Average peak height for the commonly detected metabolites in each polarity.

ZIC-HILIC gave much better average peak width and height for metabolites detected in positive ion mode in comparison with negative ion mode. This can be understood with regards to the retention mechanism where a negative surface charge on the ZIC-HILIC column will interact (by weak ion exchange) with positively charged molecules, which are observed in positive ion mode. This leads to better retention and better peak shape, in comparison with organic acids which will suffer a negative charge repulsion effect leading to early elution (see Figure.2.7.B) and broader peaks especially for acids with multiple carboxylic functions and multiple pKa values lying on both sides of the mobile phase additive Formic Acid. The Luna-HILIC column was intermediate in peak width, number of detected compounds and peak height.

Another observation is the unacceptable peak width for Obelisc_N. This is due to the long retention times on this column (mean k` = 11.4) which allows plenty of time for the longitudinal diffusion effect. Using a shorter column, although reducing theoretical plates, may give better peak shapes and will be discussed later in this chapter.

2.3.3.2.2. Dynamic Retention Range and Interference Index.

It is impossible to ensure a baseline chromatographic resolution for all metabolites in a sample, given the number of these and also because authentic standards are not available for each and every metabolite. Thus, the metabolic profiling requirement from a chromatographic column is to ensure as much separation as possible in a convenient time frame.

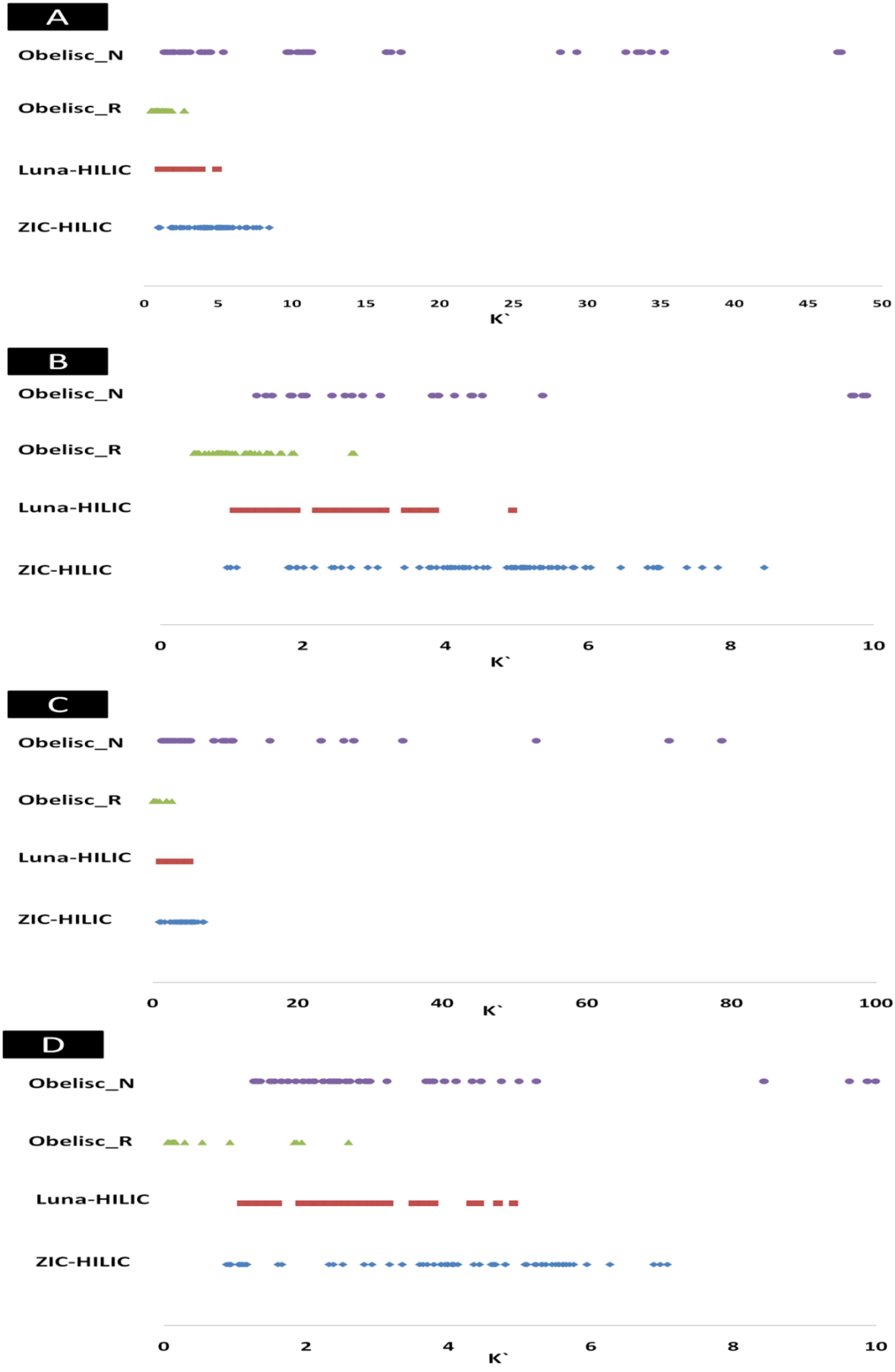


Figure.2.11 Dynamic retention range in the four HILIC SPs. (A) positive mode, , (B) positive mode retention factor(0-10), (C) negative mode, (D) negative mode retention factor(0-10).

For the very last point, we may identify a new term the “dynamic retention range” which is the time range encompassing eluting all metabolites of interest. The best value for this parameter is neither the highest nor the lowest; instead it could be the apex of a Gaussian peak which represents a value where maximum resolution is achieved within a frame of one analytical run time. The dynamic retention range for the four columns involved in this work may be visualized by examining Figure.2.11. In Panels (B, D) the scale was zoomed on the retention factor range of 0-10. Panels A and C show how Obelisc N excessively retains some metabolites with some compounds not eluting from the column at all (e.g. arginine, histidine and lysine). The other Sielc column, Obelisc R, did not give a very strong retention of the standards. Luna-HILIC and ZIC-HILIC have both good dynamic retention range, with a preference for ZIC-HILIC especially in the positive ion mode.

In order to assess which of the four HILIC columns involved in this experiment satisfies the metabolic profiling resolution requirements, the peaks obtained by individual injections were used to construct a pseudo chromatogram for each column at each polarity. The pseudo chromatogram basically records at each time point (intervals of 0.1 min) the number of co-eluting standards.Figure.2.12.

The pseudo chromatograms can be viewed on two axes. The horizontal (x) axis (time domain) shows how spread metabolites in each stationary phase are, in a similar fashion to Figure.2.11. The vertical (y) axis (pseudo intensity domain) indicates the number of co-eluting peaks at each time point. Since the compounds will elute along with their isotopes, spontaneous fragments and adducts then the higher the number of co-eluting metabolites the more the chances are of isobaric interference.

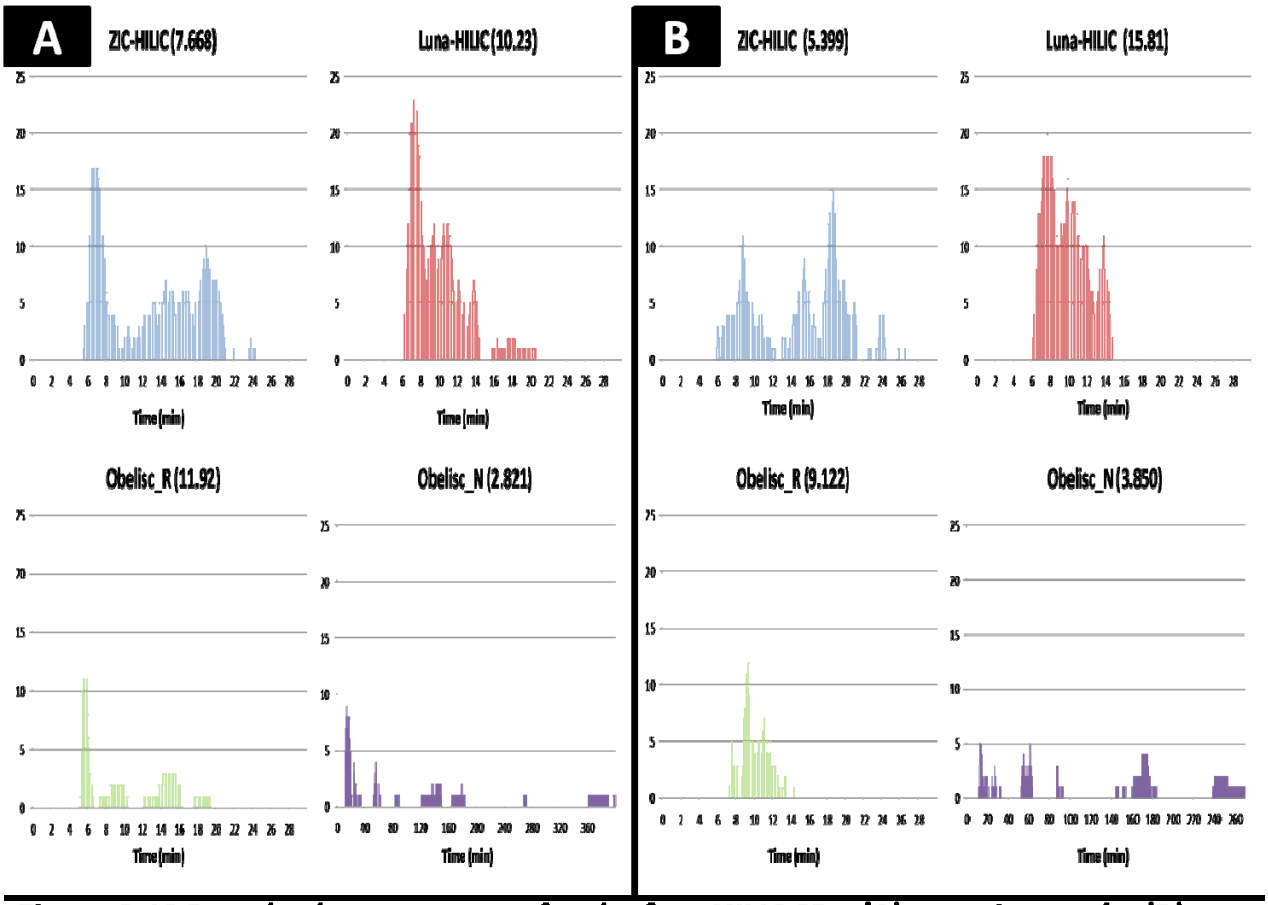


Figure.2.12 Pseudo chromatograms for the four HILIC SPs. (A) negative mode, (B) positive mode. The number between brackets is the interference index.

The value in brackets is the interference index (*G*, best low), a numerical value which enables a quick comparison of the different stationary phases. The value is derived by equation (1)

$$G = \frac{100 * F}{A * N} \qquad (1)$$

Where:

F: total frequency of co-elution points.

A: number of interfering points in the pseudo chromatogram

N: number of detected metabolites.

The part of equation (1) is optimal at a low values where the minimum number of interferences happen, and is spread all over the run time, resembling a low and flat shape for the pseudo chromatogram.

The number of detected metabolites *N* is included in calculation because of the differences in number of detected metabolites for each column. Each additional metabolite will have a probability to interfere with any of the other *N* metabolites. This effect was studied taking ZIC-HILIC in positive mode (highest *N* value) as a test case. Removing a random peak at a time and monitoring the value {*F/A*} gave a linear curve with r=0.988212 (Figure.2.13) concluding that a simple division by the value *N* is sufficient to compensate for the number of metabolites included in each analytical method.

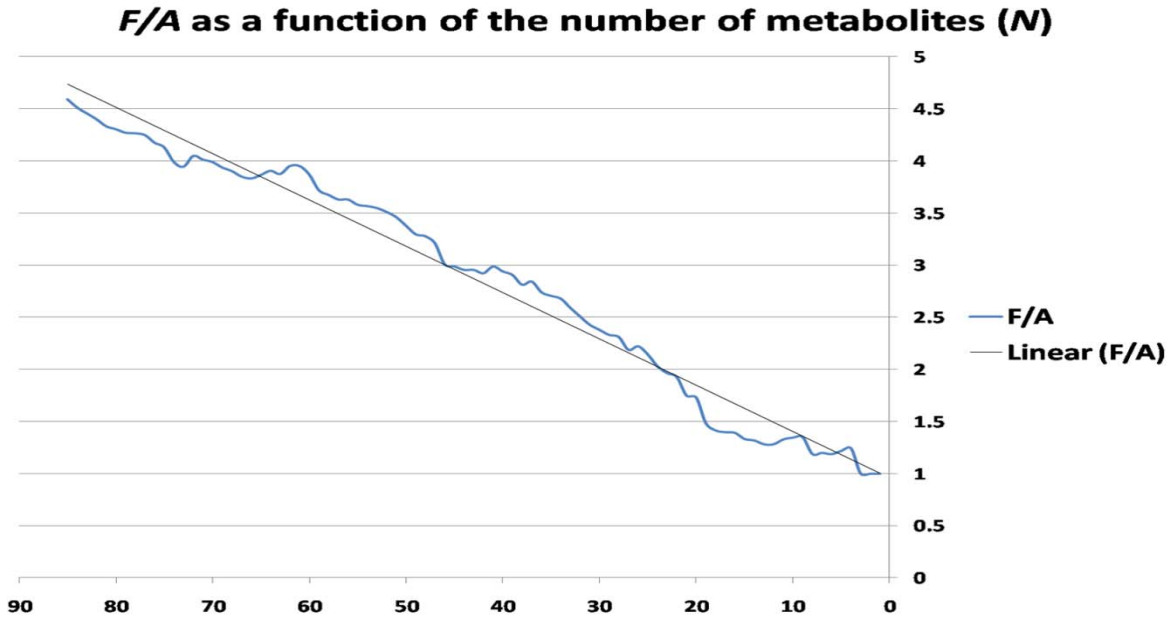


Figure.2.13 Plotting the {F/A} part against number of metabolites. A linear relationship is observed (r=0.988).

Luna-HILIC and Obelisc R gave high interference indices in both polarities, a result visually observed earlier in the compact retention charts for these columns. ZIC-HILIC showed good and intermediate interference indices in positive and negative polarities, respectively. The poorer performance for negative polarity compounds was discussed above. Obelisc N again seems to give the lowest level of interfering compounds. This comes at no surprise as only 47 and 56 compounds were

detected and were spread over 260 and 400 minutes in positive and negative polarities, respectively. The last observation may infer superiority for Obelisc N column type, however, the excessive retention of metabolites calls for very long run times (large dynamic retention range), which is inconvenient for metabolic profiling experiments from a high throughput and instrumental stability perspectives.

		ZIC-HILIC		Luna-HILIC		Obelisc_R		Obelisc_N	
No	**Chemical**	**RT**	**BW**	**RT**	**BW**	**RT**	**BW**	**RT**	**BW**
86	Alanine	18.99	1.46	11.27	0.61	9.2	1.02	59.99	2.13
100	Glycine methyl ester (HCl)	17.93	1.09	ND	ND	7.32	0.46	176.58	10.07
6	gamma-aminobutyric Acid	18.22	1.25	8.68	0.88	9.01	0.76	145.97	3.9
95	Dimethyl(N,N)glycine (HCl)	18.3	2.07	11.06	1.29	12.74	0.7	ND	ND
50	Maleic Acid	6.17	2.95	9.52	1.76	ND	ND	143.79	8.03
97	Fumaric Acid	6.15	1.39	6.64	1.74	ND	ND	47.13	ND
83	(DL) Valine	16.59	1.39	10.6	1.14	9.19	2.64	ND	ND
89	Betaine	17.98	1.33	10.88	3.98	12.53	0.47	ND	ND
58	Methyl(6)uracil	8.72	0.6	ND	ND	ND	ND	12.4	1.13
118	Thymine	8.46	0.58	6.95	0.89	ND	ND	23.62	0.86
44	(DL) Isoleucine	15.4	1.35	9.95	1.43	9.25	2.46	61.62	2.68
105	Leucine	14.91	1.47	10	1.52	9.56	2.31	60.29	2.57
5	Alanylglycine	20.26	0.56	11.38	0.9	9.09	0.07	ND	ND
141	(L) Glutamine	19.35	1.1	13.97	1.72	11.27	0.97	55.77	1.87
7	Arabinose	ND	ND	ND	ND	5.53	0.57	ND	ND
85	(D(+)) Xylose	14.97	1.75	10.08	1.23	5.67	0.92	15.53	1.23
115	Ribose	13.03	0.72	ND	ND	5.78	1.24	16.2	1.03
40	(3)Hydroxytyramine	18.73	0.5	8.46	2.17	8.08	0.01	ND	ND
64	Octopamine	18.26	0.37	8.57	1.18	ND	ND	ND	ND
17	Deoxyglucose	ND	ND	ND	ND	5.79	0.5	137.04	1.34
31	L(-)Fucose	12.5	1.43	ND	ND	ND	ND	14.77	1.41
63	(DL) Noradrenaline (HCl)	21.79	0.46	10.64	1.08	ND	ND	ND	ND
114	Pyridoxine	16.86	1.23	7.76	2.44	7.48	0.99	168.11	17.97
20	Galactose	16.29	0.46	11.31	1.56	5.63	0.61	ND	ND
22	alpha D (+) Glucose	16.29	1.37	11.13	1.7	5.68	1.03	17.29	1.9
30	Fructose	15.12	1.96	10.37	2.48	5.69	0.91	16.61	2.58
53	Mannose	15.17	0.16	10.96	1.14	5.58	0.69	16.91	1.31
103	Inositol (myo)	18.94	0.1	ND	ND	5.37	1	19.19	0.58
77	Sucrose	17.58	0.03	13.87	0.23	5.29	0.74	ND	ND
104	Lactose	18.55	0.88	14.42	1.42	5.23	0.61	19.16	0.42
119	Trehalose (Dihydrate)	ND	ND	ND	ND	5.23	0.89	19.45	1.06

Table.2.4 list of isomeric pairs (groups) with RTs and peak widths. ND=Not Detected

Using shorter columns, higher flow rate, using a more hydrophilic organic modifier such as MeOH or higher aqueous percentages in the gradient are all possible solutions for the strong retention of Obelisc N. However, loss of chromatographic efficiency, incompatibility with the resolution needed from the Orbitrap (the MS detector used in all subsequent work), loss of resolution of early eluting peaks and reduction of signal due to the higher aqueous content of the mobile phase are pitfalls associated with each of the above mentioned solutions.

2.3.3.2.3. Isomer separation:

As they have the same molecular formula, hence, the same exact mass, isomers are inseparable by MS regardless of the detector separation power. Yet, these isomers may have different biochemical roles, hence, identifying the abundance/variation of each isomer is necessary for a valid translation of a metabolic profiling experiment into observation of a biochemical event. The chromatographic challenge is not easy as demonstrated previously in Figure.2.2 and will be discussed further later in chapter 4. To test which of the four HILIC columns has the potential to separate isomers, 13 isomeric groups were included in the runs Table.2.4. These groups included structural and geometric but not enantiomeric isomerism.

Assessment was based on calculating the average resolution for isomeric pairs, concluding a separation index where the best chromatography for this purpose will have the highest value of this index. Figure.2.14 shows these indices for the four columns indicating the superiority of Obelisc N and inferiority of Luna-HILIC.

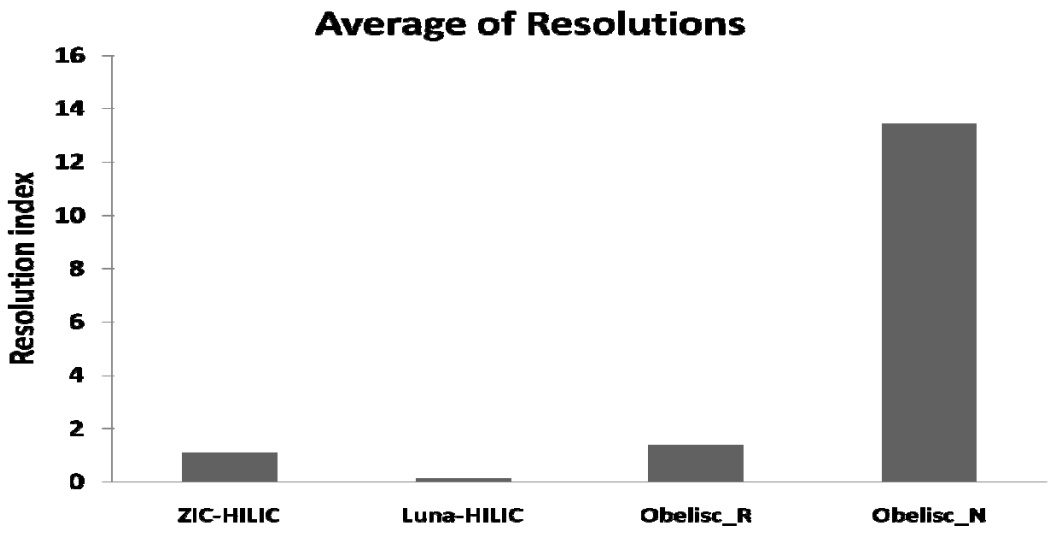


Figure.2.14 Average resolution of isomeric pairs for the four HILIC SPs.

According to the resolution equation (Kato et al., 2004) (equation.2) in chromatography, it is not surprising that Obelisc N with its strong retention of compounds will have such high resolution values.

$$R = \left(\frac{\sqrt{N}}{4}\right)\left(\frac{\alpha - 1}{\alpha}\right)\left(\frac{k` + 1}{k`}\right) \qquad (2)$$

However, the resolution of "1.5" provides a baseline separation, and the resolution value of "2.0" is desirable as it provides assurance that small tolerable variations will not affect the separation, yet without unnecessarily lengthening of run time.

2.3.3.2.4. Background level:

The dynamic range of Orbitrap is quite substantial and is in the order of 50,000. However, the ion accumulation in the C-trap controls the injection time consequently affecting this dynamic range as well as detection, quantification and mass measurement accuracy. Column bleed and high background levels from solvents may cause a space charge effect in the C-trap (or in the ion trap for LCQ) leading to loss of signal and/or deterioration of accuracy.

To test the background level for ions observed in the four columns, the sum of all ions extracted from the same blank sample , taken within 24 hours with no calibration in between, run in triplicate for each polarity (positive-blue and negative-red) was recorded and shown in Figure 2.15. The Obelisc R background is the highest in both polarities, which may explain higher limit of detection observed with this column and hence a lower number of detected metabolite standards.

There is still one more flaw in the Sielc columns. The liquid cell mechanism seems to require a long equilibrium time, the overlaid chromatograms of three identical *Drosophila* samples (blue) and three blanks (red) for Obelisc_R (upper panel) and ZIC-HILIC (lower panel) are shown in Figure.2.16. The figure indicates no unique chromatogram shape which will reflect the irreproducibility of retention time observed with this type of column compared to ZIC-HILIC for example.

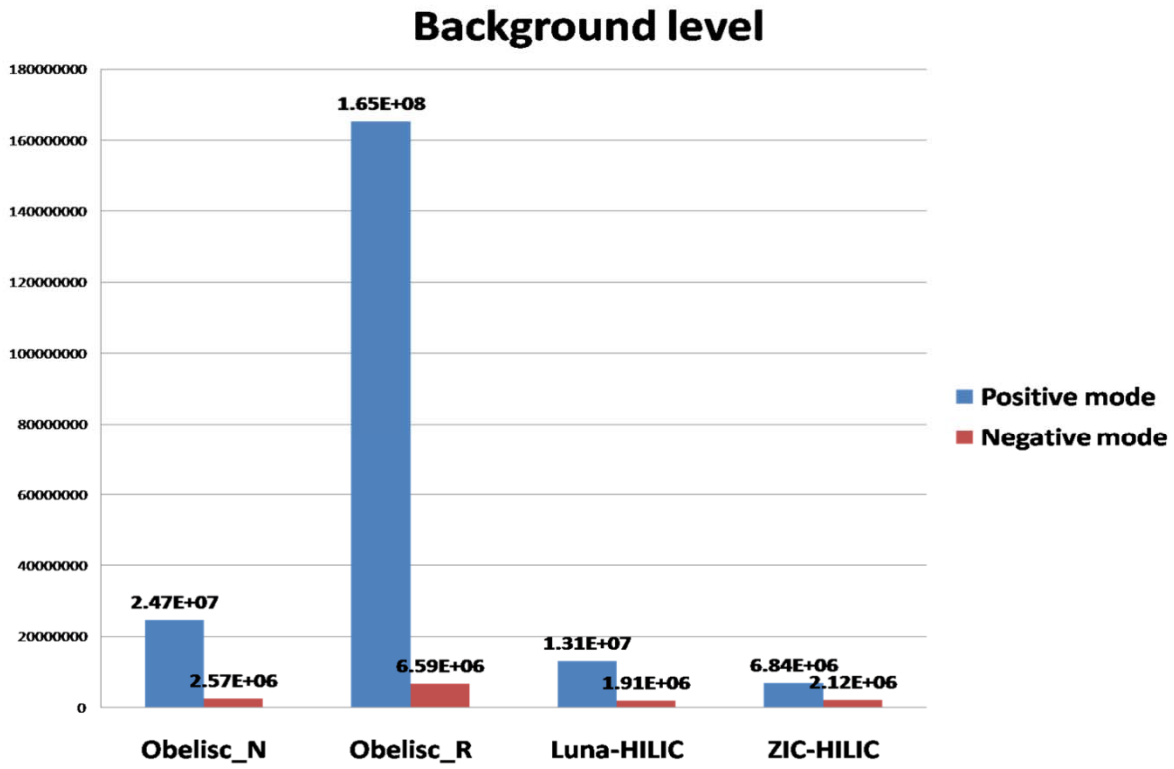


Figure.2.15 Background levels generated by the four HILIC SPs in positive and negative ion modes.

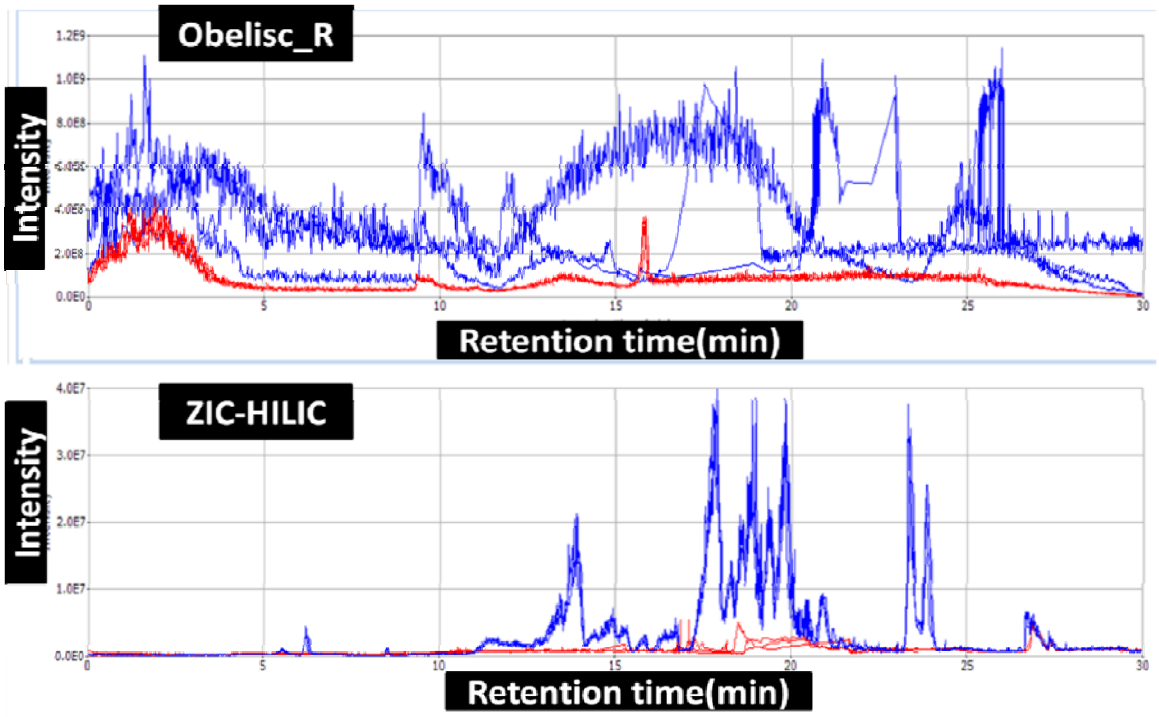


Figure.2.16 Blank (red) and sample (blue) injections TICs of Obelisc_R and ZIC-HILIC showing very long equilibrium need in Obelisc but not ZIC-HILIC.

The data above suggest both Sielc columns under these chromatographic conditions (considered optimal for our experiments) are less suitable for metabolic profiling purposes. ZIC-HILIC and Luna-HILIC seem to be a good compromise between retention, speed of analysis, efficiency, detection and resolution, although insufficient in the retention characteristics and isomer separation for negative polarity metabolites (organic acids and sugars). At this stage, a decision was made to pursue the work only with the latter two columns.

2.3.4. Impact of chromatography on the *Drosophila* Metabolic profiling experiments (ZIC-HILIC vs. Luna-HILIC).

2.3.4.1. Feature count:

The results of the test above suggest many similarities between Luna-HILIC and ZIC-HILIC columns, but as instrumental time as well as sample volume was limited, selection of one HILIC column was highly desired.

Two peak picking procedures were used, Sieve®.1.2(beta 5) a trial version of a commercially available software and VBA (visual basic application) subroutines, written in-house (MAK) to process a full peak list exported from X calibur (Qual-browser) the raw data viewer software. The pros and cons for each of the two procedures will be discussed in more detail in the next chapter, but, some necessary information are described below. Sieve® separates the aligned chromatograms into time windows and mass windows, then matches these according to the parameters specified by the user which are: time width, mass tolerance (amu) and detection threshold. MAK disregards the time domain and runs peak matching according to m/z value only, thus concatenating all isomers in the same bin.

The optimal time width for Sieve® can be extracted from average peak widthes of the authentic standards experiment which were 0.8 and 1.3 for ZIC-HILIC and Luna-HILIC in positive mode and 1.3 and 1.4 for ZIC-HILIC and Luna-HILIC in negative mode, respectively.

Nevertheless, because narrowing the time window will increase background and irrelevant peaks, leading to a partially fraudulent increase in feature number, the parameters for both columns were unified.

The number of reproducible features has been utilised in literature as a measure of the success of a certain method compared to others (Dunn et al., 2008, Michopoulos et al., 2009, Gika et al., 2008a, Kind et al., 2007). However, the mere feature count is less indicative when comparing different types of chromatography, as different background species and levels could be observed. In this work, the feature count was restricted with the filters of blank subtraction, reproducibility and signal/noise ratio. Three identical *Drosophila* samples were run on both columns in negative and positive modes (same analytical batch with no calibration in between.) The resulting data were extracted, filtered then surviving features counted and reported in Table.2.5. To pass the filter, a feature should be three times more abundant than blank level occurring at the same m/z and time and only m/z for Sieve® and MAK, respectively, and should turn up in all three *Drosophila* samples.

Furthermore, the filtered features extracted by both procedures were screened against KEGG database (11369 low molecular Weight MS compatible metabolites) recording the number of potential hits in the last column of the table. The mass tolerance for each forward search is included in the table as well.

The positive mode data indicated a five-fold increase in feature count of Luna-HILIC over ZIC-HILIC. However, this huge gain is only apparent, and is reduced dramatically to equal levels when subjected to the filter. The negative mode data suggested that ZIC-HILIC gave double the feature count for Luna-HILIC before and after filtration. Forward searching reports better hit counts for ZIC-HILIC over Luna-HILIC in both modes.

Column	polarity	Threshold	S/N	mass window	Peak width (min)	missing value	FWD searching max	FWD searching min	Original Feature count	Blank Filtered feature count	*Potential* Hits
Sieve®1.2											
ZIC-HILIC	+	100000	3	0.001 amu	1.5	0%	-5 ppm	+5 ppm	5830	1973	331
ZIC-HILIC	+	100000	3	0.001 amu	1	0%			6574	2055	
Luna-HILIC	+	100000	3	0.001 amu	1.5	0%	-5 ppm	+5 ppm	25000	2501	225
ZIC-HILIC	-	5000	3	0.001 amu	2.5	0%	-15 ppm	+5 ppm	1605	895	156
Luna-HILIC	-	5000	3	0.001 amu	2.5	0%	-15 ppm	+5 ppm	827	330	107
MAK											
ZIC-HILIC	+	NA	3	5 ppm	NA	0%	-5 ppm	+5 ppm	899	555	156
Luna-HILIC	+	NA	3	5 ppm	NA	0%	-5 ppm	+5 ppm	651	298	84
ZIC-HILIC	-	NA	3	8 ppm	NA	0%	-15 ppm	+5 ppm	1706	1392	304
Luna-HILIC	-	NA	3	8 ppm	NA	0%	-15 ppm	+5 ppm	1379	1023	285

Table2.5 Number of features obtained by analysing identical *Drosophila* samples using ZIC-HILIC or Luna-HILIC. The numbers were computed for different peak widths in Sieve® to satisfy a good comparison free from SP peak width effect

2.3.4.2. Comparison of Signal intensity:

Enhanced signal intensity is crucial in metabolic profiling for several reasons. Better signals give less analytical variance (Crews et al., 2009, Plumb et al., 2009). Furthermore, as metabolic profiling aims at the detection of "all metabolites in a biological system" (Fiehn and Kind, 2007), it follows that enhanced signal will give lower limit of detection and a higher number of detected metabolites. *i.e.* better coverage of the metabolome. Isotopic patterns can be helpful in assigning an ID to a feature, especially under automation as it considerably reduces the number of potential formulae. Natural levels of isotopes lie in a range of 0.05 to 4 % for the common biological elements, and the enhanced signal may allow for the detection of the full isotopic pattern of a metabolite, giving better confidence in identification.

To test the effect of the HILIC stationary phase on signal intensity, the filtered peak lists obtained in the process used for feature counts were matched and the signals compared by a t-test. As different stationary phases will give different retention times for the same metabolite, removing the time domain, which is one of the disadvantages of the MAK type process, was deemed suitable for this purpose. Figure.2.17 is a plot of the results of this analysis. It shows only matched metabolites which showed significant difference in signal intensity ($p<0.005$). The x-axis represents the log of the ZIC-HILIC/Luna-HILIC fold changes and the y-axis the P value between the two groups (Panel A) and the m/z (Panel B). The right side of the positive mode panels are busier and indicate better enhancement of ionization and more features in ZIC-HILIC (298 vs. only 66 for Luna-HILIC). The negative ion mode chart leads to a similar conclusion with 758 features having significantly better signals in ZIC-HILIC and 423 in Luna-HILIC.

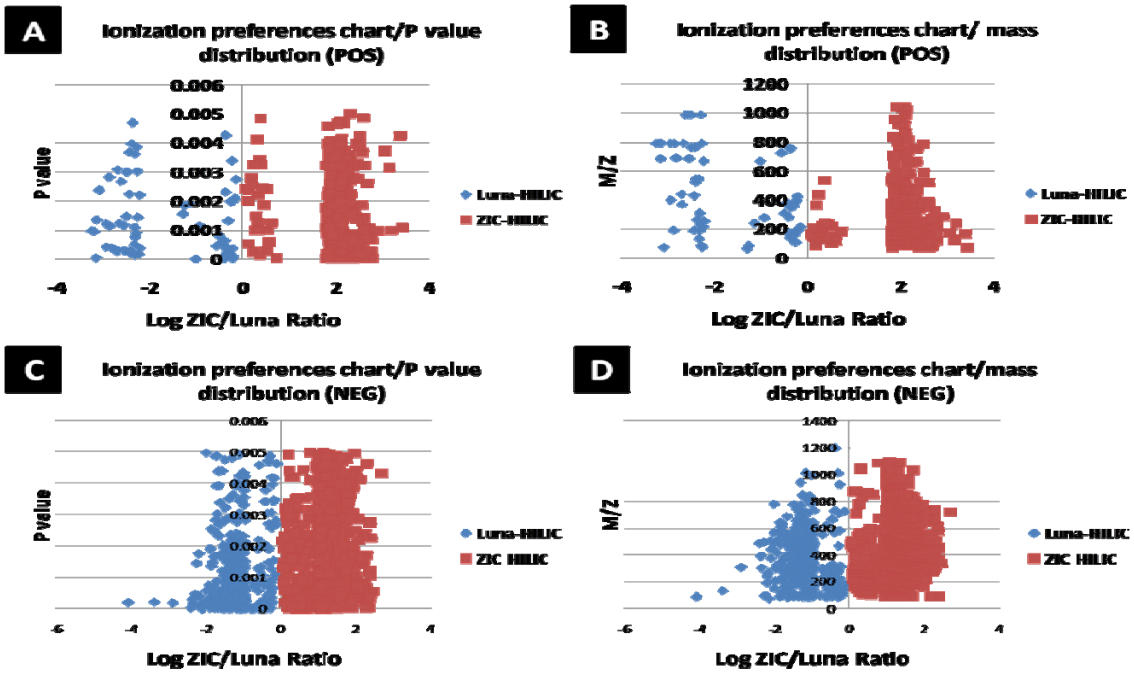


Figure.2.17 Comparison of signal intensities of features detected after ZIC-HILIC or Luna-HILIC elution in the two polarities. Panels (A) & (C) plot of log ratio against P.values of significant differences (P<0.005) and Panels (B) & (D) are plots of Log ratio against M/Z.

A second positive ion mode confirmation test was run to assure that the signal improvement was not a coincidence and showed similar results (553 for ZIC-HILIC vs. 188 for Luna-HILIC), Figure.2.18.

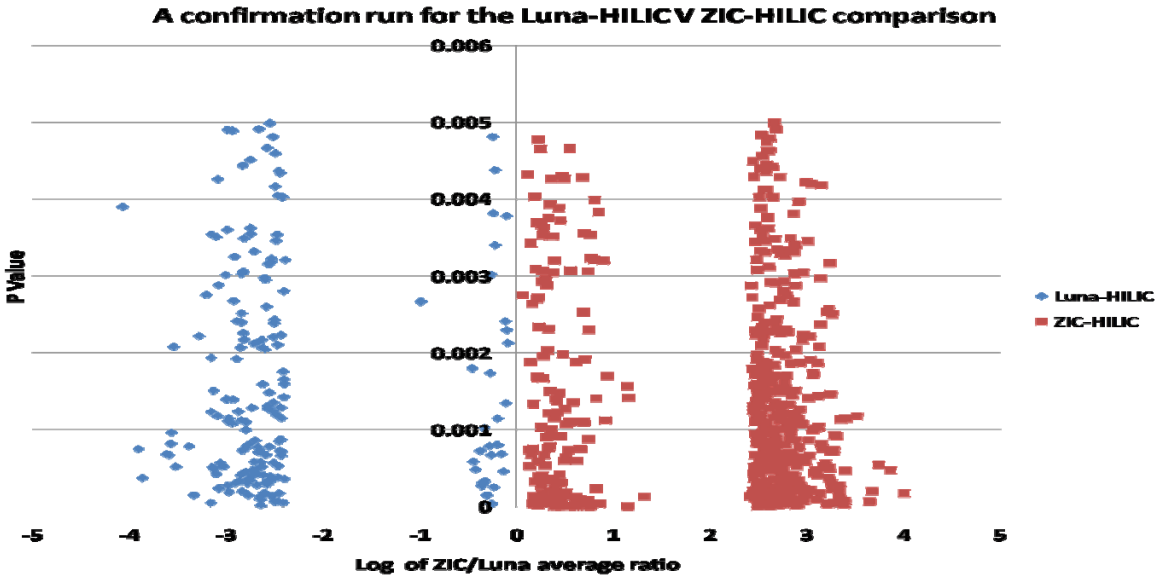


Figure.2.18 plotting the Log ratio (ZIC to Luna) against P.value of the difference between the detected features in ZIC-HILIC and Luna-HILIC.

Equation (1) under section 2.3.3.2.2 can be modified to include weights for peak height, number of extremely retained metabolites, background level, average limit of detection , number of reproducible features and dynamic retention range each weighted to suit the application and/or the effect of the function on the quality of data (linear, quadratic....etc).

2.3.5. Selection of mobile phase additive:

The latest test strengthened the preceding findings pointing to the superiority of ZIC-HILIC for *Drosophila* metabolic profiling experiments over the other tested HILIC columns. However, negative ion detection metabolites did not meet the standards required in terms of peak widths, intensity, selectivity and isomer separation. In all the analysis so far 0.1% formic acid was the only additive and other additives were tested in order to observe any improvement on the peak parameters of negative ion metabolites (39 metabolites listed in Table.2.6) and to observe the effect on retention profile for positive ion mode metabolites as well.

0.1% Formic Acid in water: 0.1% Formic Acid in Acetonitrile [60:40] in isocratic mode was compared to 20 mM ammonium formate in water (pH 7.6): Acetonitrile [60:40] and 0.1% Formic Acid in water: 0.1% Formic Acid in Acetonitrile [90:10]. The ionic strength of all the three solutions were similar (0.1% FA ~= 20 mM). The selected pH complied with the highest pH permitted for the ZIC-HILIC type column (SEQUANT, 2010). These tests were run on the LCQ-Deca ion trap MS.

None of the parameters measured were significantly improved by changing the mobile phase additive. The peak areas/peak heights were poorer for most of the compounds with the exception of coumaric acid, ferulic acid, lactic acid, pyruvic acid levulinic acid, malonic acid and glutaric acid Figure2.19 (A). The first five acids are all oxy (hydroxy- or keto-) mono-carboxylic acids while malonic and glutaric are symmetric di-carboxylic acids. The signal enhancement seems to relate to the ionization process rather than peak shape as only glutaric acid among all the acids showed a significantly narrower peak width. Peak width at higher pH was significantly improved for sugars (probably due to more rapidly equilibration

between the α- and β-forms), but it was poor for many of the organic acids Figure.2.19 (B).

Standard	Formula	MW (amu)
Aconitic Acid	C6H6O6	174.01589
Arabinose	C5H10O5	150.05227
Ascorbic Acid	C6H8O6	176.03154
Caffeic Acid	C9H8O4	180.04171
Cinnamic Acid	C9H8O2	148.05188
Coumaric Acid	C9H8O3	164.0468
Deoxyglucose	C6H12O5	164.06792
Galactose	C6H12O6	180.06284
Galacturonic Acid	C6H10O7	194.0421
Glucose	C6H12O6	180.06284
Fructose	C6H12O6	180.06284
Hydroxyisovaleric Acid(alpha)	C5H10O3	118.06245
Hydroxy(β)methyl(β)glutaric Acid	C6H10O5	162.05227
Isocitric Acid Trisodium Salt	C6H8O7	192.02645
Ketoglutaric Acid(alpha)	C5H6O5	146.02097
Lactic Acid (Sodium Salt)	C3H6O3	90.03115
Maleic Acid	C4H4O4	116.01041
Mannose	C6H12O6	180.06284
Oxaloacetic Acid	C4H4O5	132.00532
Oxalic Acid	C2H2O4	89.99476
Pyruvic Acid	C3H4O3	88.0155
Rhamnose	C6H12O5	164.06792
Succinic Acid	C4H6O4	118.02606
Sucrose	C12H22O11	342.11566
Taurocholate Hydrate	C26H45NO7S	515.29112
Xylose	C5H10O5	150.05227
Dimethyl(3,3)glutaric Acid	C7H12O4	160.07301
Ferulic Acid	C10H10O4	194.05736
Fumaric Acid	C4H4O4	116.01041
Glutaric Acid	C5H8O4	132.04171
Inositol(myo)	C6H12O6	180.06284
Lactose	C12H22O11	342.11566
Levulinic Acid	C5H8O3	116.0468
Malonic Acid	C3H4O4	104.01041
Palmitic Acid	C16H32O2	256.23968
Ribose	C5H10O5	150.05227
Tartaric Acid	C4H6O6	150.01589
Trehalose dihydrate	C12H22O11	342.11566

Table.2.6 Negative ion metabolites analysed for comparison of formic acid to ammonium formate buffer as additives.

