Development and Application of Isotope Labeling Method and Data Processing Program for Liquid Chromatography Mass Spectrometry Based Metabolome Profiling

by

Chiao-Li Tseng

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry
University of Alberta

Abstract

The objective of this thesis focuses on establishing a software tool (IsoMS) used in the quantification and qualification of metabolites in complex biological samples, developing robust methods to extract metabolites from different plant samples, and applying a differential ^{13}C-/^{12}C-isotope dansylation labeling technique with liquid chromatography-mass spectrometry (LC-MS) for the analysis of amine- and phenol-containing metabolites in plant and human urine samples. This approach, a differential ^{13}C-/^{12}C-isotope dansylation labeling method combined with LC-MS, has successfully been applied to the relative quantification, peak pair identification, and putative metabolite identification in a study of metabolic profiling.

First of all, IsoM has been developed to handle the raw LC-MS data generated by a chemical isotope labeling or isotope coded derivatization (ICD) metabolomics platform by peak picking, peak pairing, peak-pair filtering and peak-pair intensity ratio calculation. Secondly, a protocol for rapid extraction, derivatization and detection of amine- and phenol-containing metabolites in mg-scale samples of flax (*Linum usitiatissimum*) bast fibers has been developed and can typically be completed in under 13 hours for up to 8 samples. Thirdly, the development of a robust metabolite extraction method from ginseng roots tailored to the analysis of amine- and phenol-containing metabolites and the use of the dansylation labeling method to gauge the detectability of the LC-MS technique for the ginseng metabolome has been used on the application of spatial distribution

of metabolites in ginseng roots. Fourthly, a microwave-assisted extraction method of *Arabidopsis thaliana* has been optimized and used prior to a differential ^{13}C-/^{12}C-isotope dansylation labeling technique with LC-FT-MS for the study of metabolic profiling of different *TIFY* gene expression plants under methyl jasmonate treatment. Finally, a highly sensitive dansylation isotope labeling LC-MS method was used to determine the urine metabolomes before and after a moderate amount of drinking Goji tea, considered as a phyto-nutritional supplement drink. The result clearly showed that the consumption of Goji tea does not affect the urine metabolome significantly for the studies of the short term (<3 hr) and longer term (12 hr) effects of drinking Goji tea.

Preface

A version of Chapter 2 has been submitted for publication as Ruokun Zhou, Chiao-Li Tseng, Tao Huan and Liang Li, "IsoMS: Automated Processing of LC-MS Data Generated by a Chemical Isotope Labeling Metabolomics Platform", Analytical Chemistry, 86, 4675-4679. R. Zhou and I were responsible for design of data processing program for FTICR-MS version and data validation for each steps. R. Zhou also designed another version for TOF-MS and draft the initial manuscript. T. Huan designed a method of peak pair alignment. L. Li supervised the project and edited the manuscript.

A version of Chapter 6 has been published as Chiao-Li Tseng, and Liang Li, "High-Performance Isotope-Labeling Liquid Chromatography Mass Spectrometry for Investigating the Effect of Drinking Goji Tea on Urine Metabolome Profiling", Science China Chemistry, 57, 1-8. I was responsible for design of experiments, data collection and analysis, as well as manuscript preparation. L. Li supervised the project and edited the manuscript.

Acknowledgements

First of all, I would like to express my deepest gratitude and appreciation to my supervisor, Professor Liang Li, for his skillful guidance, valuable suggestion, financial support, encouragement and patience throughout my graduate studies. Without his persistent help this dissertation would not have been possible.

I would also like to extend my sincere thanks to my dissertation committee members, Professor Derrick L. J. Clive, Professor James Harynuk, Proferror Richard L. McCreery and Professor Wojciech Gabryelski for the time reviewing my thesis and for insightful suggestions on my thesis work.

I would like to show my gratitude to Dr. Ruokun Zhou and Dr. Xianmin Hu for their great and wonderful contributions on the data processing program. I am very grateful to Professor Michael K. Deyholos and Dr. Dan Zhu for providing me flax and *Arabidopsis thaliana* samples, as well as much advice. I would like to acknowledge Dr. Randy Whittal and the entire staff from the Mass Spectrometry Laboratory for their support and excellent advices. My deep gratitude is given to all group members in Dr. Liang Li's research group for their helpful discussions and assistance throughout these years. Working with them made my graduate experience enjoyable. During my graduate program, I have collaborated with many colleagues for whom I have great regard, and I wish to extend my warmest gratefulness to all those who helped me with my work, as well as expressing my apology that I could not mention personally one by one.

At the end, a special thanks to my family. Words fail to express my gratitude and indebtedness to my parents and sisters for their blessings and affection, who were a constant source of inspiration and always with me as a great support and motivation which made me realize the dream.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACN	Acetonitrile
APCI	Atmospheric Pressure Chemical Ionization
API	Atmospheric Pressure Ionization
APPI	Atmospheric Pressure Photoionization Ionization
BPC	Base Peak Chromatogram
CID	Collision-Induced Dissociation
CE	Capillary Electrophoresis
CRM	Charge Residue Model
Da	Dalton
DC	Direct Current
DIL	Differential Isotopic Labeling
DnCl	Dansyl Chloride
EML	Evidence-based Metabolome Library
ESI	Electrospray Ionization
nM	Nano molarity
nm	Nano meter
FID	Free Inductive Decay
FT-ICR	Fourier-Transform Ion Cyclotron Resonance
FT-ICR-MS	Fourier Transform Ion Cyclotron Resonance Mass Spectrometry
GC	Gas Chromatography
GC-MS	Gas Chromatography Mass Spectrometry
HMDB	Human Metabolome Database
HPLC	High Performance Liquid Chromatography
ICR	Ion Cyclotron Resonance
IEM	Ion Evaporation Model
JA	Jasmonate
LC-MS	Liquid Chromatography Mass Spectrometry
LC	Liquid Chromatography
m/z	Mass to Charge
MAE	Microwave-assisted Extraction
MALDI	Matrix-assisted Laser Desorption Ionization
MCP	Microchannel Plate
MeJA	Methyl Jasmonate

MeOH	Methanol
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MW	Molecular Weight
NMR	Nuclear Magnetic Resonance
NPC	Normal Phase Chromatography
PLS-DA	Partial Least Square Discriminant Analysis
PCA	Principal Component Analysis
PDA	Photo Diode Array
ppm	part(s) per million
QC	Quality control
QTOF	Quadrupole Time-Of-Flight
RF	Radio Frequency
RT	Retention Time
RPLC	Reversed Phase Liquid Chromatography
RT	Retention Time
RSD	Relative standard derivation
S/N	Signal to noise ratio
SIL	Stable Isotopic labeling
SIDT	Single Ion in Droplet Theory
TCM	Traditional Chinese Medicine
TOF	Time-Of-Flight
UAE	Ultrasonic-assisted Extraction
UPLC	Ultra Performance Liquid Chromatography
UV	Ultra- violet
VIP	Variable Importance on the Projection
W	Watts
μM	Micro molarity

Chapter 1

Introduction

1.1 Metabolomics

Metabolomics is a scientific study focused on the identification and quantification of endogenous and exogenous metabolites, whose molecular weight is normally smaller than 1500 daltons, in given biological samples. The term 'metabolome' was first used by Olivier in 1998 and defined as the entirety of metabolites present in cells or organisms which participate in particular physiological or developmental metabolic reactions.[1] The definition of metabolomics was described as the comprehensive, qualitative, and quantitative study of all the small molecules that could help in the understanding of biosystems and the revealing of metabolome in cells, tissues, biological fluids or organisms.[2-5]

Metabolomics is a relatively new member of the 'omics' family which includes genomics, transcriptomics, proteomics, and so on. At the endpoint of the "omics" cascade (see Figure 1-1[6]), concentrations of metabolites can be affected by proteins, genes, environmental influences, diseases and drug exposure. Due to the complexities of the metabolites with diverse chemical and physical properties, three different analytical platforms are needed to analyze the metabolites in a biological system, namely, metabolic target analysis, metabolic fingerprinting and metabolic profiling. The objective of a metabolic target analysis is to establish the quantitative measurement of some selected metabolites.[7,8] Metabolic fingerprinting is an approach often used to quickly compare patterns of metabolites among different samples according to those of global metabolites instead of quantification of selected metabolites. The aim of metabolic fingerprinting is not to identify all the involved metabolites, but to discover putative biomarkers.[9-11] Metabolic profiling

focuses on qualification and quantification of a group of related metabolites or metabolites belonging to a specific metabolic or biological pathway.[6,12]