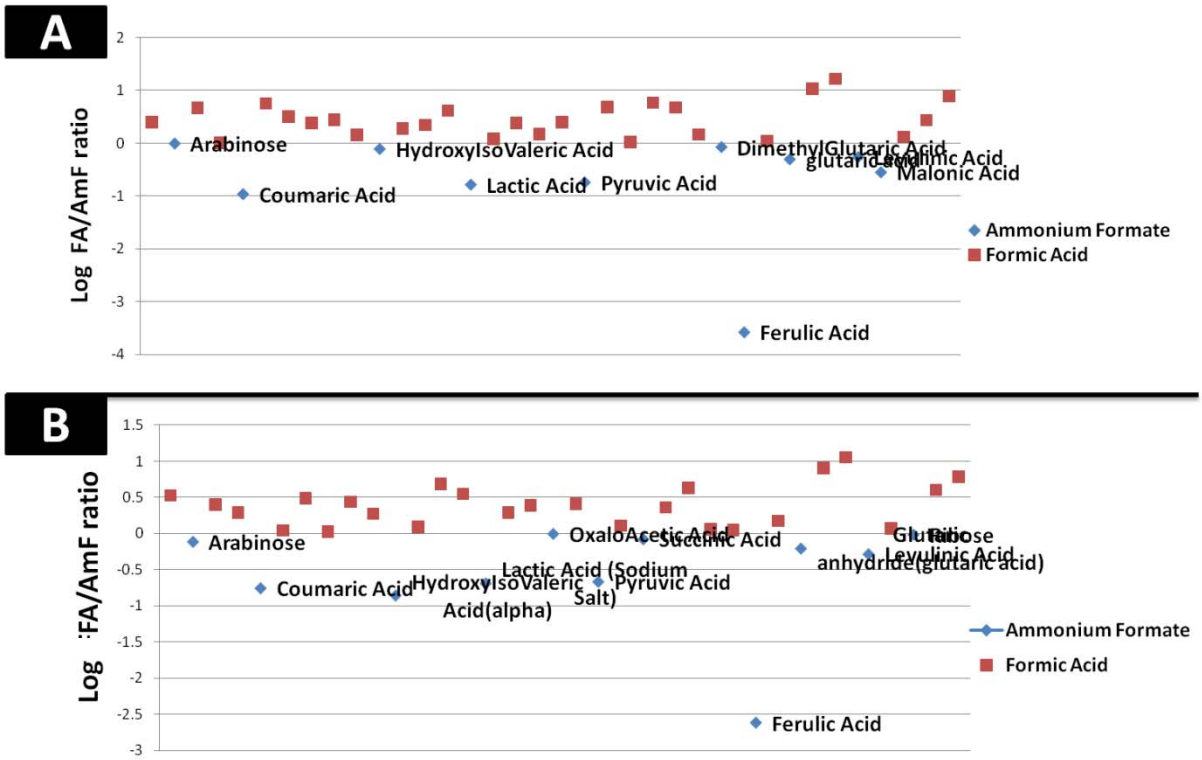


Figure.2.19 Effect of mobile phase additive on peak area (A) and peak height (B) of negative ion metabolites in Formic Acid (FA) or Ammonium Formate (Amf). The Y axis is the Log FA/Amf ratio.

In terms of retention and selectivity, the compounds run in this test can be divided into two major categories, sugars and organic acids. The change in pH neither improved the weak retention capacity nor changed the selectivity for the negative ion metabolites. Oxalate and malonate, the two small dicarboxylic acids, seem to benefit most from the high pH with regards to retention on column (Figure 2.20).Moreover, the addition of the buffer did not change the base peak which remained the formate adduct for most of these negative ion metabolites.

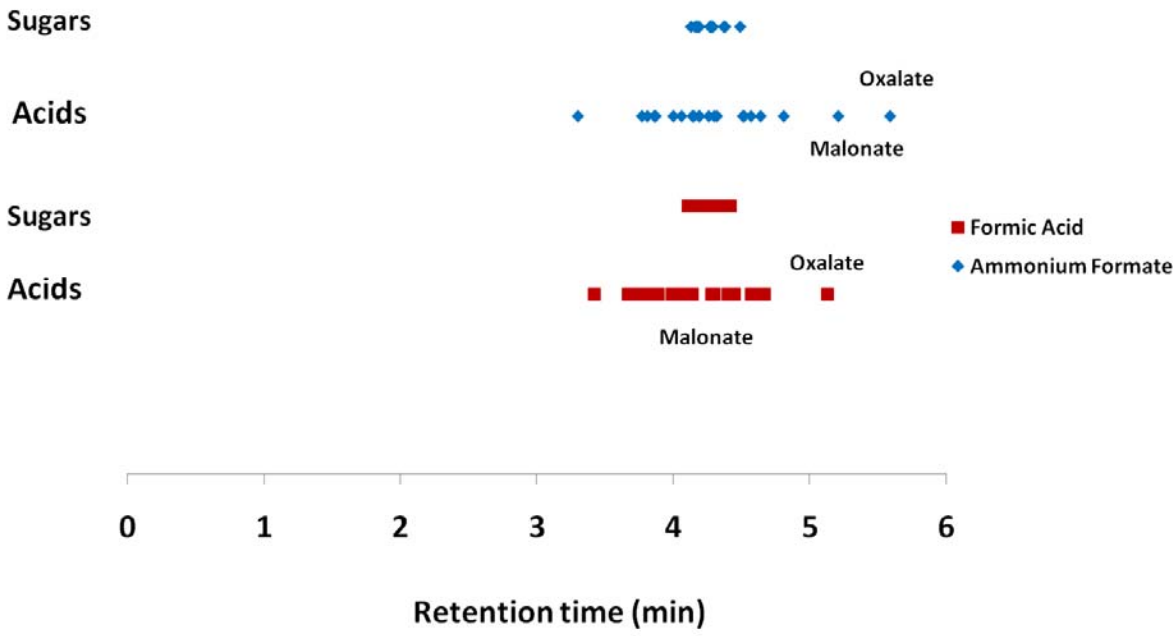


Figure.2.20 Effect of mobile phase additives on negative ion metabolite retention.

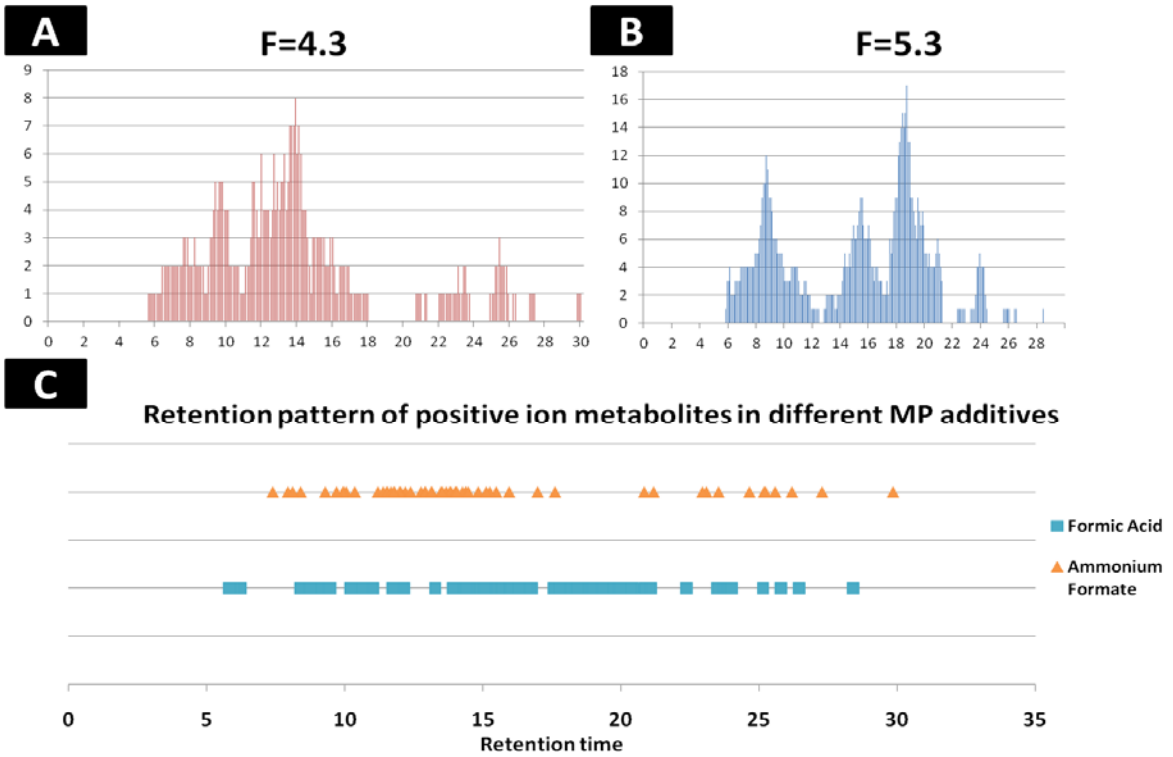


Figure.2.21 Effect of mobile phase additives on positive ion metabolites. Pseudochromatogams of AmF pH=7.8 (A) and FA (B) and retention pattern (C).

For the positive ion mode metabolites the addition of the ammonia buffer did not have a great impact on the general retention profile, although there were some changes in selectivity. The retention profiles of positive ion metabolites with formic acid or ammonium buffer along with the pseudochromatograms and interference indices are shown in Figure.2.21. This work was not oriented towards a mechanistic investigation of HILIC columns, but rather towards their suitability for analytical purposes, especially analysis of metabolomes. In this regard, the addition of ammonium formate (AmF) buffer with the highest permitted pH did not significantly improve the outcome of the experiment.

2.3.6. Selection of organic modifier for ZIC-HILIC metabolic profiling:

The composition of the mobile phase affects the mobile phase/stationary phase interaction, leading to different elution patterns of metabolites. Acetonitrile is the most widely used organic modifier in HILIC mode because of its favourable characteristics (low viscosity, cheap, no smell, excellent miscibility with water and relatively low polarity). However, the acetonitrile shortage made it necessary to also consider alternative modifiers.

In order to assess whether methanol or ethanol could work as a replacement for the then hard to get Acetonitrile, three Drosophila samples from two pooled groups with different metabotypes (pooled_1= wild type and pooled_2 = a mixture of xanthine oxidase deficient genotypes of *Drosophila*) were run using the same HILIC compatible gradient, using the three modifiers. The runs were achieved on the same day with only column conditioning taking place between solvent switching.

The three methods were compared for the number of features attainable by each, represented by the number of reproducible features using the beta version of the commercial software Sieve®.1.2(beta-5) and the in-house VBA macro (MAK). The initial number of features observed was in the order of EtOH, Acetonitrile then MeOH. However, a simple filtration for reproducibility per group (CV<20%) reversed the order. The order of number of features obtained by MAK was consistent before and after CV filtration, however, the percentage increase in the number of features was significantly reduced. The number of potential metabolites based on

accurate m/z forward screening of the KEGG database also shows the favourability for using MeOH or ACN (Table.2.7).

	ACN	MeOH	EtOH
Sieve® initial	21196	13823	25000
Sieve® filtered	5987	6911	4834
MAK initial	1035	1950	552
MAK filtered	540	711	404
MAK-KEGG	165	184	105
average peak width	0.873	1.351	1.064
	ACN-EtOH	**ACN-MeOH**	**EtOH-MeOH**
Pair-wise t-test (PW)	0.059	3.80E-05	0.003

Table.2.7 Effect of organic modifier on the number of features and the peak parameters.

Comparing the effect of organic modifier on the signal intensity for commonly detected features (MAK) gave variable results with no favour for any organic modifier. In a further manipulation of the data obtained the peak parameters of 37 features, known to match authentic standards of metabolites (accurate mass and retention time), were manually extracted. In contrast to providing the highest number of features, MeOH gave the worst peak width among the three solvents while the difference in peak width for ACN and EtOH was not significant (paired t-test).

Moreover, the spread of the peaks over the run time was again in favour of EtOH and ACN [6.3-24.89] and [8.98-25.98], respectively, over MeOH [5.99-24.25]. Although it though the difference between the maximum and minimum time was not particulary large for the three methods, a close look at the elution pattern shows 33/37 metabolites eluting within the space of eight minutes for MeOH whereas the same number of metabolites is spread over twelve and eleven minutes for EtOH and ACN, respectively (Figure.2.22).

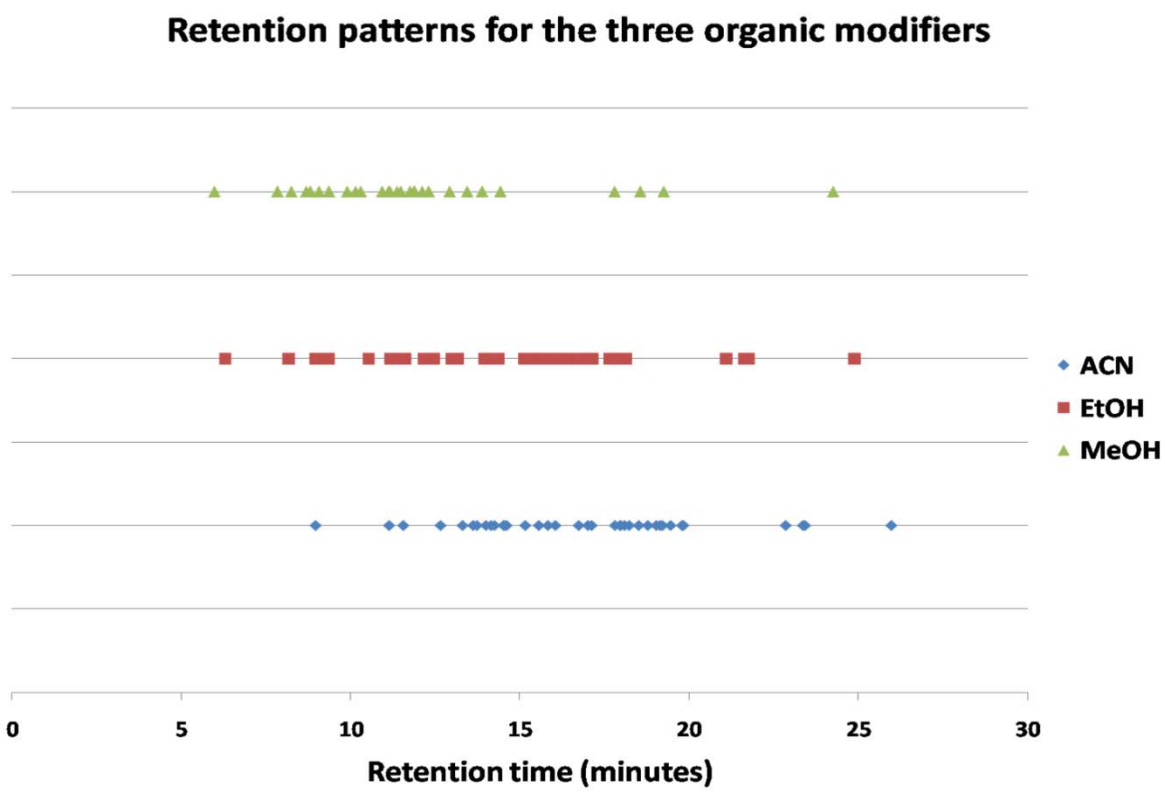


Figure.2.22 retention pattern of 33 metabolites on ZIC-HILIC when run using different organic modifiers.

For a given metabolite, the retention time was, not surprisingly, commensurate with the polarity of the solvent ACN>EtOH> MeOH. The metabolites responded differently to changing the organic modifier. For six metabolites (carnitine, tryptophan, guanosine, inosine, urate and xanthine) switching from ACN to MeOH reduced the retention time by 2.9-3.3 minutes, while switching from ACN to EtOH resulted in a RT reduction of 0-2.8 minutes. However, for the other metabolites (e.g. methionine, glutamic acid) the effect of organic modifier is more prominent, with RT reduction about six minutes when switching from ACN to MeOH compared to less than three minutes when switching from ACN to EtOH Figure.2.23.

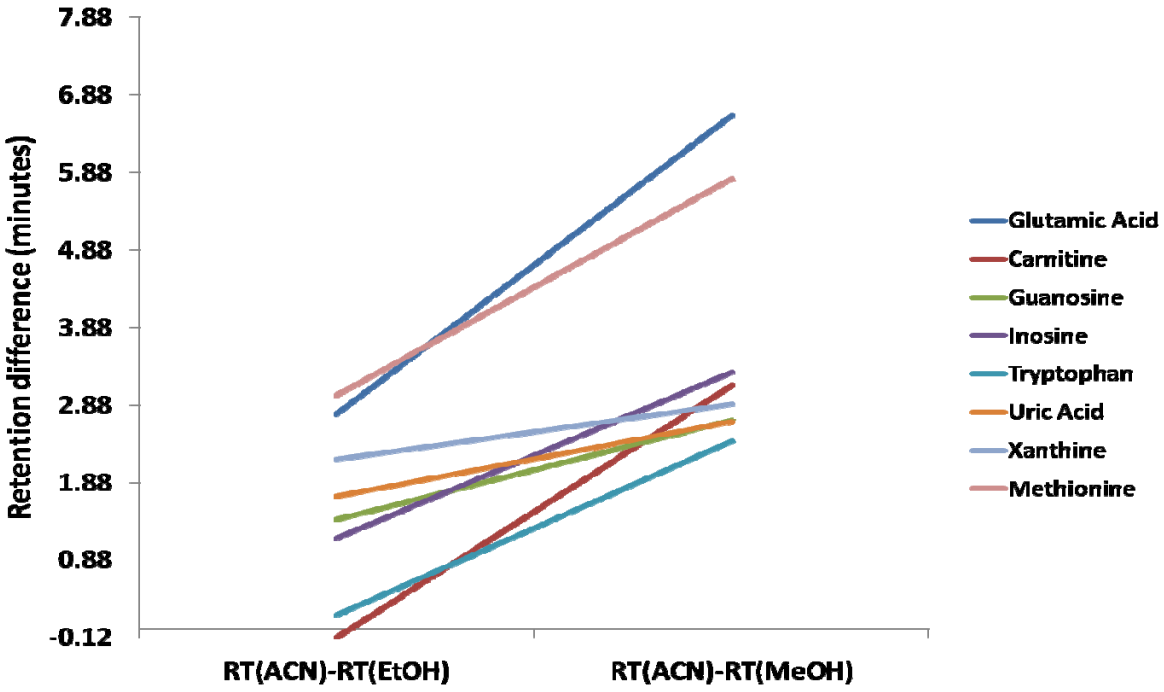


Figure.2.23 Effect of the organic modifier on retention of selected metabolites. The plotted values indicate relative retention of ACN/ETOH (left) and ACN/MeOH (right).

Furthermore, a change of the degree of separation between isomers was observed. No single method was ideal for separating all isomers. For example, although acetonitrile elution was the only method producing some separation for leucine/isoleucine, it failed to show separation for sugar-phosphates which occurred as a single peak Figure.2.24. The separation of the positional isomeric pairs alanine/β-alanine, valine/betaine was enhanced with the alcohol elution plans (Figure.2.25). Moreover, the elution pattern for MeOH was the only solution to resolve the lysine peak from arginine (Figure.2.26).

These observations indicate that each of the three methods has its distinct merits, and they can be interchangeable when required. However, as this specific experiment was done only when acetonitrile was in short supply, which happened in mid 2008 where a good portion of the work in this thesis was already complete, ACN kept its position as the solvent of choice, especially now that its supply is no longer an issue. Overall ACN produces better retentivity, peak shape and lower back pressure.

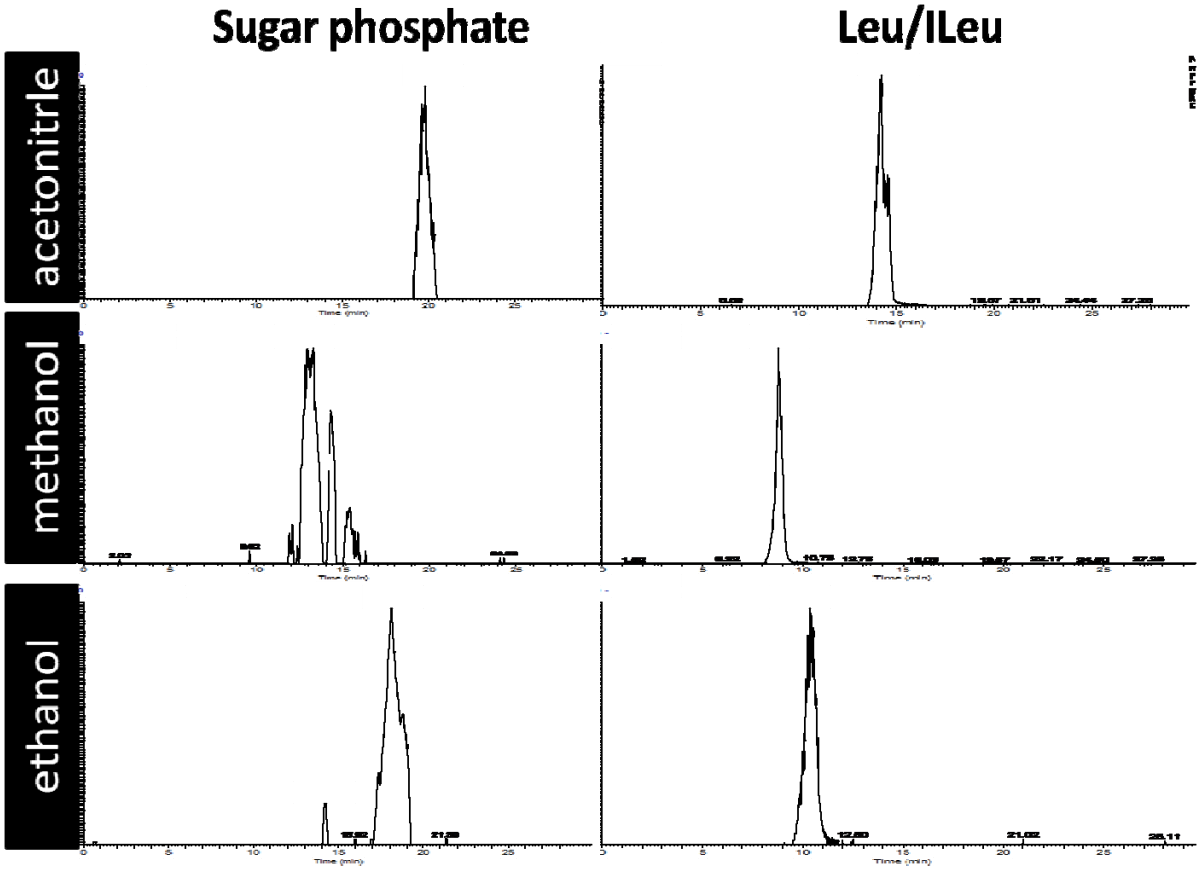


Figure.2.24 Examples of cases where one organic modifier gives better separation between isomers than the others. Sugar phosphates and leucin/isoleucine run on a ZIC-HILIC column using acetonitrile , methanol or ethanol as an organic modifier.

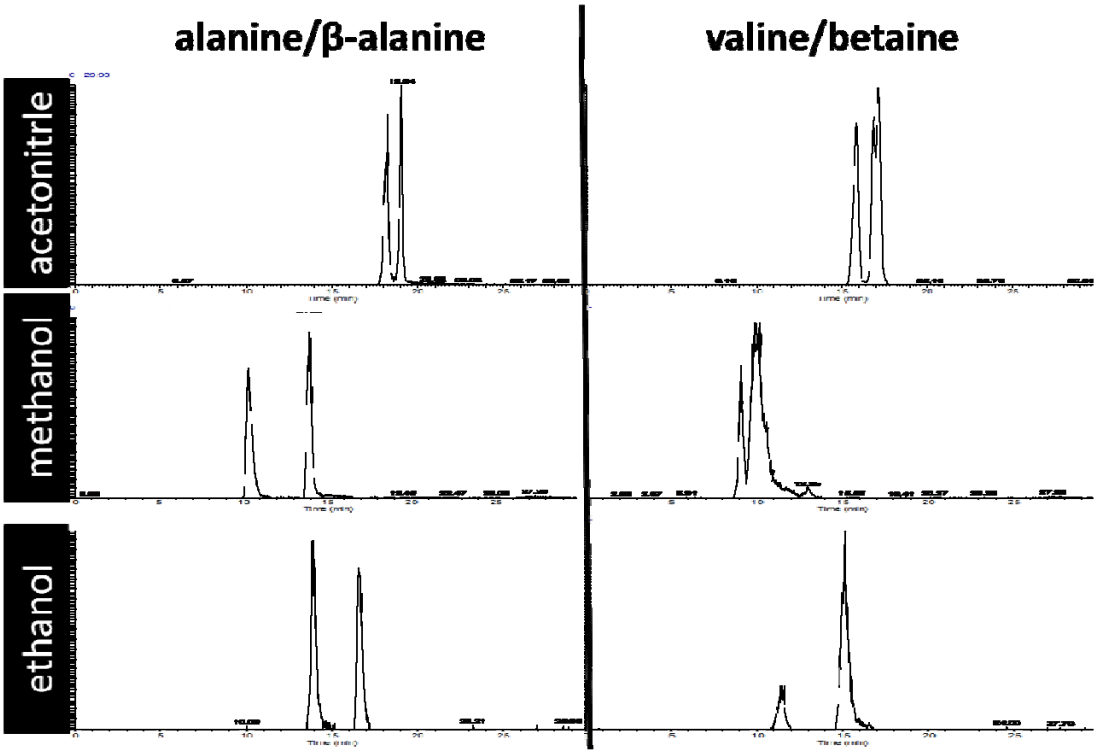


Figure.2.25 Separation of positional isomers on a ZIC-HILIC column when using acetonitrile , methanol or ethanol as an organic modifier.

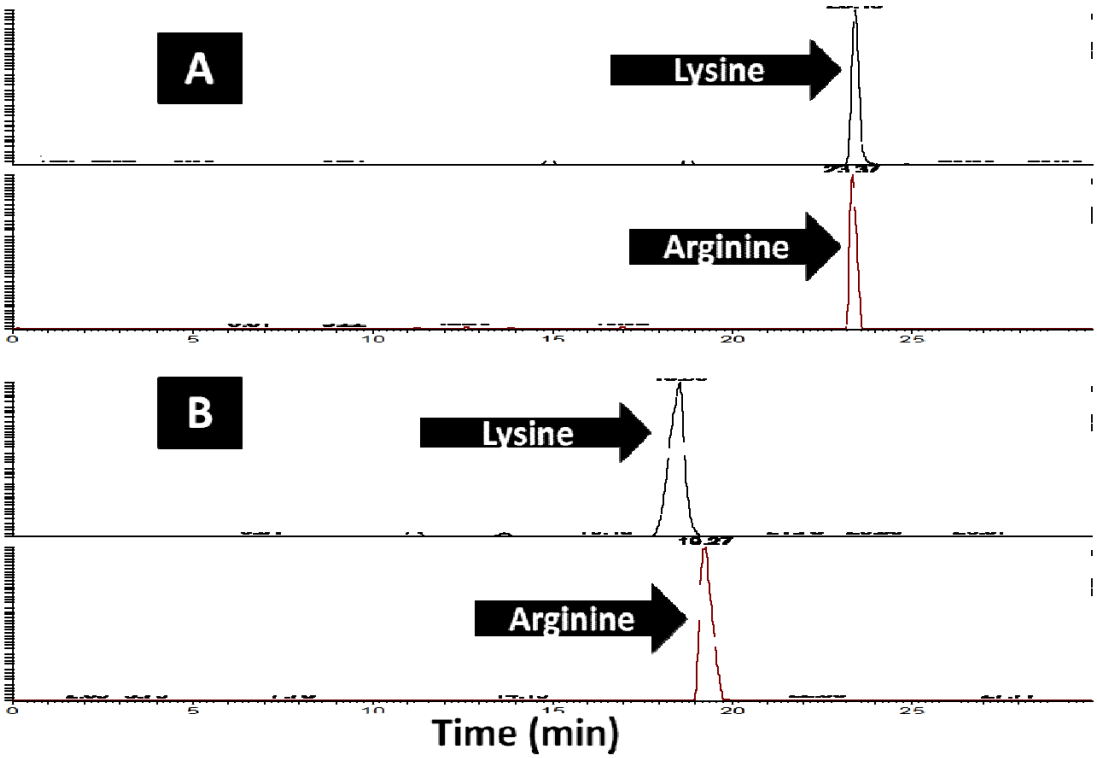
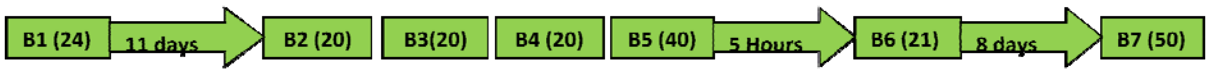


Figure.2.26 Separation of lysine/arginine on a ZIC-HILIC column when using (A) acetonitrile or (B) methanol as an organic modifier.

2.3.7. Retention time repeatability/reproducibility:

The repeatability and reproducibility of retention time (RT) is important in metabolic profiling, especially for automated peak picking, identification and database construction. The Metabolomics Society Standards Initiative requires a match of two orthogonal methods for the identification of a metabolite. With inconjunction with accurate mass from the Orbitrap, a retention time match provides the second orthogonal identification method. However, this will not be valid unless the stability of RT for the metabolites in their matrix is proved, otherwise, closely eluting features of close m/z values could be wrongly matched. Problems with retention time reproducibility on ZIC-HILIC were reported previously (Idborg et al., 2005a). However, the problem in that early work could be specific to the first few batches of the column, or to the mobile phase used in those studies.

To assess the repeatability and reproducibility of RT over a group of samples, 195 Drosophila samples were monitored for RT stability. The design of the runs was as shown in scheme.2.1. This run design was suitable for assessing both repeatability and reproducibility of RT in a *Drosophila* sample matrix. The samples used in these analyses were not identical QC samples, and they were not even all an identical matrix as 20 of the samples were extracted tubules, and 20 were extracted heads of the flies. This introduces a type of robustness check of RT with this slight change of the matrix. Still, it is worth mentioning, that in some matrices higher content of protein and salts and may have some effect on the retention pattern, and the results presented here may not be applicable to other matrices such as plasma or liver extracts. Other robustness checks used were system maintenance, blank runs and increased chromatographic equilibration time.



Scheme.2.1 Retention time repeatability and reproducibility test design.

The literature has provided multiple assessment criteria for RT stability assessment. Dunn *et al* (2008) 3% (111 samples over 40 h) , Gika *et al* (2008a)

considered a CV of ±1% (116 injections/16 QC) and Guy *et al* (2008) (7 injections over 70 samples) reported a ±1.3% value compared to 1% for Michopoulos *et al* (2009) (63 injections/7 QC).(Gika et al., 2008a, Michopoulos et al., 2009, Dunn et al., 2008, Guy et al., 2008)

All these criteria were constructed upon either standard QCs or pooled QCs which will give rise to no significant variation in peak height, contrary to the current work where this test was attained on top of a biological experiment. Moreover, the number and diversity of monitored peaks (36 compounds) in the current work is greater than numbers used in the above mentioned four experiments (7, 7, 11 and 13, respectively). For these reasons, more variance in RT is expected to be seen in the test provided in this thesis.

In the work presented here, 35 metabolites, identified by matching to accurate mass and authentic standards retention time, were traced. These metabolites were chosen to cover the full range of the run time (10 to 26 minutes) and to cover a diversity of chemical groups (amino acids, nucleosides, nucleotides, sugar phosphates ,etc).There is currently no available software to give a list of peak information for Orbitrap data, hence, a tedious manual extraction was used for this test.

The middle batch consisted of 100 runs divided into four subsets of samples (B2, B3, B4 and B5). This big batch was used to assess the repeatability of RT. Across the whole set of 35 metabolites, only one (alanine) failed the criteria of CV < 1.4%. Alanine`s failure was only marginal with a CV value of 1.57% which was reduced to 0.76% when removing six samples showing very low abundance of the metabolite. The three remaining batches (B1, B6 and B7) were used for RT reproducibility checks and showed a very good RT stability for all the metabolites. None of the 35 metabolites had a CV value more than 2% with only four metabolites with CV > 1.4% across the batches run in a span of a whole month. The four failing metabolites were xanthine, hypoxanthine, inosine and NAD which were at low levels, and this may explain the RT instability in the light of a variability in response. Within any of the seven batches, the repeatability was very good with rare occasions of CV values >1.4% and many occurrences of a CV< 0.5%. The assessment of RT

stability over the whole set of batches did not exceed the 2% upper limit for any metabolite (Table.2.8).

	B1	B2	B3	B4	B5	B6	B7	B2-B5	B1,B6 and B7	B1-B7
(+-) Carnitine (HCl)	0.240	1.107	0.433	0.390	0.689	0.370	0.437	0.903	0.914	1.548
Glucose-6-Phosphate	0.157						0.466	0.901	0.933	0.928
(DL) Valine	0.190	1.210	0.165	0.171	0.643	0.406	0.448	0.813	1.080	1.211
Betaine	0.132	1.120	0.131	0.160	0.617	0.431	0.498	0.858	0.917	1.158
(DL)Kynurenine	0.185	1.235	0.160	0.176	0.838	0.460	0.497	0.919	0.730	1.196
(L) Citrulline	0.157	0.844	0.212	0.125	0.373	0.363	0.374	0.662	0.688	0.902
(L) Glutamine	0.129	0.905	0.134	0.090	0.412	0.319	0.375	0.681	0.757	1.009
(L) Lysine	0.124	0.672	0.086	0.071	0.204	0.136	0.144	0.740	0.509	0.888
Adenine	0.228	1.093	0.218		0.687	1.521		0.812	1.136	1.194
Adenosine	0.162	1.121	0.263	0.542	0.950	0.493	0.458	0.883	1.049	1.343
Alanine	0.097	1.071	0.072	0.113	0.474	0.385	0.431	1.576	0.785	1.725
β-alanine	0.220	1.061	0.049	0.105	0.461	0.429	0.353	0.754	0.478	1.130
Arginine	0.164	0.599	0.095	0.112	0.365	0.243	0.221	0.909	0.511	0.903
Asparagine	0.084	0.859	0.099	0.171	0.365	0.343	0.343	0.677	0.847	1.040
Aspartic Acid	0.124	1.074	0.163	0.115	0.479	0.374	0.386	0.735	0.854	1.043
Choline (base)	0.259	1.158	0.213	0.217	0.740	0.530	0.494	1.280	1.200	1.679
GABA	0.063	1.301	0.366		0.528	0.415	0.373	0.953	0.750	1.041
Glutamic Acid	0.110	0.295	0.083	0.161	0.401	0.348	0.387	0.741	0.723	1.097
Guanine	0.259				1.024			0.434	0.259	0.391
Guanosine	0.209	1.042	0.210	0.208	0.715	0.452	0.581	0.876	1.344	1.392
Histamine	0.209	0.579	0.101	0.257	0.537	0.113	0.271	0.485	1.131	1.008
Hypoxanthine	0.293	1.119	0.330	0.264	1.628	0.544	0.724	1.177	1.571	1.611
Inosine	0.132	1.207	0.256	0.298	1.133	0.513	0.712	1.068	1.638	1.653
L(-)Histidine	0.255	0.654	0.090	0.086	0.339	0.267	0.288	0.548	0.446	0.634
Leucine	0.153	1.297	0.185	0.181	0.865	0.401	0.465	0.996	0.836	1.263
Methionine	0.133	1.197	0.206	0.147	0.694	0.375	0.469	0.889	0.916	1.172
NAD	0.178	1.147	0.743	0.287	0.952	0.368	0.588	1.274	1.899	1.754
Phenylalanine	0.181	1.255	0.214	0.190	0.873	0.419	0.522	0.939	0.843	1.248
Proline	0.082	1.106	0.164	0.173	0.677	0.444	0.548	0.868	0.961	1.263
Serine	0.076	0.871	0.101	0.126	0.380	0.274	0.348	0.673	0.804	1.006
Taurine	0.107	0.992	0.124		0.495	0.461	0.510	0.779	1.241	1.214
Tryptophan	0.165	1.226	0.115	0.164	0.743	0.402	0.472	0.879	0.604	1.120
Tyrosine	0.342	1.138	0.171	0.114	0.553	0.339	0.447	0.794	0.863	1.117
Uric Acid	0.262	1.093	0.159	0.245	0.845	0.458	0.627	0.889	1.385	1.420
Xanthine	0.237	1.262	0.279	0.292	2.442	0.527	0.857	1.337	1.589	1.661

Table.2.8 RT repeatability and reproducibility for the different batches shown as RSD of retention times. Highlighted cells indicate values exceeding the "1.4" selected threshold.

Although the ZIC-HILIC RT range operated under our conditions does not look as precise as the BEH ACQUITY results shown in the work by Michopoulos *et al (2009)*, the reproducibility can be considered to be excellent taking into account the differences in experimental design and instrumentation. In the above mentioned study the authors used only seven samples with good retention, signal and peak shapes whereas 35 metabolites were monitored in the current work. Furthermore, the UPLC used gives peaks of the order of 5-10 seconds in width, compared to the average peak width reported previously on HPLC-HILIC being in the range of 40-60 seconds. The difference in peak width is crucial, as variation of 2.5 seconds will not affect the peak picking algorithm in HPLC (4-5% of peak position) but may have some effect on the UPLC data (25-50% of peak position).Normalising the reproducibility to the peak width may see ZIC-HILIC achieving comparable results to the BEH ACQUITY. One last factor in support for the reproducibility of ZIC-HILIC is the method of data extraction which relied on automatic apex positioning. At the apex of high abundance peaks, the C-trap may reach the injection threshold quicker giving short accumulation time, hence , less ion counts and consequently, a shift of the peak`s "most abundant scan" 3-4 scans at the expense a greater RT variation.

2.4. Conclusion:

In this chapter, the importance of a separation step for metabolic profiling experiments was investigated and it was shown, by simulating a part of the *Drosophila* metabolome, that even for a small set of metabolites the required resolution for separating isobaric masses is beyond the current capability of the currently available mass spectrometric technology. Following on, four HILIC stationary phases were tested and the results discussed in terms of the suitability of the SPs for metabolic profiling studies, concluding ZIC-HILIC and Luna-HILIC were the most satisfactory, with a preference for the former. Then, ZIC-HILIC was further tested with different mobile phase compositions and organic modifiers, and a general method was suggested, for which repeatability and reproducibility of retention times for 35 test compounds were studied and found to be satisfactory.

3.Data Pre-Processing:

3.1. Introduction:

The MS profiles of biological samples are complex. The necessary prerequisite of a separation step adds much to the complexity of the data which become positioned in three dimensions i.e. time, m/z and intensity each of which has its own alignment problems. The raw MS files provide data which can be easily manipulated for individual sample visualization, but not for comparison of a cohort of samples. For a cohort of samples, the raw files should be properly aligned in the time and m/z dimensions in order to facilitate the comparison of intensities. The higher the resolution of the instrument, the more complex the data expected and the analytical technology seems to run faster than the software development technology in the field of metabolic profiling.There are a number of available software packages from academic or commercial origins. Some of these packages have been available for several years, and some have been recently released, with a common feature of being able to process Orbitrap data only recently (basically mid-2008).

SIEVE® is commercially available software for evaluation of differential expression supplied by Thermo Scientific. Its algorithm is not well described either by the vendor or in scientific literature. Its web page shows no applications and reflects the lack of reports for the use of this package, Sieve® was used for MS data processing only twice, both come from the work in this thesis (Kamleh et al., 2008, Kamleh et al., 2009). Nevertheless, the vendors claim the software can do differential expression analysis for treatment vs. control (2 classes) or trend analysis (more than two classes), is able to do chromatographic alignment by a newly introduced, but not well described tile algorithm, but relies on calculating the best correlation between a master sample and the other samples. The next step in the work flow in Sieve® is finding m/z – RT matrix of features (the term used here is frame) followed by simple statistical analyses (t-test) and chart generation. The last step is searching databases (Chemspider for metabolomics and SEQUEST for proteomics) (Thermofisher, 2010a, Thermofisher, 2010b).

XCMS is a freely available package for analysing complex MS data written in R environment (Smith et al., 2006) with XCMS2 adding the feature of analysing

MS/MS data (Benton et al., 2008) and new versions including functions designed specifically for high resolution data analysis (Tautenhahn et al., 2008). It basically executes peak picking from the raw mass spectrometry data files (converted to XML or netCDF) making use of the MEND denoising algorithm for recognising genuine peaks from background ions. In brief, MEND, matches the width of the peak with a Gaussian peak of similar width and the final matrix will retain only peaks with a specific pass/fail cut-off of the correlation value between the two peaks (data peak and dummy Gaussian) (Andreev et al., 2003). It then corrects for retention time shift using the Correlation optimised Warping (COW) algorithm (Nielsen et al., 1998). Database searching is restricted to METLIN for the original XCMS code with a web link generated for each feature.

The package was used for processing data for several plant (Arbona et al., 2009), animal (Fraga et al., 2010, Lin et al., 2009, Dai et al., 2010) clinical (Nordstram et al., 2006) as well as method optimisation (Want et al., 2005, Nordstrom et al., 2007, Dunn et al., 2008) metabolic profiling papers.

Among all these applications, only one paper described the use of XCMS to process LC-Orbitrap data (Dunn et al., 2008). In this paper, the XCMS function parameters were optimised, and these parameters will be used for XCMS processing later, edited only to cope with the chromatographic difference between the reported work and the current work.

MZMINE is another freely available tool for processing and visualising MS data with many functions of de-isotoping, quantification and database searching (Katajamaa et al., 2006). The toolbox was shown to be suitable for metabolic profiling (Katajamaa and Oresic, 2005) and showed similar output when compared to XCMS (Kind et al., 2007). Unfortunately, it did not accommodate .raw thermo files until very recently and applications to show how successful it is with Orbitrap data have yet to appear in literature. MZMINE will not be included in the discussions in this chapter.

Database searching has established itself in metabolic profiling and specifically MS-based metabolic profiling. The density of information and complexity of finding

a simple explanation in the most cases of MS-based metabolic profiling has dictated the necessity for automated database searching, assignment and comparison tools. This is manifested in the fact that all the commercial and academic processing packages are linked to a database search resource. Still, improvement of the database search is desirable and led to the release of further tools which address this important need. For example, XCMS features can be intelligently filtered from artefacts and screened across KEGG database (Gipson et al., 2008). Mass lists generated by any type MS instrument were used for metabolic network construction for the objective of understanding biological process (Jourdan et al., 2010). These network tools provide excellent graphical representation of the available metabolome, and are often linked with a web database as KEGG or biocyc in Masstrix and Metexplore, respectively (Suhre and Schmitt-Kopplin, 2008, Cottret et al., 2010). However, both these packages rely on mass only assignment (assuming all positive and negative mode m/z are $M+H^+$ and $M-H^-$, respectively), with no isotopic or retention information. Moreover, to use these network models, any statistical information should be calculated and subsequently metabolites filtered accordingly, offline before loading the mass list with no option, for example, for colouring or detecting a lesion position in the metabolome.

3.2. Materials and methods:

Reagents, *Drosophila* stocks, extraction and LC-MS methods were as mentioned in chapter 2.

3.3. Data processing:

3.3.1. Description of the macros

At the time when the work in this thesis commenced there were limited options for processing the data from the LTQ-Orbitrap. Among the packages tested, the first (in chronological order) to be used for analyzing data was a collection of Excel Visual Basic for application (VBA) macros. The whole set is given the name “MAK” which is a short of “Metabolomics Aid Kit”.

This sub-section of the chapter will be dedicated for describing the algorithms, outcomes and rational for using the macros. The description covers MAK functions

that will be assessed or discussed in this chapter. Other functions will be either discussed in the later chapters, or not discussed at all if the usage is not represented by a biological experiment within the scope of this thesis.

3.3.1.1. The basic macro:

The idea of using Excel macros originated from the need of a quick method to process the large volume of data generated by the high resolution MS because of the lack of commercial or academic packages for this purpose. Ihe only feasible way to process the Orbitrap data was to learn this very basic programming language which makes use of the built-in mathematical and statistical functions in Microsoft Excel.

The first macro used had been already in the Department before the work in this thesis commenced and improving on it to suit the needed tasks required many programming hours and over 200 different codes were written, many of them were improved versions of the main task (m/z alignment) in the domains of quality, simplicity and speed while the remaining were a vast array of functions which will be discussed only if relevant to the content of this thesis.

The full mass list which can be obtained from Xcalibur disregards the time domain and sums the intensity of a predefined m/z range for the whole chromatogram. The list is then sorted either by mass or by intensity and is restricted to one RAW file at a time. The basic task of the macro work was to produce a matrix with rows of intensities (variables) and columns of samples (observations) from the full mass lists of a set of samples. This was achieved by doing a vertical search of matching m/z(s) (closest match) of each m/z from a feed sample and the constructing a matrix row with this information. A second round of filling missing values was automatically run at the end of the analysis and m/z values which were not found in the feed sample were added as new rows. This step was important to avoid any loss of information due to m/z deviation. The default m/z width was set to 3 ppm for positive and 5 ppm for negative mode but was editable according to the run conditions.

The full mass list for one sample generally included over 500,000 m/z binary values of (m/z, intensity) most of them are background spikes with very low intensity. As no number of scans per feature or chromatographic information was available, this number was cut down by setting a signal/noise cutoff level. The noise level for each peak was calculated by the Xcalibur software for each peak, but could be set to a single threshold value for all peaks in the macros. An S/N value of 3 is the default parameter for the macro. Missing peaks were replaced with 0.01 as a value to avoid zero values in the final matrix as it could hinder further mathematical processing e.g. normalisation, t-test or correlation.

These were the basic features required from the macros which would generate a matrix of feature rows (variables) and sample columns (observations) compatible with exporting to SIMCA-P for multivariate analysis. However, most important features were colour coded according to a statistical test in order to help the analyst in extracting preliminary biochemical indicators from the datasets.

3.3.1.2. Blank filtration :

Many of the peaks selected, even after S/N filtering were background artefacts. The blank contains these very same peaks, with intensity values close to those occurring in the samples. The exact m/z match and the intensity were employed to pick out these background ions and further refine the data matrix to contain only biochemically sound information. The m/z width was set to 5 ppm and the intensity range to at least 0.5 fold (sample/blank) of the average of repetitive blank measurements for each sample set. The value 0.5 was selected considering ion suppression or enhancement, yet not too low so as to avoid possible carryover from a previous sample to a blank (usually less than 0.1%).

3.3.1.3. Deisotoping/ adduct detection:

The outcome of a MS experiment has been thoroughly simplified with the introduction of soft ionisation techniques such as ESI and APPI. However, the close study of MS indicates this is not as simple as the theory denotes. Isotopes, adducts, dimers, trimers and a mix of these could all be related to one single compound and

these may overlap in the m/z dimension with other metabolites, especially when the time domain is disregarded, as in the current case.

To minimize this false positive possibility, and to further filter the content of the matrix produced, a code was compiled to identify a pre-defined sets of adducts and isotopes. The macro made the best of the high accuracy mass measurement attained by the Orbitrap (below 2 ppm) over a wide range of intensities. Nevertheless, other parameters could be used to ascertain a correct assignment of an isotope or an adduct.

For isotope detection (de-isotoping) two criteria were used, in addition to accurate mass, for considering a feature as an isotope of another feature. These were relative intensity and a correlation of the two intensities. An isotopomer of a specific compound occurs at a very constant ratio relative to the mono-isotopic compound, for example, ^{13}C always occurs at a 1.1% of the corresponding ^{12}C compound for each carbon atom in the structure. However, the measurement on the MS may not be as accurate as the ratio produced by nature. So ,building from the accurate mass, a Pearson`s correlation coefficient across the whole sample set between the two rows is constructed, and as the ratios are fixed (within a margin of error) the correlation coefficient should be positive and high. Experimental data suggested a cut-off value of +0.6. Moreover, the measured ratio itself should be an indication for elemental composition, but for an unknown compound/metabolite, how would one estimate the number of carbons? Simple mathematics can provide a simple but only partial solution. By taking the integer of the mass and divided it by 14 (the fully saturated hydrocarbon) one could get the maximum possible number of carbons in a given compound, hence, the maximum permitted average $^{13}C/^{12}C$ ratio and excludes any masses which passed the first two criteria but not this third. The isotope peaks behave generally in the same way as the original compounds in terms of ionisation and ion suppression, so these effects are not a worrying issue in this specific situation. Furthermore, as the concentration of the isotopomers considered in our analysis are always lower than the original compound, the isotope may fall below the detection limit or the picking threshold (S/N=3) so it is of less importance to assign a lower permitted ratio. The deisotoping task becomes simpler when forward searching

(discussed later) is used. It was found that above certain threshold, the Orbitrap can give relative isotope abundance with error rate less than 20% (Xu et al., 2010).

In the case of adducts, no ratio information is available, especially for unknown compounds. However, accurate mass coupled to correlation may still make a good estimation of adduct genuineness. Of course adding RT matching will provide the indisputable evidence in the case of both isotopes and adducts. Unfortunately, this domain was beyond the capabilities of these macros. All isotopic data for forward searching (see later) were preserved in a separate spreadsheet.

3.3.1.4. Forward searching:

In spite of the filtration of the data in the noise, blank and deisotoping similarity dimensions, the data may still contained a very large number of features, many of which could not be readily assigned to a compound. Metabolite assignment is one of the most difficult tasks in metabolic profiling and the Metabolomics standard initiative requires two orthogonal identification methods for this purpose. An accurate mass and a match of RT to an authentic standard suffice, in the latter initiative rules. In the case of Orbitrap, the high accuracy mass measurement is only one method which is not sufficient for assignment, but, could be of a pivotal importance for a preliminary screening. This serves to indicate whether or not a specific metabolite is potentially in a list of metabolites or not and shows which metabolites are potentially of importance for discriminating between different groups of samples. This method is exactly the reverse to the usually implemented procedure of picking peaks then looking them up in a database. The hypothesis for this procedure is to make a quick assessment of missing metabolites and those potentially captured by the analytical method.

In the forward searching, two metabolite lists were used both derived from the Kyoto Encyclopaedia of Genes and Genomes (KEGG). We will call the first the full MS compatible list, and the second the metabolic pathway list. To construct the full compatible list, the full KEGG metabolite database (16666 entries) was imported and treated offline in order to accomplish the following tasks:

- Render salts into counter-ion free form.
- Convert the metabolites into a matrix, easily readable by macros.
- Remove species incompatible with MS detection such as organometallics electrolytes and inorganic metabolites or those which lie outside the used mass range (proteins and macro molecules).
- Remove drug molecules.

The list eventually contained 11371 entries the exact monoisotopic mass of each along with $M+H^+$ $M-H^-$ was calculated and aligned in the proper row.

All the filtering processes (including accurate mass calculation) were done using VBA written macros in order to facilitate an offline update of the list on a regular basis. To construct the metabolic pathway list, all the *Drosophila* pathways were imported from KEGG and arranged into one spreadsheet with KEGG identifier for metabolites in each column. These identifiers were looked up in the full compatible list and isolated to a separate list (562 metabolites) (Figure.3.1).

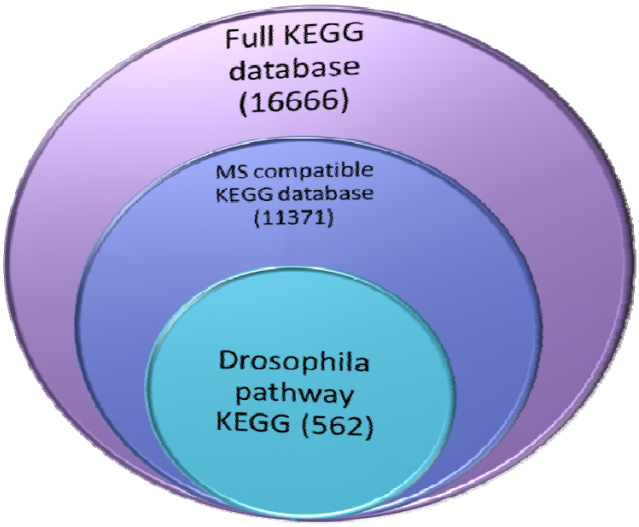


Figure..3.1 A Venn diagram showing the number of metabolites in the full, MS compatible and *Drosophila* pathway from KEGG database.

The driver for this process was the ultimate objective of the *Drosophila* metabolic profiling work which was basically tracking changes in metabolic pathways in response to different perturbations (naturally occurring pheno- or geno-types, drug treatment, genetic manipulations ...etc). Hence, it was plausible to focus

our efforts on metabolites occurring in these pathways rather than those outside the *Drosophila* metabolome.

3.3.1.5. Additional functions:

3.3.1.5.1. Deviation matrix:

The deviation matrix is a matrix generated in parallel to the last forward searching matrix showing how far each metabolite is from corresponding to the feature for each sample. This will help track any mass drift across the samples, as well as indicate how confident we are with the presence of each potential metabolite even before examining the raw data.

3.3.1.5.2. Statistical analysis:

In order to perform a simple t-test and fold change measurement for two groups in an Excel spreadsheet is fairly an easy task once the data are aligned properly, even when the number of variables is large. However, this same task starts to become tedious in case of more than two groups, especially when ANOVA is required. ANOVA is not a readily available function in Microsoft Excel since it has to imported via the statistical package. Moreover, Excel ANOVA can only be applied to one variable at a time. For the ease of analysis, the macro contains functions to do the ANOVA test for all variables based on a pre-selected confidence interval of 0.95, 0.99 or 0.999. In case of more than two groups, the macro will calculate the permutated t-test or Tukey test for all binary combinations of groups. Finally, the macro gives a colour to those metabolites which have a significant change (P value below a user pre-defined threshold) with different colours in case of up or down regulation (purple and red, respectively). An example of the outcome is shown in Figure.3.2.

A	B	C	D	E	F	G	H	I	J	K	L	M
Mass List	MG067	MG079	MG091	MG103	MG115	MG076	MG088	MG100	MG112	MG124	P.value	Fold chang
65.0604	0.71509	0.58395	0.77346	0.66752	0.8409	0.60005	0.6212	0.41724	0.59215	0.55919	0.02523	1.28356
70.0651	0.54734	0.76199	0.52665	0.68048	0.95534	0.4879	0.48071	0.6996	0.71007	0.77363	0.53824	1.10149
71.0604	2.76621	2.88897	2.65276	4.01111	3.58483	2.82419	3.7012	4.43957	5.51124	5.11031	0.08326	0.73675
73.0648	0.61687	0.52832	0.50325	1.35996	1.0014	0.8269	0.56345	1.37614	1.54796	1.20329	0.25335	0.72671
74.0964	1.49018	2.00981	0.01	0.01	0.01	1.63393	0.75476	0.01	0.01	0.01	0.69248	1.45946
76.0393	0.63829	0.97798	0.55555	0.89293	0.93292	1.3075	1.25784	1.3674	1.53355	1.45108	0.00078	0.57792
113	5.53406	4.94957	4.34209	6.0866	5.189	2.6952	2.67201	2.93189	1.99482	2.26626	0.00014	2.0781
118.086	32.8885	24.9479	27.5155	32.4028	35.1104	48.0876	41.0701	41.5364	50.831	44.339	0.00058	0.6768
137.04	0.37851	0.50485	0.63913	0.26748	0.32077	0.01	0.01	0.01	0.01	0.01	0.00356	42.2147
137.046	81.8272	101.224	112.751	83.8853	68.1926	19.705	13.9797	14.4541	10.085	10.48	0.00044	6.519
137.05	0.30767	0.43366	0.52103	0.20458	0.26115	0.01	0.01	0.01	0.01	0.01	0.0044	34.5616
137.051	0.33635	0.50507	0.57538	0.22718	0.31373	0.01	0.01	0.01	0.01	0.01	0.00405	39.1543
138.043	0.95504	1.22977	1.40452	1.0265	0.82109	0.01	0.01	0.01	0.01	0.01	0.00048	108.738
138.049	4.23426	5.28691	5.95131	4.33229	3.5405	0.89144	0.60031	0.6126	0.41404	0.44624	0.00049	7.8746
139.05	0.92281	1.83095	0.54547	0.348	0.25065	1.06772	0.34829	0.39313	0.39162	0.22671	0.3979	1.60574
140.011	0.52597	1.23464	0.42031	0.59184	0.72367	0.79642	0.95429	0.74182	0.74113	0.73224	0.55702	0.88162
148.004	0.2033	0.23155	0.17143	0.24597	0.20469	0.33597	0.40042	0.36512	0.40378	0.3998	1.6E-05	0.5548
148.061	32.3647	25.8002	22.7603	27.8934	22.2778	32.0101	34.6455	34.8485	28.5175	27.1291	0.06475	0.83421
148.08	1.3735	0.9437	1.42649	1.1239	1.02176	1.13762	1.11498	1.06547	1.04878	1.08657	0.41611	1.07994
148.139	0.38429	0.01	0.21149	0.25312	0.01	0.34418	0.30296	0.33431	0.29599	0.28673	0.12842	0.55551
149.023	0.65743	3.67455	1.61766	1.5018	1.40759	0.80815	1.28064	1.80407	1.83923	1.57358	0.58915	1.21262

raw_data | positive | negative | reaction | gr_pos | gr_neg

Figure.3.2. The output matrix and colour code for significant features from MAK. Red code is down regulated and purple is up regulated in the second group.

3.3.1.5.3. Normalisation:

At the very start of the analysis, the macro will ask for a choice between normalisation to total intensity or no normalisation at all. In both cases, there must be a normalisation to the number of scans. This is necessary because Xcalibur sums intensity across the whole chromatogram and divides it by the number of scans. A change in the number of scans of 10% leads to a 10% false positive (or false negative) change in intensity for the outcome. The number of scans should be added in the first row and second column for each sample, otherwise, the macro will crash (multiplying by text) for the old macros or will report an error indicating the missing information in the most recent macros. Neither XCMS nor the preliminary versions of Sieve® include a normalisation function while later versions of Sieve® (Sieve®1.3) include simple normalisation to a pre-defined manually dialled value.

3.3.1.5.4. Concatenating polarities:

This provides an automated way to combine data from positive and negative polarities for a maximal view of the metabolome. The macro was able to carry out this process on feature picking or forward searching, reporting negative mode data with a green text colour for simplification of raw data revision.

3.3.2. Reporting system for untargeted metabolomics (RESUME):

The whole run of the macro can be summarized finally in one sheet viewed in accordance to the objectives of the analysis. The spreadsheet is composed of KEGG pathway names each linked with three scores:

- Score A (analytical method scope): percentage of potential hits in each pathway.
- Score B (pathway contribution to the statistical model) : a cumulative weight of the size of variability in each pathway (reciprocal of multiplication of the t-test values).
- Score C (pathway lesion potential): number of significantly changed metabolites in each pathway.

The macro automatically produces a spreadsheet for each binary group combination (where more than two groups are involved).

Additionally, the macro will search the KEGG reaction database, and lookup enzymes which have significant changes in opposite directions on both sides of a specific enzyme. This last step may be useful for understanding drug mechanism of action or mechanism of toxicity, exact lesion detection and functional genomics.

The last two functions, albeit highly important, are hampered by the fact that by removing the chromatographic dimension all isomers of a certain compound will be counted in the score which will reduce the quality of the scores (A,B and C). However, tweaking the macro to accommodate retention time information could extend its usefulness to processing outcome from other packages (discussed below).

3.4. A test run for RESUME (comparing Rosy (ry) and wild type (WT) *Drosophila*:

RESUME was given a test run for the sample sets mentioned above. The biochemical lesion is known in this case (purine metabolism, xanthine oxidase specifically), which gives indication of what should be observed, and thus assess if the macros are working appropriately.

Figure3.3 and Figure.3.4 show the style of the output as well as the results of interrogating the ry/WT drosophila sample set. Figure3.3 shows the table of score A and B for each *Drosophila* pathway. The first column is the pathway name, the second is the total number of metabolites indicated by KEGG in this specific pathway, the third table is Score A which is the method scope. Score A depicts the number of metabolites potentially and putatively identified in the sample list. Had the analyst desired to focus on a specific pathway, this score might play a role in a preliminary screening of different analytical methods. In this case, purine metabolism was of a special importance, and we were satisfied to see a high turn up of purine pathway metabolites in the samples. Score B is a cumulative P value score and operates to indicate which metabolic pathways are most affected by the biological perturbation. In this context, purine metabolism was correctly identified as the major interest in the ry/WT comparison. Purine metabolism was followed by phosphotransferase systems and then Tryptophan metabolism. The last two pathways are involved in osmolyte and pigment formation, respectively. The interpretation of the last observations will be discussed in chapter 4.

	A	B	C	D
1	Pathway	Pathway Total Metabolite	Pathway simple score	Cumulative P value
2	Purine metabolism	91	56	1.26625E+70
3	Phosphotransferase system (PTS	40	32	5.69056E+45
4	Tryptophan metabolism	77	42	4.90155E+42
5	Glycine, serine and threonine me	57	36	5.06926E+36
6	Pyrimidine metabolism	60	36	2.88587E+36
7	Tyrosine metabolism	82	29	9.49023E+33
8	Phenylalanine metabolism	42	24	5.38194E+33
9	ABC transporters - General	78	45	3.31842E+33
10	Glutamate metabolism	29	26	9.16033E+28
11	Vitamin B6 metabolism	32	19	1.81157E+28
12	Aminoacyl-tRNA biosynthesis	63	38	8.86834E+27
13	Pentose phosphate pathway	34	17	4.62159E+27
14	Nicotinate and nicotinamide meta	44	25	2.02511E+27
15	Lysine biosynthesis	33	29	1.11566E+27
16	Alanine and aspartate metabolisr	32	21	1.562E+26
17	Arginine and proline metabolism	53	35	9.66059E+25

Figure.3.3 Score A and score B tables as an output from RESUME function in MAK. Score B identified purine, phosphotransferase and tryptophan as the most affected pathways.

Figure.3.4 shows the score C which is another parameter for identifying affected pathways but with disregard for P values and focusing on the number of affected peaks. MAK will additionally generate a web link which upon pasting in an internet browser uniform resource locator (URL) will give the KEGG chart of the corresponding pathway, with the significantly changing metabolites coloured in red.

Pathway	pathway ID	Graphical link	Number of significant variables					
Glycolysis / Gluconeogene	00010	http://www.genome.jp/dbget	2	C00022	C00074			
Citrate cycle (TCA cycle)	00020	http://www.genome.jp/dbget	3	C00022	C00042	C00074		
Pentose phosphate pathw	00030	http://www.genome.jp/dbget	7	C00022	C00198	C00257	C00345	C00672
Pentose and glucuronate i	00040	http://www.genome.jp/dbget	1	C00022				
Fructose and mannose me	00051	http://www.genome.jp/dbget	2	C00392	C00644			
Galactose metabolism	00052	http://www.genome.jp/dbget	2	C00089	C05401			
Ascorbate and aldarate me	00053	http://www.genome.jp/dbget	2	C00022	C00433			
Biosynthesis of steroids	00100	http://www.genome.jp/dbget	2	C00022	C00418			
Ubiquinone biosynthesis	00130	http://www.genome.jp/dbget	1	C05817				
Oxidative phosphorylation	00190	http://www.genome.jp/dbget	2	C00008	C00042			
Photosynthesis	00195	http://www.genome.jp/dbget	1	C00008				
Urea cycle and metabolisr	00220	http://www.genome.jp/dbget	1	C00437				
Purine metabolism	00230	http://www.genome.jp/dbget	8	C00008	C00035	C00144	C00301	C00366
Caffeine metabolism	00232	http://www.genome.jp/dbget	2	C00385	C01762			
Pyrimidine metabolism	00240	http://www.genome.jp/dbget	3	C00015	C00475	C00672		
Glutamate metabolism	00251	http://www.genome.jp/dbget	3	C00042	C00144	C00352		
Alanine and aspartate met	00252	http://www.genome.jp/dbget	3	C00022	C00042	C01042		
Glycine, serine and threon	00260	http://www.genome.jp/dbget	4	C00022	C00109	C00430	C06231	
Methionine metabolism	00271	http://www.genome.jp/dbget	1	C00109				
Cysteine metabolism	00272	http://www.genome.jp/dbget	1	C00022				
Valine, leucine and isoleu	00290	http://www.genome.jp/dbget	3	C00022	C00109	C02631		

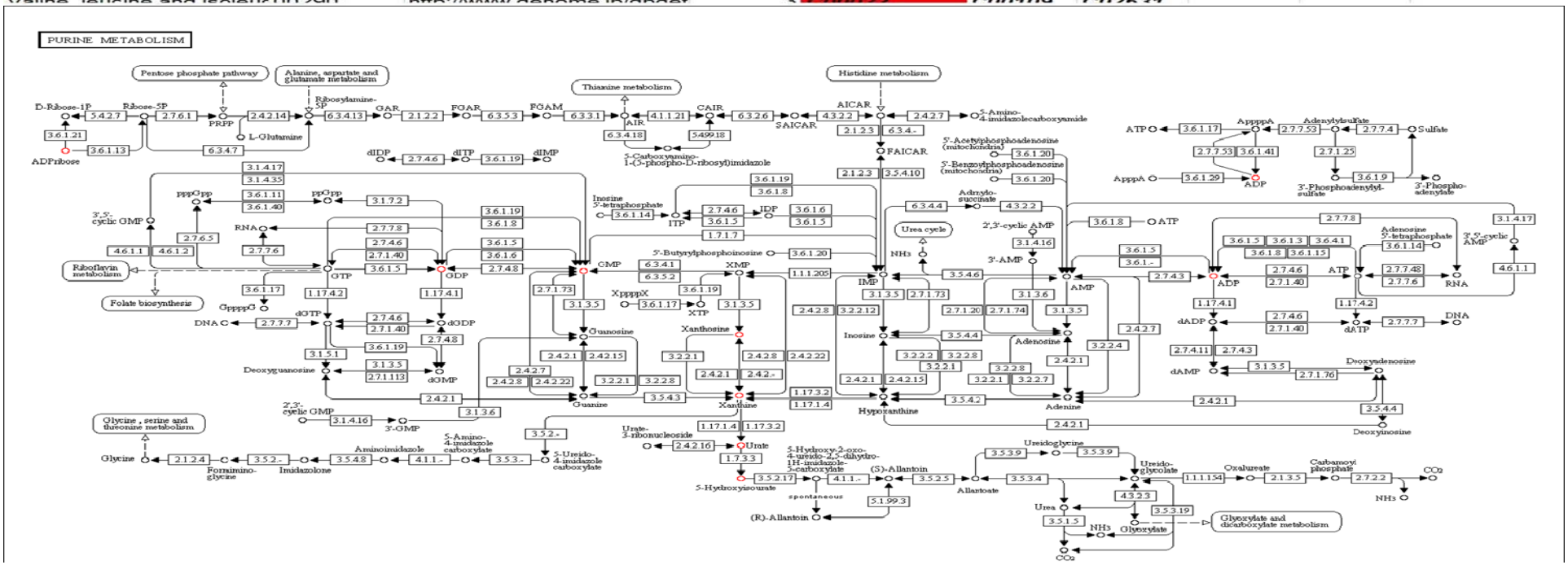


Figure.3.4 Score C of RESUME and the web link of the KEGG graphical chart of the pathway with the significant metabolites as red circles (inset).

3.5. Comparing MAK to Sieve® and XCMS:

3.5.1. Processing parameters:

The older versions of XCMS were not designed to accommodate LC-FTMS data, hence, the software crashed when decreasing the mass window to 0.001 amu to suit the high resolution of Orbitrap. The XCMS parameters were chosen as in the work of (Dunn et al., 2008) apart from band width which was kept as default to compensate for the HPLC/UPLC differences. Dunn *et al* (2008) tested all the XCMS parameters and indicated that the most important of these were band width (time window) and signal/noise threshold. The signal to noise (snthresh) argument was crucial for the success of the processing and the value was set to 1. The step (step) was set to 0.001 amu and the band width (bw) to 30 seconds.

The use of Sieve® is very straightforward with only three parameters amenable for manipulation. These are the m/z window size which was set as m/z width = 0.001 amu (lowest possible). The other parameters are the time window (set to 1 minute for positive mode and 1.5 minutes for negative mode) and the lower intensity threshold (set to 100,000 for positive mode and 50000 for negative mode). The time width was based on average peak width concluded in positive and negative modes for the same chromatographic method on a collection of 140 authentic standards (Chapter 2) and the intensity thresholds were extracted from a prior optimisation step on similar data (data not shown).

3.5.2. Assessment criteria:

3.5.2.1. Number of reproducible peaks:

The number of features has been employed on several occasions for comparing different analytical methods, provided identical quality control (QC) samples were injected repeatedly with a judgment criterion based on the number of features which pass a coefficient of variation (CV) filter. However, due to the unavailability of identical QC samples, the filtration step was based on the number of missing values (MV) per group. A 0% MV was allowed within one group. The MV filter was satisfactory since the ability of the processing software rather than the instrument for picking peaks was being tested. The number of features in the positive

ion mode analysis was highest in Sieve®, yet the MV filter led to a remarkable deterioration of the number of features in Sieve® but not MAK or XCMS. This does not come as a surprise as Sieve® does not take into consideration the peak shape, and reports any m/z in a continuum in the time domain as illustrated in Figure.3.5.

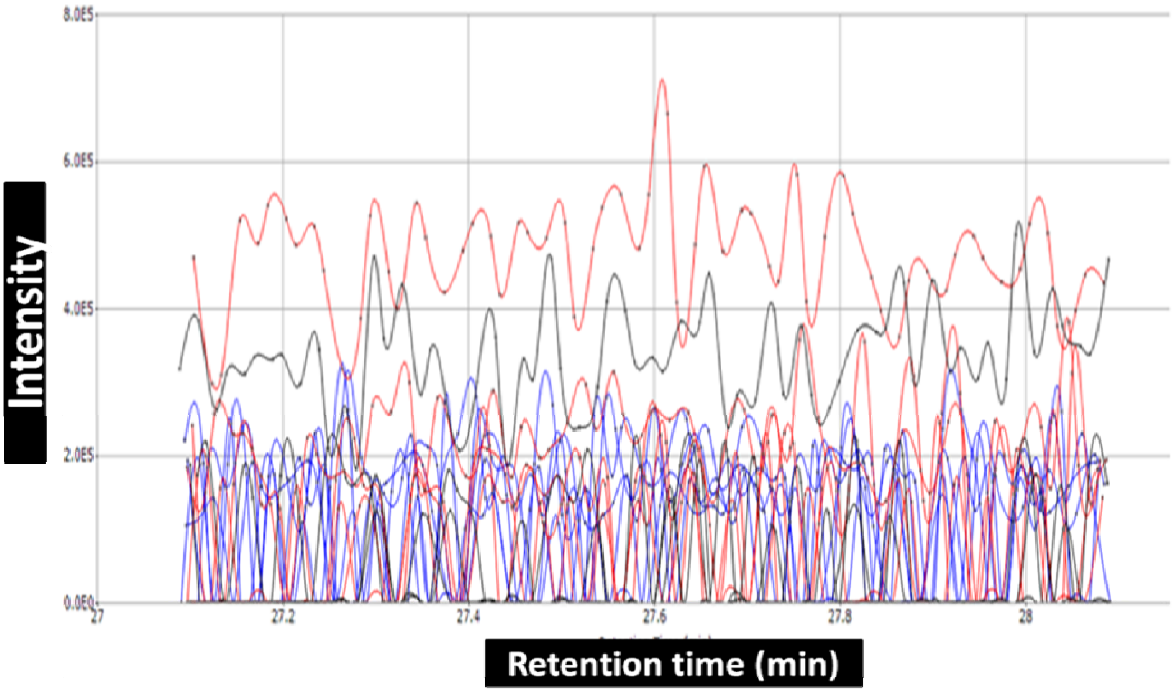


Figure3.5 An example of a feature (frame) picked by Sieve® indicating the lack of a Gaussian peak matching.

A second filter was applied based on a repetitive runs of blank samples (matching the sample solvent). The importance of this filter stems from the high background in the LC-Orbitrap, which may incidentally cause the report of false positive hits in the forward searching algorithm. These hits could even show some variance between groups in cases where ion suppression is encountered in its time region which itself is variable between biological groups. In order to avoid confusion with carryover, or incidental peaks, a feature is deemed to belong to the background ions if its average intensity (n=4) is at least twice the average intensity of the ion in the biological samples.

3% of the features in Sieve® or XCMS had matching blank features, while MAK seems to filter noise properly in its algorithm leading to an insignificant number of

blank matched features. Nevertheless, this happened at the expense of information on metabolites as the final number of features was only 10% and 30% of the number of features extracted from Sieve® and XCMS, respectively. Another factor explaining the low number of features in MAK is the loss of isomeric separation upon disregarding the chromatographic information i.e. each m/z will be represented by a maximum of one feature regardless of the real number of chromatographic peaks showing up, a drawback which is partially resolved in the other two packages.

The negative mode data did not follow a similar trend. XCMS showed a higher number of features, and kept its rank despite a loss of 25% of the features by the MV filter. MAK again was more resistant to blank based filtration but the total number of features remained lower than the other packages (Table 3.1).

	Positive			**Negative**		
	Total Features	**Reproducible features**	**Blank filtered**	**Total Features**	**Reproducible features**	**Blank filtered**
Sieve® 1.1	25000	4840	4703	1324	958	921
XCMS	1781	1640	1596	2930	2208	2191
MAK	20000*	482	480	20000*	728	724

Table.3.1 the number of features extracted by each processing method prior and post filtration

* the number is about half a million peaks per sample, cut to the most intense 20000 as a first filtering step.

The MAK FWD function can search a pre-defined list across the mass list (e.g. KEGG database) of both polarities. The picked peaks match the m/z of metabolites in KEGG, with the final matrix composed of potential metabolite/intensity probably with data structure of more biological relevance. The master mass list used here was the MS compatible KEGG metabolites (11371) and the MAK FWD found 1674 potential hits which occurred in 100% of at least one group and was not in the corresponding blank runs. 145 hits were found in both polarities, and 504 were isomers of other metabolites. The last two numbers were subtracted from the number of detected features in order to have an idea about unique ***potential*** metabolites as MAK does not distinguish isomers, which makes the record drop down to 1102 potential metabolites.

3.5.2.2. Mass accuracy reservation:

15 metabolites were selected to assess the differences between the three processing packages. The chosen metabolites ensure a gradient of: intensity, RT, peak shape and variance between the two groups. Table 3.2 describes the characteristics of the chosen metabolites highlighting the maximum, median and minimum of each feature. Additionally, the ^{13}C and ^{34}S isotopes of one of the metabolites (Methionine) were added to the list in order to provide an example of isotopic pattern detection by any package. Moreover, to make this subset of metabolites more representatives, the metabolites were chosen to fall in one of the following cases:

I. Group (I) incorporating time domain would be highly desired: adenine, hypoxanthine, xanthine and GABA occur in overlapping chromatographic peaks originating from the biological metabolites themselves and neutral losses from other metabolites, adenosine, guanosine, inosine and glutamic acid, respectively. Nonetheless, a baseline separation is achieved for all the previously mentioned peaks. Dihydropterin is associated with a second isomer happening far away in the chromatogram and with a weak signal.
II. Group (II) incorporating the time domain may or may not improve the outcome: the chromatographic method is not able to separate positional isomers leucine and isoleucine with their combined signal showing as one peak on the shoulder of another. See Figure.2.24.
III. Group (III) incorporating the time domain is not necessary: all other peaks occur as one single peak with varying inter-group and total average intensities.

metabolite/sample	Average	m/z	RT	peak width	ry-WT P.value
Adenine	27107063	136.06187	14.1	1.04	0.071
Arginine	1489330576	175.11901	22.65	0.52	0.011
cholinegrlycerphosphate	734588610	258.11034	18.25	0.78	0.0002
Citrulline	63907222	176.10312	17.6	0.71	0.049
dihydropterine	174982447	240.10942	13.33	0.95	0.061
drosopteridine	275642281	369.15322	17.6	0.73	0.001
GABA	94494371	104.07059	14.48	1.25	0.001
Glutamine	1523094466	147.07651	16.96	1.1	0.083
Hypoxanthine	2757843274	137.04584	10.68	0.72	0.006
Leucine/Isoleucine	2884612286	132.10195	12.8	1.88	0.073
Methionine	372970142	150.05846	13.73	0.87	0.722
Methionine_C13	15975472	151.06186	13.63	0.44	0.756
Methionine_S34	12412669	152.05684	13.63	0.49	0.791
Serine	264712787	106.04987	17.41	0.88	0.002
Uric Acid	339500433	169.03573	12.92	0.31	0.0000004
Xanthine	1014294108	153.04075	10.26	1.16	0.009
Table.3.2 characterstics of the metabolites chosen as asssessment standards.					
Index :	Maximum	Minimum	Median		

The peak parameters of these metabolites were manually extracted from 10 samples belonging to two distinct biological groups as will be extensively discussed in the next chapter. In brief 5 samples were of the ry and 5 were of the CS wild type of *Drosophila melanogaster*. The selection was based on a known genetic lesion in ry which helps to illustrate which packages are able to pick real biological information from the raw LC-MS data.

It is highly desirable for the output of any package to preserve the mass accuracy. The next step in processing the raw data is assigning identities to features, and high mass accuracy helps in reducing multiple false positive hits. The peak picking algorithm, in one way or the other, uses a bucketing functions (sum of peaks within an m/z width and a time width). The larger bucket size may cause the inclusion of nearby signals from neighbouring peaks, thus, reducing the final reported m/z accuracy. The mass accuracy measurement from the three test packages was determined and the difference from the manually extracted accuracies are listed in Table.3.3

metabolite/sample	Original accuracy (ppm)	MAK(ppm)	Sieve®(ppm)	XCMS(ppm)
arginine	0.343	0.037	0.779	0.053
glutamine	0.602	0.081	0.965	0.021
xanthine	0.427	0.534	0.926	0.138
uric acid	0.833	0.047	0.486	0.174
hypoxanthine	0.201	0.284	0.284	0.177
citrulline	0.820	0.090	0.828	0.029
adenine	0.740	0.118	0.779	0.074
GABA	-0.099	0.087	0.183	0.054
leucine/isoleucine	0.413	0.105	0.862	0.074
glycerphosphocholine	0.930	0.183	0.106	0.032
dihydrobiopterine	1.318	0.028	1.069	0.054
drosopterin	0.446	0.026	0.270	0.037
serine	0.026	0.066	0.726	0.095
methionine	0.869	0.078	0.211	0.025
methionine_34S	1.312	16.463	0.308	16.496
methionine_13C	1.013	missing	0.062	0.140
Average	0.637	1.215	0.553	1.105

Table.3.3 Difference of accuracy measurement between actual and recorded m/z in the three data processing packages.

The average of differences indicated that Sieve® was the best tool in terms of preserving accuracy. However, a closer look in the mass deviations shows that MAK and XCMS performed better in all samples but the low intensity methionine ^{34}S isotope, which seems to be mixed with a neighbouring peak in these two packages. Nevertheless, the average of differences is less than 1.3 ppm for any package which, in combination with the excellent accuracy achieved on the Orbitrap (average of 0.6 ppm) is still within a very acceptable range of searching any database with a window of less than 3 ppm.

3.5.2.3. Accuracy of integration:

For the same reason discussed above with regards to affecting mass accuracy measurement, bucketing can negatively affect the true integration of the metabolite. To measure how close to the actual measurement, the output of each package was compared to the manually extracted data for the 17 compounds in the list. As each package applies a different integration function, scale or divisor, direct comparison

will produce confusing results, thus, signals instead were normalised to the total areas of each feature in each set converting each individual compound into a proportion of the specified compounds in the set. Any value (*X'*) in table 3.4 will be given by the formula:

$$X' = abs(\frac{y}{\sum_{i=1}^{i=n} yi} - \frac{x}{\sum_{i=1}^{i=n} xi})$$

Where:

y: the manually integrated value for a given sample and a given metabolite.

X: *the* value reported by a given package for the same sample and same metabolite (*y*)

The values in the table 3.4 are differences of percentages so they are reported without units. The average of differences from the actual shares was used to compare the three software packages. 10 samples allow for a maximum of average 10% variation per sample. Each variation will be compensated for in other samples but in the opposite direction, which means that each difference value represents twice as much its real variation An average value of 0.5 was considered significant as it represents 2.5% variation in one sample (2.5% of the maximum variation of 10% multiplied by 2 to account for forced variation in the other direction). Table 3.4 reports at the end the average difference for each peak from each package. MAK, as expected, performed badly with peaks of group (I) apart from xanthine. For peaks with low intensity, MAK seems to deviate as well, with ^{34}S methionine varying significantly from the actual values and ^{13}C methionine missing entirely. The high consistency between MAK output and xanthine peak probably originates from the close biological correlation between xanthine and inosine in this sample set, where this can be considered a chance false positive fit. Dihydropterin integration was minimally affected by the MAK output as the intensity of the peak is much higher than its isomer, leading to a very small share in the MAK output, which may even be going in the same direction as dihydropterin peak. All the members of group (III) and even isoleucine/leucine peak were reported with the correct intensity by MAK.

metabolite/sample	MAK	Sieve®	XCMS
arginine	0.005	0.019	0.032
glutamine	0.004	0.022	0.03
xanthine	0.014	0.417	0.708
uric acid	0.032	0.041	0.072
hypoxanthine	1.507	5.051	2.748
citrulline	0.017	0.031	0.027
adenine	5.394	3.217	3.264
GABA	8.307	1.941	11.55
leucine/isoleucine	0.295	0.755	0.118
glycerphosphocholine	0.009	0.012	0.017
dihydropterin	0.142	0.044	0.14
drosopterin	0.009	0.015	0.039
serine	0.009	0.023	0.069
methionine	0.176	0.018	0.073
methionine_S34	2.886	0.18	1.799
methionine_C13	missing	0.162	0.154
average	1.254	0.747	1.303

Table 3.4 Mean differences of the normalised area from the .raw manually extracted area of the studied peaks. Highlighted cells are those which exceeded the suggested cut-off.

While the shortcomings of MAK were expected, the surprise came from the other two packages. Sieve® managed to avoid the integration problems of the weak peaks, but showed inappropriate integration of metabolites belonging to group (I). Furthermore, the integration of the Isoleucine/Leucine peak was deviated. Tracking back into the processed data, Sieve® seemed to have split the peak into two features (frames), however the split position was on the side of the peak not at the shoulder of the overlapping two peaks Figure.3.6.

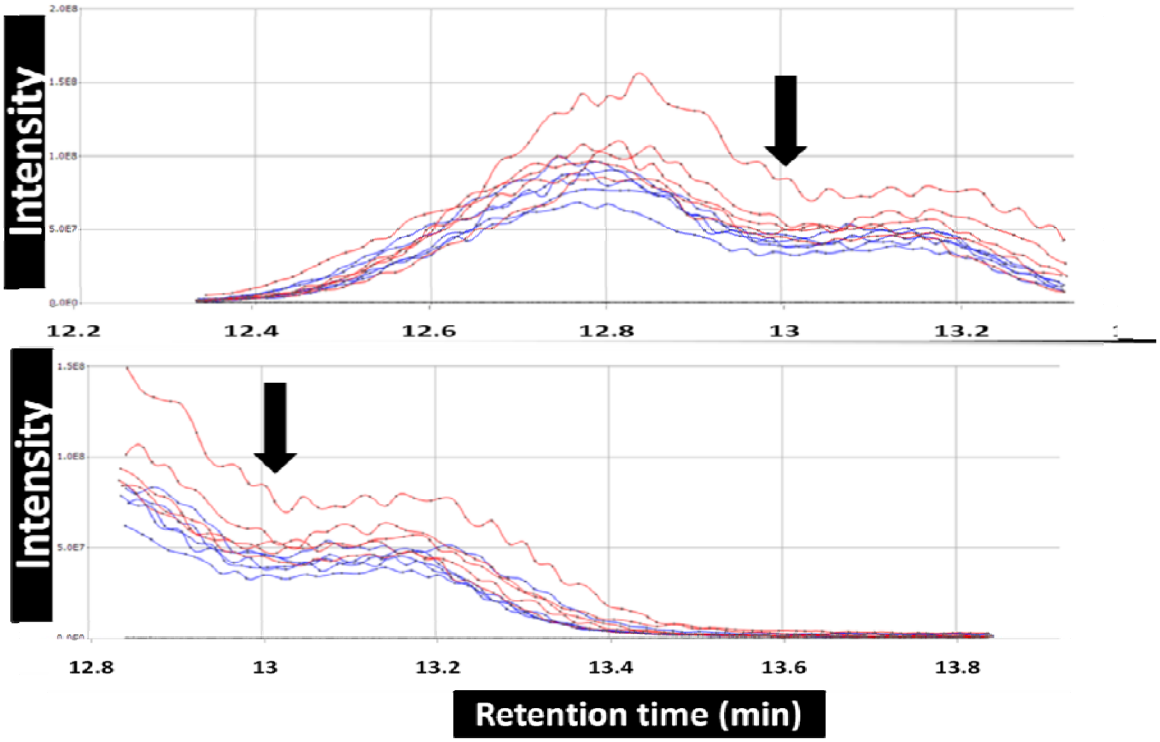


Figure.3.6 Leucin/Isoleucin split peak by Sieve®. The arrow indicates the shoulder position.

In fact, the manual check of the Sieve® generated peaks reveals many similar incidences. XCMS showed significant deviation from actual signals of group (I) peaks as well with GABA showing a huge difference (11%) from its original values. Tracking the XCMS generated Extracted Ion Chromatograms (EICs), the GABA peak is correctly assigned and the limits are set correctly as well. However, strangely, in the signal only one sample was integrated from the set of neighbouring peaks (Figure3.7). For hypoxanthine, one peak was misaligned and the three most intense peaks in the set were truncated at mid height from one side. The hypoxanthine peak was awkwardly missed with the neighbouring peaks in only half of the samples. GABA appeared to be mixed with its neighbouring isomers (e.g. a fragment of glutamic acid) ensuing a wrong integration of the real peak (Figure3.8).

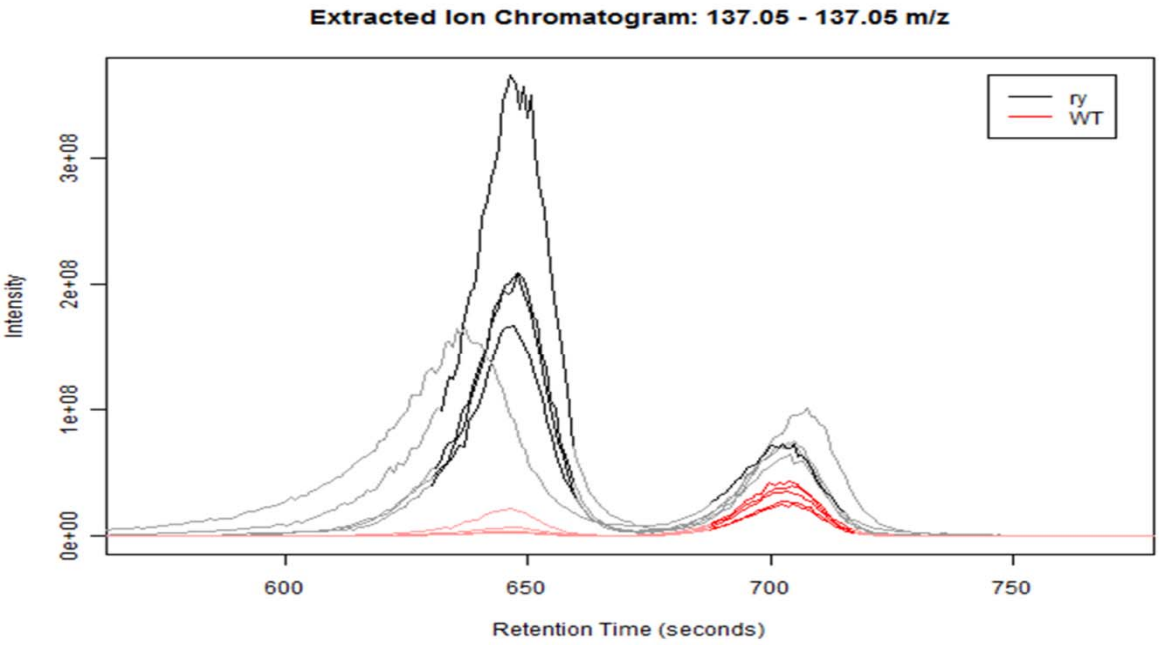


Figure.3.7 EIC of hypoxanthine showing the incorrect integration by xcms. The reported integration corresponds to the dark lines in the chromatogram.

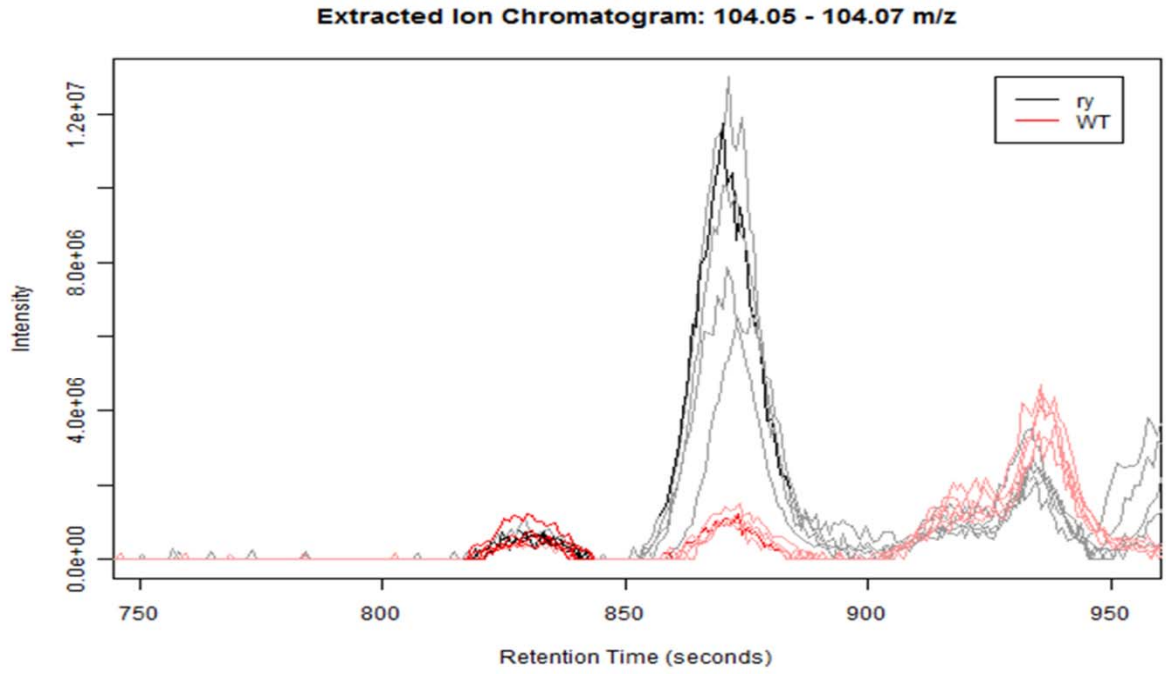


Figure.3.8 EIC of GABA showing the incorrect integration by xcms. The reported integration corresponds to the dark lines in the chromatogram.