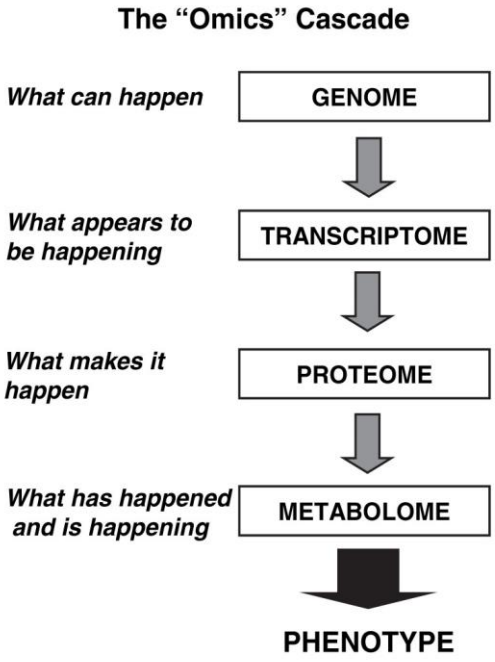


Figure 1-1 The components of the "omics" cascade.[6]

In metabolomics studies, nuclear magnetic resonance (NMR) spectroscopy and/or mass spectrometry (MS) are commonly used for identification and quantification of molecules based on either spectral interpretation or mass to charge ratio. NMR is a useful technique for qualifying and quantifying metabolites in samples with minimal sample preparation.[13,14] Because NMR is a non-destructive technique, it is very suitable for valuable samples. However, the sensitivity of NMR is relative low compared to MS so it is not a good technique for detecting low abundance or low concentration metabolites. In addition, compared to MS analysis, NMR analysis requires large sample volumes.[15] Furthermore, it is challenging to couple NMR detection with separation techniques.[15] Therefore, it is hard to use NMR for the analysis of complex biological samples because of the lack of separation prior to analysis. MS has been successfully coupled with

separation techniques such as liquid chromatography (LC), gas chromatography (GC) and capillary electrophoresis (CE).[16,17] GC offers chromatographic separation, and is capable of identification and quantification of known and unknown volatile metabolites. However, a large proportion of metabolites are polar and non-volatile so chemical derivatization is often required to reduce polarity and increase thermal stability and volatility of metabolites before being analyzed by GC. Although CE requires small sample volumes and offers high resolution, direct coupling of CE with MS is still more challenging relative to high performance liquid chromatography (HPLC) because of the high electric field and low flow rate.[18] Besides, other disadvantsges of CE-MS include the relatively poor concentration sensitivity and the fluctuation in analyte migration time.[18,19] LC plays an important role in metabolomic analysis for separating complex biological samples prior to analysis. The function of LC is to separate a mixture of compounds with the purpose of purifying, identifying and quantifying each individual component in the mixture. LC can not only decrease ion suppression effect in MS caused by co-elution of compounds, but also reduce isobaric interferences in cases of using low-resolving mass analyzers. With MS to be used for detection, the improved separation and ion production through LC can offer highly selective and sensitive quantitative analysis, and provide the potential to identify unknown metabolites. Hence LC-MS has become an analytical tool to be chosen in the area of untargeted and unbiased metabolomics studies.

Metabolomics is a critically important technique in systems biology, and has been applied in diverse fields such as nutrition, drug development, agriculture, clinical diagnostics, environmental chemistry, toxicology, etc.[20-24] Metabolomics is the latest episode in the “omics” era (Figure 1-1), so the metabolome is believed to be the closest representative and the best reporter of the phenotype. In addition, metabolomics should ultimately reflect an overall picture of the metabolic pathways involved in the interaction of proteins, encoded by the genes and influenced by

environmental conditions or drug exposure.[25-27] Furthermore, small changes taking place in genes, transcripts, or proteins may be enlarged in the structures or concentrations of metabolites which can be discovered as potential biomarkers.[28,29] Thus, metabolomics plays an important role in the "omics" family. The challenges in metabolomics are that the chemical and physical properties as well as structures of metabolites are extremely diverse and there is a wide concentration range of metabolites in biological samples. For example, the human metabolome contains more than 8,000 metabolites[30,31] and the concentration ranges of metabolites are approximately 7-9 orders of magnitude.[32,33] Until now, no single analytical platform or technique is capable of analyzing all metabolites at once. Thus, the field of metabolomics is still in its early stage of development and facing many analytical and data processing challenges.

1.2 Technique of Liquid Chromatography Mass Spectrometry (LC-MS)

LC-MS is an anytical technique which is widely used for metabolomics studies because of its capabilities to perform chemical separation and mass analysis. In my research, reversed phase liquid chromatography (RPLC), the most commonly used form of LC, is coupled with mass spectrometry. The interface used to couple RPLC with MS is electrospray ionization (ESI) which is the most popular atmospheric pressure ionization interface. RPLC, ESI and mass analyzers are briefly introduced in the following sections.