Similar explanation applies for other members of group (I). Low intensity peaks were affected in XCMS as well. Manipulating the bandwidth for both Sieve® and XCMS yielded similar cases of incorrectly integrated peaks (data not shown). It is important to mention, that from applicability and convenience points of view, the traceability of integration error is much easier in XCMS and Sieve® owing to the feature EICs generated by both packages, compared to manual integration for validating MAK output. This is of special value for saving the analyst`s time in the case of runs of tens of samples per set.

3.5.2.4. Retention time alignment (XCMS, Sieve®)

MAK throws away the retention time domain, a process which corrects for any potential retention time shift when the case is one peak per the assigned m/z width. However, as discussed before, in the case of biological or occurring isomers or isobaric compounds even when clear chromatographic separation is achieved, the macro is not able to differentiate any deviation and will simply and improperly sum all the intensities of these features.

Both Sieve® and XCMS do retention time alignment with different algorithms. While XCMS uses correlation optimised warping (COW), Sieve® uses a local tiling technique followed by best correlation score for time correction. However, as seen above, the complexity of the data forces the packages to make mistakes, and although Sieve® and XCMS surpass the base MAK for taking the time domain into account, all the three reported mistakes when multiple chromatographic peaks share one m/z, and whenever the processing is checked, a final manual check of the .raw files is strongly recommended.

Table.3.5 shows the difference between the real RT of the chosen metabolites and the matching peaks from Sieve® and XCMS output. The Sieve® process seems to offer better consistency with original RTs than XCMS. While this has no direct impact on the interpretation of the results, it may be of special importance for selecting parameters of forward searching of databases incorporating a RT element.

metabolite/sample	Sieve®	XCMS
Arginine	0.027	0.141
Glutamine	0.011	0.129
Xanthine	0.041	0.130
Uric Acid	0.062	0.180
Hypoxanthine	0.421	0.561
Citrulline	0.001	0.127
Adenine	0.044	0.156
GABA	0.031	0.298
Leucine/Isoleucine	0.031	0.226
cholinegrlycerphosphate	0.009	0.123
dihydrobiopterine	0.009	0.084
Drosopterin	0.026	0.119
Serine	0.025	0.135
Methionine	0.074	0.105
Methionine_34S	0.067	0.107
Methionine_13C	0.074	0.106
Average	0.060	0.171

Table.3.5 Difference of RT reported by Sieve® and XCMS from the apex RT extracted manually

3.5.2.5. Convenience of use:

In terms of convenience of use, Sieve® is the friendliest among the three options. The graphical User Interface (GUI) is very simple and straightforward with a very limited number of parameters to control (m/z width, RT width and intensity threshold). The manipulations are windows based with control of events by button clicking. Second in ease of use come the macros (MAK) which have a user interface (Figure 3.9) as well with control of events happening via buttons and input boxes prompting the entry of values when appropriate. Nevertheless, unforeseen errors can happen, leading to a crash of the macro. Avoiding errors is attained by the careful reading of the instructions attached to each type of processing, but confusion is likely to occur with the many functions embedded in the macro, and the varying instruction for each function.

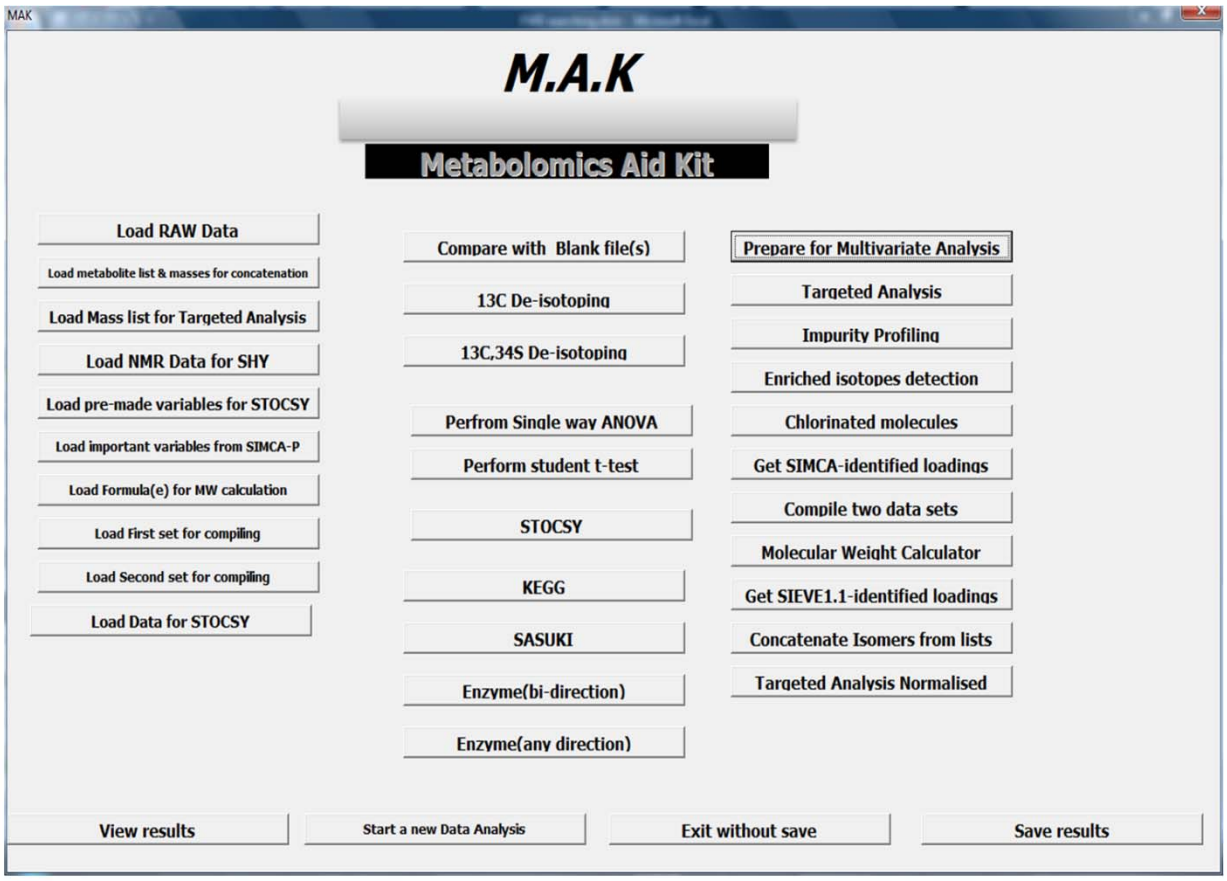


Figure.3.9 User interface of MAK showing most of the available functions.

XCMS is an R based package with a command line mode of control. it does not work on .RAW Thermo files, which should be converted to .CDF before commencing the processing. The command line has always been a problem for computer users as it requires careful attention for the dialled commands. Any typo in the command may lead to error message or, most dangerously, a whole set of defective outcomes. However, the basic functions of XCMS are easy to use and familiarity comes with practice. In a very personal opinion, due to the huge advances in data generation, I think proficiency in computer work has become essential for the 21st century mass spectroscopist.

An additional non-trivial advantage of the academic packages (XCMS, MZMINE, MAK....etc) is the availability. These packages are free to use, and as academics can learn the programming environment in which it was compiled, they

can be edited according to the need of each group which can vary considerably. Commercial packages have to be accepted as they are.

With regards to computing time, although MAK takes considerably less time than the other packages, it is fair to mention that the MAK follow up from the m/z peak picking of the Thermo software Xcalibur which itself takes about 2 minutes per sample. The transfer of data from Xcalibur to spreadsheets may cause additional user errors. Sieve® and XCMS roughly take the same analysis time when both operating at a m/z width of 0.001.

A drawback of the commercial software is the slow progress of versions, and the profit orientation of the production company. For example, Sieve® is well established in all its parameters for proteomics research but not for the niche market of metabolic profiling. Moreover, Sieve® refinement was advanced more slowly than Thermo technology advances, as for example only the latest (1.3) version of Sieve® can process the .raw data generated by Exactive (the new benchtop Orbitrap).

Taking all the factors into consideration, XCMS, despite the command line requirement, seems to offer a balanced output in terms of reporting RT, peak integration, number of reproducibly occurring features in both positive and negative modes and its ability to filter background features.

3.6. Combining XCMS with macros (added value)

Forward searching is a bright idea for data reduction by narrowing down the scope of information into what may be of real biological relevance. It has been widely deployed in GC-MS but not LC-MS metabolic profiling, which is beset by the lack of consistency including inconsistent retention information due to the use of different stationary phases, column dimensions and elution programmes in the different laboratories. LC-MS database construction has been the responsibility of each laboratory on its own, with only one database available (Brown et al., 2009) which is applicable only when exactly the same stationary , mobile phases and gradient programmes are used.

There has been a recent awareness of the importance of forward searching for LC-MS processing software makers (e.g. Metabolynx) which, unfortunately, cannot process LC-Orbitrap data. Only the most recent versions of Sieve® (Sieve® 1.3) allows a seed file composed of binary values of m/z and RT to be screened across the output. XCMS has a built in function to screen for a user defined database. However, as with other XCMS functions, it is not easy to use and does not provide statistical test options apart from t-test.

Since macro functions are easily editable to cope with different output styles, and in an attempt to find a compromise between the tedious Metlin link generated by XCMS and the lack of chromatographic information by forward searching in MAK, a sub-routine was developed to screen XCMS output, positive and negative and combine all the data then execute the other useful MAK functions (statistical test, colouring and reporting), all with a click of a button.

To this purpose, the metabolite/RT list generated in chapter 2 was used. The manually observed metabolites compose a collection of 111 metabolites representing 93 unique masses. Using the XCMS-MAK combination, 95 (60 unique metabolites) metabolites from both polarities could be identified correctly in *Drosophila* samples. The missed metabolites can be divided into two categories of true misses (true negative) for metabolites which did not turn up in the samples, and wrong misses (false negative)where metabolites were in the samples but XCMS could not pick them. Most of the metabolites in the second category have closely eluting isomers.

A matching test was performed on MAK forward searching function. Due to its lack of chromatographic information, the maximum number of unique masses to be identified is only 93 and the macro could pick 116 peaks representing 75 unique masses (Table 3.6).

	Masses	theoretical	practical	single polarity	dual polarity	coverage
Maximum	111	222	160	62	49	111
X-calibur full mass list	93	186	116	38	37	75
XCMS peak list output	111	222	95	29	31	60

Table.3.6 Applying MAK FWD to X-calibur mass list or XCMS output

Although MAK was able to cover more metabolites than XCMS, the incorporation of the RT information in the latter makes the assignment more credible and compliant with the minimum requirements of Metabolomics reporting standards initiative (van der Werf et al., 2007).

3.7. Conclusion:

A description of a new in-house built tool for analyzing LC-Orbitrap data (MAK) was reported, assessed, and criticized. Following on, a comparison between MAK and two other options (the commercial Sieve® and the freely available XCMS software) was made on the basis of the number of features, m/z accuracy, peak integration accuracy, retention time alignment and convenience of use.

The lack of chromatographic information favours XCMS and Sieve® over MAK but the three tools make good candidates for initial screening of the metabolite profiles, with more or less similar credibility for mass accuracy, peak intensity but easier visual assessment of data in XCMS and Sieve®. Overall, it is highly recommended to re-check any interesting biomarkers in the .raw files, especially for MAK.

The results of the XCMS-MAK integration looks promising, and with improvement of the confidence and coverage of the database could make a powerful tool for the fast and reliable discovery of affected pathways and even potential lesion position.

At the very end of this chapter, it is important to re-emphasise that at the time the instrumental analyses forming the bulk of the content of this thesis was carried out, only MAK was available to processes the data and served as a useful tool when there was no alternative and may still have some useful unique functions despite the advances in other processing packages.

4. Metabolic Profiling of *Drosophila* Extracts by ZIC-HILIC-FTMS:

4.1. Introduction:

Metabolic profiling has proved a valuable approach in simple unicellular organisms (Oliver et al., 2002, Allen et al., 2003), and metabolomes have also been described for humans (Wishart et al., 2007). While human metabolic profiling is necessarily observational, studies of simpler organisms offer the prospect of reconciling experimental genetic lesions with their impact on the metabolomes, both to understand existing pathways and to elucidate new networks. Of the metazoans, *Drosophila melanogaster* offers perhaps the best trade-off between genetic tractability, availability of well-characterised genetic mutant stocks, and organismal complexity (Dow, 2007, Dow and Davies, 2003) for such metabolomic approaches. In particular, some mutations in metabolic pathways have been studied for nearly a century (Morgan, 1910); and interactions with, and epistatic interactions between, similar mutations, established. It was thus of interest to see if *Drosophila* is amenable to the powerful analytical technologies that have been established over the last few years.

There are numerous number of examples in which *Drosophila* has been used for hypothesis testing in molecular biology, owing to the ease of manipulating its genetic composition in a fast and inexpensive (Berg et al., 2010, Renn et al., 2010, Shemesh and Spira, 2010, Sellami et al., 2010). However, the modern analytical technology has not been extensively utilised to investigate the biochemical consequences in these manipulations with most of the work reliant on simple clinical chemistry ki(Jumbo-Lucioni et al., 2010, Ristow and Zarse, 2010, Mamolen et al., 2010). Small molecule detection in *Drosophila* is restricted to a handful of applications using NMR on extracts (Pedersen et al., 2008) , in-vivo NMR (Null et al., 2008) profiling of sphingadienes by ESI-MS (Fyrst et al., 2008, Fyrst et al., 2004) and cuticular pheromones sampled from the surface of awake flies by direct analysis real time TOF-MS (Yew et al., 2008).

As a test-case, the second mutation ever characterised in *Drosophila* (around 1916) was selected. *ry*, originally identified as an eye colour mutation with reddish brown eyes, was found to encode xanthine oxidase / dehydrogenase, an enzyme in the

purine metabolism pathway (Keith et al., 1987). Through the foresight of the *Drosophila* research community, the original mutations (together with hundreds more) are still extant in stock centres. *ry* mutants accumulate the enzyme's substrates, xanthine and 2-amino-4-hydroxypteridine as larvae plus hypoxanthine in the adult; xanthine forms calculi in the Malpighian tubules, bloating and distorting them (Bonse, 1967). The *ry* mutants lack the corresponding enzyme products uric acid and isoxanthopterin (Mitchell and Glassman, 1959). Mutants are also sensitive to administration of purine to the medium (Glassman, 1965). Mutations are semilethal, but viability can be increased on low-purine diets, and the *rosy* eye colour can be phenocopied by the anti-gout drug allopurinol (an isomer of hypoxanthine) (Boni and Parisi, 1967). Remarkably, *ry* mutations thus exactly recapitulate xanthinuria type I, the renal disease caused by mutation of the human homologue of *ry*, across over 400M years of divergent evolution (Dent and Philpot, 1954).

The *ry* corrupted pathway thus provides a good test of modern analytical technology, both because results can be reconciled with painstaking classical findings, and because there is scope to discover new impacts of the mutation that may have escaped the best available analytical chemistry of its day. The *ry* mutation still attracts interest and recently work was published indicating that the xanthine oxidase gene, which is absent from *ry*, is necessary for mediating the effects of juvenile hormone (JH). In the absence of xanthine oxidase the cuticle of the insect did not develop properly when stimulated by JH (Zhou and Riddiford, 2008).

Mass spectrometry has been applied to metabolic profiling (Wilson et al., 2005, Southam et al., 2007, Plumb et al., 2006, Williams et al., 2005, Michaud et al., 2008, Kind et al., 2007). High mass accuracy mass spectrometry offers great potential for exploration of cellular metabolomes. Two approaches have been used, based on either time-of-flight or Fourier transform instruments (Underwood et al., 2006, Breitling et al., 2006b).The FT-MS studies reported initially involved the use of magnetic ion cyclotron resonance instruments often using direct infusion of the sample. However, the use of Orbitrap-FTMS has become widespread in the last four years. Direct infusion is conducive to high throughput analyses but ion suppression offers a significant challenge (Breitling et al., 2006b) and the absence of

chromatography makes it impossible to distinguish isomers and isobaric compounds, as discussed in chapter 2. The Orbitrap Fourier transform mass spectrometer offers similar performance to classical FT-MS but without the requirement for a high strength magnetic field (Qizhi et al., 2005). The Orbitrap is readily coupled to chromatographic systems and the use of chromatography reduces the risk of dominant metabolites suppressing the ionisation of more minor species. The main difficulty with the technique is in extracting information from the massive, complex data sets generated; currently there is no ideal method for data mining. Principal component analysis (PCA) among other types of multivariate analysis, has been widely used in NMR studies to differentiate one set of samples from another (Rezzi et al., 2007, Ebbels et al., 2007, Kochhar et al., 2006, Skordi et al., 2007). There are some difficulties in using it with high resolution MS data in that the loading plots generated contain many data points necessitating application of further data interrogation methods for extracting the most out of the metabolic profiling experiment. In this chapter, it will be demonstrated that global metabolic profiling is readily achievable with a small amount of starting material (at most 10 flies per sample). The technology did not only verify earlier findings of changes induced by *ry* in the purine metabolism and eye pigment synthesis pathways, but also found that metabolic perturbations can be traced further from the molecular lesion than was previously possible. In addition, entirely unexpected changes in a range of apparently unrelated metabolic pathways, some of which can be rationalised, were observed.

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. Drosophila stocks

Drosophila melanogaster were kept in vials of standard medium (Ashburner, 1989) at 25ºC and 55% r.h. on a 12:12 photoperiod. Stocks used were Canton S wild-type flies, and ry^{506} homozygotes (Bloomington stock Center).

4.2.3. Extraction of Flies

The extraction of the flies was carried out as mentioned in chapter 2 with male and female flies extracted separately.

4.2.4. LC-MS Method

The LC-MS method used here was the same as mentioned in chapter 2.

4.2.5. Data Processing

Data pre-treatment and preparations for multivariate analysis were performed using Excel-Visual basic application (VBA) subroutines (Microsoft Excel-2007) developed in-house. Multivariate analysis was carried out using SIMCA-P version 11.0.0.0 Umetrics AB.

4.3. Results and discussion:

4.3.1. High resolution mass spectrometric variable extraction:

Figure.4.1 shows the total ion current (TIC) profiles of a methanol/water extract of whole wild type and *ry* fruit flies, these profiles are not informative. However, ion ranges can be extracted from the data set for individual components as shown in Figure. 4.2 where extracted ion traces are shown for some metabolites associated with purine catabolism (Kanehisa et al., 2002): uric acid and its metabolites allantoin, xanthine, hypoxanthine, guanine, guanosine and guanosine monophosphate.

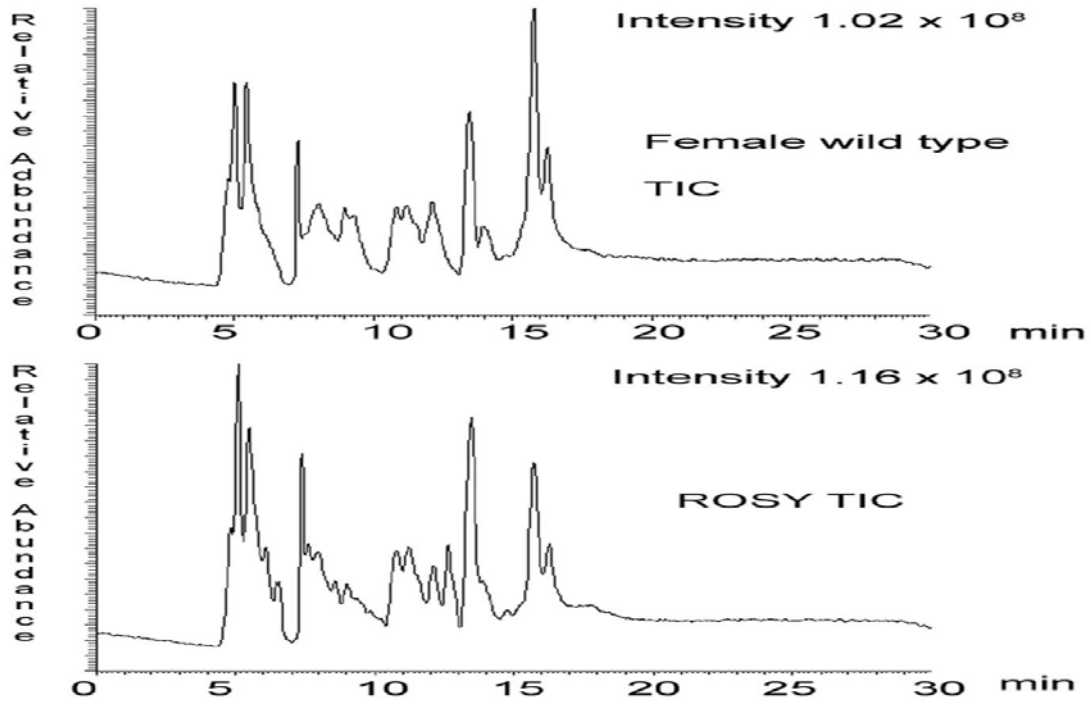


Figure.4.1 Total ion current traces (TIC) for extracts from wild type and ry flies run on a ZICHILIC column in positive ion ESI mode.

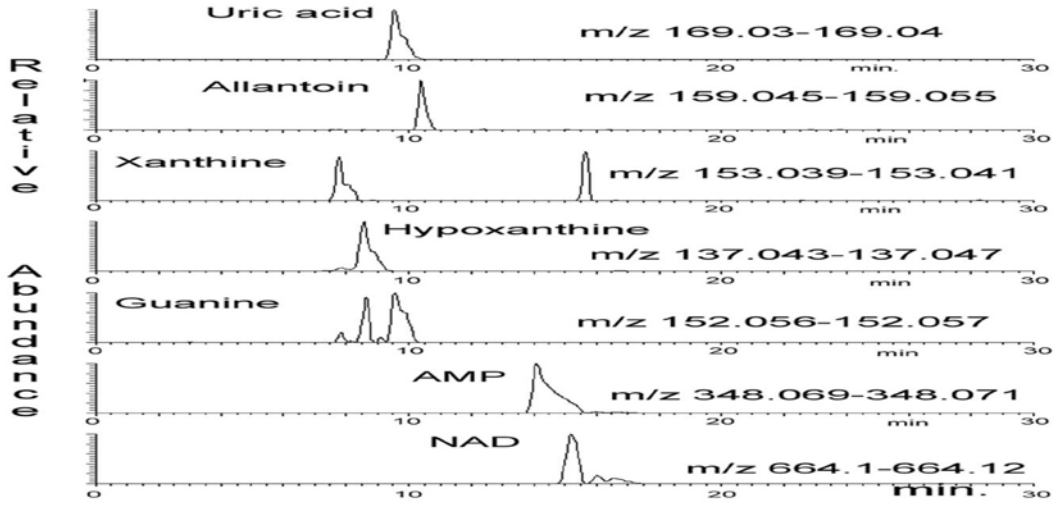


Figure.4.2 Narrow-range extracted ion traces extracted across a range of 0.02 amu showing the metabolites related to uric acid biosynthesis from female wild type *Drosophila*.

Table 4.1 shows the elemental compositions for compounds related to pathways upstream from uric acid formation with an indication of the closeness of the mass match to the predicted elemental composition, the deviation in ppm and the elemental composition for the nearest matching rival composition. Often the other elemental compositions provided by the software (Xcalibur Qualbrowser) are biologically unlikely or not close enough to be a good match; with a freshly tuned instrument it would be expected that mass matching would be within 2 ppm of the true elemental composition.

Compound	Elemental Composition [M+H]+	Deviation ppm	Nearest match	Deviation ppm
Uric acid	C5H4N4O3	0.552	-	-
Allantoin	C4H7N4O3	0.083	-	-
Xanthine	C5H4N4O2	-0.34	-	-
Hypoxanthine	C5H4N4O	-0.783	-	-
Guanine	C5H5N5O	-0.22	-	-
Guanosine	C10H13N5O5	0.298	C11H13N6O2Na	-0.669
Biopterin Isomer	C9H11N5O3	-0.025	C10H11N6Na	-1.179
Biopterin Isomer	C9H11N5O3	-0.109	C10H11N6Na	-1.263
Dihydrobiopterin	C9H13N5O3	0.266	C10H11N6Na	-0.878
Inosine	C10H12N4O5	-0.088	C11H20ONNaS2	0.802
AMP	C10H14N5O7P	0.02	C15H11O4N5Na	0.128

Table.4.1 Mass spectrometric data for metabolites in close proximity to the xanthine oxidase lesion in ry.

Thus we can be very confident of the accuracy of a particular elemental composition selected by the software. Another point to observe from Figure 4.2 is that the retention times make sense in terms of the structures of the compounds. Thus uric acid, xanthine, guanine and guanosine all elute at similar retention times from the HILIC column but are, importantly, removed from the lipid peaks which all elute before 7 minutes. Phospholipid peaks generally overwhelm data obtained in direct infusion mode because of their high abundance and the efficiency with which they are ionised in ESI. Adenosine monophosphate elutes much later than the rest of the metabolites because of its stronger interaction with the hydrophilic layer associated with the column surface, this was observed for all mono and di-phosphorylated metabolites, however, the current method was not able to reliably observe triphosphates.

4.3.2. Identification of metabolites on the basis of elemental composition

Data on the relative levels of various metabolites can be extracted directly from the runs using a targeted compound approach such as that shown in Figure 4.2 but the major problem with the data sets obtained from FT-MS is that they are huge and there was no software package readily available to extract peaks and compare variations between groups analyzed from LC-Orbitrap data at the time this work was done.. The benefit of FT-MS is that one can be very sure of the identity of a compound with regard to its elemental composition. However, a particular elemental composition can often match several structures so some judgement and conformational analytical experiments have to be exercised as to what are the likely structures. To this purpose, the metabolomics standards initiative permitted annotations of metabolites where two orthogonal identification methods have been provided. Accurate mass and retention time of authentic standards will be used through this chapter. The established knowledge of the studied lesion adds as well to our confidence in assigning identities to some of the metabolites which have multiple potential elemental compositions. The METLIN data base (Smith et al., 2005) contains over 15000 metabolite structures and provides useful additional confirmation of compound identity and the compound hits discussed below mainly have only a single match within this data base. This is not to say that isomers cannot occur and this is the case with biopterin (Table 4.1) where a number of isomers are possible .The case is further complicated by the isomers resulting from analytical artefacts such as adduct peaks. In addition, another type of isomer may arise from the spontaneous fragmentation of metabolites in the ion-sources. The elemental composition $[M+H]^+$= C5H6N5 corresponds to a unique metabolite, adenine, in the METLIN database. The accurate mass did not coincide in the same database with any other metabolites even when permitted deviation was increased to 7 ppm. However, the chromatogram of the matching mass shows two peaks. The second peak for adenine was confirmed to be a fragment of adenosine (Figure.4.3). To cope with this hurdle, correlation matrices combined with visual inspections were used.

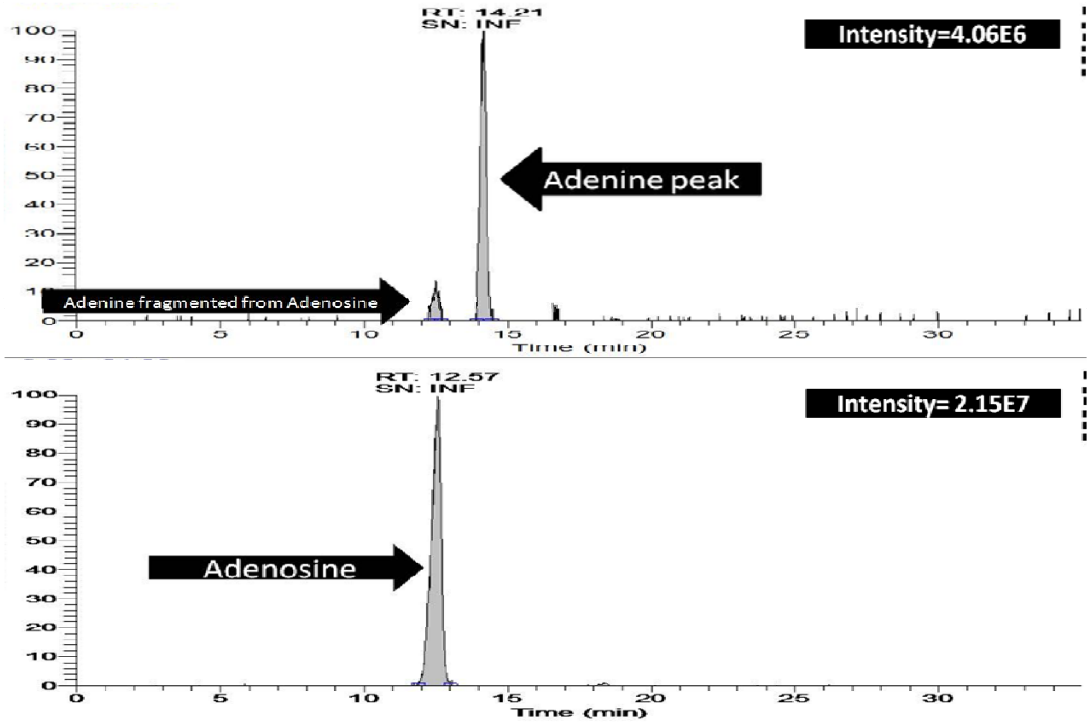


Figure.4.3 An example of occurring isomers. The upper EIC of adenine shows two peaks, the first of which was found to be a spontaneous fragment of adenosine (lower EIC.)

4.3.3. Data processing:

In order to carry out multivariate analysis we had to dispense with the chromatographic dimension and sum the ions across the whole chromatographic range. Although this may seem to be discarding chromatographic information, once a hit is generated the extracted ion trace can be checked back in the raw data. However, with the summing approach compounds will be missed due to the necessary filtering step, particularly those with low intensity ions, but as we demonstrate below a huge amount of statistically valid information can be collected quite quickly. Two different analysis approaches were used which will be described as being either forward searching (metabolites expected to be affected by the lesion) or global (all features amenable to detection by the analytical system).

4.3.3.1. Comparison of samples using forward (targeted) searching

In this approach a VBA sub routine was developed to look up the most common mass forms (M+H, M+Na and M+2/2) of 59 metabolites directly involved in the purine / pyrimidine metabolism pathway (Kanehisa et al., 2002). The intensities of the different mass forms of the same metabolite were summed using the VBA routine. The analytical method (positive polarity) was able to capture 23 of these metabolites, which were introduced into SIMCA-P+11 for PCA figure.4.4 in which *ry* and WT flies were clearly separated, as expected, with uric acid (in the wild type region) and xanthine and hypoxanthine (in the *ry* region) playing a dominant role in clustering the sample types. Other metabolites found to be highly variable between the two groups were adenine, AMP, adenosine and deoxyadenosine, which were up-regulated in *ry* mutants; and allantoin, uridine, uracil and thymine, which were down-regulated in *ry*.

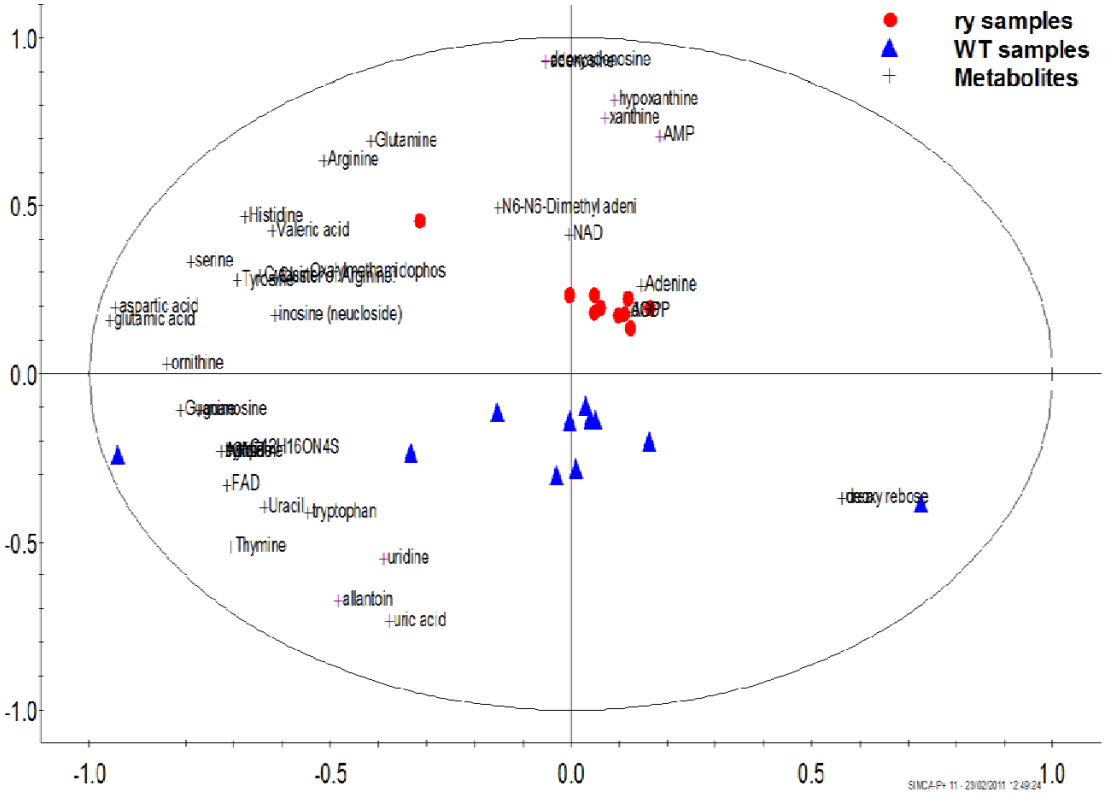
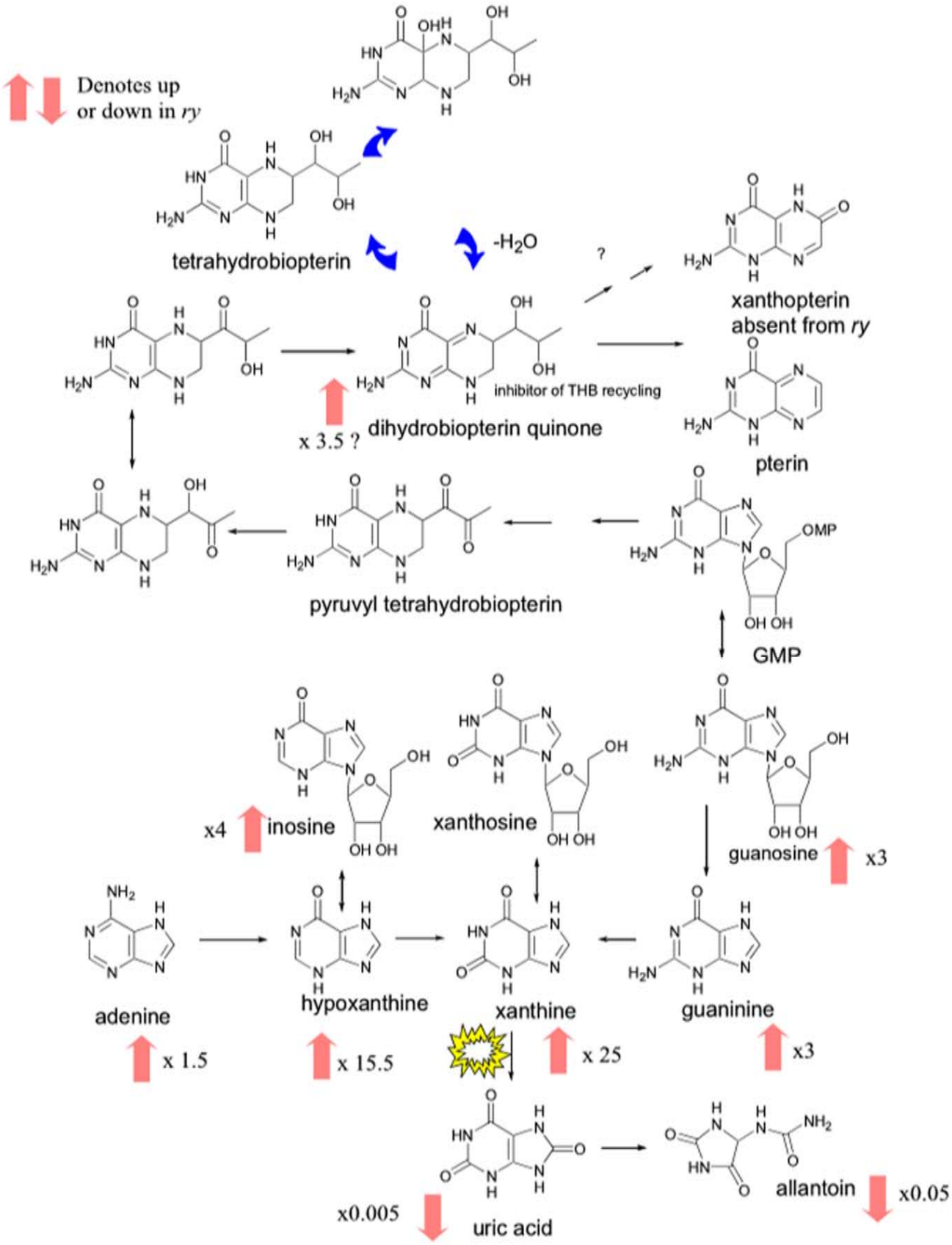


Figure.4.4 Bi-plot of *Drosophila* extracts with metabolites associated with purine/pyrimidine metabolic pathways. The bi-plot can show simultaneously the axis of separation between the groups and the metabolites forcing the separation as they position closer to the group in which they are highly expressed.

There were some differences between male and female flies, but broadly the picture was similar. Scheme 4.1 shows the metabolic pathways most closely associated with the conversion of xanthine into uric acid taken from the KEGG data base (Kanehisa et al., 2002). Significant effects related to the absence of xanthine oxidase were observed for many of the metabolites shown in Scheme 4.1. Most of these effects are probably related to feedback inhibition where metabolites accumulate upstream from the point where xanthine or hypoxanthine would normally be converted into uric acid. Since xanthine oxidase is absent in *ry* then the xanthine accumulation must derive from guanine since xanthine oxidase is also required to convert hypoxanthine into xanthine. It is well known that the absence of xanthine oxidase affects pigmentation in *ry* flies and for instance xanthopterin which is produced via the action of xanthine oxidase was undetectable in *ry* flies. So the mass spectrometric profile rapidly confirms what is known from classical work (Bonse, 1967, Mitchell and Glassman, 1959), although it clearly detects perturbations further from the original lesion than was previously possible. It is also able to identify effects in the guanosine/guanine arm of the purine metabolism pathway, had previously not been described. However, a further strength of this metabolic profiling approach is that global, as well as targeted changes can be obtained from the same data sets. The potential features were assigned IDs after matching with confirmational runs of authentic standards.



Scheme.4.1 Some of the metabolic network in close proximity to the ry lesion.

4.3.3.2. Effects on unrelated metabolic pathways (global screening)

In the non-targeted approach to the determination of metabolic effects full spectrum lists of all the samples were extracted by summing them across the full chromatographic range using MAK to eliminate artefacts and background ions and to correct and to align the samples in the m/z direction and thus generate a matrix with rows as variables (the intensities of the same m/z values in all the samples) and columns as observations (the intensities for all the ions recorded in one sample). Full descriptions of these routines are discussed in chapter 3. The generated matrix was then exported into SIMCA-P for multivariate analysis. In the untargeted procedure, we moved from pathways known to be affected by genotyping to tracking the whole system in order to capture knock on effects further from the metabolic perturbations in the region of the lesion. PCA of the global profiles of both sexes showed a clear clustering of wild type compared with *ry* flies (Figure.4.5).

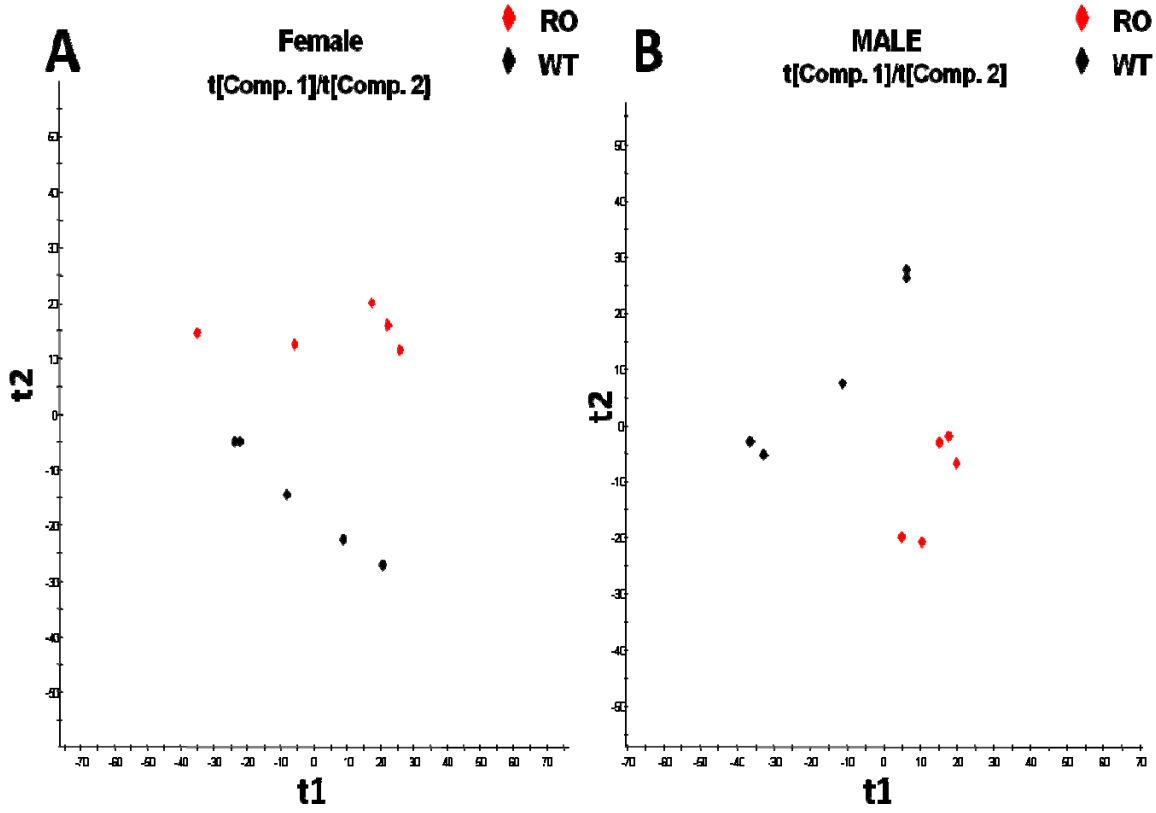


Figure.4.5 PCA scores plot (Par scaled) of female (A) and male (B) *drosophila* extracts showing good distinct metabotypes of ry (red) and WT (black).

Metabolites which were found to vary significantly between ry and WT were confirmed up by entering extracted ion chromatograms into Xcalibur 2.0.6 and determining the elemental composition for the observed chromatographic peaks with an allowed deviation from the exact mass of ± 2 ppm. The elements included in the queries were (C, H, N, O, S, P and Na). In the case of zero hits, other elements were added (starting with calcium, then magnesium and finally silicon) in order to elucidate the structure and generate a list of potential adducts/background ions. Compounds not matching elemental compositions containing these elements were disregarded for the present. In case of more than one hit, any chemically impractical formulae were excluded. (These are formulae like $C_{16}H_{19}O_2N_2P_3$ where phosphorus has to be linked directly to carbon which is not likely to happen in a biological system, or $H_{16}N_2O_6$ where there is no carbon in the formula.) Isotopic patterns were additionally used to prune the number of suggested formulae and confirm the elemental composition (Kind and Fiehn, 2006).

A mass was considered unequivocally linked to one specific formula if:

1. Only one hit for the mass was shown within 2 ppm of the predicted elemental composition.
2. All but one hit of the mass were not "chemically reasonable" formulae.