1.2.1 Reversed Phase Liquid Chromatography (RPLC)

During the 1940s and 1950s, Archer John Porter Martin and Richard Laurence Millington Synge, who were awarded the Nobel Prize in 1952, established the principles of partition chromatography, and also developed the plate theory. Based on their principle and theory, the development of several chromatographic techniques, including paper chromatography, gas chromatography, and liquid chromatography, have advanced rapidly. Nowadays, the family of LC

has become larger and larger. The main principle of LC is the separation of analytes based on the interactions among analytes, stationary phase, and mobile phase. There are many types of LC including normal phase chromatography (NPC), reverse phase chromatography (RPC), ion exchange chromatography (IEC), and size exclusion chromatography (SEC).

Reverse phase liquid chromatography (RPLC) is one of the most important and specific separation methods used in chemistry, biology, and many other fields because it can provide premium resolution for most medium polarity and small non-polar molecules. For metabolome analysis, thousands to tens of thousands of analytes may be present in a given biological sample. High resolution separation prior to MS analysis can deal with ion suppression effect and reduce MS complexity in the time dimension. RPLC is not only compatible with electrospray ionization (ESI) but can also handle water-soluble biological samples. Owing to its highly resolved separation and reproducible compound specific retention, the retention factor of analytes can be used as a secondary criterion for identification by comparing with that of authentic standards. Thus, RPLC coupled with MS via ESI can provide quantitative and qualitative analyses with high selectivity and sensitivity, as such possesses great potential to be the mainstream technique for metabolomics.[6,34,35]

RPLC consists of a non-polar stationary phase inside a column and a moderately polar mobile phase composed of water with an organic solvent. The mobile phase is often a mixture of two miscible solvents with differing polarities. The most typical is water combined with a less polar solvent such as acetonitrile (ACN) or methanol (MeOH) and it can be further mixed with buffers or other additives in relatively low concentrations depending on the specific analysis. The stationary phase is commonly packed with microporous silica particles which have been surface-modified with hydrocarbon chains, e.g. C18, C8, or C4. In metabolomic studies, RPLC using C18 or C8 narrow-bore columns with particle sizes of 3-5 μm has been the most popular equipment for

separation of moderately polar and non-polar metabolites. Once a sample is injected, the mobile phase carries the different solute analytes through the column. These analytes will interact and partition differently between the mobile and stationary phases so each analyte will pass through the column with a different migration rate. For relatively polar analytes, they will interact less with the non-polar stationary phase and more with the polar mobile phase so they will migrate through the column with a higher velocity compared with less polar analytes and elute earlier. On the other hand, the relatively non-polar analytes will interact more strongly with the stationary phase and less with the polar mobile phase. Separating a simple sample with only a few analytes is routine; however, for complex biological samples it is not possible for all the analytes to be separated. Thus, the analytes with same polarity will begin to co-elute with one another and form a band during the migration. Isocratic separations refer to the composition of mobile phase fixed in a constant ratio throughout the entire run while gradient separations means the composition of mobile phase is changed during separation process. Generally gradient separations are much more powerful with complex biological samples as largely diverse analytes can be separated in short amount of time.

In metabolomics studies, RPLC is the most used method of separation prior to MS analysis. Conventional HPLC systems with a variety of conventional RP columns are used in most of metabolomic studies for analyzing target metabolites. However, for profiling metabolomics, those columns may not be sufficient to separate thousands to ten thousands of metabolites in a given biological sample. In order to separate and detect more metabolites in a chromatographic run, peak capacity needs to be increased. One approach is to pack a column with smaller size particles. Recently, sub-2 μm particle columns were developed and used in ultra-performance liquid chromatography (UPLC) for much faster and efficient separations.[36] Compared to HPLC, the advantages of UPLC include improvement of sample processing throughput and reduction of solvent consumption without compromising separation efficiency. On the downside, using sub-

2μm particle columns increases the high back-pressure (10,000 to 15,000 psi) and generates column heat, which warms up a fraction of the mobile phase against the stationary phase.[37,38] The dissipation of the heat causes radial temperature gradients, and the temperature difference between the center and the column wall can be as much as 6 K. The radial temperature gradients affect the heterogeneous distribution of the mobile phase linear velocity, viscosity and density along the column, as well as the spactial distribution of analytes loading on the head of a column. Thus, temperature gradients have a negative effect on column efficiency.[38,39] In order to minimize the radial temperature gradient, the maximum diameter of sub-2 μm particle columns has to be kept narrow (2.1 mm). Most sub-2 μm particle columns show outstanding performance in UPLC, and they also work well in traditional HPLC (up to 6000 psi) when running at a moderate flow rate.

1.2.2 Electrospray Ionization (ESI)

Liquid chromatography mass spectrometry (LC-MS) coupled with an electrospray ionization (ESI) interface is one of techniques chosen for a robust metabolomics study. The most challenging task in this method is the transfer of ions from liquid solution to gas phase at atmospheric pressure, so the development of atmospheric pressure ionization (API) interface is neccesary. The function of the API interface is to rapidly evaporate the liquid eluting from LC and to ionize sample molecules prior to analysis by mass spectrometry. Several types of API interface have been developed and coupled with LC, including electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), and atmospheric pressure photoionization ionization (APPI). Among these, ESI is the most widely used in metabolomics studies, especially for medium polar and small polar molecules. The rest of this section will discuss ESI in more detail.

Electrospray occurs when a liquid flows through a capillary tube biased with a high applied voltage, and this potential induces liquid drops at the capillary tip to disperse into a fine aerosol.

This physical phenomenon has been known for hundreds of years and was first described by Zeleny in 1914.[40] In 1968, Dole and his coworkers used electrospray to ionize intact chemical species and the technique of ESI was invented.[41] In 1984, ESI was coupled with mass spectrometry by Fenn and Yamashita, and used to ionize high mass biologically important compounds.[42] In the late 1980's, Fenn and his coworkers successfully described the basic experimental principles and methodologies of ESI, such as soft ionization of non-volatile and thermally labile compounds, and intact ionization of complexes.[43] In this thesis, ESI was chosen as the ionization method, due to its high sensitivity of detecting small polar molecules and readiness to interface with LC.

The technique of ESI-MS involves transferring ions from solution to the gas phase and then subjecting them to mass spectrometry analysis.[44] To date, although the mechanism of ESI has been extensively studied, the mechanism on how analytes in tiny charged droplets are converted into gas phase ions is still not fully understood.[45,46] Simply speaking, the production of gas phase ions from ions in solution can be divided into four major stages including: 1) generation of charged droplets at the ESI capillary tip; 2) shrinkage of charged droplets due to solvent evaporation from droplets; 3) repeated droplet shrinkage and fission; and 4) the formation of gas phase ions via ion evaporation model (IEM) or charge residue model (CRM).[41,47,48] To simplify the discussion, only the positive ion mode is considered here and described as following.

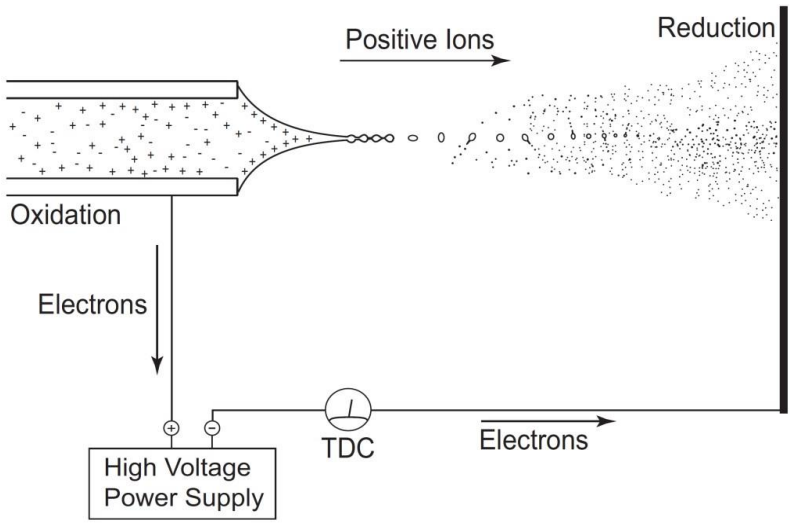


Figure 1-2 Generation of charged droplets in the positive ESI process.[45]