If any of the previous two criteria were not met, all other potential formulae were used for searching Metlin.

4.3.3.3. Statistical total correlation spectroscopy (STOCSY)

STOCSY has recently been applied in NMR-based metabonomics approaches (Cloarec et al., 2005a, Keun et al., 2008, Wang et al., 2008, Coen et al., 2007b, Smith et al., 2007, Holmes et al., 2007, Holmes et al., 2006). It allows the detection of signal areas which correlate highly across different samples. This may indicate either, peaks which originate from the same metabolite, thus facilitating correct elucidation, or of another metabolite which shows similar response upon perturbation. A mass spectrum of a metabolite obtained under ESI conditions should "theoretically" turn up as a single peak, but this is not the case in practice, as many chemical compounds turn up in multiple mass forms (adducts, multiply charged,

dimers, clusters, fragments etc) which makes STOCSY potentially useful for detecting peaks with multiple mass forms as well as those with similar metabolic responses. STOCSY was applied to our MS data with correlation cut-off values set to 0.8 for positive correlation and -0.8 for negative correlation. In NMR, the correlation is calculated between a "driver peak" and the rest of the peaks , very high correlations are deemed likely to derive from the same metabolite, and intermediate positive, or negative correlations (intermediate to high) of different metabolites with similar or opposite profiles, respectively. MS is not as repeatable as NMR, and signals show more fluctuations. However, the same principles of STOCSY can be edited to accommodate MS data as well. Highly correlated peaks only are considered and those with similar retention times are more likely to be from the same metabolites. Similar approaches have been reported in the literature about the same time or shortly after I used this method for data mining (Mohamed et al., 2009, Ivosev et al., 2008, Werner et al., 2008).

A loadings plot of the PCA of all features was used to extract driver peaks, which were investigated for correlations with other peaks in the matrix in order to extract more latent information. To illustrate the application of STOCSY, the correlation list for a selected driver peak glycerophosphoethanolamine (GPE), which was initially picked up by PCA, is considered. GPE gave 84 strong correlations across the mass/intensity matrix. With the confirmation procedure at least twenty masses were identified. These were either other mass forms of GPE or of other metabolites which had the similar behaviour across the samples (Table 4.2).

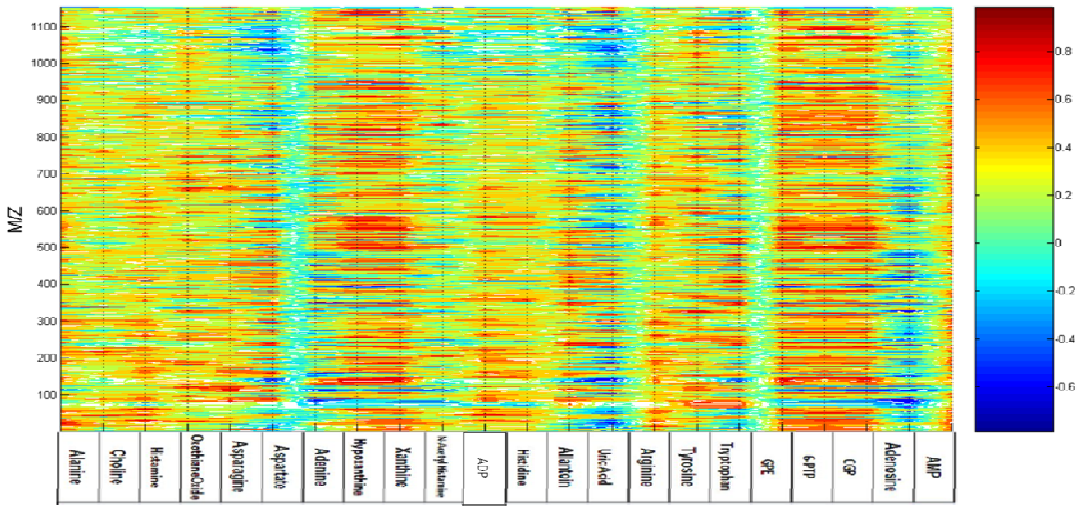
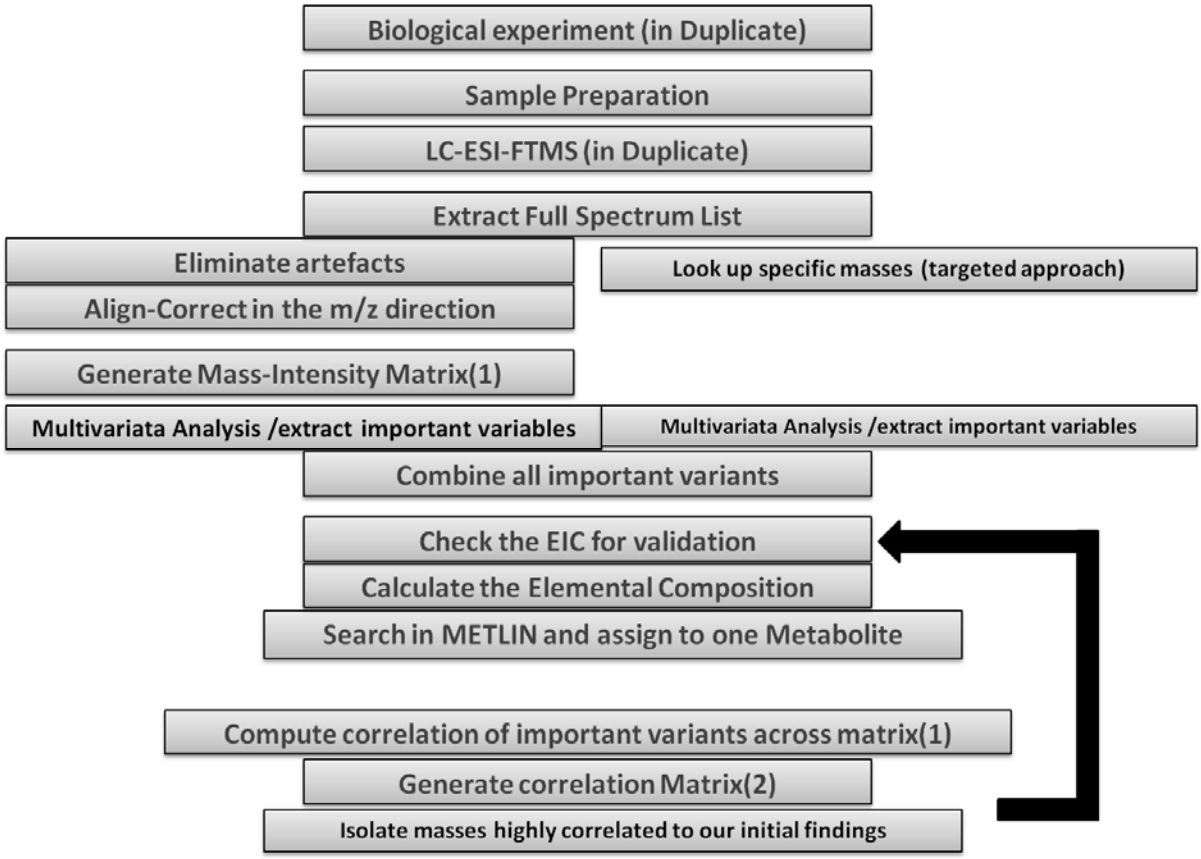


Figure.4.6 A correlation heat map for some of the metabolites identified to be affected by the lesion in the PCA loadings plot.

The correlation matrix of some variant metabolites (driver metabolites) with the whole features set is shown in (figure 4.6). All masses correlated to our previously identified potential biomarkers (urate, xanthine, hypoxanthine, uridine etc) were listed for confirmation by using Xcalibur This correlation approach extracted 35 additional variant features in male, and 57 variant features in female flies. The total work flow is summarized in (scheme.4.2).

M/Z	Correlation Coefficient	Compound
137.0457	0.888	HypoXanthine
138.0496	0.886	Hypoxanthine (13C isotope)
139.0502	0.884	Urocanate
148.0219	0.803	Amino-Alanine
153.0413	0.927	Xanthine
154.0441	0.839	Xanthine (13C isotope)
159.0278	0.969	HypoXanthine (sodium adduct)
164.0567	0.93	Pterin
186.0387	0.906	Pterin (Sodium Adduct)
198.0535	0.999	GPE (Loss of H2O)
216.0632	1	GPE
238.0452	0.995	GPE (Sodium Adduct)
239.0484	0.803	GPE (Sodium Adduct) (13C isotope)
240.1091	0.947	DihydroBiopterin
241.1127	0.982	DihydroBiopterin (13C isotope)
254.019	0.985	GPE(Potassium Adduct)
260.028	0.933	GPE (Di-Sodium Adduct)
431.1208	0.982	GPE*2 (cluster)
453.1009	0.987	GPE*2 (cluster) (Na adduct)

Table.4.2 STOCSY for a selected example GPE (GlycerPhosphoEthanolAmine) which shows strong positive correlation with Xanthine/Hypoxanthine mass forms , Urocanic Acid and multiple mass forms of GPE itself.



Scheme.4.2 Work flow of the experiment conducted to investigate *Drosophila* metabolome.

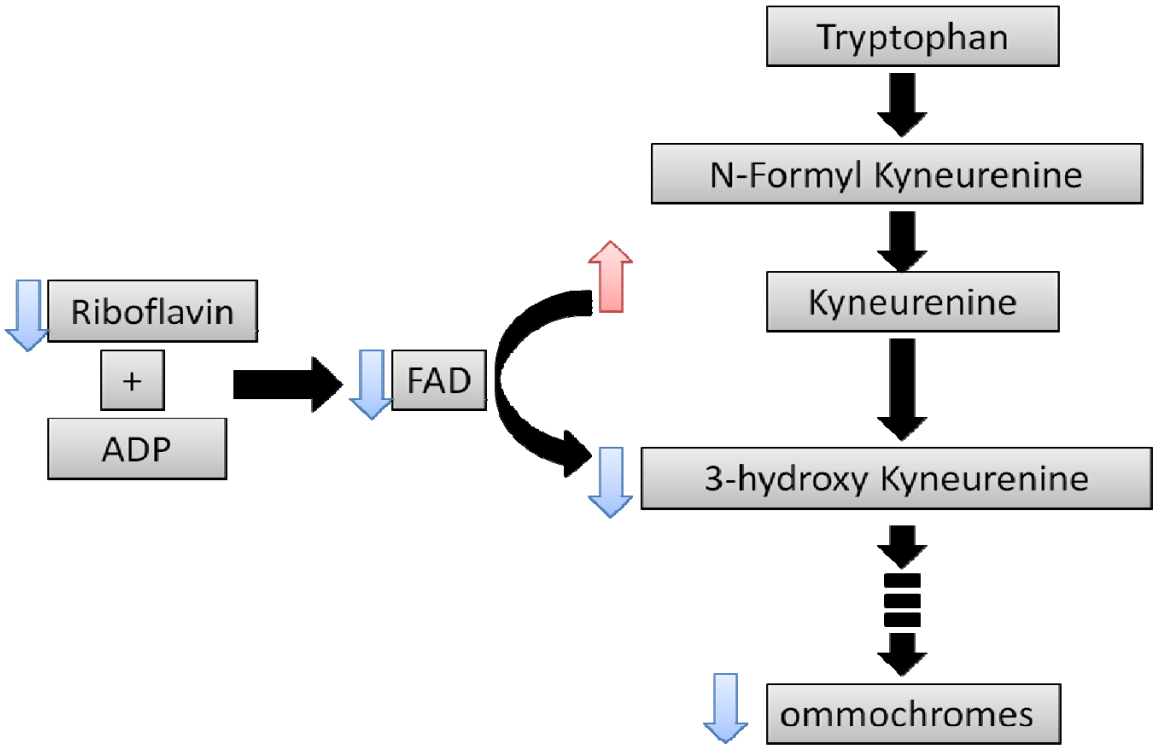
Features deemed to be of importance, whether by targeted, untargeted PCA, or STOCSY analysis were re-examined using the full spectrum list, and a two-tailed Student's *t* test was used to calculate the P value of each mass / metabolite which was reported side by side with the fold changes. A difference was considered to be significant if the P value was below 0.05.

The metabolites can be grouped and the first grouping consists of metabolites directly related to the absence of xanthine oxidase. However, several other pathways were found to be affected, either directly or indirectly. At the moment the benefits of the rapid metabolic profiling experiments described in this chapter are that they can suggest novel hypotheses that can be tested by further careful biological and more detailed analytical work to produce quantitative rather than comparative results.

4.3.4. Biochemical interpretation of variance outside the ry lesion:

4.3.4.1. Tryptophan metabolism:

It was found that 3-hydroxykynurenine levels were lower in *ry* (tables 4.3 and 4.4) and kynurenine was correspondingly elevated in *ry*, suggesting that the enzyme responsible for 3-hydroxylating kynurenine displays reduced activity. This should have an effect on pigmentation since the other pigments formed in *Drosophila*, ommochromes such as xanthommatin are formed via this pathway, scheme 4.3. It is possible to observe that xanthommatin is absent from *ry* flies but is present at low levels in all the extracts from WT flies, possibly because ommochromes are quite insoluble and difficult to extract. The reason for the down regulation of the enzyme is not immediately obvious; however, it was observed that levels of riboflavin which is necessary for the biosynthesis of the FAD co-factor required for hydroxylation of kynurenine were much lower in *ry*. This may be purely due to malabsorption of the vitamin in *ry* mutants. The ability to return to the original data to test a hypothesis *post hoc* illustrates another strength of this experimental approach.



Scheme 4.3 The effect of rosy absence on Tryptophan metabolism and mechanism of corruption of pigment production.

4.3.4.2. Osmolyte biosynthesis

Another observation from the chemometric analysis is that glycerophosphotidylcholine (GPC) and GPE levels are far higher in *ry* (tables 4.4 and 4.5). GPC (and probably GPE) is a counteracting osmolyte and has been found to accumulate in mammalian renal medullary cells, in response to high extracellular concentrations of NaCl or urea, maintaining osmotic balance without denaturing proteins within the cell which would occur in the presence of high concentrations of inorganic ions (Zablocki et al., 1991, Gallazzini et al., 2006). *Inebriated* mutants of *Drosophila*, which lack an osmolyte transporter in the Malpighian tubule, are hypersentive to NaCl (Huang et al., 2002). Why might *ry* mutants require elevated levels of osmolytes? Terrestrial insects are generally considered to be uricotelic; like birds, they eliminate nitrogenous waste as insoluble uric acid crystals, thus conserving water (Yokota and Shoemaker, 1981). As *ry* mutants are unable to make uric acid, elimination would be likely to occur in the form of hypoxanthine and xanthine. Although some xanthine is deposited as tubular calculi in *ry* mutants (Bonse, 1967), both xanthine and particularly hypoxanthine are more water soluble than uric acid (Merck, 2007) thus nitrogen excretion in *ry* may require higher rates of water loss. Accumulation of GPC (and GPE) in tissues may thus help to compensate for this additional osmotic stress. In humans, the high ratio of phosphocholine/GPC was found to be associated with elevated breast cancer malignancies (Aboagye and Bhujwalla, 1999).

4.3.4.3. Arginine metabolism:

A third unexpected impact of the *ry* mutation is on arginine metabolism. This is more subtle, and was observed as citrulline was lower in ry than wild-type flies. A possible explanation of this is that the accumulation of dihydrobiopterin causes an inhibition of tetrahydrobiopterin recycling (scheme 4.1). One of the isomers of DHBT is known to be an inhibitor of THBT regeneration (Thony et al., 2000) and THBT is required as a co-factor in arginine oxidation which results in the formation of NO and citrulline. It has been reported that accumulation of DHBT in vascular tissue may have a role in insulin resistance where it inhibits vasodilation through inhibiting NO production (Shinozaki et al., 1999).

4.3.4.4. Additional metabolic variations:

Tables 4.4 and 4.5 list some additional metabolites that are significantly altered in *ry* mutants. Again, there are several interesting findings, including a possible impact on pyrimidine (by contrast with purine) metabolites, such as cytosine and uracil. At the present time it is not possible to rationalize these effects but improved data extraction and automated development of metabolic networks should enable the metabolic effects to link up. Thus it is evident that a single mutation can have wide ranging metabolic effects. This data helps to explain not only why *ry* flies are relatively sickly, but also suggests a number of compensatory mechanisms to allow for their short term survival. The results also provide a salutary lesson for *Drosophila* genetics; *ry* has been widely used as a selectable marker in mutant stocks (the Flybase database lists 2155 such stocks), and on early synthetic P-element constructs (Rubin and Spradling, 1983). It is clear that the impact of *ry* on the fly ramifies widely across the metabolic landscape; and so *ry*, (like ebony), is a particularly poor choice of marker because of its pleiotropy.

Metabolite/Metabolic pathway	Formula	Recorded Mass	Retention time	Hits in METLIN	Hits in METLIN (Filtered)	Male	
						ratio ry/WT	P.value
Uric Acid related							
Uric Acid	C5H4N4O3	169.0356	9.62	1	1	2.49E-06 *	6.00E-04
Allantoin	C4H6N4O3	159.013	10.57	2	1	0.018	0.002
Hypoxanthine	C5H4N4O	137.0458	8.76	2	1	10.844	4.86E-05
Xanthine	C5H4N4O2	153.0407	7.85	2	1	20.665	5.5E-07
Guanine	C5H5N5O	152.0567	9.05	3	3	1.325	0.073
Guanosine	C10H13N5O5	284.0989	9.86	1	1	1.119	0.527
6-pyruvoyltetrahydropterin (6-PTP)	C9H11N5O3	238.0935	7.6	5	5	1.576	5.00E-04
biopterin	C9H11N5O3	238.0935	9.42	5	5	1.576	5.00E-04
Dihydrobiopterin	C9H13N5O3	240.1091	9.42	3	3	5.188	1.2E-06
Dihydropterin	C7H9N5O2	196.0829	7.89	1	1	0.214	0.006
Inosine	C10H12N4O5	269.088	8.67	4	2	2.023	0.011
AMP	C10H14N5O7P	348.0704	12.43	3	2	1.075	0.43
Tryptophan pathway Metabolites							
Tryptophan	C11H12N2O2	205.0972	9.89	2	2	0.36	6.39E-07
HydroxyTryptophan	C11H12N2O3	221.0921	9.43	1	1	0.509	0.004
kynurenine	C10H12N2O3	209.921	9.51	1	1	0.64	0.004
Hydroxykynurenine	C10H12N2O4	225.0921	10.69	3	1	0.345	7.11E-07
Osmolytes							
Choline glycerophosphate	C8H21NO6P	258.1101	13.94	1	1	9.56	1.82E-11
Glycerophosphoethanolamine	C5H14NO6P	216.1101	12.66	1	1	9.672	8.71E-08
Choline	C5H14NO	104.107	13	1	1	0.792	0.09
Pyrimidines							
Cytosine	C4H5N3O	112.0505	13.48	1	1	2.571	0.016
Uracil	C4H4N2O2	113.0346	7.96	1	1	0.402	0.002
uridine	C9H12N2O6	245.0768	8	2	2	6.1E-5 *	2.00E-04
Arginine Pathway Metabolites							
Citrulline	C6H13N3O3	176.103	13.37	1	1	0.348	0.042
Ornithine	C5H12N2O2	133.0972	17.56	1	1	0.128	0.036
L-Homocitrulline	C7H15N3O3	190.1186	17.82	1	1	1.4E-6 *	3.00E-04
Misc Metabolites							
Alanine	C3H7NO2	90.055	13.2	2	2	1.259	0.075
glutamic acid	C5H9NO4	148.0604	12.22	4	4	0.871	0.094
Glutamine	C5H10N2O3	147.0764	12.85	5	5	1.11	0.233
Threonine	C4H9NO3	120.0655	12.51	2	2	0.578	2.00E-04
Nicotinamide	C6H6N2O	123.0553	7.81	1	1	0.732	0.081
valine	C5H11NO2	118.0861	12.02	5	5	0.76	0.016

Table 4.4 Metabolite changes identified in ry male flies.

*** The metabolite was completely missed from one genotype and intensity wasreplaced by the value 0.01 for automation purposes.**

Metabolite/Metabolic pathway	Formula	Recorded Mass	Retention time	Hits in METLIN	Hits in METLIN (Filtered)	Female	
						ratio ry/WT	P.value
Uric Acid related							
Uric Acid	C5H4N4O3	169.0356	9.62	1	1	5.6E-4*	7.6E-04
Allantoin	C4H6N4O3	159.013	10.57	2	1	0.047	1.2E-04
Hypoxanthine	C5H4N4O	137.0458	8.76	2	1	15.66	1.9E-04
Xanthine	C5H4N4O2	153.0407	7.85	2	1	25.198	3.2E-04
Guanine	C5H5N5O	152.0567	9.05	3	3	2.828	0.001
Guanosine	C10H13N5O5	284.0989	9.86	1	1	2.891	0.003
6-pyruvoyltetrahydropterin (6-PTP)	C9H11N5O3	238.0935	7.6	5	5	3.034	4.1E-09
biopterin	C9H11N5O3	238.0935	9.42	5	5	3.034	4.1E-09
Dihydrobiopterin	C9H13N5O3	240.1091	9.42	3	3	5.465	1.2E-06
Dihydropterin	C7H9N5O2	196.0829	7.89	1	1	0.13	0.013
Inosine	C10H12N4O5	269.088	8.67	4	2	4.693	0.005
AMP	C10H14N5O7P	348.0704	12.43	3	2	1.458	0.055
Tryptophan pathway Metabolites							
Tryptophan	C11H12N2O2	205.0972	9.89	2	2	0.529	0.005
HydroxyTryptophan	C11H12N2O3	221.0921	9.43	1	1	1.874	0.028
kynurenine	C10H12N2O3	209.921	9.51	1	1	3.352	5.5E-08
Hydroxykynurenine	C10H12N2O4	225.0921	10.69	3	1	0.405	2.5E-07
Osmolytes							
Choline glycerophosphate	C8H21NO6P	258.1101	13.94	1	1	14.574	4.1E-11
Glycerophosphoethanolamine	C5H14NO6P	216.1101	12.66	1	1	16.157	2.0E-10
Choline	C5H14NO	104.107	13	1	1	1.333	0.006
Pyrimidines							
Cytosine	C4H5N3O	112.0505	13.48	1	1	14.305	0.002
Uracil	C4H4N2O2	113.0346	7.96	1	1	0.921	0.6
uridine	C9H12N2O6	245.0768	8	2	2	6.3E-5 *	0.002
Arginine Pathway Metabolites							
Citrulline	C6H13N3O3	176.103	13.37	1	1	0.2999	0.004
Ornithine	C5H12N2O2	133.0972	17.56	1	1	0.376	0.147
L-Homocitrulline	C7H15N3O3	190.1186	17.82	1	1	0.011	1.9E-07
Misc Metabolites							
Alanine	C3H7NO2	90.055	13.2	2	2	1.929	8.4E-08
glutamic acid	C5H9NO4	148.0604	12.22	4	4	1.746	3.1E-04
Glutamine	C5H10N2O3	147.0764	12.85	5	5	1.319	5.6E-04
Threonine	C4H9NO3	120.0655	12.51	2	2	0.526	3.2E-05
Nicotinamide	C6H6N2O	123.0553	7.81	1	1	1.546	1.6E-04
valine	C5H11NO2	118.0861	12.02	5	5	1.06	0.494

Table.4.5 Metabolite changes identified in ry Female flies

*** The metabolite was completely missed from one genotype and intensity wasreplaced by the value 0.01 for automation purposes.**

4.4. The scope of ESI-MS:

ESI-MS cannot see every metabolite and this will present a challenge when it comes to developing complete metabolic networks and techniques such as LC-NMR and GC-MS will be complementary (Michaud et al., 2008). *Drosophila* provides an excellent model for metabolic studies. Stable isotope labelling studies could also potentially be carried out in *Drosophila* and will assist greatly in defining the metabolome.

A key goal of functional genomics is the elucidation of functions for all genes and genetic elements encoded in the genome (Dow and Davies, 2003). Given that perhaps a third of genes in a typical metazoan genome have neither obvious functions based on their encoded protein structures nor evocative tissue distributions, it has been considered that reverse genetics, coupled to painstaking analysis of informative functional phenotypes, is the best approach to identify function in both plants and animals (Adams and Sekelsky, 2002, Ballinger and Benzer, 1989, Bargmann, 2001, Kaiser, 1990, Landel et al., 1990, Lichtenstein and Barrena, 1993, Orkin, 1986). However, such studies are critically limited by the lack of informative phenotypes to study in genetic model organisms – the so-called phenotype gap (Brown and Peters, 1996, Dow, 2003, Dow and Davies, 2003) – so the ability of metabolic profiling to produce rapid, informative metabolic fingerprints from reasonable quantities of starting material, suggests an alternative method of adding value to the many thousands of unstudied mutant stocks available today. Metabolic profiling has already been employed in a few pioneering studies in both *Drosophila* (Feala et al., 2007) and non-model insects (Michaud et al., 2008), but the combination of *Drosophila* and ZICHILIC-FTMS, with the downstream computational methods described here, may provide a balance of cost, throughput and comprehensiveness that allows rapid future progress.

4.5. Conclusion

The informative capabilities of the ZIC-HILIC-FTMS (Orbitrap) combination were investigated using an animal model unusually suited for metabolic profiling studies. The *Drosophila* system provides a model, sufficiently complex to test the analytical techniques with no trade-off in controllability. There is a large collection of mutations in the *Drosophila* world, many of them well studied over a long period of time. We chose one of the most studied *Drosophila* mutations (ry) where documented and speculated metabolic changes allows the monitoring of the value of our selected system for endorsing the already known information. The ZIC-HILIC-Orbitrap system was able to indicate the expected variation in the lesion position (xanthine oxidase in purine metabolism pathway) but also proved capable, when combined with sound data manipulations, of shedding the light on an extended perturbation stemming from the lesion and other variations occurring far away in the metabolic map. The biological interpretations presented here are conjectural and were provided only as examples of how metabolic profiling can provide additional valuable information, hidden before, and the ability to drive the research in the direction of new hypotheses.

5. Metabolic Profiling of *Trypanosoma* Extracts:

5.1. Introduction

Following on the successful application of HILIC-Orbitrap FTMS for the analysis of *Drosophila* metabolome, widening the scope of this method to encompass other matrices and biological species was of interest.

We have been interested in learning about metabolism in *Trypanosoma brucei* subspecies which are the causative agents of human African trypanosomiasis and a number of animal diseases. Trypanosomes depend on their hosts to provide a source of carbon for energy metabolism (Bringaud et al., 2006). The bloodstream form of *T. brucei* exclusively uses glucose for its energy metabolism(Fairlamb and Opperdoes, 1986, Cannata and Cazzulo, 1984). However, the procyclic form of *T.brucei*, which colonises the mid-gut of the tsetse fly, relies on L-proline as its energy source since this amino acid is a major constituent of the haemolymph of the fly (Bursell, 1981). In a glucose rich medium, however, they preferentially utilise glucose as a carbon source, which is converted into succinate, acetate and smaller amounts of lactic acid and CO_2 (Besteiro et al., 2002, van Weelden et al., 2003). The classical culture medium for procyclic trypanosomes is rich in glucose (Brun R, 1979). However, these cells can alternatively be cultured in glucose depleted, proline rich media (Lamour et al., 2005, Furuya et al., 2002, James C. Morris et al., 2002).

Here we show that metabolic differences can be observed between procyclic trypanosomes grown with glucose or proline as carbon sources by using hydrophilic interaction (HILIC) chromatography and Fourier transform mass spectrometry.

Trypanothione is an anti-oxidant oligopeptide not available in humans (Ariyanayagam and Fairlamb, 2001). It is an unusual form of Glutathione, as two glutathione molecules (reduced form) are linked with a spermidine molecule (figure.5.1). The absence of this molecule from humans makes its metabolic pathway a potential target for anti-trypanosomiasis drug discovery (Gildardo Rivera, 2009, Spinks et al., 2009) with many attempts having been made to disrupt the trypansome parasite system at this site (Iribarne et al., 2009, Patterson et al., 2009, Richardson et al., 2009). Trypanothione synthesis was genetically engineered in *E.coli* and showed radioprotective properties (Richardson et al., 2009, Fitzgerald et al., 2010). In

addition to its role in metabolism, this metabolite makes a viability indicator for Trypanosomatids cell growth, which is usually evaluated by a visual subjective microscopic count of viable parasites.

The previous attempts to identify and perform qualitative analysis on trypanothione (and other thiols) by mass spectrometry used a lengthy procedure involving bromobimane derivitization, preparative chromatography and a MALDI-MS analysis (Ondarza et al., 2006b, Ariyanayagam and Fairlamb, 2001, Ondarza et al., 2005, Shim and Fairlamb, 1988). These approaches were able to provide reliable qualitative information about thiol products but obviously were too specific to extract information about other metabolites such as those involved in the polyamine metabolic pathway. The aim of this work was to provide qualitative information about trypanothione and glutathione as major targets, as well as secondary targets of other metabolites, with priority for those involved in the polyamine synthesis pathway. The method used in the approach described here involves much less sample preparation than the derivatisation approaches.

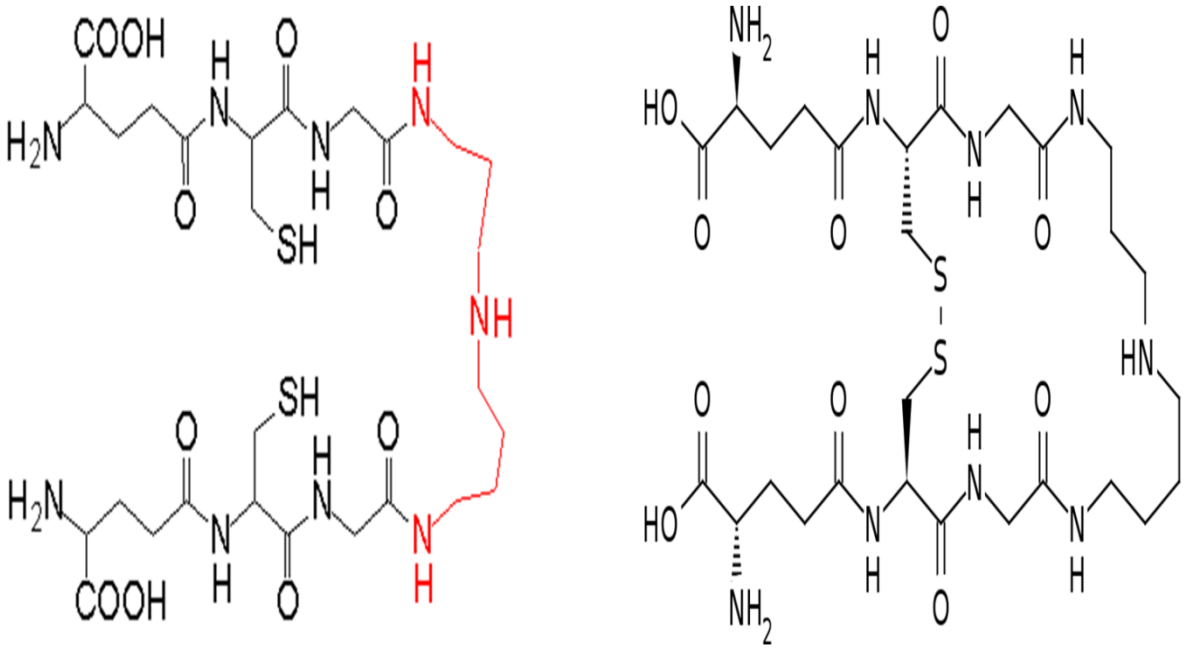


Figure.5.1 Oxidised (right side panel) and reduced (left side panel) forms of trypanothione. The reduced form showing the spermidine moiety.

5.2. Materials and methods:

5.2.1. 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.). Ethanol was from Sigma (>99% pure). The following standards were obtained from Sigma: O-acetylserine, NAD, glutathione, succinic acid. A mixture of the 20 common aminoacids was obtained from Phenomenex (Macclesfield, U.K.). Oxidised trypanothione was a gift from Professor Graham Coombs. Culture medium was produced with chemicals from Sigma-Aldrich at the highest purity available.

5.2.2. Culture and Extraction of trypanosomes

Procyclic form trypanosomes (strain 427) were cultivated at 28ºC in SDM80 (Furuya et al., 2002), supplemented with 10% heat inactivated foetal calf serum that had been dialysed by ultrafiltration against 0.15 M NaCl. The medium was also supplemented with either D-glucose or L-proline at 10 mM. Once the culture had reached a density of $2x10^6$ cells/ml, cells were spun down and re-suspended in serum free medium to a density of 10^9 cells/ml. After 30 min, during which time they re-equilibrated their metabolism, the cells (0.2 ml aliquots) were extracted, by directly pipetting, into hot ethanol (0.8 ml of 80% ethanol in 20 mM HEPES pH 7.2). The procedure lysed the cells, and the lysate was spun to remove insoluble material. The supernatant was frozen in a dry-ice ethanol water bath and stored at -80ºC until analysis. The optimisation of the extraction solvent is discussed below.

5.2.3. HILIC LC-MS

LC-MS data were acquired using a Finnigan LTQ Orbitrap instrument (Thermo Fisher, Hemel Hempstead, U.K.). Sample analysis was carried out under positive and negative ion modes of detection. The mass scanning range was 50-1200 m/z, while the temperature was 200ºC and the sheath and auxiliary gas flow rates were 30 and 10 respectively (units not specified by the manufacturer). 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 × 4.6 mm (HiChrom, Reading U.K.) was used for all

analyses. The gradient was as follows: 80% B (0 min.) to 50% B at 12 min. and stabilize at this composition until 26 min then to 80% B at 28 min and stabilize until 35 min back to 20% B and equilibrate until 42 min.

The gradient was slightly modified for two reasons. The first is to separate the late eluting Trypanothione peak from background clusters of ions and the second was to obtain a better retention of sugars in the negative mode runs.

The raw data for proline and glucose grown trypanosomes was processed and compared using Sieve® version 1.1.0 (Thermo Fisher Scientific Inc., San Jose, USA). ANOVA or t-rest on manually processed data was carried out using Microsoft Excel® 2007.

5.3. Results and discussion:

Earlier work investigating the metabolome of *T. brucei* using FT-MS identified over one thousand individual peaks (Breitling et al., 2006b). Those experiments which were based on direct infusion, failed to detect numerous key metabolites known to be present in trypanosomes. The identified metabolites were also very lipid rich. A number of parameters influenced the output in those experiments. Firstly, the samples represented pellets of *T. brucei* that had been separated from suspension by centrifugation. This procedure induced numerous stresses including anoxia and nutrient depletion. Alternative methods were therefore investigated in order to quench metabolism and optimise metabolite capture. Direct addition of various cell types that grow in suspension (e.g. yeast and various bacterial species) to a solution of freezing methanol quenches metabolism and leads to metabolites being retained in cells (Bolten et al., 2007, Schaefer et al., 1999, Koning and Dam, 1992). Unfortunately, in *T. brucei* substantial leakage of metabolites was apparent (by comparing retained radiolabelled D-glucose in cells treated in this way as compared to a centrifugation through oil technique has been routinely employed for rapid transport analysis (Barrett et al., 1995). Therefore boiling buffered ethanol was tested as a medium in which to quench metabolism. It was shown that direct addition of trypanosomes in their medium to 80% ethanol in 20 mM HEPES pH7.2 enabled the successful identification of numerous metabolites. The selection of this extraction system was based on a pilot experiment, where the same grown cells were lysed and

extracted with nine different solvent systems (Chloroform, ACN, Acetone, hot MeOH, cold MeOH, Hot EtOH, cold EtOH, hot EtOH with HEPES and cold EtOH with HEPES). For untargeted metabolic profiling, it is difficult to judge which solvent is the optimum in terms of metabolite yield, as each solvent serves to give maximal signal for a set of metabolites. However, in this work, trypanothione and glutathione along with the associated metabolites are of the greatest importance. These metabolites seem to give the best intensity with an ethanol based solvent, with hot EtOH+HEPES and hot EtOH ranking first and second, respectively (Figure 5.2).

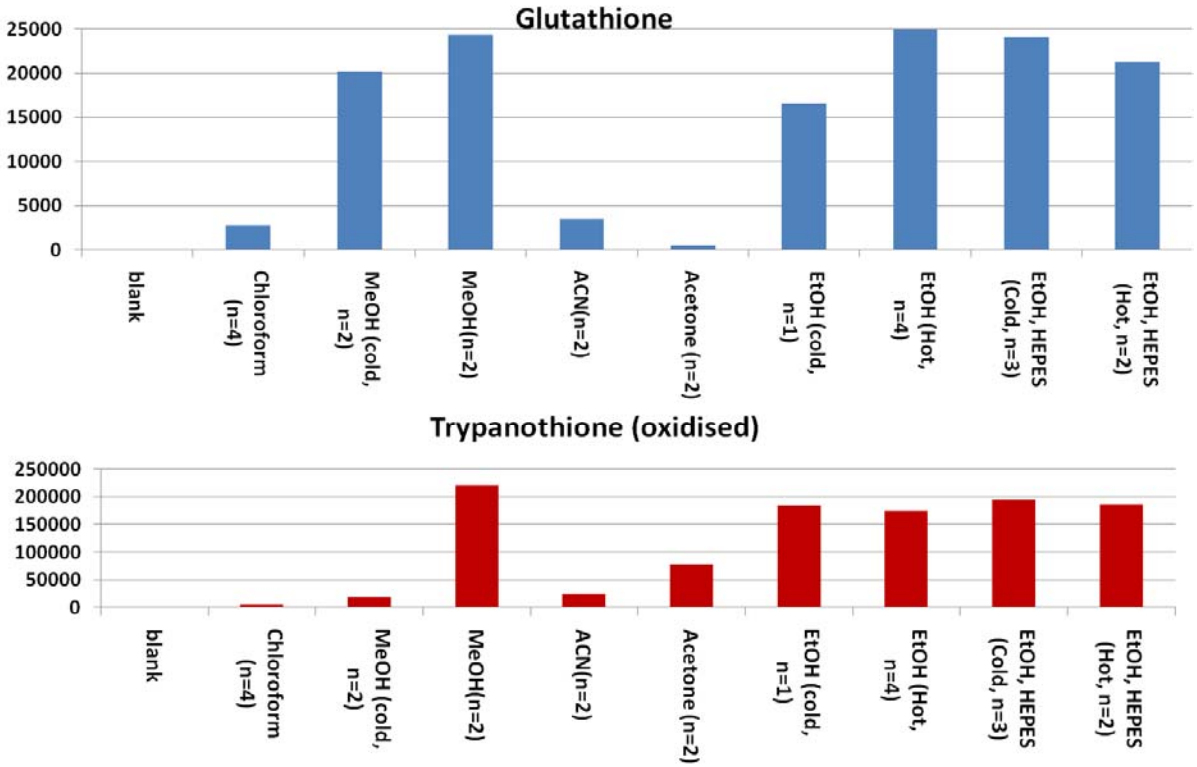


Figure.5.2 Intensity profiles of glutathione (upper panel) and trypanothione (lower panel) upon extraction with different solvent systems.

Preliminary data using direct infusion of the trypanosome lysate into the Orbitrap failed to provide a peak corresponding to trypanothione, a trypanosome specific metabolite the presence of which we considered critical in validating the integrity of extracts. Earlier work using FT-ICR-MS with direct infusion of trypanosome lysates also failed to identify this metabolite, possibly due to ion suppression induced by abundant lipids in the sample (Breitling et al., 2006b). The

utility of HILIC chromatography in retaining hydrophilic compounds while allowing hydrophobic species to flow-through rapidly, prompted us to use this technique to enhance our ability to identify trypanothione and numerous other polar species. From our experience polar compounds such as small amino acids, glutathione and trypanothione elute in the void volume from C18 reversed-phase columns unless a gradient is tailored for retaining a specific compound which is counter-productive to the purpose of metabolic profiling. Figure.5.3.A shows a typical chromatographic trace obtained from an extract of trypanosomes run on a ZicHilic column with the mass spectrometer operated in positive ion mode and Figure.5.3.B shows a trace obtained for the same sample in negative ion mode.

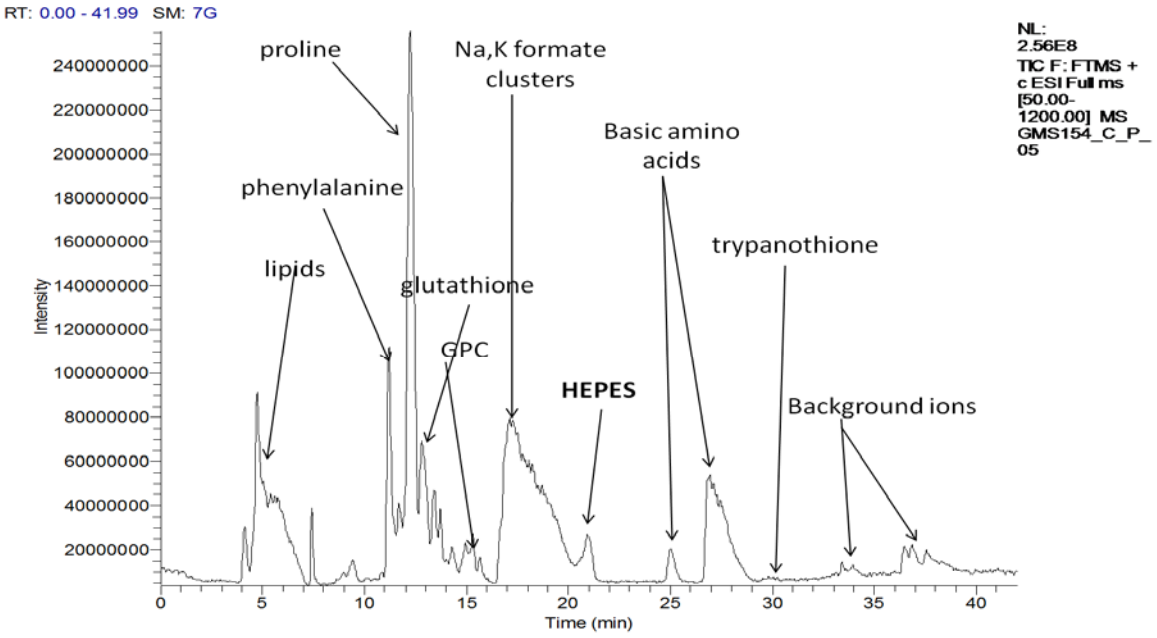


Figure.5.3 (A) A total ion chromatogram (TIC) of a *Trypanosoma brucei* extract under positive ion mode, acquired by ZIC-HILIC-FTMS showing some identified metabolites.

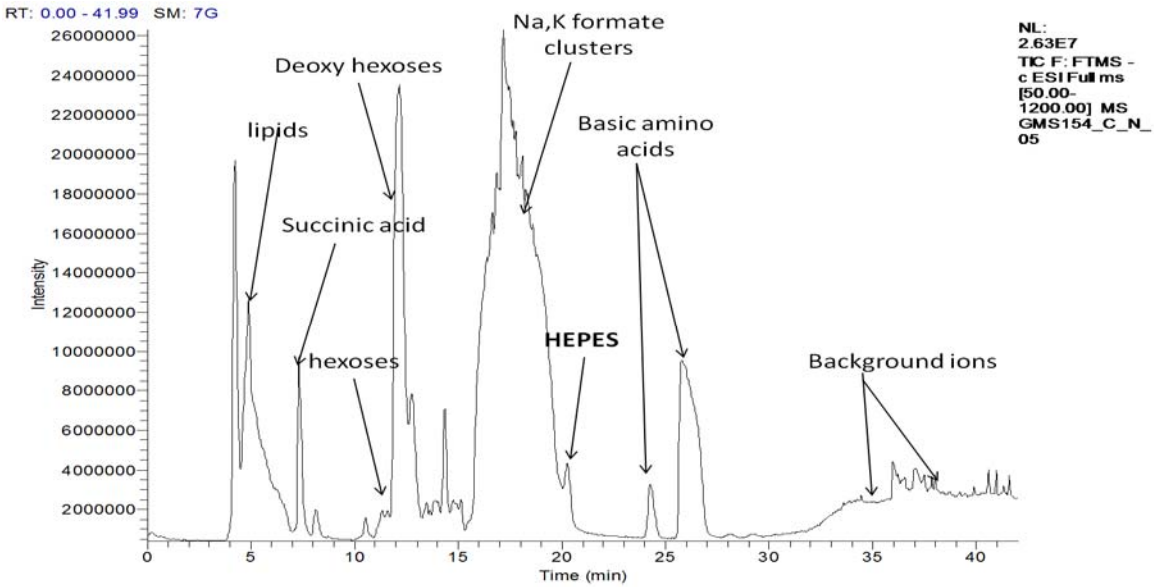


Figure.5.3 (B) A total ion chromatogram (TIC) of a *Trypanosoma brucei* extract under negative ion mode, acquired by ZIC-HILIC-FTMS acquisition showing some identified metabolites.