First of all, a strong electric field is applied to generate charged droplets at the capillary tip (shown in Figure 1-2[45]). In positive ion mode, a strong electric field (3-6 kV) is applied on the capillary tip (+) and the counter electrode (-) when HPLC eluent flows through. Because the spray capillary tip is very narrow, the electric field ($E_c \approx 106$ V/m) induces a charge accumulation on the droplet surface at the capillary tip.[45] As a result, the solvent near the meniscus of the liquid will be polarized and positive and negative electrolyte ions in solution will be separated under this electric field. Positive ions will move toward the liquid surface and negative ions will drift away from the surface. Due to the accumulation of positive charges at the liquid surface, the surface becomes destabilized and the electron repulsion forces the surface to elongate and form a conical shape, referred to as the Taylor cone.[45,46] A fine jet is emitted from its tip when the Taylor cone becomes unstable. Then, because of the surface of the fine jet with an excess of positive ions, the downstream jet becomes unstable. When the repulsion among the positive charges overcomes surface tension of the solvent (Rayleigh limit), the jet breaks into a mist of highly charged droplets. Under the pressure of the electric field, the released positively charged droplets fly to the counter-

electrode close to the mass spectrometer at atmospheric pressure. In the positive mode, the leaving droplets contain positive charges and the negative charges will accumulate in bulk liquid. In order to stabilize the ESI region close to the capillary tip, an electrochemical reaction occurs to balance the negative charges in the solvent.[45] Although ESI is a concentration dependent source since the droplet current linearly increases with the concentration of analytes, the current will be saturated when the total concentration of analytes is greater than 10^{-5} M.[45] For high flow rates (>10 μL/min) of HPLC eluent, a coaxially nebulizing gas is used to form a more stable spray and to limit the dispersion in space as well.

The next step of ESI mechanism is the shrinkage of charged droplets due to solvent evaporation. When the charged droplets move to the counter-electrode, a counter flow of heated dry gas, typically nitrogen, passes over the source at moderate temperature to speed up the solvent evaporation from the initially formed droplets while maintaining the charges constant. As solvents evaporate, the droplets become smaller.

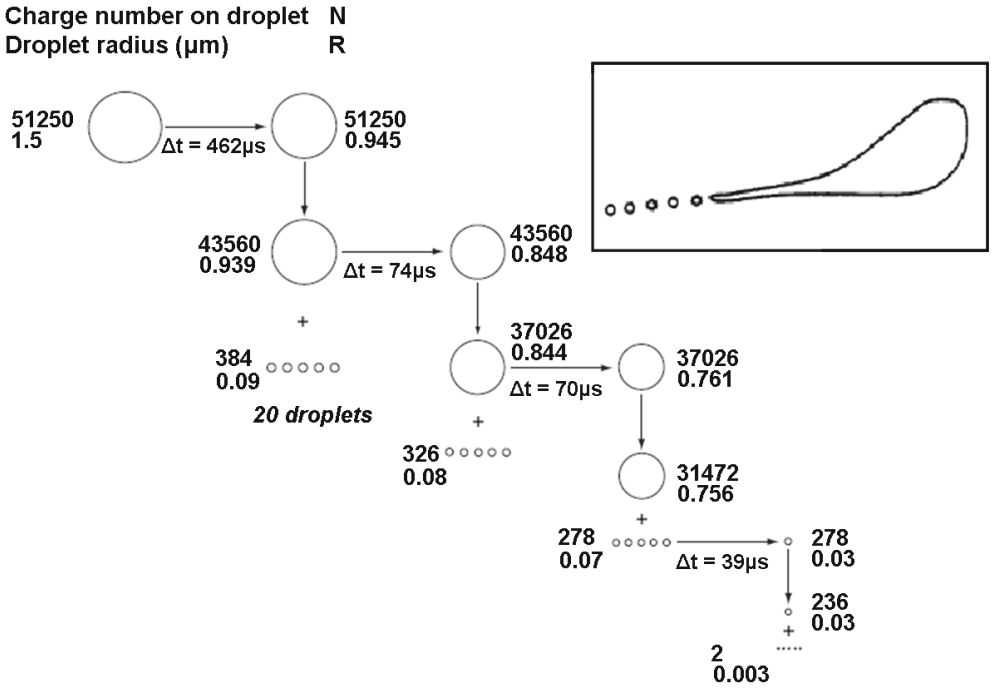


Figure 1-3 Schematic diagram of time history of production of parent and offspring droplets.[45]

The following step is the repeated droplet shrinkage and the fission of parent droplets (see Figure 1-3[45]). The charged droplets continually shrink at a constant rate due to solvent evaporation but the charge of droplets remains constant. This results in increasing charge density of droplets and repulsive forces between the charges at the surface of the drops. The droplet keeps on shrinking further until the droplet radius is close to or at the Rayleigh limit where the repulsive forces of the charges become sufficient to overcome the surface tension holding the droplet together. The parent droplets become unstable, form a new Taylor cone, and then release smaller charged droplets from the tip when the electric repulsion is higher than the Rayleigh limit. The resulting fission undergoes an uneven fission process as proposed by Kebarle and his coworkers.[44-46] A diameter of the first generation of droplets from the capillary is typically around 1.5 μm in the ESI processing. Offspring droplets typically carry about 2% of the volume of the parent droplets, but only have 15% of the parent charge.[46] Therefore, the offspring droplets are much smaller and have a much higher charge-to-mass ratio. Figure 1-3 illustrates the first three fissions of a parent droplet. In each fission process, about 20 smaller offspring droplets are produced from a parent droplet. The solvent evaporation and fission cycles are repeated until the very small and highly charged droplets are capable of producing gas phase ions.[44-46]

The last step of ESI is the formation of gas phase ions. Two different mechanisms have been proposed to elucidate the conversion of analytes from highly charged droplets into gas phase ions. The first mechanism was proposed by Dole et al. in the late 1960s and is known as the charge residue model (CRM), also referred to as single ion in droplet theory (SIDT).[41] CRM describes that the consecutive evaporation and fission process eventually generates some tiny droplets containing only one analyte molecule and multiple charges on the surface of each droplet. This is believed to be what happens for very large, multiply charged species, such as proteins.[49] For a large molecule, further evaporation will get rid of the remaining solvent molecules and lead to a

gas phase ion. The charges of the gas phase ion would come from the original charges of the droplets. CRM is typically used to demonstrate the generation of multiple charged ions of macromolecules in ESI. The other mechanism named ion evaporation model (IEM) was proposed by Iribarne and Thomson in 1976.[48] IEM differs in that ions are emitted from the charged droplets by coulombic repulsion. IEM proposes that ions can be expelled directly from surface of droplets into the surrounding gas when the droplets reach a certain radius and the electric field at the surface has become sufficiently high.[48] As the radius of a droplet shrinks to less than 10 nm, the electric field generated by ions at the droplet surface would overcome the surface tension, and the ions can be directly desorbed from the highly charged droplet into the gas phase. The theory of IEM also explains that Na^+ and Na^+ hydrates are main ion species produced by a droplet, and proposed an equation to predict the condition for droplet ionization. IEM is experimentally well-supported for the ion generation of small molecules.[50,51]

1.2.3 Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR-MS)

Fourier Transform Ion Cyclotron Resonance MS (FT-ICR-MS) is a trapping mass analyzer for determining *m/z* of ions based on the cyclotron frequency of the ions in a fixed magnetic field, which was derived from the cyclotron principles first introduced by Lawrence in the 1930s.[52] It was initially developed by Comisarow and Marshall in 1974 and has been an important platform in metabolic analysis.[53] FT-ICR-MS is an analytical technique that offers an ultrahigh resolution of 100,000 – 1,000,000 (routinely 50,000 – 100,000) so masses of ions can be measured with very high accuracy (0.01-1 mDa). Since a superconducting magnet is much more stable than RF voltage, FT-ICR-MS is able to achieve higher levels of resolution than other mass analyzers and can distinguish the signals of two ions with similar *m/z* as distinct. Unlike other trapping mass analyzers that scan one frequency at a time to generate one spectrum, the ICR cell is able to collect

an entire spectrum at once. In addition, it can exhibt a low detection limit in the attomole to femtomole range and possesses MS^n capabilities.[54,55]

FT-ICR-MS is fundamentally based on the phenomenon known as ion cyclotron resonance (ICR), and the measurement of an ion's cyclotron frequency is inversely proportional to its mass-to-charge ratio. An ICR analyzer cell (known as a Penning trap) is located within a strong magnetic field and requires a high vacuum to minimize ion collisions with residual gas. When an ion perpendicularly moves into a spatially uniform magnetic field, it would face a force, the Lorentz force, given by Eq. 1.1 where m, q, and v represent ionic mass, charge and velocity.

$$Force = mass \times acceleration = m\frac{d\boldsymbol{v}}{dt} = q(\vec{v} \times \vec{B}) \qquad (1.1)$$

The vector cross product term "× B" indicates that the direction of the magnetic component of the Lorentz force is perpendicular to the plane established by the ion velocity and the magnetic field. If the ion moves at a constant speed without any collision, the magnetic field will bend the ion path into a circle, of which the radius is r (see Figure 1.4). Because the angular acceleration in the plane perpendicular to B is defined as $dv/dt = v_{xy}^2/r$, and the angular velocity (ω) is expressed as $\omega = v_{xy}/r$, Eq. 1.1 can be simplified to give Eq. 1.2.