Signal to noise ratios in negative ion mode were generally not as good as in the positive ion mode. Lipophilic components such as sphingolipids and phosphatidylcholine lipids eluted quickly from the column. This is a significant advantage since on reverse-phase columns lipids are difficult to elute and can accumulate and cause a background bleed, which is well known to produce ion suppression problems (Ismaiel et al., 2008, Alzweiri et al., 2008) . Phosphoplipids are quite difficult to remove from reversed-phase columns even with strong organic solvents (Ismaiel et al., 2008),which prolongs run times, quite apart from the ventilation problems which are produced when less polar organic solvents such ethyl acetate or tetrahyrofuran are used to clean the column in conjunction with ESI-MS. Table.5.1 gives some examples of mass matches and retention times for a selection of tentatively identified metabolites found in a trypanosome lysate. Many of the prominent peaks, up to 10 minutes, are due to phospholipids and lipophilic amino acids.

Retention time (min)	m/z (amu)	Formula	Deviation from Exact Mass (ppm)	Identity	Peak Area
5.08	288.2896	$C_{17}H_{38}O_2N$	-0.23	Sphingosine	1.2×10^{10}
5.03	316.3212	$C_{19}H_{42}O_2N$	-0.8	Sphingosine	6.4×10^{9}
7.15	772.6211	$C_{44}H_{87}O_7NP$	-0.41	Unsaturated phosphatidyl choline ether lipid	3.0×10^{9}
7.07	830.5691	$C_{48}H_{81}O_8NP$	-0.41	Unsaturated diacyl phosphatidyl choline lipid	2.9×10^{9}
7.91	703.5741	$C_{39}H_{80}O_6N_2P$	-1.1	Unsaturated sphingomyelin	2.8×10^{9}
8.53	166.0861	$C_9H_{12}O_2N$	-1	Phenylalanine	6.6×10^{7}
9.94	780.5381	$C_{40}H_{79}O_{11}NP$	-0.56	Phosphatidyl serine lipid	8.8×10^{8}
9.94	150.0583	$C_5H_{12}O_2NS$	-0.98	Methionine	5.9×10^{10}
10.67	182.081	$C_9H_{12}O_3N$	-0.71	Tyrosine	7.0×10^{9}
14.72	76.0392	$C_2H_6O_2N$	-0.86	Glycine	1.7×10^{9}
17.28	258.1099	$C_8H_{21}O_6NP$	-0.7	Phosphatidyl glycerol	1.0×10^{10}
18.15	722.2957	$C_{27}H_{48}O_{10}N_9S_2$	-0.4	Trypanothione	9.8×10^{7}
19.09	348.0702	$C_{10}H_{15}O_7N_5P$	-0.58	AMP	7.72×10^{9}
24.8	184.0731	$C_5H_{15}O_4NP$	-1.1	Phosphocholine	3.2×10^{9}
25.16	664.1159	$C_{21}H_{28}O_{14}N_7P_2$	-0.74	NAD	1.2×10^{8}

Table.5.1 A selection of metabolites observed in a *T.brucei* pellet extract analysed by FT-MS with HILIC chromatography in positive ion ESI mode.

Much of the time an elemental composition was sufficient to get an identification of a compound since often there were no competing sensible elemental matches within 3 ppm of the nominal mass. When elemental compositions are considered in conjunction with the Metlin database (Smith et al., 2005) there is often only one metabolite out of the 15000 in the data base which matches the exact mass. Late eluting peaks include polar compounds such as adenosine monophosphate (AMP) and nicotinamide adenosine diphosphate (NAD). Although it was possible to

observe these phosphates, triphosphates such as ATP were not readily observed in the biological samples probably because of ion suppression effects or possibly due to the instability of ATP when absence of tri-phosphates include low ionization affinity (in positive mode) and strong retention on the Zwitterionic stationary phase. Figure.5.4 shows the positive ion mass spectrum of oxidized trypanothione, which was detected as both single and double charged forms of the molecule (it can potentially carry up to three charges). Reduced Trypanothione was not observed, which could bear different possible explanations. The use of trypanothione by the parasites to combat oxidative stress during extraction is one possibility. Analytical considerations are more likeley with oxidation during the hot EtOH extraction being a strong suspect; although cold extraction did not improve the yield.Insufficient resolution can affect the detection as well, especially when recognizing the isotopic pattern of trypanothine. Trypanothione has two sulphur atoms leading to a strong 34S isotope peak with m/z of $[M+H^+]$ = 724.291806 and $[M+2H^+]/2$ = 362.6498155 compared to the reduced Trypanothione of m/z $[M+H^+]$ = 724.31166 and $[M+2H^+]/2$ = 362.65947. Thus, if the reduced and oxidized forms co-eluted MS resolution power required to identify these two peaks simultaneously is calculated to be 37000 to separate the reduced form from the ^{34}S peak of the oxidized form. The instrument was operated on a resolution of 30000, and with a perfect co-elution, MS will not be powerful enough to enable discrimination between the two peaks. This emphasises the importance of higher resolutions (discussed in chapter 2).

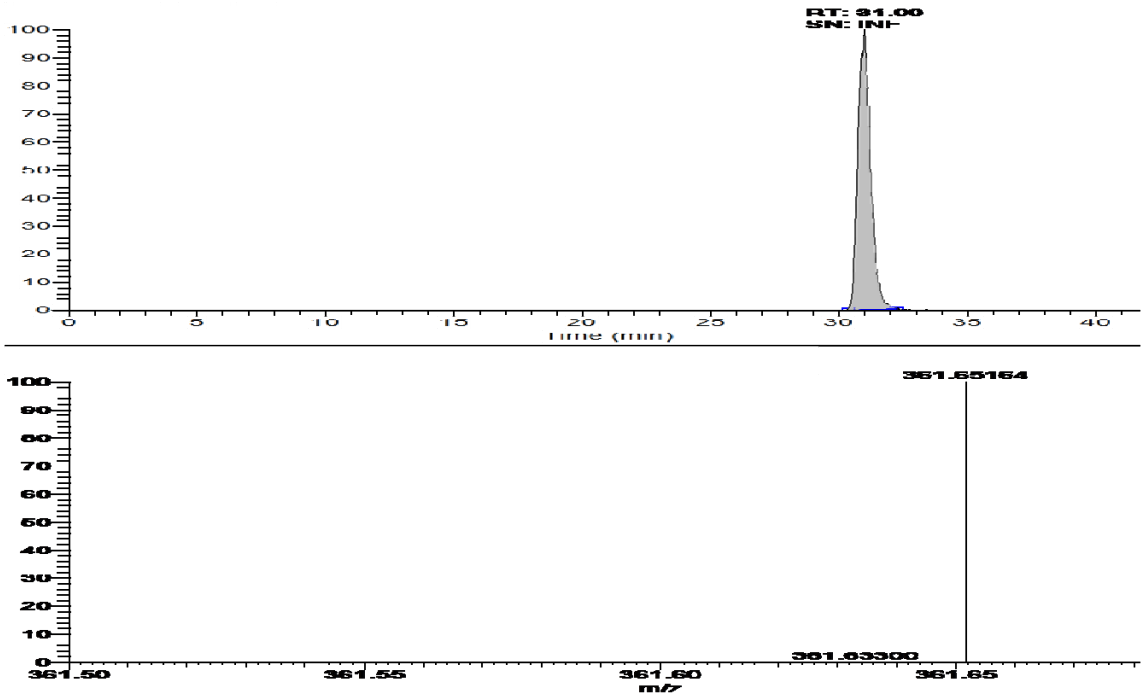


Figure.5.4 spectrum (bottom panel) and Extracted Ion Chromatogram (EIC, top panel) of oxidized Trypanothione eluted through ZIC-HILIC and analyzed by FTMS under positive ion mode.

Resolution was not the only problem we encountered during this work. The extraction with HEPES rich solvents lead to a very large peak of the buffer, along with its contaminants. HEPES is a buffer used in biological culture media and was shown to be necessary for trypanosome homeostasis (Yabu Y et al., 1989). Ion suppression by this peak is the major, but not the only, problem. HEPES manifested as a horde of related peaks, of different forms (Table.5.2). There is as well a set of unusual related peaks were a set of peaks with equal distances and symmetrical distribution both sides of the major peak.Figure 5.5.

	m/z	RT	HEPES related mass
1	238.03709	15.97	Instrument specific
2	238.69876	15.97	Instrument specific
3	239.06	15.96	Instrument specific
4	239.06412	15.96	Instrument specific
5	239.06905	15.96	Instrument specific
6	239.07294	15.96	Instrument specific
7	239.07734	15.96	Instrument specific
8	239.08236	15.96	Instrument specific
9	239.08565	15.96	Instrument specific
10	239.08885	15.96	Instrument specific
11	239.0921	15.96	Instrument specific
12	239.09664	15.96	Instrument specific
13	239.10585	15.96	HEPES
14	239.10596	15.96	Instrument specific
15	239.11493	15.96	Instrument specific
16	239.11811	15.96	Instrument specific
17	239.12219	15.96	Instrument specific
18	239.12607	15.96	Instrument specific
19	239.13158	15.96	Instrument specific
20	239.13654	15.96	Instrument specific
21	239.14031	15.96	Instrument specific
22	239.14541	15.96	Instrument specific
23	239.15089	15.96	Instrument specific
24	240.10917	15.96	HEPES (13C)
25	241.10159	15.96	HEPES (34S)
26	261.08787	15.58	HEPES + Na
27	477.20469	15.96	2*HEPES
28	478.20792	15.96	Instrument specific
29	478.20792	15.96	2*HEPES (13C)
30	478.35584	15.96	Instrument specific
31	479.19953	15.96	2*HEPES (34S)
32	480.20373	15.96	2*HEPES (34S and 13C)
33	499.18627	15.58	2*HEPES + Na
34	715.3037	15.96	3*HEPES
35	716.30624	15.96	3*HEPES (13C)
36	717.29857	16.01	3*HEPES (34S)
37	737.28532	15.67	3*HEPES+Na
38	738.28832	15.67	3*HEPES +Na (13C)
39	953.40248	15.96	4*HEPES
40	953.90394	15.96	Instrument specific
41	954.40446	15.96	4*HEPES (13C)
42	954.90352	15.96	Instrument specific
43	955.40152	15.96	4*HEPES (34S)
44	975.3842	15.67	4*HEPES + Na
45	976.3871	15.67	4*HEPES +Na (13C)
46	1033.4354	15.96	unknown

Table.5.2 HEPES related peaks in HILIC-FTMS analysis, positive ion mode.

This seems to happen with all peaks of high intensity (> 20 million units). Again care should be taken upon peak to metabolite assignment, and the correlation matrices (as described in chapter 4) can reveal this type of feature.

This type of artefact was reported later and given the name "instrument-specific artefacts " (Brown et al., 2009). It was recommended for any further analysis to avoid using HEPES (or any other buffer) during extraction. This was later tested and found to remarkably improve the quality of the samples (data not shown).

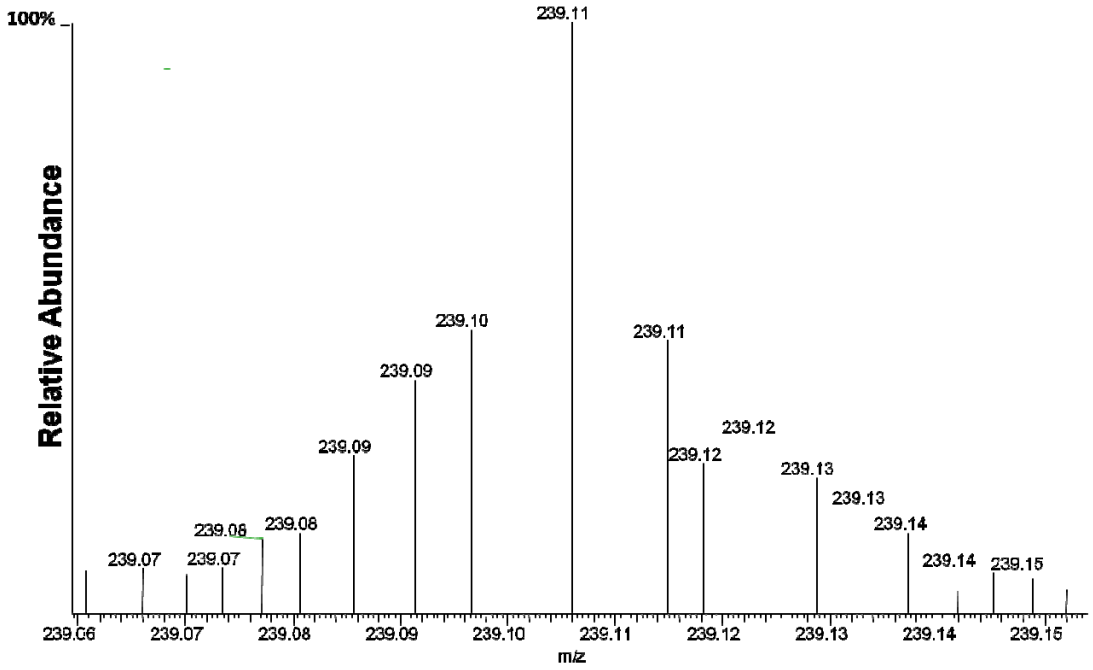


Figure.5.5 Instrument specific artefacts with symmetrical distribution either side of the major HEPES peak.

Figure5.6 shows extracted ion traces for trypanothione and glutathione extracted from trypanosomes grown in proline and glucose rich media. The signal to noise level on these peaks is infinite (i.e. the noise is too low to measure at this signal intensity) since the mass window used to display them is very narrow. The peak areas are listed in the individual traces indicating about a 2.5 fold difference in glutathione levels between the proline rich and glucose rich cultures. The high signal to noise level means that, if required, glutathione and trypanothione could be

quantified in a very small number of cells using this technique. In the case of trypanothione the calculated exact mass is 722.29600 and over the whole run the mean deviation from the exact mass was 1.1 ± 0.8 ppm. In the case of glutathione the calculated exact mass is 308.09108 with mean deviation from the exact mass being 1.3±0.5 ppm indicating the high mass accuracy of the Orbitrap.

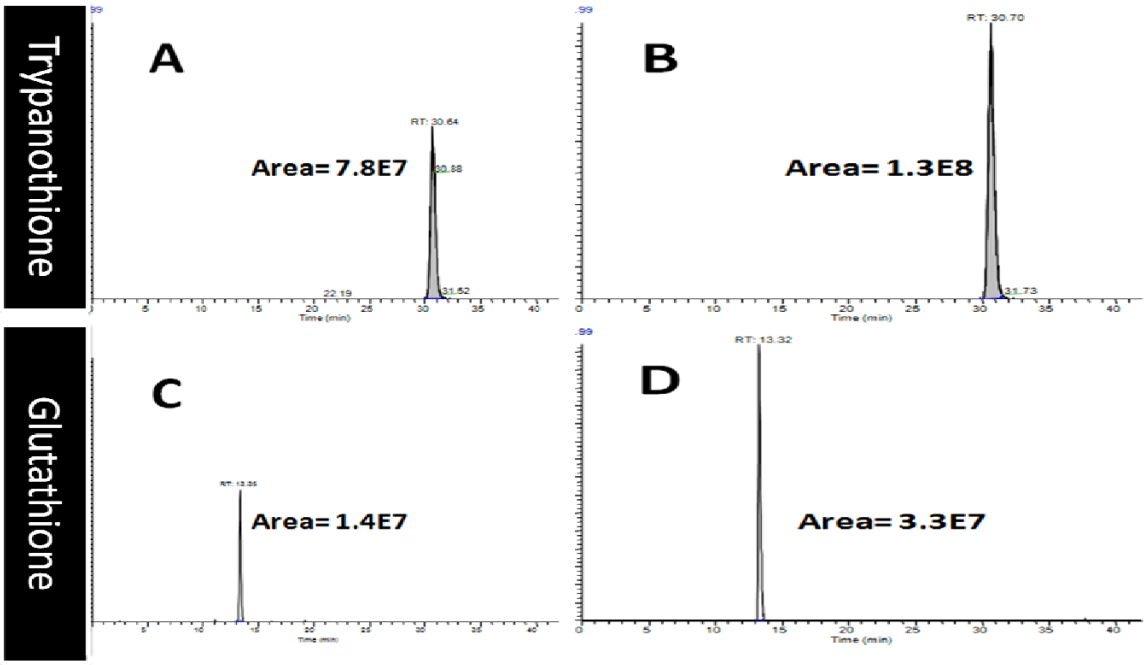


Figure.5.6 Comparison of EICs of Trypanothione and Glutation in *Trypanosoma Brucei* grown with glucose (A,C) or proline (B,D) as a carbon source.

Next the ability of Sieve® 1.1 software to differentiate between trypanosomes grown in related media (using glucose or proline as a carbon source) was tested. Other than including proline or glucose as the carbon source the media composition was identical. Sieve® 1.1 software was used to identify individual peaks that differed significantly between glucose and proline grown samples by univariate statistics (student`s t-test) as an alternative to the more commonly used principal component analysis. A number of differences were highlighted and the most significant of which are shown in Table5.3 where it can be seen that some high abundance metabolites give rise to several peaks e.g. glutamic acid gives rise to peaks due to the sodium adduct of its sodium salt and also to a peak due to neutral loss of water from the molecular ion.

m/z	Formula	Identity	P value	Ratio
70.06496	C_4H_8N	Pyrroline neutral loss of formic acid from proline	4.10E-06	25.8
160.0344	$C_5H_8O_2NNa_2$	Na adduct of proline Na	2.06E-04	12.6
205.0818	$C_7H_{13}O_5N_2$	Glutamate-glycine	6.5E-6	9.63
148.0602	$C_5H_{10}O_4N$	Glutamic acid	6.7E-09	8.44
192.0242	$C_5H_8O_4NNa_2$	Na adduct of glutamate Na	7.0E-8	4
130.0496	$C_5H_8O_3N$	Neutral loss of H_2O from glutamic acid	1.2E-8	2.9
146.1173	$C_7H_{16}O_2N$	Acetyl choline	2.2E-5	7.45
87.04384	$C_4H_7O_2$	Loss of trimethylammonium from acetyl choline	4.0E-4	5.8
308.0908	$C_{10}H_{18}O_6N_3S$	Glutathione	1.0E-4	4.9
189.1232	$C_8H_{17}O_3N_2$	Acetyl lysine	4.8E-6	3.2
126.0911	$C_7H_{12}ON$	Loss of formic acid and ammonia from acetyl lysine	3.7E-6	3.1
104.1068	$C_5H_{14}ON$	Choline	9.1E-6	2.3
166.0531	$C_5H_{12}O_3NS$	Oxomethionine	7.2E-4	1.9
149.0265	$C_5H_9O_3S$	Neutral loss of ammonia from oxomethionine	0.01	1.8

Table.5.3 Components in *T.brucei* pellet extracts detected in positive ion mode highlighted as differing between glucose (n=7) and proline (n=5) rich cultures by Sieve® 1.1. A high ratio indicates a high amount in the cultures grown in proline rich medium.

Thus it is obvious that chromatographic separation simplifies the data interpretation in comparison with direct infusion FT-MS analyses where it is not possible to see which ions are related on the basis of chromatographic retention time, especially when combining the peak picking with a correlation investigation as discussed in chapter 4. The high level of proline present in the trypanosomes grown in proline rich medium was not detected due to the fact that Sieve® is set to detect chromatographic peaks below a certain width threshold and the very broad peak for proline was thus excluded, however, proline will be eventually identified as an important variable when performing the manual detection of related features (e.g. ^{13}C isotopomer of proline). In Chapter 3 the importance of manual checking as a confirmatory step for each dataset was discussed. Following automated Sieve® analysis, the raw data was examined by manual processing of the raw signal data in

Xcalibur. Mass ranges of 0.01 amu or less were entered into Xcalibur to confirm the exact mass of some of peaks identified by Sieve®. The exact mass of the peak was taken from its apex and recorded as reported by Xcalibur and the peak was integrated manually. The peak areas for the individual compounds were normalised in each case against the peak area response (the sum of the singly and doubly charged ions) for trypanothione in each culture. This process was meant to be a compensation for differences in viability between different growth flasks. Table5.4 shows some of the components which exhibit clear differences between glucose and proline grown trypanosomes most of which relate to the pathway of proline metabolism.

Metabolite	**Elemental composition**	**Mass match ppm (n=12)**	**Normalised abundance glucose grown types ± RSD**	**Normalised abundance proline grown types ± RSD**	**ANOVA P value**
Proline	C5H10NO2	1.6±0.5	4.43 ± 4.04	1217.35 ± 92.1	5.22 E-7
5-Carboxypyrroline	C5H6NO2	*1.1±0.6	0	57.3±0.0051	-
Glutamic Acid	C5H10NO4	1.3±0.9	47.8±0.16	13.6±1.61	1.5E-08
Glutathione	C10H16N3O6S	1.3±0.5	87.1±0.04	76.9±0.236	0.018
Glutamate-glycine	C7H13N2O5	1.5±1.0	71.7±0.003	92.5±0.052	0.021
Alanine	C3H8NO2	1.6±0.7	18.6±0.02	27.6±0.01	0.006
‡Succinic Acid	C4H5O4	5.2±0.8	42.0±3.63	160±1.39	0.065

Table.5.4 Manually processed data for significant compounds in *T.brucei* pellet extracts from proline (n=5) and glucose (n=7) cultures.

*n=5. The peak area for individual components in a pellet extract have been divided by the peak are for trypanothione pbtained for that pellet to reduce variations due to cell numbers and extraction effecienty. ‡ Determined in negative ion mode.

The retention time of each of these compounds was matched to those of standards where available. In addition, it was sometimes possible to confirm the identities of the compounds without carrying out further experiments since there is always a small amount of fragmentation of the main peak. Identification of some of the compounds in the trypanosome metabolome on the basis of their fragmentation pattern is shown in table 5.5. As stated above often there is only one metabolite, or one likely metabolite, in the Metlin data base matching a particular exact mass. A problem can arise if there are isomers of a particular compound. Glutamic acid

provides a good example since three other metabolites (although glutamic acid is by far the most abundant) are listed in the Metlin database as being isomers of glutamic acid: *N*-methylaspartic acid and *O*- and *N*-acetylserine. The following ions can be observed to arise from glutamic acid: m/z 131 due to loss of ammonia (eliminates *N*-methyl aspartic acid and *N*-acetyl serine), m/z 130 due to loss of water, m/z 102 due to loss of formic acid and m/z 84 due loss of formic acid followed by loss of water (eliminates *O*-acetylserine, the standard of which elutes before glutamic acid anyway). Extracted ion traces for the fragment ions all give chromatographic peaks corresponding to the peak of the parent compound. Examples of observed spontaneous fragmentations of other metabolites are listed in table5.5.

Compound	**Confirmatory Ions (nominal mass + i.d. + ppm deviation)**
Acetyl choline	$C_7H_{16}O_2N$ (148 M^+ -0.72), $C_4H_7O_2$ (87 M^+ - $N(CH_3)_3$ -1.2), $C_5H_{12}N$ (86 M^+ - CH_3COOH -1.2)
Acetyllysine	$C_8H_{17}O_3N_2$ (189 MH^+ -1.0), $C_7H_{15}ON_2$ (143 MH^+ - HCOOH-1.3), $C_7H_{12}ON$ (126 M^+ - HCOOH-NH_3 -1.4)
Carboxypyrroline	$C_5H_8O_2N$ (114 M^+ -1.3), C_4H_8ON(86 M^+ - CO -2.3)
Choline	$C_5H_{14}ON$ (104 M^+ -0.87), $C_5H_{12}N$ (82 M^+ - H_2O -1.0).
Glutamic acid	$C_5H_{10}O_4N$ (148 MH^+ -0.77), $C_5H_7O_4$ (131 MH^+ - NH_3 -1.49), $C_5H_8O_3N$ (130 MH^+ - H_2O -1.0), $C_4H_8O_2N$ (102 MH^+ -HCOOH -0.84), C_4H_6ON (84 MH^+ -HCOOH- H_2O -1.4)
Glutathione	$C_{10}H_{18}O_6N_3S$ (308 MH^+ -0.18), $C_8H_{15}O_5N_2S$ (251 MH^+ - glyc-1.3), $C_5H_{11}O_3N_2S$ (179 MH^+ - glut -0.51), $C_5H_8O_3N$ (130 MH^+ -cys-gly -1.2)
Glycine-glutamate	$C_7H_{13}O_5N_2$ (205 MH^+ -0.33), $C_5H_{11}O_3N_2$ (147 MH^+ - glyc-0.33),
Oxomethionine	$C_5H_{12}O_3NS$ (166 MH^+ -0.37), $C_5H_9O_3S$ (149 MH^+ - NH_3 - 1.3), C_4H_7OS (MH^+ - NH_3 -HCOOH -0.45).
Succinic acid	$C_4H_5O_4$ (117 MH^- +5.3), $C_4H_3O_3$ (99 MH^- - H_2O +6.9),

Table.5.5 Spontaneous fragmentation patterns for some of the metabolites listed in tables 5.3 and 5.4.

One drawback of Sieve® is that it disregards peaks below a certain signal-to-noise ratio. For instance 5-carboxypyrroline, the product of the first step in proline catabolism, can be observed in the proline grown but not in glucose grown trypanosomes (figure5.7), but only by manual processing of the ion current trace.

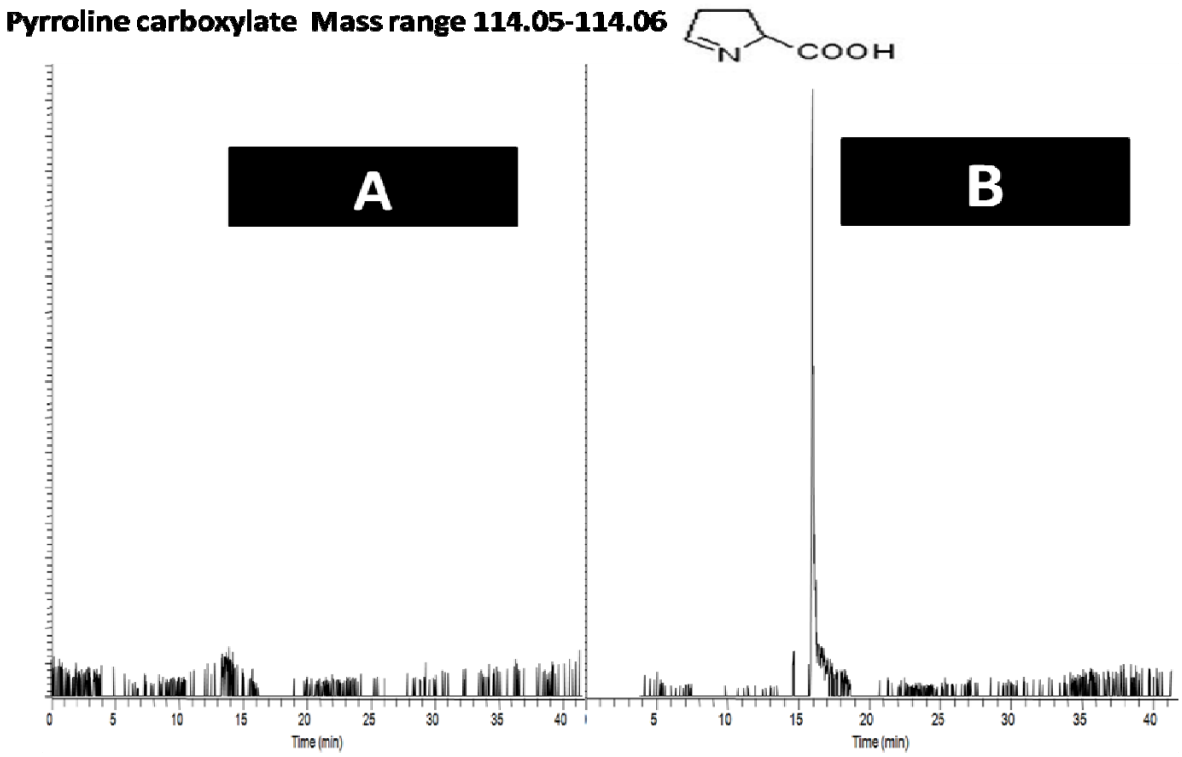


Figure.5.7 Extracted ion current for pyrroline carboxylate (deviation from exact mass -0.92 ppm) in positive ion mode in (A) glucose grown trypanosomes and (B) proline grown trypanosomes.

Figure.5.8 shows phospholipids from glucose and proline grown trypanosomes. There are clear differences in the abundance of various peaks between the two culture conditions. For instance a lipid peak at m/z 728 is more abundant in the glucose grown than proline grown trypanosomes (figure.5.8) but this was not noted by Sieve®.

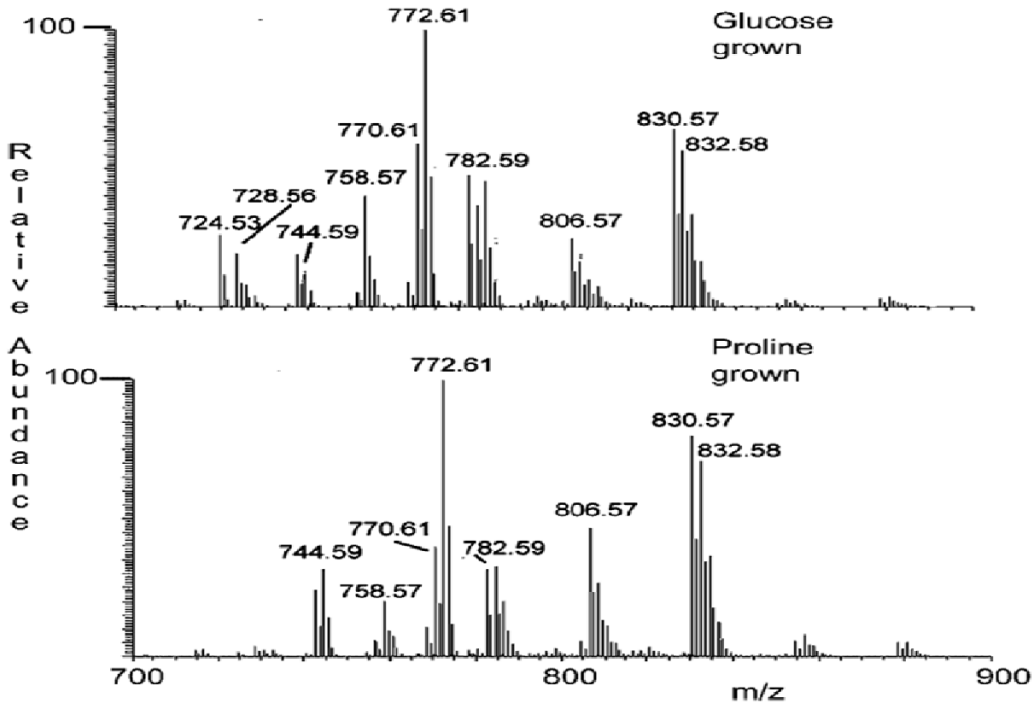


Figure.5.8 Phospholipid peak in positive ion mode from *T.brucei* pellet extracts from proline and glucose rich cultures showing variations in the pattern of phospholipid peaks.

The MS metabolome analysis of Trypanosomatides and other cell culture based organisms has its own challenges. The metabolite rich media culture places these cells in a very high homeostasis system which may mask any metabolite variance, and contribute to the metabolic fingerprint detected by the MS. The use of labelled isotope precursors in this regard may bear a special importance.

5.4. Conclusion:

It was shown that coupling hydrophilic interaction chromatography (HILIC) to an Orbitrap Fourier transform mass spectrometer (FT-MS) provided an excellent tool for metabolic profiling of trypanosome extracts, principally due to rapid elution of lipids in advance of most other metabolites prior to entering the ion source. We used in vitro cultivated procyclic forms of the protozoan parasite *T.brucei* grown with glucose or proline as a carbon source for testing the performance of the HILIC column and the mass accuracy of MS and we could identify a large number of metabolites in these samples. It is clearly shown here that many metabolites from trypanosomes, where sample processing is difficult, can be captured by rapid

quenching of the cells by adding them to hot ethanol and metabolites can then be quantified in a relative sense between samples grown in different media. It was, for example, possible to identify the trypanosome's signature metabolite, trypanothione, and also glutathione which were well retained by the HILIC column. HILIC columns are very robust because the high organic content used to elute them keeps them free from highly lipophilic materials and the high organic content in the mobile phase is also beneficial with respect to the efficiency of the ESI process. The ZIC-HILIC column used in these studies appears to be particularly effective. The mass accuracy achieved fell within 3 ppm for most of the metabolites identified within the samples. By comparing trypanosomes grown in two different media we were able to clearly distinguish the samples in terms of the relative abundance of a number of metabolites using Sieve® 1.1 software. However, at this stage the intention is to present the methodology rather than discuss the biological implications of the observations. In the high resolution environment of mass analyzers like Orbitrap, and unless a specific task is required from the analysis, It would seem that spending time optimising different chromatographic gradients is less important than selecting a "fit for purpose" type of chromatography run under a universal gradient. This was clearly shown in Chapter 2 where the preference of ZIC-HILIC over Luna-HILIC was minimal. As mentioned in chapter 3, at the time of this work, the data pre-processing tools did not match the performance of the Orbitrap, and an imperative need for such tools was a mark of that stage, is expressed in many times within this thesis.

6.Summary and Future Work:

6.1. Summary and final conclusion:

The work presented work in this thesis discussed the potential for combining HILIC with FTMS for metabolic profiling or pharmaceutical analyses. The work was focused on metabolic profiling of biological systems, and the selected systems were *D.melanogaster* and *T.brucei*. The work started with a computer simulation of the data to emphasise the importance of time (chromatographic) separation over the physical separation (mass resolution) and concluded that even a hypothetical non-existent instrument with a resolution of 250000 cannot resolve all the peaks even for a small set of metabolites. The work was pursued the testing of different chromatographic separation platforms concluding that complementary of data between HILIC and reversed-phase chromatography gave the best coverage of the metabolome. Nevertheless, one of the methods could be chosen over the other where there was a priori knowledge of the research question. The HILIC approach was further studied and four different commercially available columns were compared for the suitability for the analysis goals. The columns tested were three Zwitterionic (ZIC-HILIC, Obelisc N and Obelisc R) phases and one neutral HILIC phase (diol column, Luna-HILIC). The evaluation relied on multiple criteria calculated from data generated by running 140 authentic standards through the four different columns with the same mobile phase composition and gradient. The Obelisc columns required a long equilibrium time and Obelisc N retained the standards of the metabolites strongly showing wider peaks and lower signals compared to the other phases, while Obelisc R showed a high background level in the mass spectrometer especially in negative ion mode. ZIC-HILIC and Luna-HILIC columns were comparable in terms of the peak widths and signals of the authentic standards. However, ZIC-HILIC showed a better retention profile for standards compared to Luna-HILIC especially in the positive ion mode and this led to a better signals being observed when analyzing quality control *Drosophila* samples on both columns. Further optimisation was carried out on ZIC-HILIC by studying the optimal organic modifier and mobile phase additives for the best chromatographic.. Acetonitrile provided the best elution profile of the three organic modifiers and increasing the mobile phase pH did not improve the peak

shapes of acidic metabolites. The selected chromatographic method was tested for repeatability and reproducibility of retention times in biological samples over a span of about a month and found to be consistent with a general coefficient of variation in retention times below 2% for all the tested metabolites (n=35) and below 1.2% of many of these metabolites. The data generated by metabolic profiling using high resolution mass spectrometry is rich and complex, and with the advances of technology seeming to be faster than the processing software, a part of the work of this thesis was dedicated for composing computer sub-routine to pre-process the mass spectrometry data including alignment, filtration and background subtraction, screening the m/z features against KEGG and screening the KEGG pathways and classifying the contribution of each pathway to the distance between the pairs of studied groups. The proposed collection of sub routines (metabolomics aid kit MAK) was tested using *Drosophila* samples with a known lesion (ry compared to WT *Drosophila*) and could successfully isolate the lesion pathway (purine metabolism) as the strongest discriminating pathway. MAK was then compared with the later versions of software Sieve® and XCMS, which are compatible with LC-Orbitrap data, testing the ability of the three packages to correctly report the m/z and intensity of 14 metabolites and 2 additional isotope peaks of one of these metabolites. The MAK method, although missing the retention time domain could report correct m/z and intensity for peaks with no isomers. The two other packages reported correct intensities for peaks with no isomers but mistakes happened in the integration of some of the peaks when a neighbouring isomer peak occurred. Processing of the XCMS output with MAK based tools was tested and it was shown to to be complementary to XCMS leading to more confidence in peak assignment and discriminating pathway detection.

During the course of my PhD studies I applied HILIC-FTMS to the study of several genotypes and phenotypes of the fruit fly *Drosophila melanogaster* and the work in this thesis includes only one example of these studies. The studied example was rosy (ry) a naturally occurring genotype reported since 1910 and kept in stock until the present time. Ry flies were compared to wild type flies and differences in metabolic profiles acquired via the HILIC-FTMS combination

analysis and this confirmed the already reported lesion which affects xanthine oxidase activity but extended the knowledge to identify an extended effect of this single perturbation to include arginine, tryptophan and pyrimidine metabolic pathways. A significant effect could be related to osmosis and osmolytic factors glycerphosphocholine and glycerphosphoethanoalamine which was explained in the light of the inability of the flies to excrete nitrogen in the form of uric acid crystals which would lead to increased water excretion and higher osmalility.

The the HILIC-FTMS method was also applied to analysis of the metabolome of *Trypanosoma brucei*. It was demonstrated that the rich information of metabolic profiles could be used to simplify the interpretation of results through the correlation of co-eluting peaks across the spectra which potentially originate from the same metabolite in combination with the assignment elemental compositions to the parent peak and its minor fragments gving increasing confidence in peak identification. This is valuable in the cases where databases such Metlin, report multiple hits for the same elemental composition. The parasites were grown in media rich with proline or rich with glucose. The two metabolic profiles were distinct and different from each other and the method could pick different species of metabolites including amino acids, lipids and polyamines, especially the very polar anti-oxidant trypanothione which has been reported as a target for anti-trypanosomal drug discovery.

6.2. Future work:

The work on this thesis forms only the beginning of a wide range of applications of the HILIC-FTMS combination for studying biological systems. It is considered in the line of developmental work where a show case has been presented. Hence, more validation work should be carried out, especially with regard to the testing the validity of *Drosophila* as an animal model to study human diseases. The extraction procedures, metabolite coverage and the long term concatenation of data need to be further investigated. The analytical method could generate a database of the part of the *Drosophila* metabolome amenable to detection with this method and this can be used to screen the metabolome of specific organs within *Drosophila* where a phenotype has been observed and is of

special interest. For example, the ry malphigian tubules (fly kidney) which have been noticed to be swollen compared to WT. Some of the features of the proposed software MAK have been incorporated in other packages, but some of them, for example across sample set correlation and background filtration have not been described yet, and converting the code from VBA to the R package and incorporating these features in the open source XCMS package could add to the quality of this tool. The chromatographic methods presented here are suitable for the indicated applications, but need to be validated for suitability to other, more complex, biological samples, for example human plasma.

6.3. Publications:

The work on metabolic profiling of biological systems has generated the following scientific publications:

Kamleh, A., Barrett, M. P., Wildridge, D., Burchmore, R. J. S., Scheltema, R. A., & Watson, D. G. (2008). Metabolomic profiling using Orbitrap Fourier transform mass spectrometry with hydrophilic interaction chromatography: a method with wide applicability to analysis of biomolecules. Rapid Communications in Mass Spectrometry, 22(12), 1912-1918.

Kamleh, M. A., Hobani, Y., Dow, J. A. T., & Watson, D. G. (2008). Metabolomic profiling of *Drosophila* using liquid chromatography Fourier transform mass spectrometry. FEBS Letters, 582(19), 2916-2922.

Kamleh, M. A., Hobani, Y., Dow, J. A. T., Zheng, L., & Watson, D. G. (2009). Towards a platform for the metabonomic profiling of different strains of *Drosophila melanogaster* using liquid chromatography–Fourier transform mass spectrometry. FEBS Journal, 276(22), 6798-6809.

Scheltema, R. A., Kamleh, A., Wildridge, D., Ebikeme, C., Watson, D. G., Barrett, M. P., et al. (2008). Increasing the mass accuracy of high-resolution LC-MS data using background ions – a case study on the LTQ-Orbitrap. PROTEOMICS, 8(22), 4647-4656.

Vincent, I. M., Creek, D., Watson, D. G., Kamleh, M. A., Burchmore, R. J. S., & Barrett , M. P. (2010). A Molecular Mechanism for Eflornithine Resistance in African Trypanosomes. PloS Pathogens.

During the time of my PhD, I participated as well in the following reviews:

Hamoudeh, M., Kamleh, M. A., Diab, R., & Fessi, H. (2008). Radionuclides delivery systems for nuclear imaging and radiotherapy of cancer. Advanced Drug Delivery Reviews, 60(12), 1329-1346.

Kamleh, M. A., Dow, J. A. T., & Watson, D. G. (2009). Applications of mass spectrometry in metabolomic studies of animal model and invertebrate systems. Briefings in Functional Genomics & Proteomics, 8(1), 28-48.

Additional publications in the pipeline include:

Kamleh, M. A. (2010). Cholinergic Drugs. In D. G. Watson (Ed.), Pharmaceutical Chemistry (1st ed., Vol. 1, pp. 307-333). London: Elsevier.

Kamleh, M. A., Spagou, K., & and Want, E. J. (2010). Metabolic Profiling in Disease Diagnosis, Toxicology and Personalized Healthcare.

Current Pharmaceutical Biotechnology.

7. References:

ABOAGYE, E. O. & BHUJWALLA, Z. M. (1999) Malignant Transformation Alters Membrane Choline Phospholipid Metabolism of Human Mammary Epithelial Cells. *Cancer Research,* 59**,** 80-84.

ABU B. KANU, PRABHA DWIVEDI , MAGGIE TAM , LAURA MATZ & HERBERT H. HILL, J. (2008) Ion mobility-mass spectrometry. *Journal of Mass Spectrometry,* 43**,** 1-22.

ADAHCHOUR, M., BEENS, J. & BRINKMAN, U. A. T. (2008) Recent developments in the application of comprehensive two-dimensional gas chromatography. *Journal of Chromatography A,* 1186**,** 67-108.

ADAMS, M. D. & SEKELSKY, J. J. (2002) From sequence to phenotype: reverse genetics in Drosophila melanogaster. *Nat Rev Genet,* 3**,** 189-98.

AL-ASMARI, A. I., ANDERSON, R. A. & APPELBLAD, P. (2010) Direct Determination of Ethyl Glucuronide and Ethyl Sulfate in Postmortem Urine Specimens Using Hydrophilic Interaction Liquid ChromatographyElectrospray Ionization-Tandem Mass Spectrometry. *Journal of Analytical Toxicology,* 34**,** 261-272.

ALLEN, J., DAVEY, H. M., BROADHURST, D., HEALD, J. K., ROWLAND, J. J., OLIVER, S. G. & KELL, D. B. (2003) High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat Biotechnol,* 21**,** 692-6.

ALPERT, A. J. (1990) Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *Journal of Chromatography A,* 499**,** 177-196.

ALZWEIRI, M., WATSON, D. G., ROBERTSON, C., SILLS, G. J. & PARKINSON, J. A. (2008) Comparison of different water-miscible solvents for the preparation of plasma and urine samples in metabolic profiling studies. *Talanta,* 74**,** 1060-1065.

AMADOR-MUÑOZ, O. & MARRIOTT, P. J. (2008) Quantification in comprehensive two-dimensional gas chromatography and a model of quantification based on selected summed modulated peaks. *Journal of Chromatography A,* 1184**,** 323-340.

ANDREEV, V. P., REJTAR, T., CHEN, H.-S., MOSKOVETS, E. V., IVANOV, A. R. & KARGER, B. L. (2003) A Universal Denoising and Peak Picking Algorithm for LC-MS Based on Matched Filtration in the Chromatographic Time Domain. *Analytical Chemistry,* 75**,** 6314-6326.

ANTHONY, M. L., SWEATMAN, B. C., BEDDELL, C. R., LINDON, J. C. & NICHOLSON, J. K. (1994) Pattern recognition classification of the site of nephrotoxicity based on metabolic data derived from proton nuclear magnetic resonance spectra of urine. *Molecular Pharmacology,* 46**,** 199-211.

ARBONA, V., IGLESIAS, D. J., TALOÌ N, M. & GOÌ MEZ-CADENAS, A. (2009) Plant Phenotype Demarcation Using Nontargeted LC-MS and GC-MS Metabolite Profiling. *Journal of Agricultural and Food Chemistry,* 57**,** 7338-7347.

ARIYANAYAGAM, M. R. & FAIRLAMB, A. H. (2001) Ovothiol and trypanothione as antioxidants in trypanosomatids. *Molecular and Biochemical Parasitology,* 115**,** 189-198.

ASHBURNER, M. (1989) *Drosophila: a laboratory manual*, Cold Spring Harbour Laboratory Press.

ATHERTON, H. J., JONES, O. A. H., MALIK, S., MISKA, E. A. & GRIFFIN, J. L. (2008) A comparative metabolomic study of NHR-49 in Caenorhabditis elegans and PPAR-[alpha] in the mouse. *FEBS letters,* 582**,** 1661-1666.

BAJAD, S. U., LU, W., KIMBALL, E. H., YUAN, J., PETERSON, C. & RABINOWITZ, J. D. (2006) Separation and quantitation of water soluble cellular metabolites by hydrophilic

interaction chromatography-tandem mass spectrometry. *Journal of Chromatography A,* 1125**,** 76-88.
BAKER, D. L., MORRISON, P., MILLER, B., RIELY, C. A., TOLLEY, B., WESTERMANN, A. M., BONFRER, J. M. G., BAIS, E., MOOLENAAR, W. H. & TIGYI, G. (2002) Plasma Lysophosphatidic Acid Concentration and Ovarian Cancer. *JAMA,* 287**,** 3081-3082.
BALES, J. R., BELL, J. D., NICHOLSON, J. K. & SADLER, P. J. (1986) ^{1}H NMR studies of urine during fasting: Excretion of ketone bodies and acetylcarnitine. *Magnetic Resonance in Medicine,* 3**,** 849-856.
BALLINGER, D. G. & BENZER, S. (1989) Targeted gene mutations in *Drosophila*. *Proc. Natl Acad. Sci. USA,* 86**,** 9402-6.
BARGMANN, C. I. (2001) High-throughput reverse genetics: RNAi screens in Caenorhabditis elegans. *Genome Biol,* 2**,** REVIEWS1005.
BARRETT, M. P., TETAUD, E., SEYFANG, A., BRINGAUD, F. & T., B. (1995) Functional expression and characterization of the Trypanosoma brucei procyclic glucose transporter, THT2. 312**,** 687-691.
BARTON, S., HOWE, F., TOMLINS, A., CUDLIP, S., NICHOLSON, J., ANTHONY BELL, B. & GRIFFITHS, J. (1999) Comparison of in vivo 1-H MRS of human brain tumours with 1-H HR-MAS spectroscopy of intact biopsy samples in vitro. *Magnetic Resonance Materials in Physics, Biology and Medicine,* 8**,** 121-128.
BAUGHMAN, T. M., WRIGHT, W. L. & HUTTON, K. A. (2007) Determination of zanamivir in rat and monkey plasma by positive ion hydrophilic interaction chromatography (HILIC)/tandem mass spectrometry. *Journal of Chromatography B,* 852**,** 505-511.
BEAUDRY, F. & VACHON, P. (2006) Electrospray ionization suppression, a physical or a chemical phenomenon? *Biomedical Chromatography,* 20**,** 200-205.
BECKWITH-HALL, B. M., BRINDLE, J. T., BARTON, R. H., COEN, M., HOLMES, E., NICHOLSON, J. K. & ANTTI, H. (2002) Application of orthogonal signal correction to minimise the effects of physical and biological variation in high resolution 1H NMR spectra of biofluids. *Analyst,* 127**,** 1283-1288.
BEILMANN, B., LANGGUTH, P., HÄUSLER, H. & GRASS, P. (2006) High-performance liquid chromatography of lactose with evaporative light scattering detection, applied to determine fine particle dose of carrier in dry powder inhalation products. *Journal of Chromatography A,* 1107**,** 204-207.
BENGTSSON, J., JANSSON, B. & HAMMARLUND-UDENAES, M. (2005) On-line desalting and determination of morphine, morphine-3-glucuronide and morphine-6-glucuronide in microdialysis and plasma samples using column switching and liquid chromatography/tandem mass spectrometry. *Rapid Communications in Mass Spectrometry,* 19**,** 2116-2122.
BENTON, H. P., WONG, D. M., TRAUGER, S. A. & SIUZDAK, G. (2008) XCMS2: Processing Tandem Mass Spectrometry Data for Metabolite Identification and Structural Characterization. *Analytical Chemistry,* 80**,** 6382-6389.
BERG, I., NILSSON, K. P. R., THOR, S. & HAMMARSTROM, P. (2010) Efficient imaging of amyloid deposits in Drosophila models of human amyloidoses. *Nat. Protocols,* 5**,** 935-944.
BESTEIRO, S. B., BIRAN, M., BITEAU, N., COUSTOU, V., BALTZ, T. O., CANIONI, P. & BRINGAUD, F. D. R. (2002) Succinate Secreted by Trypanosoma brucei Is Produced by a Novel and Unique Glycosomal Enzyme, NADH-dependent Fumarate Reductase. *Journal of Biological Chemistry,* 277**,** 38001-38012.
BIJ, K. E., HORVÁTH, C., MELANDER, W. R. & NAHUM, A. (1981) Surface silanols in silica-bonded hydrocarbonaceous stationary phases : II. Irregular retention behavior and effect of silanol masking. *Journal of Chromatography A,* 203**,** 65-84.

BOERSEMA, P. J., DIVECHA, N., HECK, A. J. R. & MOHAMMED, S. (2007) Evaluation and Optimization of ZIC-HILIC-RP as an Alternative MudPIT Strategy. *Journal of Proteome Research,* 6**,** 937-946.

BOLLARD, M. E., STANLEY, E. G., LINDON, J. C., NICHOLSON, J. K. & HOLMES, E. (2005) NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. *NMR in Biomedicine,* 18**,** 143-162.

BOLTEN, C. J., KIEFER, P., LETISSE, F., PORTAIS, J.-C. & WITTMANN, C. (2007) Sampling for Metabolome Analysis of Microorganisms. *Analytical Chemistry,* 79**,** 3843-3849.

BONFIGLIO, R., KING, R. C., OLAH, T. V. & MERKLE, K. (1999) The effects of sample preparation methods on the variability of the electrospray ionization response for model drug compounds. *Rapid Communications in Mass Spectrometry,* 13**,** 1175-1185.

BONI, P. & PARISI, G. (1967) Abnormalities of the eye pigments (pteridins and ommochromes) induced in Drosophila melanogaster by the inhibitor of xanthine dehydrogenase 4-hydroxypyrazole (3,4 d) pyrimidine. *Experientia,* 23**,** 1020-1.

BONSE, A. (1967) [Studies on the chemical nature and formation of the urinary conglomerate in the Malpighian vessels of the rosy mutant of Drosophila melanogaster]. *Z Naturforsch B,* 22**,** 1027-1029.

BREITLING, R., PITT, A. R. & BARRETT, M. P. (2006a) Precision mapping of the metabolome. *Trends in Biotechnology,* 24**,** 543-548.

BREITLING, R., RITCHIE, S., GOODENOWE, D., STEWART, M. L. & BARRETT, M. P. (2006b) Ab initio prediction of metabolic networks using Fourier transform mass spectrometry data. *Metabolomics,* 2**,** 155-164.

BRERETON, R. G. (2002) *Chemometrics : data analysis for the laboratory and chemical plant / Richard Brereton,* New York :, J. Wiley.

BRINDLE, J. T., ANTTI, H., HOLMES, E., TRANTER, G., NICHOLSON, J. K., BETHELL, H. W. L., CLARKE, S., SCHOFIELD, P. M., MCKILLIGIN, E., MOSEDALE, D. E. & GRAINGER, D. J. (2002) Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. *Nat Med,* 8**,** 1439-1445.

BRINDLE, J. T., NICHOLSON, J. K., SCHOFIELD, P. M., GRAINGER, D. J. & HOLMES, E. (2003) Application of chemometrics to 1H NMR spectroscopic data to investigate a relationship between human serum metabolic profiles and hypertension. *Analyst,* 128**,** 32-36.

BRINGAUD, F., RIVIÈRE, L. & COUSTOU, V. (2006) Energy metabolism of trypanosomatids: Adaptation to available carbon sources. *Molecular and Biochemical Parasitology,* 149**,** 1-9.

BROWN, M., DUNN, W. B., DOBSON, P., PATEL, Y., WINDER, C. L., FRANCIS-MCINTYRE, S., BEGLEY, P., CARROLL, K., BROADHURST, D., TSENG, A., SWAINSTON, N., SPASIC, I., GOODACRE, R. & KELL, D. B. (2009) Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. *Analyst,* 134**,** 1322-1332.

BROWN, S. D., WHITE, C. A. & BARTLETT, M. G. (2002) Hydrophilic interaction liquid chromatography/electrospray mass spectrometry determination of acyclovir in pregnant rat plasma and tissues. *Rapid Communications in Mass Spectrometry,* 16**,** 1871-1876.

BROWN, S. D. M. & PETERS, J. (1996) Combining mutagenesis and genomics in the mouse--closing the phenotype gap. *Trends Genet,* 12**,** 433-5.

BRUCE, S. J., TAVAZZI, I., PARISOD, V. R., REZZI, S., KOCHHAR, S. & GUY, P. A. (2009) Investigation of Human Blood Plasma Sample Preparation for Performing

Metabolomics Using Ultrahigh Performance Liquid Chromatography/Mass Spectrometry. *Analytical Chemistry,* 81**,** 3285-3296.
BRUN R, S. (1979) Cultivation and in vitro cloning or procyclic culture forms of Trypanosoma brucei in a semi-defined medium. Short communication. *Acta Trop,* 36**,** 289-292.
BURSELL, E. (Ed.) (1981) *The role of proline in energy metabolism,* New York, Plenum Press,.
BUSCHER, J. M., CZERNIK, D., EWALD, J. C., SAUER, U. & ZAMBONI, N. (2009) Cross-Platform Comparison of Methods for Quantitative Metabolomics of Primary Metabolism. *Analytical Chemistry,* 81**,** 2135-2143.
CANELAS, A., RAS, C., TEN PIERICK, A., VAN DAM, J., HEIJNEN, J. & VAN GULIK, W. (2008) Leakage-free rapid quenching technique for yeast metabolomics. *Metabolomics,* 4**,** 226-239.
CANNATA, J. & CAZZULO, J. (1984) The aerobic fermentation of glucose by Trypanosoma cruzi. *Comp Biochem Physiol B,* 79**,** 297-308.
CHALMERS, M. J. & GASKELL, S. J. (2000) Advances in mass spectrometry for proteome analysis. *Current Opinion in Biotechnology,* 11**,** 384-390.
CHANG, W., JUN, Y. & JIHUA, N. (2009) Plasma phospholipid metabolic profiling and biomarkers of rats following radiation exposure based on liquid chromatography-mass spectrometry technique. *Biomedical Chromatography,* 23**,** 1079-1085.
CHEN, J., WANG, W., LV, S., YIN, P., ZHAO, X., LU, X., ZHANG, F. & XU, G. (2009) Metabonomics study of liver cancer based on ultra performance liquid chromatography coupled to mass spectrometry with HILIC and RPLC separations. *Analytica Chimica Acta,* 650**,** 3-9.
CHIRITA, R.-I., WEST, C., FINARU, A.-L. & ELFAKIR, C. (2010) Approach to hydrophilic interaction chromatography column selection: Application to neurotransmitters analysis. *Journal of Chromatography A,* 1217**,** 3091-3104.
CLARIDGE, T. D. W. (1999) High-Resolution NMR Techniques in Organic Chemistry. pp 384 Comment: ISBN 0080427995.].
CLAYTON, T. A., BAKER, D., LINDON, J. C., EVERETT, J. R. & NICHOLSON, J. K. (2009) Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. *Proceedings of the National Academy of Sciences,* 106**,** 14728-14733.
CLAYTON, T. A., LINDON, J. C., CLOAREC, O., ANTTI, H., CHARUEL, C., HANTON, G., PROVOST, J. P., LE NET, J. L., BAKER, D., WALLEY, R. J., EVERETT, J. R. & NICHOLSON, J. K. (2006) Pharmaco-metabonomic phenotyping and personalized drug treatment. *Nature,* 440**,** 1073 - 7.
CLOAREC, O., DUMAS, M.-E., CRAIG, A., BARTON, R. H., TRYGG, J., HUDSON, J., BLANCHER, C., GAUGUIER, D., LINDON, J. C., HOLMES, E. & NICHOLSON, J. (2005a) Statistical Total Correlation Spectroscopy- An Exploratory Approach for Latent Biomarker Identification from Metabolic ^{1}H NMR Data Sets. *Analytical Chemistry,* 77**,** 1282-1289.
CLOAREC, O., DUMAS, M. E., CRAIG, A., BARTON, R. H., TRYGG, J., HUDSON, J., BLANCHER, C., GAUGUIER, D., LINDON, J. C., HOLMES, E. & NICHOLSON, J. (2005b) Statistical Total Correlation Spectroscopy: An Exploratory Approach for Latent Biomarker Identification from Metabolic ^{1}H NMR Data Sets.
COEN, M., HONG, Y. S., CLAYTON, T. A., ROHDE, C. M., PEARCE, J. T., REILY, M. D., ROBERTSON, D. G., HOLMES, E., LINDON, J. C. & NICHOLSON, J. K. (2007a) The Mechanism of Galactosamine Toxicity Revisited; A Metabonomic Study. *Journal of Proteome Research,* 6**,** 2711-2719.
COEN, M., HONG, Y. S., CLOAREC, O., RHODE, C. M., REILY, M. D., ROBERTSON, D. G., HOLMES, E., LINDON, J. C. & NICHOLSON, J. K. (2007b) Heteronuclear ^{1}H-^{31}P

statistical total correlation NMR spectroscopy of intact liver for metabolic biomarker assignment: application to galactosamine-induced hepatotoxicity. *Anal Chem,* 79**,** 8956-66.

COEN, M., WANT, E. J., CLAYTON, T. A., RHODE, C. M., HONG, Y. S., KEUN, H. C., CANTOR, G. H., METZ, A. L., ROBERTSON, D. G., REILY, M. D., HOLMES, E., LINDON, J. C. & NICHOLSON, J. K. (2009) Mechanistic Aspects and Novel Biomarkers of Responder and Non-Responder Phenotypes in Galactosamine-Induced Hepatitis. *Journal of Proteome Research,* 8**,** 5175-5187.

COTTRET, L., WILDRIDGE, D., VINSON, F., BARRETT, M. P., CHARLES, H., SAGOT, M.-F. & JOURDAN, F. (2010) MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. *Nucl. Acids Res.***,**312.

CREWS, B., WIKOFF, W. R., PATTI, G. J., WOO, H.-K., KALISIAK, E., HEIDEKER, J. & SIUZDAK, G. (2009) Variability Analysis of Human Plasma and Cerebral Spinal Fluid Reveals Statistical Significance of Changes in Mass Spectrometry-Based Metabolomics Data. *Analytical Chemistry,* 81**,** 8538-8544.

CROCKFORD, D. J., HOLMES, E., LINDON, J. C., PLUMB, R. S., ZIRAH, S., BRUCE, S. J., RAINVILLE, P., STUMPF, C. L. & NICHOLSON, J. K. (2006) Statistical Heterospectroscopy, an Approach to the Integrated Analysis of NMR and UPLC-MS Data Sets: Application in Metabonomic Toxicology Studies.

CUBBON, S., BRADBURY, T., WILSON, J. & THOMAS-OATES, J. (2007) Hydrophilic Interaction Chromatography for Mass Spectrometric Metabonomic Studies of Urine. *Analytical Chemistry,* 79**,** 8911-8918.

CWIK, M. J., WU, H., MUZZIO, M., MCCORMICK, D. L. & KAPETANOVIC, I. (2010) Direct quantitation of glucoraphanin in dog and rat plasma by LC-MS/MS. *Journal of Pharmaceutical and Biomedical Analysis,* 52**,** 544-549.

DAI, Y., LI, Z., XUE, L., DOU, C., ZHOU, Y., ZHANG, L. & QIN, X. (2010) Metabolomics study on the anti-depression effect of xiaoyaosan on rat model of chronic unpredictable mild stress. *Journal of Ethnopharmacology,* 128**,** 482-489.

DANIEL, M., JOSÉ MANUEL, M., ANTONIO, B., JOSE ANTONIO, L., CARLOS, V. & BERNARDO, C. (2009) Metabolite profiling of fecal water extracts from human colorectal cancer. *NMR in Biomedicine,* 22**,** 342-348.

DE HOFFMANN, E. & STROOBANT, V. (2001) Mass spectrometry: Principles and applications.

DELL'AVERSANO, C., HESS, P. & QUILLIAM, M. A. (2005) Hydrophilic interaction liquid chromatography-mass spectrometry for the analysis of paralytic shellfish poisoning (PSP) toxins. *Journal of Chromatography A,* 1081**,** 190-201.

DENT, C. E. & PHILPOT, G. R. (1954) Xanthinuria: an inborn error of metabolism. *Lancet,* I**,** 182-185.

DESMET, G. (2008) Comparison techniques for HPLC column performance. *LC-GC Europe,* 21**,** 310-320.

DOW, J. A. T. (2003) The Drosophila phenotype gap - and how to close it. *Brief Funct Genomic Proteomic,* 2**,** 121-7.

DOW, J. A. T. (2007) Model organisms and molecular genetics for endocrinology. *Gen. Comp. Endocrinol.,* 153 3–12.

DOW, J. A. T. & DAVIES, S. A. (2003) Integrative physiology and functional genomics of epithelial function in a genetic model organism. *Physiol. Reviews,* 83**,** 687-729.

DUNN, W. B., BAILEY, N. J. C. & JOHNSON, H. E. (2005) Measuring the metabolome: current analytical technologies. *The Analyst,* 130**,** 606-625.

DUNN, W. B., BROADHURST, D., BROWN, M., BAKER, P. N., REDMAN, C. W. G., KENNY, L. C. & KELL, D. B. (2008) Metabolic profiling of serum using Ultra Performance Liquid

Chromatography and the LTQ-Orbitrap mass spectrometry system. *Journal of Chromatography B,* 871**,** 288-298.
EBBELS, T., KEUN, H., BECKONERT, O., ANTTI, H., BOLLARD, M., HOLMES, E., LINDON, J. & NICHOLSON, J. (2003) Toxicity classification from metabonomic data using a density superposition approach: `CLOUDS'. *Analytica Chimica Acta,* 490**,** 109-122.
EBBELS, T. M. D., KEUN, H. C., BECKONERT, O. P., BOLLARD, M. E., LINDON, J. C., HOLMES, E. & NICHOLSON, J. K. (2007) Prediction and Classification of Drug Toxicity Using Probabilistic Modeling of Temporal Metabolic Data: The Consortium on Metabonomic Toxicology Screening Approach. *J. Proteome Res.,* 6**,** 4407-4422.
EL RASSI, Z. (1996) Recent progress in reversed-phase and hydrophobic interaction chromatography of carbohydrate species. *Journal of Chromatography A,* 720**,** 93-118.
ERIC, C. Y. C., SWEE-LEE, Y., AIK-JIANG, L., PAY-CHIN, L., DING-FUNG, T. & HWEE-LING, K. (2007) Ultra-performance liquid chromatography/time-of-flight mass spectrometry based metabolomics of raw and steamed *Panax notoginseng*. *Rapid Communications in Mass Spectrometry,* 21**,** 519-528.
ERVE, J. C. L., BEYER, C. E., MANZINO, L. & TALAAT, R. E. (2009) Metabolite identification in rat brain microdialysates by direct infusion nanoelectrospray ionization after desalting on a ZipTip and LTQ/Orbitrap mass spectrometry. *Rapid Communications in Mass Spectrometry,* 23**,** 4003-4012.
EWING, G. W. (1997) *Analytical instrumentation handbook / edited by Galen Wood Ewing,* New York :, M. Dekker.
FAIRLAMB, A. & OPPERDOES, F. (Eds.) (1986) *Carbohydrate Metabolism in Cultured Cells*, Plenum Publishing Corporation.
FAN, T., LANE, A., HIGASHI, R., FARAG, M., GAO, H., BOUSAMRA, M. & MILLER, D. (2009) Altered regulation of metabolic pathways in human lung cancer discerned by 13C stable isotope-resolved metabolomics (SIRM). *Molecular Cancer,* 8**,** 41.
FARDET, A., LLORACH, R., MARTIN, J.-F., X, OIS, BESSON, C., LYAN, B., PUJOS-GUILLOT, E. & SCALBERT, A. (2008) A Liquid Chromatography-Quadrupole Time-of-Flight (LC-QTOF)-based Metabolomic Approach Reveals New Metabolic Effects of Catechin in Rats Fed High-Fat Diets. *J. Proteome Res.,* 7**,** 2388-2398.
FEALA, J. D., COQUIN, L., MCCULLOCH, A. D. & PATERNOSTRO, G. (2007) Flexibility in energy metabolism supports hypoxia tolerance in Drosophila flight muscle: metabolomic and computational systems analysis. *Mol Syst Biol,* 3**,** 99.
FENN, J. B., MANN, M., MENG, C. K., WONG, S. F. & WHITEHOUSE, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science,* 246**,** 64-71.
FIEHN, O. (2001) "Combining Genomics, Metabolome Analysis, and Biochemical Modelling to Understand Metabolic Networks," *Comparative and Functional Genomics,* 2**,** 155-168.
FIEHN, O. (2002) Metabolomics – the link between genotypes and phenotypes. *Plant Molecular Biology,* 48**,** 155-171.
FIEHN, O. & KIND, T. (2006) Metabolite profiling in blood plasma. IN WECKWERTH, W. (Ed.) *Metabolomics: Methods and protocols.* Totowa NJ, Humana Press.
FIEHN, O. & KIND, T. (2007) Metabolite Profiling in Blood Plasma. *Metabolomics.*
FITZGERALD, M. P., MADSEN, J. M., COLEMAN, M. C., TEOH, M. L. T., WESTPHAL, S. G., SPITZ, D. R., RADI, R. & DOMANN, F. E. (2010) Transgenic Biosynthesis of Trypanothione Protects Escherichia coli from Radiation-Induced Toxicity. *Radiation Research*.
FOURDINIER, M., BOSTYN, S., DELÉPÉE, R. & FAUDUET, H. (2010) Interest of a chemometric approach in understanding the retention behaviour of three columns in hydrophilic

interaction liquid chromatography: Application to the separation of glycerol carbonate, glycerol and urea. *Talanta,* 81**,** 1281-1287.

FRAGA, C. G., CLOWERS, B. H., MOORE, R. J. & ZINK, E. M. (2010) Signature-Discovery Approach for Sample Matching of a Nerve-Agent Precursor Using Liquid Chromatographyâˆ'Mass Spectrometry, XCMS, and Chemometrics. *Analytical Chemistry*.

FURUYA, T., KESSLER, P., JARDIM, A., SCHNAUFER, A., CRUDDER, C. & PARSONS, M. (2002) Glucose is toxic to glycosome-deficient trypanosomes. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 14177-14182.

FYRST, H., HERR, D. R., HARRIS, G. L. & SABA, J. D. (2004) Characterization of free endogenous C14 and C16 sphingoid bases from Drosophila melanogaster. *J. Lipid Res.,* 45**,** 54-62.

FYRST, H., ZHANG, X., HERR, D. R., BYUN, H. S., BITTMAN, R., PHAN, V. H., HARRIS, G. L. & SABA, J. D. (2008) Identification and characterization by electrospray mass spectrometry of endogenous Drosophila sphingadienes. *J. Lipid Res.,* 49**,** 597-606.

GALASSO, D., CARNUCCIO, A. & LARGHI, A. (2010) Pancreatic cancer: diagnosis and endoscopic staging. *EUR REV MED,* 14**,** 375-385

GALLAZZINI, M., FERRARIS, J. D., KUNIN, M., MORRIS, R. G. & BURG, M. B. (2006) Neuropathy target esterase catalyzes osmoprotective renal synthesis of glycerophosphocholine in response to high NaCl. *Proceedings of the National Academy of Sciences,* 103**,** 15260-15265.

GARCÍA-PÉREZ, I., VALLEJO, M., GARCÍA, A., LEGIDO-QUIGLEY, C. & BARBAS, C. (2008) Metabolic fingerprinting with capillary electrophoresis. *Journal of Chromatography A,* 1204**,** 130-139.

GARROD, S., HUMPFER, E., SPRAUL, M., CONNOR, S., POLLEY, S., CONNELLY, J., LINDON, J., NICHOLSON, J. & HOLMES, E. (1999) High-resolution magic angle spinning ^{1}H NMR spectroscopic studies on intact rat renal cortex and medulla. *Magnetic Resonance in Medicine,* 41**,** 1108-1118.

GEORGAKAKOU, S., KAZANIS, M. & PANDERI, I. (2010) Hydrophilic interaction liquid chromatography/positive ion electrospray ionization mass spectrometry method for the quantification of perindopril and its main metabolite in human plasma. *Analytical and Bioanalytical Chemistry,* 397**,** 2161-2170.

GERMAN, J. B., ROBERTS, M.-A. & WATKINS, S. M. (2003) Genomics and Metabolomics as Markers for the Interaction of Diet and Health: Lessons from Lipids.

GERMAN, J. B. A., GILLIES, L. A. A., SMILOWITZ, J. T. A., ZIVKOVIC, A. M. A. & WATKINS, S. M. B. (2007) Lipidomics and lipid profiling in metabolomics. *Current Opinion in Lipidology,* 18**,** 66-71.

GIKA, H. G., MACPHERSON, E., THEODORIDIS, G. A. & WILSON, I. D. (2008a) Evaluation of the repeatability of ultra-performance liquid chromatography-TOF-MS for global metabolic profiling of human urine samples. *Journal of Chromatography B,* 871**,** 299-305.

GIKA, H. G., THEODORIDIS, G., EXTANCE, J., EDGE, A. M. & WILSON, I. D. (2008b) High temperature-ultra performance liquid chromatography-mass spectrometry for the metabonomic analysis of Zucker rat urine. *Journal of Chromatography B,* 871**,** 279-287.

GIKA, H. G., THEODORIDIS, G. & WILSON, I. D. (2008c) Hydrophilic interaction and reversed-phase ultra-performance liquid chromatography TOF-MS for metabonomic analysis of Zucker rat urine. *Journal of Separation Science,* 31**,** 1598-1608.

GIKA, H. G., THEODORIDIS, G. A. & WILSON, I. D. (2008d) Liquid chromatography and ultra-performance liquid chromatography-mass spectrometry fingerprinting of human

urine: Sample stability under different handling and storage conditions for metabonomics studies. *Journal of Chromatography A,* 1189**,** 314-322.
GILDARDO RIVERA, V. B.-G., CYNTHIA ORDAZ-PICHARDO, BENJAMIN NOGUEDA-TORRES, ANTONIO MONGE (2009) New Therapeutic Targets for Drug Design Against Trypanosoma cruzi, Advances and Perspectives. *Current Medicinal Chemistry* 16**,** 3286-3293.
GIPSON, G., TATSUOKA, K., SOKHANSANJ, B., BALL, R. & CONNOR, S. (2008) Assignment of MS-based metabolomic datasets via compound interaction pair mapping. *Metabolomics,* 4**,** 94-103.
GISKEODEGARD, G. F., GRINDE, M. T., SITTER, B., AXELSON, D. E., LUNDGREN, S., FJOSNE, H. E., DAHL, S., GRIBBESTAD, I. S. & BATHEN, T. F. (2009) Multivariate Modeling and Prediction of Breast Cancer Prognostic Factors Using MR Metabolomics. *Journal of Proteome Research,* 9**,** 972-979.
GLASSMAN, E. (1965) Genetic regulation of xanthine dehydrogenase in *Drosophila melanogaster*. *Fed. Proc.,* 24**,** 1243--1251.
GOWDA, G. N. (2010) Human bile as a rich source of biomarkers for hepatopancreatobiliary cancers. *Biomarkers in Medicine,* 4**,** 299-314.
GRIFFIN, J. L. (2003) Metabonomics: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterisation of xenobiotic toxicity and disease diagnosis. *Current Opinion in Chemical Biology,* 7**,** 648-654.
GUAN, W., ZHOU, M., HAMPTON, C., BENIGNO, B., WALKER, L. D., GRAY, A., MCDONALD, J. & FERNANDEZ, F. (2009) Ovarian cancer detection from metabolomic liquid chromatography/mass spectrometry data by support vector machines. *BMC Bioinformatics,* 10**,** 259.
GUEVREMONT, R. & PURVES, R. W. (1999) Atmospheric pressure ion focusing in a high-field asymmetric waveform ion mobility spectrometer. *Review of Scientific Instruments* 70.
GUO, D., MYRDAL, P., KARLAGE, K., O'CONNELL, S., WISSINGER, T., TABIBI, S. & YALKOWSKY, S. (2010) Stability of 5-Fluoro-2′-deoxycytidine and Tetrahydrouridine in Combination. *AAPS PharmSciTech,* 11**,** 247-252.
GUO, Y. & GAIKI, S. (2005) Retention behavior of small polar compounds on polar stationary phases in hydrophilic interaction chromatography. *Journal of Chromatography A,* 1074**,** 71-80.
GUY, P. A., TAVAZZI, I., BRUCE, S. J., RAMADAN, Z. & KOCHHAR, S. (2008) Global metabolic profiling analysis on human urine by UPLC-TOFMS: Issues and method validation in nutritional metabolomics. *Journal of Chromatography B,* 871**,** 253-260.
HALKET, J. M., WATERMAN, D., PRZYBOROWSKA, A. M., PATEL, R. K. P., FRASER, P. D. & BRAMLEY, P. M. (2005) Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J. Exp. Bot.,* 56**,** 219-243.
HARDMAN, M. & MAKAROV, A. A. (2003) Interfacing the Orbitrap Mass Analyzer to an Electrospray Ion Source. *Anal. Chem.,* 75**,** 1699-1705.
HELLER, D. N. & NOCHETTO, C. B. (2008) Simultaneous determination and confirmation of melamine and cyanuric acid in animal feed by zwitterionic hydrophilic interaction chromatography and tandem mass spectrometry. *Rapid Communications in Mass Spectrometry,* 22**,** 3624-3632.
HENG-KEANG, L., JIE, C., CARLO, S., KEVIN, C. & VANGALA, S. (2007) Metabolite identification by data-dependent accurate mass spectrometric analysis at resolving power of 60 000 in external calibration mode using an LTQ/Orbitrap. *Rapid Communications in Mass Spectrometry,* 21**,** 1821-1832.

HIRAYAMA, A., KAMI, K., SUGIMOTO, M., SUGAWARA, M., TOKI, N., ONOZUKA, H., KINOSHITA, T., SAITO, N., OCHIAI, A., TOMITA, M., ESUMI, H. & SOGA, T. (2009) Quantitative Metabolome Profiling of Colon and Stomach Cancer Microenvironment by Capillary Electrophoresis Time-of-Flight Mass Spectrometry. *Cancer Res,* 69**,** 4918-4925.

HOCART, C. H., LEW, M. & HUNG-WEN, L. (2010) Mass Spectrometry: An Essential Tool for Trace Identification and Quantification. *Comprehensive Natural Products II.* Oxford, Elsevier.

HOLMES, E., CLOAREC, O. & NICHOLSON, J. K. (2006) Probing latent biomarker signatures and in vivo pathway activity in experimental disease states via statistical total correlation spectroscopy (STOCSY) of biofluids: application to HgCl2 toxicity. *J Proteome Res,* 5**,** 1313-20.

HOLMES, E., FOXALL, P. J. D., SPRAUL, M., DUNCAN FARRANT, R., NICHOLSON, J. K. & LINDON, J. C. (1997) 750 MHz 1H NMR spectroscopy characterisation of the complex metabolic pattern of urine from patients with inborn errors of metabolism: 2-hydroxyglutaric aciduria and maple syrup urine disease. *Journal of Pharmaceutical and Biomedical Analysis,* 15**,** 1647-1659.

HOLMES, E., LOO, R. L., CLOAREC, O., COEN, M., TANG, H., MAIBAUM, E., BRUCE, S., CHAN, Q., ELLIOTT, P., STAMLER, J., WILSON, I. D., LINDON, J. C. & NICHOLSON, J. K. (2007) Detection of urinary drug metabolite (xenometabolome) signatures in molecular epidemiology studies via statistical total correlation (NMR) spectroscopy. *Anal Chem,* 79**,** 2629-40.

HTTP://PUBCHEM.NCBI.NLM.NIH.GOV/.

HUANG, X., HUANG, Y., CHINNAPPAN, R., BOCCHINI, C., GUSTIN, M. C. & STERN, M. (2002) The Drosophila inebriated-Encoded Neurotransmitter/Osmolyte Transporter: Dual Roles in the Control of Neuronal Excitability and the Osmotic Stress Response. *Genetics,* 160**,** 561-569.

HUO, T., CAI, S., LU, X., SHA, Y., YU, M. & LI, F. (2009) Metabonomic study of biochemical changes in the serum of type 2 diabetes mellitus patients after the treatment of metformin hydrochloride. *Journal of Pharmaceutical and Biomedical Analysis,* 49**,** 976-982.

IDBORG, H., ZAMANI, L., EDLUND, P.-O., SCHUPPE-KOISTINEN, I. & JACOBSSON, S. P. (2005a) Metabolic fingerprinting of rat urine by LC/MS: Part 1. Analysis by hydrophilic interaction liquid chromatography-electrospray ionization mass spectrometry. *Journal of Chromatography B,* 828**,** 9-13.

IDBORG, H., ZAMANI, L., EDLUND, P.-O., SCHUPPE-KOISTINEN, I. & JACOBSSON, S. P. (2005b) Metabolic fingerprinting of rat urine by LC/MS: Part 2. Data pretreatment methods for handling of complex data. *Journal of Chromatography B,* 828**,** 14-20.

IKEGAMI, T., TOMOMATSU, K., TAKUBO, H., HORIE, K. & TANAKA, N. (2008) Separation efficiencies in hydrophilic interaction chromatography. *Journal of Chromatography A,* 1184**,** 474-503.

IRIBARNE, F., PAULINO, M., AGUILERA, S. & TAPIA, O. (2009) Assaying phenothiazine derivatives as trypanothione reductase and glutathione reductase inhibitors by theoretical docking and Molecular Dynamics studies. *Journal of Molecular Graphics and Modelling,* 28**,** 371-381.

ISMAIEL, O. A., HALQUIST, M. S., ELMAMLY, M. Y., SHALABY, A. & THOMAS KARNES, H. (2008) Monitoring phospholipids for assessment of ion enhancement and ion suppression in ESI and APCI LC/MS/MS for chlorpheniramine in human plasma and the importance of multiple source matrix effect evaluations. *Journal of Chromatography B,* 875**,** 333-343.

IVOSEV, G., BURTON, L. & BONNER, R. (2008) Dimensionality Reduction and Visualization in Principal Component Analysis. *Analytical Chemistry,* 80**,** 4933-4944.

JACKSON, A. U., WERNER, S. R., TALATY, N., SONG, Y., CAMPBELL, K., COOKS, R. G. & MORGAN, J. A. (2008) Targeted metabolomic analysis of Escherichia coli by desorption electrospray ionization and extractive electrospray ionization mass spectrometry. *Analytical Biochemistry,* 375**,** 272-281.

JAMES C. MORRIS, ZEFENG WANG, MARK E. DREW & ENGLUND, A. P. T. (2002) Glycolysis modulates trypanosome glycoprotein expression as revealed by an RNAi library. *European Molecular Biology Organization,* 21**,** 4429-4438.

JANDERA, P. & HÁJEK, T. (2009) Utilization of dual retention mechanism on columns with bonded PEG and diol stationary phases for adjusting the separation selectivity of phenolic and flavone natural antioxidants. *Journal of Separation Science,* 32**,** 3603-3619.

JONES, O. A. H., SPURGEON, D. J., SVENDSEN, C. & GRIFFIN, J. L. (2008) A metabolomics based approach to assessing the toxicity of the polyaromatic hydrocarbon pyrene to the earthworm Lumbricus rubellus. *Chemosphere,* 71**,** 601-609.

JOURDAN, F., COTTRET, L., HUC, L., WILDRIDGE, D., SCHELTEMA, R., HILLENWECK, A., BARRETT, M., ZALKO, D., WATSON, D. & DEBRAUWER, L. (2010) Use of reconstituted metabolic networks to assist in metabolomic data visualization and mining. *Metabolomics,* 6**,** 312-321.

JUMBO-LUCIONI, P., AYROLES, J., CHAMBERS, M., JORDAN, K., LEIPS, J., MACKAY, T. & DE LUCA, M. (2010) Systems genetics analysis of body weight and energy metabolism traits in Drosophila melanogaster. *BMC Genomics,* 11**,** 297.

KADDURAH-DAOUK, R., MCEVOY, J., BAILLIE, R. A., LEE, D., YAO, J. K., DORAISWAMY, P. M. & KRISHNAN, K. R. R. (2007) Metabolomic mapping of atypical antipsychotic effects in schizophrenia. *Mol Psychiatry*.

KAISER, K. (1990) From gene to phenotype in Drosophila and other organisms. *Bioessays,* 12**,** 297-301.

KAMLEH, A., BARRETT, M. P., WILDRIDGE, D., BURCHMORE, R. J. S., SCHELTEMA, R. A. & WATSON, D. G. (2008) Metabolomic profiling using Orbitrap Fourier transform mass spectrometry with hydrophilic interaction chromatography: a method with wide applicability to analysis of biomolecules. *Rapid Communications in Mass Spectrometry,* 22**,** 1912-1918.

KAMLEH, M. A., HOBANI, Y., DOW, J. A. T., ZHENG, L. & WATSON, D. G. (2009) Towards a platform for the metabonomic profiling of different strains of Drosophila melanogaster using liquid chromatography–Fourier transform mass spectrometry. *FEBS Journal,* 276**,** 6798-6809.

KANEHISA, M. & GOTO, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucl. Acids Res.,* 28**,** 27-30.

KANEHISA, M., GOTO, S., KAWASHIMA, S. & NAKAYA, A. (2002) The KEGG databases at GenomeNet. *Nucleic Acids Res,* 30**,** 42-6.

KARATAPANIS, A. E., FIAMEGOS, Y. C. & STALIKAS, C. D. (2009) HILIC separation and quantitation of water-soluble vitamins using diol column. *Journal of Separation Science,* 32**,** 909-917.

KATAJAMAA, M., MIETTINEN, J. & ORESIC, M. (2006) MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics,* 22**,** 634-636.

KATAJAMAA, M. & ORESIC, M. (2005) Processing methods for differential analysis of LC/MS profile data. *BMC Bioinformatics,* 6**,** 179.

KATO, M., KATO, H., EYAMA, S. & TAKATSU, A. (2009) Application of amino acid analysis using hydrophilic interaction liquid chromatography coupled with isotope dilution mass spectrometry for peptide and protein quantification. *Journal of Chromatography B,* 877**,** 3059-3064.

KATO, Y., NAKAMURA, K., KITAMURA, T., TSUDA, T., HASEGAWA, M. & SASAKI, H. (2004) Effect of chromatographic conditions on resolution in high-performance ion-exchange chromatography of proteins on macroporous anion-exchange resin. *Journal of Chromatography,* 1031**,** 101-105.

KEITH, T. P., RILEY, M. A., KREITMAN, M., LEWONTIN, R. C., CURTIS, D. & CHAMBERS, G. (1987) Sequence of the structural gene for xanthine dehydrogenase (rosy locus) in Drosophila melanogaster. *Genetics,* 116**,** 67-73.

KELLER, B. O., SUI, J., YOUNG, A. B. & WHITTAL, R. M. (2008) Interferences and contaminants encountered in modern mass spectrometry. *Analytica Chimica Acta,* 627**,** 71-81.

KENNY, L. C., BROADHURST, D. I., DUNN, W., BROWN, M., NORTH, R. A., MCCOWAN, L., ROBERTS, C., COOPER, G. J. S., KELL, D. B., BAKER, P. N. & ON BEHALF OF THE SCREENING FOR PREGNANCY ENDPOINTS, C. (2010) Robust Early Pregnancy Prediction of Later Preeclampsia Using Metabolomic Biomarkers. *Hypertension,* 56**,** 741-749.

KEUN, H. C., ATHERSUCH, T. J., BECKONERT, O., WANG, Y., SARIC, J., SHOCKCOR, J. P., LINDON, J. C., WILSON, I. D., HOLMES, E. & NICHOLSON, J. K. (2008) Heteronuclear ^{19}F-^{1}H Statistical Total Correlation Spectroscopy as a Tool in Drug Metabolism: Study of Flucloxacillin Biotransformation. *Anal Chem,* 80**,** 1073-1079.

KEUN, H. C., BECKONERT, O., GRIFFIN, J. L., RICHTER, C., MOSKAU, D., LINDON, J. C. & NICHOLSON, J. K. (2002a) Cryogenic Probe 13C NMR Spectroscopy of Urine for Metabonomic Studies. *Analytical Chemistry,* 74**,** 4588-4593.

KEUN, H. C., EBBELS, T. M. D., ANTTI, H., BOLLARD, M. E., BECKONERT, O., SCHLOTTERBECK, G. T., SENN, H., NIEDERHAUSER, U., HOLMES, E., LINDON, J. C. & NICHOLSON, J. K. (2002b) Analytical Reproducibility in ^{1}H NMR-Based Metabonomic Urinalysis. *Chemical Research in Toxicology,* 15**,** 1380-1386.

KIM, J. Y., PARK, J. Y., KIM, O. Y., HAM, B. M., KIM, H.-J., KWON, D. Y., JANG, Y. & LEE, J. H. (2010) Metabolic Profiling of Plasma in Overweight/Obese and Lean Men using Ultra Performance Liquid Chromatography and Q-TOF Mass Spectrometry (UPLC-Q-TOF MS). *Journal of Proteome Research,* 9**,** 4368-4375.

KIND, T. & FIEHN, O. (2006) Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics,* 7**,** 234.

KIND, T., TOLSTIKOV, V., FIEHN, O. & WEISS, R. H. (2007) A comprehensive urinary metabolomic approach for identifying kidney cancerr. *Analytical Biochemistry,* 363**,** 185-195.

KOCHHAR, S., JACOBS, D. M., RAMADAN, Z., BERRUEX, F., FUERHOLZ, A. & FAY, L. B. (2006) Probing gender-specific metabolism differences in humans by nuclear magnetic resonance-based metabonomics. *Analytical Biochemistry,* 352**,** 274-281.

KONING, W. D. & DAM, K. V. (1992) A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. *Analytical Biochemistry,* 204**,** 118-123.

KOPP, E. K., SIEBER, M., KELLERT, M. & DEKANT, W. (2008) Rapid and Sensitive HILIC-ESI-MS/MS Quantitation of Polar Metabolites of Acrylamide in Human Urine Using Column Switching with an Online Trap Column. *Journal of Agricultural and Food Chemistry,* 56**,** 9828-9834.

KULAS, J., SCHMIDT, C., ROTHE, M., SCHUNCK, W.-H. & MENZEL, R. (2008) Cytochrome P450-dependent metabolism of eicosapentaenoic acid in the nematode Caenorhabditis elegans. *Archives of Biochemistry and Biophysics,* 472**,** 65-75.

LAKSO, H.-A., APPELBLAD, P. & SCHNEEDE, J. (2008) Quantification of Methylmalonic Acid in Human Plasma with Hydrophilic Interaction Liquid Chromatography Separation and Mass Spectrometric Detection. *Clin Chem,* 54**,** 2028-2035.

LÄMMERHOFER, M., RICHTER, M., WU, J., NOGUEIRA, R., BICKER, W. & LINDNER, W. (2008) Mixed-mode ion-exchangers and their comparative chromatographic characterization in reversed-phase and hydrophilic interaction chromatography elution modes. *Journal of Separation Science,* 31**,** 2572-2588.

LAMOUR, N., RIVIRE, L., COUSTOU, V., COOMBS, G. H., BARRETT, M. P. & BRINGAUD, F. (2005) Proline Metabolism in Procyclic Trypanosoma brucei Is Down-regulated in the Presence of Glucose. *Journal of Biological Chemistry,* 280**,** 11902-11910.

LANDEL, C. P., CHEN, S. Z. & EVANS, G. A. (1990) Reverse genetics using transgenic mice. *Annu Rev Physiol,* 52**,** 841-51.

LATOCHA, M., CHODUREK, E., KURKIEWICZ, S., SWIATKOWSKA, L. & WILCZOK, T. (2000) Pyrolytic GC-MS analysis of melanin from black, gray and yellow strains of *Drosophila* melanogaster. *Journal of Analytical and Applied Pyrolysis,* 56**,** 89-98.

LAY, J. J. O., LIYANAGE, R., BORGMANN, S. & WILKINS, C. L. (2006) Problems with the "omics". *TrAC Trends in Analytical Chemistry,* 25**,** 1046-1056.

LAYNE, J. (2002) Characterization and comparison of the chromatographic performance of conventional, polar-embedded, and polar-endcapped reversed-phase liquid chromatography stationary phases. *Journal of Chromatography A,* 957**,** 149-164.

LEGIDO-QUIGLEY, C., STELLA, C., PEREZ-JIMENEZ, F., LOPEZ-MIRANDA, J., ORDOVAS, J., POWELL, J., VAN-DER-OUDERAA, F., WARE, L., LINDON, J., C., NICHOLSON, J., K. & HOLMES, E. (2010) Liquid chromatography-mass spectrometry methods for urinary biomarker detection in metabonomic studies with application to nutritional studies. *Biomedical Chromatography,* 24**,** 737-743.

LENZ, E. M., BRIGHT, J., WILSON, I. D., MORGAN, S. R. & NASH, A. F. P. (2003) A 1H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects. *Journal of Pharmaceutical and Biomedical Analysis,* 33**,** 1103-1115.

Houkai Li, Yan Ni, Mingming Su, Yunping Qiu, Mingmei Zhou, Mingfeng Qiu, Aihua Zhao, Liping Zhao, & Wei Jia.(2007) Pharmacometabonomic Phenotyping Reveals Different Responses to Xenobiotic Intervention in Rats. *Journal of Proteome Research,* 6**,** 1364-1370.

LICHTENSTEIN, C. & BARRENA, E. (1993) Prospects for reverse genetics in plants using recombination. *Plant Mol Biol,* 21**,** v-xii.

LIN, H.-M., EDMUNDS, S. J., HELSBY, N. A., FERGUSON, L. R. & ROWAN, D. D. (2009) Nontargeted Urinary Metabolite Profiling of a Mouse Model of Crohnâ€™s Disease. *Journal of Proteome Research,* 8**,** 2045-2057.

LINDEGARDH, N., HANPITHAKPONG, W., PHAKDEERAJ, A., SINGHASIVANON, P., FARRAR, J., HIEN, T. T., WHITE, N. J. & DAY, N. P. J. (2008) Development and validation of a high-throughput zwitterionic hydrophilic interaction liquid chromatography solid-phase extraction-liquid chromatography-tandem mass spectrometry method for determination of the anti-influenza drug peramivir in plasma. *Journal of Chromatography A,* 1215**,** 145-151.

LINDEGÅRDH, N., HANPITHAKPONG, W., WATTANAGOON, Y., SINGHASIVANON, P., WHITE, N. J. & DAY, N. P. J. (2007) Development and validation of a liquid chromatographic-tandem mass spectrometric method for determination of oseltamivir and its

metabolite oseltamivir carboxylate in plasma, saliva and urine. *Journal of Chromatography B,* 859**,** 74-83.
LINDEGÅRDH, N., SINGTOROJ, T., ANNERBERG, A., WHITE, N. J. & DAY, N. P. J. (2005) Development and Validation of a Solid-Phase Extraction-Liquid Chromatographic Method for Determination of Amoxicillin in Plasma. *Therapeutic Drug Monitoring,* 27**,** 503-508.
LINDON, J., HOLMES, E. & NICHOLSON, J. (2006) Metabonomics Techniques and Applications to Pharmaceutical Research & Development. *Pharmaceutical Research,* 23**,** 1075-1088.
LINDON, J. C., HOLMES, E. & NICHOLSON, J. K. (2007a) Metabonomics in pharmaceutical R & D. *FEBS Journal,* 274**,** 1140-1151.
LINDON, J. C., HOLMES, E. & NICHOLSON, J. K. (2007b) Metabonomics in pharmaceutical R& D. *FEBS Journal,* 274**,** 1140-1151.
LINDON, J. C., NICHOLSON, J. K., EVERETT, J. R. (1999) "NMR spectroscopy of biofluids" G. A. eds. *Annual Reports on NMR Spectroscopy,* 38.
LV, Y., LIU, X., YAN, S., LIANG, X., YANG, Y., DAI, W. & ZHANG, W. (2010) Metabolomic study of myocardial ischemia and intervention effects of Compound Danshen Tablets in rats using ultra-performance liquid chromatography/quadrupole time-of-flight mass spectrometry. *Journal of Pharmaceutical and Biomedical Analysis,* 52**,** 129-135.
MAJORS, R. E. & PRZYBYCIEL, M. (2002) Columns for Reversed-Phase LC Separations in Highly Aqueous Mobile Phases. *LC-GC Europe,* 15**,** 780-786.
MAKAROV, A. (1999) Mass Spectrometer. IN U.S.PATENT (Ed., 5.
MAKAROV, A. (2000) Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.,* 72**,** 1156-1162.
MAKAROV, A., DENISOV, E., KHOLOMEEV, A., BALSCHUN, W., LANGE, O., STRUPAT, K. & HORNING, S. (2006) Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. *Anal. Chem.,* 78**,** 2113-2120.
MAKAROV, A. & SCIGELOVA, M. (2010) Coupling liquid chromatography to Orbitrap mass spectrometry. *Journal of Chromatography A,* 1217**,** 3938-3945.
MAMOLEN, M., SMITH, A. & ANDRULIS, E. D. (2010) *Drosophila* melanogaster Dis3 N-terminal domains are required for ribonuclease activities, nuclear localization and exosome interactions. *Nucl. Acids Res.***,** gkq295.
MARTIJN VAN, D., JACK, V., ALBERT, T., EWOUD JAN VAN, H., JACOBUS, B., ADAM, C. & JAN VAN DER, G. (2007) Evaluation of metabolite profiles as biomarkers for the pharmacological effects of thiazolidinediones in Type 2 diabetes mellitus patients and healthy volunteers. *British Journal of Clinical Pharmacology,* 63**,** 562-574.
MARTOS, P. A., JAYASUNDARA, F., DOLBEER, J., JIN, W., SPILSBURY, L., MITCHELL, M., VARILLA, C. & SHURMER, B. (2010) Multiclass, Multiresidue Drug Analysis, Including Aminoglycosides, in Animal Tissue Using Liquid Chromatography Coupled to Tandem Mass Spectrometry *Journal of Agricultural and Food Chemistry,* 58**,** 5932-5944.
MATYSOVÁ, L., HAVLÍKOVÁ, L., HÁJKOVÁ, R., KRIVDA, A. & SOLICH, P. (2009) Application of HILIC stationary phase to determination of dimethindene maleate in topical gel. *Journal of Pharmaceutical and Biomedical Analysis,* 50**,** 23-26.
MCALISTER, G. C., PHANSTIEL, D., GOOD, D. M., BERGGREN, W. T. & COON, J. J. (2007) Implementation of Electron-Transfer Dissociation on a Hybrid Linear Ion Trap-Orbitrap Mass Spectrometer. *Anal. Chem.,* 79**,** 3525-3534.

MCDOUGALL, G., MARTINUSSEN, I. & STEWART, D. (2008) Towards fruitful metabolomics: High throughput analyses of polyphenol composition in berries using direct infusion mass spectrometry. *Journal of Chromatography B,* 871**,** 362-369.

MCLUCKEY, S. A. & WELLS, J. M. (2001) Mass Analysis at the Advent of the 21st Century. *Chem. Rev.,* 101**,** 571-606.

MERCK (2007) *Merck Index*.

METZ, T. O., PAGE, J. S., BAKER, E. S., TANG, K., DING, J., SHEN, Y. & SMITH, R. D. (2008) High-resolution separations and improved ion production and transmission in metabolomics. *TrAC Trends in Analytical Chemistry,* 27**,** 205-214.

MICHAUD, M. & DENLINGER, D. (2007) Shifts in the carbohydrate, polyol, and amino acid pools during rapid cold-hardening and diapause-associated cold-hardening in flesh flies (Sarcophaga crassipalpis): a metabolomic comparison. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology,* 177**,** 753-763.

MICHAUD, R. M., BENOIT, J. B., LOPEZ-MARTINEZ, G., ELNITSKY, M. A., LEE JR, R. E. & DENLINGER, D. L. (2008) Metabolomics reveals unique and shared metabolic changes in response to heat shock, freezing and desiccation in the Antarctic midge, Belgica antarctica. *Journal of Insect Physiology,* 54**,** 645-655.

MICHEL, M., BACZEK, T., STUDZINSKA, S., BODZIOCH, K., JONSSON, T., KALISZAN, R. & BUSZEWSKI, B. (2007) Comparative evaluation of high-performance liquid chromatography stationary phases used for the separation of peptides in terms of quantitative structure-retention relationships. *Journal of Chromatography A,* 1175**,** 49-54.

MICHOPOULOS, F., EDGE, A. M., THEODORIDIS, G. & WILSON, I. D. (2010) Application of turbulent flow chromatography to the metabonomic analysis of human plasma: Comparison with protein precipitation. *Journal of Separation Science,* 33**,** 1472-1479.

MICHOPOULOS, F., LAI, L., GIKA, H., THEODORIDIS, G. & WILSON, I. (2009) UPLC-MS-Based Analysis of Human Plasma for Metabonomics Using Solvent Precipitation or Solid Phase Extraction. *Journal of Proteome Research,* 8**,** 2114-2121.

MISHRA, J., SAXENA, A. & SINGH, S. (2007) Chemotherapy of Leishmaniasis: Past, Present and Future. *Current Medicinal Chemistry,* 14**,** 1153-1169.

MITCHELL, H. K. & GLASSMAN, E. (1959) Hypoxanthine in rosy and maroon-like mutants of Drosophila melanogaster. *Science,* 129**,** 268.

MOHAMED, R., VARESIO, E., IVOSEV, G., BURTON, L., BONNER, R. & HOPFGARTNER, G. R. (2009) Comprehensive Analytical Strategy for Biomarker Identification based on Liquid Chromatography Coupled to Mass Spectrometry and New Candidate Confirmation Tools. *Analytical Chemistry,* 81**,** 7677-7694.

MOHLER, R. E., TU, B. P., DOMBEK, K. M., HOGGARD, J. C., YOUNG, E. T. & SYNOVEC, R. E. (2008) Identification and evaluation of cycling yeast metabolites in two-dimensional comprehensive gas chromatography-time-of-flight-mass spectrometry data. *Journal of Chromatography A,* 1186**,** 401-411.

MORGAN, T. H. (1910) Sex limited inheritance in *Drosophila*. *Science,* 32**,** 120-122.

MÜLLER, C., SCHÄFER, P., STÖRTZEL, M., VOGT, S. & WEINMANN, W. (2002) Ion suppression effects in liquid chromatography-electrospray-ionisation transport-region collision induced dissociation mass spectrometry with different serum extraction methods for systematic toxicological analysis with mass spectra libraries. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences,* 773**,** 47-52.

NDAW, S., DENIS, F., MARSAN, P., D'ALMEIDA, A. & ROBERT, A. (2010) Biological monitoring of occupational exposure to 5-fluorouracil: Urinary [alpha]-fluoro-[beta]-alanine

assay by high performance liquid chromatography tandem mass spectrometry in health care personnel. *Journal of Chromatography B,* In Press, Corrected Proof.
NESTERENKO, E. P., NESTERENKO, P. N. & PAULL, B. (2009) Zwitterionic ion-exchangers in ion chromatography: A review of recent developments. *Analytica Chimica Acta,* 652**,** 3-21.
NEZIREVIC, D., ÅRSTRAND, K. & KÅGEDAL, B. (2007) Hydrophilic interaction liquid chromatographic analysis of aminohydroxyphenylalanines from melanin pigments. *Journal of Chromatography A,* 1163**,** 70-79.
NICHOLSON, J., SADLER, P., BALES, J., JUUL, S., MACLEOD, A. & SONKSEN, P. (1984) MONITORING METABOLIC DISEASE BY PROTON NMR OF URINE. *The Lancet,* 324**,** 751-752.
NICHOLSON, J. K., CONNELLY, J., LINDON, J. C. & HOLMES, E. (2002) METABONOMICS: A PLATFORM FOR STUDYING DRUG TOXICITY AND GENE FUNCTION. *Nature Reviews Drug Discovery,* 1**,** 153-161.
NICHOLSON, J. K., TIMBRELL, J. A. & SADLER, P. J. (1985) Proton NMR spectra of urine as indicators of renal damage. Mercury-induced nephrotoxicity in rats. *Molecular Pharmacology,* 27**,** 644-651.
NIELSEN, N.-P. V., CARSTENSEN, J. M. & SMEDSGAARD, J. (1998) Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *Journal of Chromatography A,* 805**,** 17-35.
NORDSTRAM, A., O'MAILLE, G., QIN, C. & SIUZDAK, G. (2006) Nonlinear Data Alignment for UPLC-MS and HPLC-MS Based Metabolomics: Quantitative Analysis of Endogenous and Exogenous Metabolites in Human Serum. *Analytical Chemistry,* 78**,** 3289-3295.
NORDSTROM, A., WANT, E., NORTHEN, T., LEHTIO, J. & SIUZDAK, G. (2007) Multiple Ionization Mass Spectrometry Strategy Used To Reveal the Complexity of Metabolomics. *Analytical Chemistry,* 80**,** 421-429.
NOVÁKOVÁ, L., SOLICHOVÁ, D., PAVLOVIČOVÁ, S. & SOLICH, P. (2008) Hydrophilic interaction liquid chromatography method for the determination of ascorbic acid. *Journal of Separation Science,* 31**,** 1634-1644.
NOVÁKOVÁ , L., SOLICHOVÁ, D. & SOLICH, P. (2009) Hydrophilic interaction liquid chromatography - charged aerosol detection as a straightforward solution for simultaneous analysis of ascorbic acid and dehydroascorbic acid. *Journal of Chromatography A,* 1216**,** 4574-4581.
NULL, B., LIU, C. W., HEDEHUS, M., CONOLLY, S. & DAVIS, R. W. (2008) High-Resolution, *In Vivo* Magnetic Resonance Imaging of *Drosophila* at 18.8 Tesla. *PLoS ONE,* 3**,** e2817.
ODUNSI, K., WOLLMAN, R. M., AMBROSONE, C. B., HUTSON, A., MCCANN, S. E., TAMMELA, J., GEISLER, J. P., MILLER, G., SELLERS, T., CLIBY, W., QIAN, F., KEITZ, B., INTENGAN, M., LELE, S. & ALDERFER, J. L. (2005) Detection of epithelial ovarian cancer using 1H-NMR-based metabonomics. *International Journal of Cancer,* 113**,** 782-788.
OLIVER, D. J., NIKOLAU, B. & WURTELE, E. S. (2002) Functional genomics: high-throughput mRNA, protein, and metabolite analyses. *Metab Eng,* 4**,** 98-106.
OLSEN, B. A. (2001) Hydrophilic interaction chromatography using amino and silica columns for the determination of polar pharmaceuticals and impurities. *Journal of Chromatography A,* 913**,** 113-122.
OLSEN, J. V., DE GODOY, L. M. F., LI, G., MACEK, B., MORTENSEN, P., PESCH, R., MAKAROV, A., LANGE, O., HORNING, S. & MANN, M. (2005) Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap. *Mol Cell Proteomics,* 4**,** 2010-2021.

OLSEN, J. V., MACEK, B., LANGE, O., MAKAROV, A., HORNING, S. & MANN, M. (2007) Higher-energy C-trap dissociation for peptide modification analysis. *Nat Meth,* 4**,** 709-712.

ONDARZA, R., HURTADO, G., ITURBE, A., HERNÁNDEZ, E., TAMAYO, E. & WOOLERY, M. (2005) Identification of trypanothione from the human pathogen Entamoeba histolytica by mass spectrometry and chemical analysis. *Biotechnol Appl Biochem,* 42**,** 175-181.

ONDARZA, R. N., HURTADO, G., TAMAYO, E., ITURBE, A. & HERNANDEZ, E. (2006a) Naegleria fowleri: A free-living highly pathogenic amoeba contains trypanothione/trypanothione reductase and glutathione/glutathione reductase systems. *Experimental Parasitology,* 114**,** 141-146.

ONDARZA, R. N., HURTADO, G., TAMAYO, E., ITURBE, A. & HERNÁNDEZ, E. (2006b) Naegleria fowleri: A free-living highly pathogenic amoeba contains trypanothione/trypanothione reductase and glutathione/glutathione reductase systems. *Experimental Parasitology,* 114**,** 141-146.

ORKIN, S. H. (1986) Reverse genetics and human disease. *Cell,* 47**,** 845-50.

PASIKANTI, K. K., HO, P. C. & CHAN, E. C. Y. (2008) Gas chromatography/mass spectrometry in metabolic profiling of biological fluids. *Journal of Chromatography B,* 871**,** 202-211.

PATTERSON, S., JONES, D., SHANKS, E., FREARSON, J., GILBERT, I., WYATT, P. & FAIRLAMB, A. (2009) Synthesis and Evaluation of 1-(1-(Benzo[b]thiophen-2-yl)cyclohexyl)piperidine (BTCP) Analogues as Inhibitors of Trypanothione Reductase. *ChemMedChem,* 4**,** 1341-1353.

PAVEL, J. (2008) Stationary phases for hydrophilic interaction chromatography, their characterization and implementation into multidimensional chromatography concepts. *Journal of Separation Science,* 31**,** 1421-1437.

PEDERSEN, K. S., KRISTENSEN, T. N., LOESCHCKE, V., PETERSEN, B. O., DUUS, J. O., NIELSEN, N. C. & MALMENDAL, A. (2008) Metabolomic Signatures of Inbreeding at Benign and Stressful Temperatures in Drosophila melanogaster. *Genetics,* 180**,** 1233-1243.

PHAPALE, P. B., KIM, S. D., LEE, H. W., LIM, M., KALE, D. D., KIM, Y. L., CHO, J. H., HWANG, D. & YOON, Y. R. (2010) An Integrative Approach for Identifying a Metabolic Phenotype Predictive of Individualized Pharmacokinetics of Tacrolimus. *Clin Pharmacol Ther,* 87**,** 426-436.

PLUMB, R., GRANGER, J., STUMPF, C., WILSON, I. D., EVANS, J. A. & LENZ, E. M. (2003) Metabonomic analysis of mouse urine by liquid-chromatography-time of flight mass spectrometry (LC-TOFMS): detection of strain, diurnal and gender differences. *Analyst,* 128**,** 819-823.

PLUMB, R. S., GRANGER, J. H., STUMPF, C. L., JOHNSON, K. A., SMITH, B. W., GAULITZ, S., WILSON, I. D. & CASTRO-PEREZ, J. (2005) 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,* 130**,** 844-849.

PLUMB, R. S., JOHNSON, K. A., RAINVILLE, P., SHOCKCOR, J. P., WILLIAMS, R., GRANGER, J. H. & WILSON, I. D. (2006) The detection of phenotypic differences in the metabolic plasma profile of three strains of Zucker rats at 20 weeks of age using ultra-performance liquid chromatography/orthogonal acceleration time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry,* 20**,** 2800-2806.

PLUMB, R. S., RAINVILLE, P. D., POTTS, W. B., JOHNSON, K. A., GIKA, E. & WILSON, I. D. (2009) Application of Ultra Performance Liquid Chromatography-Mass

Spectrometry to Profiling Rat and Dog Bile. *Journal of Proteome Research,* 8**,** 2495-2500.
PLUSKAL, T., NAKAMURA, T., VILLAR-BRIONES, A. & YANAGIDA, M. (2009) Metabolic profiling of the fission yeast S. pombe: quantification of compounds under different temperatures and genetic perturbation. *Molecular BioSystems,* 6**,** 182-198.
PREINERSTORFER, B., SCHIESEL, S., LÄMMERHOFER, M. & LINDNER, W. (2009) Metabolic profiling of intracellular metabolites in fermentation broths from [beta]-lactam antibiotics production by liquid chromatography-tandem mass spectrometry methods. *Journal of Chromatography A,* In Press, Corrected Proof.
PUCCI, V., GIULIANO, C., ZHANG, R., KOEPLINGER, K. A., LEONE, J. F., MONTEAGUDO, E. & BONELLI, F. (2009) HILIC LC-MS for the determination of 2′-C-methyl-cytidine-triphosphate in rat liver. *Journal of Separation Science,* 32**,** 1275-1283.
QIAN, W. L., KHAN, Z., WATSON, D. G. & FEARNLEY, J. (2008) Analysis of sugars in bee pollen and propolis by ligand exchange chromatography in combination with pulsed amperometric detection and mass spectrometry. *Journal of Food Composition and Analysis,* 21**,** 78-83.
QIZHI, H., ROBERT, J. N., HONGYAN, L., ALEXANDER, M., MARK, H. & COOKS, R. G. (2005) The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry,* 40**,** 430-443.
QIZHI HU, ROBERT J. NOLL, HONGYAN LI , ALEXANDER MAKAROV , MARK HARDMAN & COOKS, R. G. (2005) The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry,* 40**,** 430-443.
RENN, S., MACHADO, H., JONES, A., SONEJI, K., KULATHINAL, R. & HOFMANN, H. (2010) Using comparative genomic hybridization to survey genomic sequence divergence across species: a proof-of-concept from *Drosophila*. *BMC Genomics,* 11**,** 271.
REZZI, S., RAMADAN, Z., MARTIN, F. P. J., FAY, L. B., VANBLADEREN, P., LINDON, J. C., NICHOLSON, J. K. & KOCHHAR, S. (2007) Human Metabolic Phenotypes Link Directly to Specific Dietary Preferences in Healthy Individuals. *J. Proteome Res.,* 6**,** 4469-4477.
RICHARDSON, J., NETT, I., JONES, D., ABDILLE, M., GILBERT, I. & FAIRLAMB, A. (2009) Improved Tricyclic Inhibitors of Trypanothione Reductase by Screening and Chemical Synthesis. *ChemMedChem,* 4**,** 1333-1340.
RIMMER, C. A., SIMMONS, C. R. & DORSEY, J. G. (2002) The measurement and meaning of void volumes in reversed-phase liquid chromatography. *Journal of Chromatography A,* 965**,** 219-232.
RISLEY, D. S. & PACK, B. W. (2006) Simultaneous Determination of Positive and Negative Counterions Using a Hydrophilic Interaction Chromatography Method. *LCGC NORTH AMERICA,* 24**,** 777-785.
RISTOW, M. & ZARSE, K. (2010) How increased oxidative stress promotes longevity and metabolic health: The concept of mitochondrial hormesis (mitohormesis). *Experimental Gerontology,* 45**,** 410-418.
ROBERTSON, D. G. (2005) *Metabonomics in Toxicity Assessment*, Taylor & Francis.
ROBINSON, J. W. & FRAME, E. M. S. F. G. M. (2005) *Undergraduate instrumental analysis / James W. Robinson, Eileen M. Skelly Frame, George M. Frame II,* New York :, M. Dekker.
ROBOSKY, L. C., WELLS, D. F., EGNASH, L. A., MANNING, M. L., REILY, M. D. & ROBERTSON, D. G. (2005) Metabonomic Identification of Two Distinct Phenotypes in Sprague-Dawley (Crl:CD(SD)) Rats.
RUBIN, G. M. & SPRADLING, A. C. (1983) Vectors for P element-mediated gene transfer in Drosophila. *Nucleic Acids Res,* 11**,** 6341-51.

RYAN, D. & ROBARDS, K. (2006) Analytical Chemistry Considerations in Plant Metabolomics. *Separation & Purification Reviews,* 35**,** 319 - 356.

SCARAFFIA, P. Y., TAN, G., ISOE, J., WYSOCKI, V. H., WELLS, M. A. & MIESFELD, R. L. (2008) Discovery of an alternate metabolic pathway for urea synthesis in adult Aedes aegypti mosquitoes. *Proceedings of the National Academy of Sciences,* 105**,** 518-523.

SCARAFFIA, P. Y., ZHANG, Q., WYSOCKI, V. H., ISOE, J. & WELLS, M. A. (2006) Analysis of whole body ammonia metabolism in Aedes aegypti using [15N]-labeled compounds and mass spectrometry. *Insect Biochemistry and Molecular Biology,* 36**,** 614-622.

SCHAEFER, U., BOOS, W., TAKORS, R. & WEUSTER-BOTZ, D. (1999) Automated Sampling Device for Monitoring Intracellular Metabolite Dynamics. *Analytical Biochemistry,* 270**,** 88-96.

SCHIESEL, S., LÄMMERHOFER, M. & LINDNER, W. (2010) Multitarget quantitative metabolic profiling of hydrophilic metabolites in fermentation broths of β-lactam antibiotics production by HILIC–ESI–MS/MS. *Analytical and Bioanalytical Chemistry,* 396**,** 1655-1679.

SEGER, S. & STURM, S. (2007) Analytical Aspects of Plant Metabolite Profiling Platforms: Current Standings and Future Aims. *J. Proteome Res.,* 6**,** 480-497.

SELLAMI, A., ISABEL, G. & VEENSTRA, J. A. (2010) Expression of the mu opioid receptor in *Drosophila* and its effects on trehalose and glycogen when expressed by the AKH neuroendocrine cells. *Peptides,* 31**,** 1383-1389.

SEQUANT (2010) A Practical Guide to HILIC
A TUTORIAL AND APPLICATION BOOK.

SERVICES, S. I. (2010). Scientific Instrument Services

SHEMESH, O. & SPIRA, M. (2010) Hallmark cellular pathology of Alzheimer's disease induced by mutant human tau expression in cultured Aplysia neurons. *Acta Neuropathologica*.

SHIM, H. & FAIRLAMB, A. H. (1988) Levels of Polyamines, Glutathione and Glutathione-Spermidine Conjugates during Growth of the Insect Trypanosomatid Crithidia fasciculata. *J Gen Microbiol,* 134**,** 807-817.

SHINOZAKI, K., KASHIWAGI, A., NISHIO, Y., OKAMURA, T., YOSHIDA, Y., MASADA, M., TODA, N. & KIKKAWA, R. (1999) Abnormal biopterin metabolism is a major cause of impaired endothelium-dependent relaxation through nitric oxide/O2- imbalance in insulin-resistant rat aorta. *Diabetes,* 48**,** 2437-2445.

SHIRO, U., WEI, Z., KINDRA, B. & VLADIMIR, T. Comprehensive mass spectrometry based metabolic profiling of blood plasma reveals potent discriminatory classifiers of pancreatic cancer. *Rapid Communications in Mass Spectrometry,* 24**,** 613-620.

SIELC (2010a) SIELC-Obelisc columns. Sielc.

SIELC (2010b) Sielc-Obelisc columns.

SKERÍKOVÁ, V. & JANDERA, P. (2010) Effects of the operation parameters on Hydrophilic Interaction Liquid Chromatography separation of phenolic acids on zwitterionic monolithic capillary columns. *Journal of Chromatography A,* In Press, Corrected Proof.

SKORDI, E., YAP, I. K. S., CLAUS, S. P., MARTIN, F., XE, OIS-PIERRE, J., CLOAREC, O., LINDBERG, J., SCHUPPE-KOISTINEN, I., HOLMES, E. & NICHOLSON, J. K. (2007) Analysis of Time-Related Metabolic Fluctuations Induced by Ethionine in the Rat. *J. Proteome Res.,* 6**,** 4572-4581.

SMITH, C. A., O'MAILLE, G., WANT, E. J., QIN, C., TRAUGER, S. A., BRANDON, T. R., CUSTODIO, D. E., ABAGYAN, R. & SIUZDAK, G. (2005) METLIN: a metabolite mass spectral database. *Ther Drug Monit,* 27**,** 747-51.

SMITH, C. A., WANT, E. J., O'MAILLE, G., ABAGYAN, R. & SIUZDAK, G. (2006) XCMS:a Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Analytical Chemistry,* 78**,** 779-787.

SMITH, L. (2002) A tutorial on Principal Component Analysis,: .

SMITH, L. M., MAHER, A. D., CLOAREC, O., RANTALAINEN, M., TANG, H., ELLIOTT, P., STAMLER, J., LINDON, J. C., HOLMES, E. & NICHOLSON, J. K. (2007) Statistical correlation and projection methods for improved information recovery from diffusion-edited NMR spectra of biological samples. *Anal Chem,* 79**,** 5682-9.

SOUTHAM, A. D., PAYNE, T. G., COOPER, H. J., ARVANITIS, T. N. & VIANT, M. R. (2007) Dynamic Range and Mass Accuracy of Wide-Scan Direct Infusion Nanoelectrospray Fourier Transform Ion Cyclotron Resonance Mass Spectrometry-Based Metabolomics Increased by the Spectral Stitching Method. *Anal. Chem.,* 79**,** 4595-4602.

SPARAGANO, O. (1998) Trypanosomiasis and Leishmaniasis – Biology and Control; Edited by G. Hide, J. C. Mottram, G. H. Coombs and P. H. Holmes Sparagano. *Tropical Animal Health and Production,* 30**,** 190-190.

SPINKS, D., SHANKS, E., CLEGHORN, L. A., MCELROY, S., JONES, D., JAMES, D., FAIRLAMB, A., FREARSON, J., WYATT, P. & GILBERT, I. (2009) Investigation of Trypanothione Reductase as a Drug Target in Trypanosoma brucei. *ChemMedChem,* 4**,** 2060-2069.

SREEKUMAR, A., POISSON, L. M., RAJENDIRAN, T. M., KHAN, A. P., CAO, Q., YU, J., LAXMAN, B., MEHRA, R., LONIGRO, R. J., LI, Y., NYATI, M. K., AHSAN, A., KALYANA-SUNDARAM, S., HAN, B., CAO, X., BYUN, J., OMENN, G. S., GHOSH, D., PENNATHUR, S., ALEXANDER, D. C., BERGER, A., SHUSTER, J. R., WEI, J. T., VARAMBALLY, S., BEECHER, C. & CHINNAIYAN, A. M. (2009) Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature,* 457**,** 910-914.

STEFFENS, D. C., WEI, J., KRISHNAN, K. R. R., KAROLY, E. D., MITCHELL, M. W., O'CONNOR, C. M. & KADDURAH-DAOUK, R. (2010) Metabolomic Differences in Heart Failure Patients With and Without Major Depression. *Journal of Geriatric Psychiatry and Neurology,* 23**,** 138-146.

STERN, M., THOMPSON, K. S. J., ZHOU, P., WATSON, D. G., MIDGLEY, J. M., GEWECKE, M. & BACON, J. P. (1995) Octopaminergic neurons in the locust brain: morphological, biochemical and electrophysiological characterisation of potential modulators of the visual system. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology,* 177**,** 611-625.

STREGE, M. A. (1998) Hydrophilic Interaction Chromatography-Electrospray Mass Spectrometry Analysis of Polar Compounds for Natural Product Drug Discovery. *Analytical Chemistry,* 70**,** 2439-2445.

SUHRE, K. & SCHMITT-KOPPLIN, P. (2008) MassTRIX: mass translator into pathways. *Nucl. Acids Res.,* 36**,** W481-484.

SUNSERI, J. D., COOPER, W. T. & DORSEY, J. G. (2003) Reducing residual silanol interactions in reversed-phase liquid chromatography: Thermal treatment of silica before derivatization. *Journal of Chromatography A,* 1011**,** 23-29.

T'KINDT, R., ALAERTS, G., VANDER HEYDEN, Y., DEFORCE, D. & VAN BOCXLAER, J. (2007) Broad-spectrum separations in metabolomics using enhanced polar LC stationary phases: A dedicated evaluation using plant metabolites. *Journal of Separation Science,* 30**,** 2002-2011.

T'KINDT, R., DE VEYLDER, L., STORME, M., DEFORCE, D. & VAN BOCXLAER, J. (2008a) LC-MS metabolic profiling of Arabidopsis thaliana plant leaves and cell cultures: Optimization of pre-LC-MS procedure parameters. *Journal of Chromatography B,* 871**,** 37-43.

T'KINDT, R., MORREEL, K., DEFORCE, D., BOERJAN, W. & VAN BOCXLAER, J. (2009) Joint GC-MS and LC-MS platforms for comprehensive plant metabolomics: Repeatability and sample pre-treatment. *Journal of Chromatography B,* 877**,** 3572-3580.

T'KINDT, R., STORME, M., DEFORCE, D. & VAN BOCXLAER, J. (2008b) Evaluation of hydrophilic interaction chromatography versus reversed-phase chromatography in a plant metabolomics perspective. *Journal of Separation Science,* 31**,** 1609-1614.

TAKEGAWA, Y., DEGUCHI, K., ITO, H., KEIRA, T., NAKAGAWA, H. & NISHIMURA, S. I. (2006a) Simple separation of isomeric sialylated N-glycopeptides by a zwitterionic type of hydrophilic interaction chromatography. *Journal of Separation Science,* 29**,** 2533-2540.

TAKEGAWA, Y., DEGUCHI, K., KEIRA, T., ITO, H., NAKAGAWA, H. & NISHIMURA, S.-I. (2006b) Separation of isomeric 2-aminopyridine derivatized N-glycans and N-glycopeptides of human serum immunoglobulin G by using a zwitterionic type of hydrophilic-interaction chromatography. *Journal of Chromatography A,* 1113**,** 177-181.

TAUTENHAHN, R., BOTTCHER, C. & NEUMANN, S. (2008) Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics,* 9**,** 504.

THERMOFISHER, S. (2010a) Product display SIEVE software for Differential Expression.

THERMOFISHER, S. (2010b) Sieve Brochure.

THONY, B., AUERBACH, G. & BLAU, N. (2000) Tetrahydrobiopterin biosynthesis, regeneration and functions. *Biochem. J.,* 347**,** 1-16.

TOLSTIKOV, V. V. & FIEHN, O. (2002) Analysis of Highly Polar Compounds of Plant Origin: Combination of Hydrophilic Interaction Chromatography and Electrospray Ion Trap Mass Spectrometry. *Analytical Biochemistry,* 301**,** 298-307.

TRYGG, J., HOLMES, E. & LUNDSTEDT, T. (2007) Chemometrics in Metabonomics.

UMETRICS (Ed.) (2009) *SIMCA-P+ 11 Tutorial*.

UNDERWOOD, B. R., BROADHURST, D., DUNN, W. B., ELLIS, D. I., MICHELL, A. W., VACHER, C., MOSEDALE, D. E., KELL, D. B., BARKER, R. A., GRAINGER, D. J. & RUBINSZTEIN, D. C. (2006) Huntington disease patients and transgenic mice have similar pro-catabolic serum metabolite profiles. *Brain,* 129**,** 877-886.

VAN DER WERF, M., TAKORS, R., SMEDSGAARD, J., NIELSEN, J., FERENCI, T., PORTAIS, J., WITTMANN, C., HOOKS, M., TOMASSINI, A., OLDIGES, M., FOSTEL, J. & SAUER, U. (2007) Standard reporting requirements for biological samples in metabolomics experiments: microbial and in vitro biology experiments. *Metabolomics,* 3**,** 189-194.

VAN DOORN, M., VOGELS, J., TAS, A., VAN HOOGDALEM, E. J., BURGGRAAF, J., COHEN, A. & VAN DER GREEF, J. (2007) Evaluation of metabolite profiles as biomarkers for the pharmacological effects of thiazolidinediones in Type 2 diabetes mellitus patients and healthy volunteers. *British Journal of Clinical Pharmacology,* 63**,** 562-574.

VAN WEELDEN, S. W. H., FAST, B., VOGT, A., VAN DER MEER, P., SAAS, J., VAN HELLEMOND, J. J., TIELENS, A. G. M. & BOSHART, M. (2003) Procyclic Trypanosoma brucei Do Not Use Krebs Cycle Activity for Energy Generation. *Journal of Biological Chemistry,* 278**,** 12854-12863.

WAKSMUNDZKA-HAJNOS, M., PETRUCZYNIK, A. & HAWRYL, A. (2001) Comparison of chromatographic properties of cyanopropyl-, diol- and aminopropyl- polar-bonded stationary phases by the retention of model compounds in normal-phase liquid chromatography systems. *Journal of Chromatography A,* 919**,** 39-50.

WANG, Y., CLOAREC, O., TANG, H., LINDON, J. C., HOLMES, E., KOCHHAR, S. & NICHOLSON, J. K. (2008) Magic Angle Spinning NMR and (1)H-(31)P Heteronuclear Statistical Total Correlation Spectroscopy of Intact Human Gut Biopsies. *Anal Chem,* 80**,** 1058-1066.

WANT, E. J., O'MAILLE, G., SMITH, C. A., BRANDON, T. R., URITBOONTHAI, W., QIN, C., TRAUGER, S. A. & SIUZDAK, G. (2005) Solvent-Dependent Metabolite Distribution, Clustering, and Protein Extraction for Serum Profiling with Mass Spectrometry. *Analytical Chemistry,* 78**,** 743-752.

WANT, E. J., WILSON, I. D., GIKA, H., THEODORIDIS, G., PLUMB, R. S., SHOCKCOR, J., HOLMES, E. & NICHOLSON, J. K. (2010) Global metabolic profiling procedures for urine using UPLC-MS. *Nat. Protocols,* 5**,** 1005-1018.

WATSON, A. D. (2006) Thematic review series: Systems Biology Approaches to Metabolic and Cardiovascular Disorders. Lipidomics: a global approach to lipid analysis in biological systems. *J. Lipid Res.,* 47**,** 2101-2111.

WATSON, D. G. (1994) *Chemical derivatisation in gas chromatography. In:Gas Chromatography a Practical Approach*, IRL Press (P. Baugh ed.)

WATSON, D. G. (1999) *Pharmaceutical Analysis,* Glasgow, Churchil-Livingstone.

WERNER, E., CROIXMARIE, V., UMBDENSTOCK, T., EZAN, E., CHAMINADE, P., TABET, J.-C. & JUNOT, C. (2008) Mass Spectrometry-Based Metabolomics: Accelerating the Characterization of Discriminating Signals by Combining Statistical Correlations and Ultrahigh Resolution. *Analytical Chemistry,* 80**,** 4918-4932.

WILLIAMS, D. K., MCALISTER, G. C., GOOD, D. M., COON, J. J. & MUDDIMAN, D. C. (2007) Dual Electrospray Ion Source for Electron-Transfer Dissociation on a Hybrid Linear Ion Trap-Orbitrap Mass Spectrometer. *Anal. Chem.,* 79**,** 7916-7919.

WILLIAMS, R. E., LENZ, E. M., LOWDEN, J. S., RANTALAINEN, M. & WILSON, I. D. (2005) The metabonomics of aging and development in the rat: an investigation into the effect of age on the profile of endogenous metabolites in the urine of male rats using 1H NMR and HPLC-TOF MS. *Molecular BioSystems,* 1**,** 166-175.

WILLY MORELLE, K. C., FRÉDÉRIC CHIRAT ,VALEGH FAID ,JEAN-CLAUDE MICHALSKI (2006) The use of mass spectrometry for the proteomic analysis of glycosylation. *PROTEOMICS,* 6**,** 3993-4015.

WILSON, I. D. & NICHOLSON, J. K. (1988) Revised method of proton NMR urinalysis for detecting inborn errors of metabolism: a critique. *Clin Chem,* 34**,** 213-215.

WILSON, I. D., PLUMB, R., GRANGER, J., MAJOR, H., WILLIAMS, R. & LENZ, E. M. (2005) HPLC-MS-based methods for the study of metabonomics. *Journal of Chromatography B,* 817**,** 67-76.

WINNIKE, J. H., LI, Z., WRIGHT, F. A., MACDONALD, J. M., O'CONNELL, T. M. & WATKINS, P. B. (2010) Use of Pharmaco-Metabonomics for Early Prediction of Acetaminophen-Induced Hepatotoxicity in Humans. *Clin Pharmacol Ther,* 88**,** 45-51.

WISHART, D. S. (2008) Quantitative metabolomics using NMR. *TrAC Trends in Analytical Chemistry,* 27**,** 228-237.

WISHART, D. S., KNOX, C., GUO, A. C., EISNER, R., YOUNG, N., GAUTAM, B., HAU, D. D., PSYCHOGIOS, N., DONG, E., BOUATRA, S., MANDAL, R., SINELNIKOV, I., XIA, J., JIA, L., CRUZ, J. A., LIM, E., SOBSEY, C. A., SHRIVASTAVA, S., HUANG, P., LIU, P., FANG, L., PENG, J., FRADETTE, R., CHENG, D., TZUR, D., CLEMENTS, M., LEWIS, A., DE SOUZA, A., ZUNIGA, A., DAWE, M., XIONG, Y., CLIVE, D., GREINER, R., NAZYROVA, A., SHAYKHUTDINOV, R., LI, L., VOGEL, H. J. & FORSYTHE, I. (2009) HMDB: a knowledgebase for the human metabolome. *Nucl. Acids Res.,* 37**,** D603-610.

WISHART, D. S., TZUR, D., KNOX, C., EISNER, R., GUO, A. C., YOUNG, N., CHENG, D., JEWELL, K., ARNDT, D., SAWHNEY, S., FUNG, C., NIKOLAI, L., LEWIS, M., COUTOULY, M. A., FORSYTHE, I., TANG, P., SHRIVASTAVA, S., JERONCIC, K., STOTHARD, P., AMEGBEY, G., BLOCK, D., HAU, D. D., WAGNER, J., MINIACI, J., CLEMENTS, M., GEBREMEDHIN, M., GUO, N., ZHANG, Y., DUGGAN, G. E., MACINNIS, G. D., WELJIE, A. M., DOWLATABADI, R., BAMFORTH, F., CLIVE, D., GREINER, R., LI, L., MARRIE, T., SYKES,

B. D., VOGEL, H. J. & QUERENGESSER, L. (2007) HMDB: the Human Metabolome Database. *Nucleic Acids Res,* 35**,** D521-6.

WREN, S. A. C. & TCHELITCHEFF, P. (2006) Use of ultra-performance liquid chromatography in pharmaceutical development. *Journal of Chromatography A,* 1119**,** 140-146.

WU, H., XUE, R., LU, C., DENG, C., LIU, T., ZENG, H., WANG, Q. & SHEN, X. (2009) Metabolomic study for diagnostic model of oesophageal cancer using gas chromatography/mass spectrometry. *Journal of Chromatography B,* 877**,** 3111-3117.

XU, Y., HEILIER, J.-F. O., MADALINSKI, G., GENIN, E., EZAN, E., TABET, J.-C. & JUNOT, C. (2010) Evaluation of Accurate Mass and Relative Isotopic Abundance Measurements in the LTQ-Orbitrap Mass Spectrometer for Further Metabolomics Database Building. *Analytical Chemistry,* 82**,** 5490-5501.

YABU Y, TAKAYANAGI T & S., S. (1989) Long-term culture and cloning system for Trypanosoma brucei gambiense bloodstream forms in semi-defined medium in vitro. *Parasitol Res.,* 76**,** 93-97.

YANG, F., YAN, S., HE, Y., WANG, F., SONG, S., GUO, Y., ZHOU, Q., WANG, Y., LIN, Z., YANG, Y., ZHANG, W. & SUN, S. (2008) Expression of hepatitis B virus proteins in transgenic mice alters lipid metabolism and induces oxidative stress in the liver. *Journal of Hepatology,* 48**,** 12-19.

YANG, J., XU, G., ZHENG, Y., KONG, H., WANG, C., ZHAO, X. & PANG, T. (2005) Strategy for metabonomics research based on high-performance liquid chromatography and liquid chromatography coupled with tandem mass spectrometry. *Journal of Chromatography A,* 1084**,** 214-221.

YEW, J. Y., CODY, R. B. & KRAVITZ, E. A. (2008) Cuticular hydrocarbon analysis of an awake behaving fly using direct analysis in real-time time-of-flight mass spectrometry. *Proceedings of the National Academy of Sciences,* 105**,** 7135-7140.

YOKOTA, S. D. & SHOEMAKER, V. H. (1981) Xanthine excretion in a desert scorpion,Paruroctonus mesaensis. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology,* 142**,** 423-428.

YOSHIDA, T. & OKADA, T. (1999) Peptide separation in normal-phase liquid chromatography: Study of selectivity and mobile phase effects on various columns. *Journal of Chromatography A,* 840**,** 1-9.

ZABLOCKI, K., MILLER, S. P. F., GARCIA-PEREZ, A. & BURG, M. B. (1991) Accumulation of Glycerophosphocholine (GPC) by Renal Cells: Osmotic Regulation of GPC: Choline Phosphodiesterase. *Proceedings of the National Academy of Sciences,* 88**,** 7820-7824.

ZHANG, B. L. (2003) Compound danshen's effective material and mechanism of action. *World Sci. Technol.: Mod. Tradit. Chin. Med. Mater. Med. ,* 5**,** 14-17.

ZHANG, X., WU, H., LIAO, P., LI, X., NI, J. & PEI, F. (2006) NMR-based metabonomic study on the subacute toxicity of aristolochic acid in rats. *Food and Chemical Toxicology,* 44**,** 1006-1014.

ZHENG, L., T'KIND, R., DECUYPERE, S., VON FREYEND, S. J., COOMBS, G. H. & WATSON, D. G. (2010) Profiling of lipids in Leishmania donovani using hydrophilic interaction chromatography in combination with Fourier transform mass spectrometry. *Rapid Communications in Mass Spectrometry,* 24**,** 2074-2082.

ZHOU, X. & RIDDIFORD, L. M. (2008) rosy Function Is Required for Juvenile Hormone Effects in Drosophila melanogaster. *Genetics,* 178**,** 273-281.