$$\omega_c = \frac{qB}{m} \qquad (1.2)$$

Eq. 1.2, known as the cyclotron equation, explains the motion of unperturbed ions within the magnetic field and also illustrates that all ions of a given *m/z*, or *m/q* in this case have the same ICR frequency that is independent of their velocity. This means that *m/z* and cyclotron frequency are in inverse proportion, and this is the essential idea on which ICR-MS can generate its unique type of data. The first ICR-MS was developed in the 1960s by Varian. The instrument applied a sweep radio frequency (RF) potential on trapped ions to measure cyclotron frequency of the ions. When the frequency of the RF field was coincident with cyclotron frequency of an ion, this ion

was excited to enlarge the radius of ion cyclotron motion until it reached the collector to give a signal. In 1974, Comisarow and Marshall applied the concept of Fourier transform on ICR-MS and successfully developed the FT-ICR-MS.[53] Nowadays, FT-ICR-MS is one of the most important mass analyzers used in metabolomics due to its unsurpassed resolution, highest mass accuracy, and excellent sensitivity.

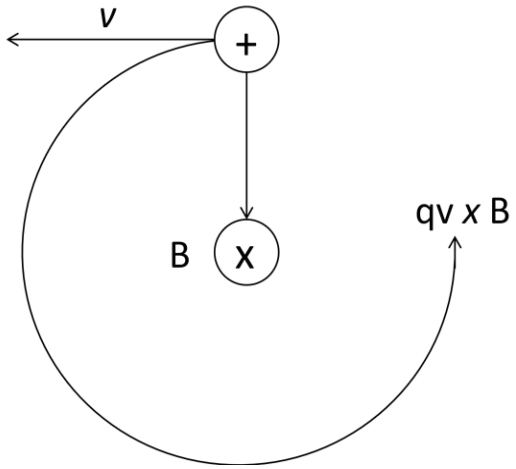


Figure 1-4 Ion cyclotron motion. Moving path of the positive ion in the plane is bent into a circle by the Lorentz magnetic force generated by a homogenous magnetic field perpendicular to the plane.[54]

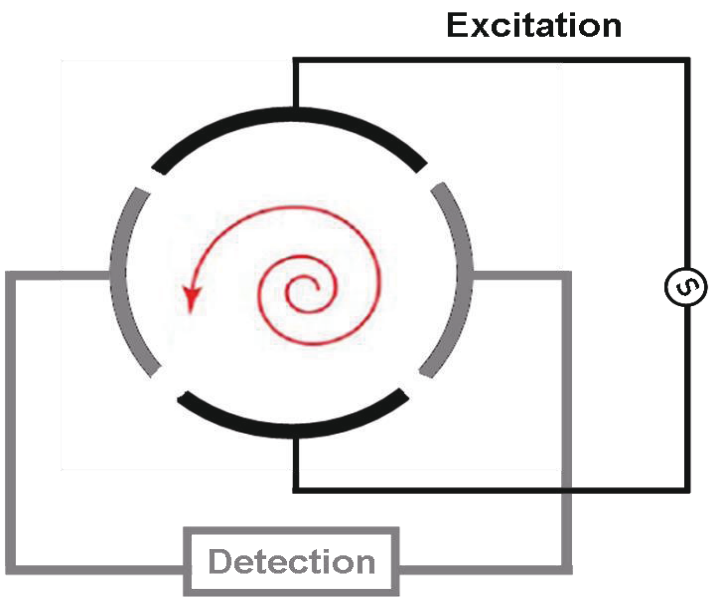


Figure 1-5 Schematic diagram of an ICR analyzer cell in which ions are excited and the image current of the orbiting ions is detected.[56]

In FT-ICR-MS, the ion signal is related to cyclotron frequency so it is important to get the accurate measurement of cyclotron frequency. Ions in ICR cell must be excited to detectable radii by applying a RF potential to two excitation plates at the resonant frequency of the ions of interest (Figure 1-5).[54,56] An excited ion of 100 Da can travel up to ~30 kilometers during a 1 second observation time[54] so ICR as a mass analyzer can provide a much higher resolution than the meter-long time-of-flight or magnetic-sector instruments. The RF potential is often a frequency sweep, known as an "RF chirp" that is used to quickly scan the range of frequencies corresponding to the *m/z* range of interest. Ions of different *m/z* with different cyclotron frequencies are excited to orbits of the same radius during this excitation process, which would induce the coherence of ion packets. As the coherently orbiting ion packet passes two detection plates, the potential change between the detector plates is measured and generates a sinusoidal signal containing the cyclotron frequency and magnitude of the ion in the time domain. The sinusoidal signal is known as time domain data,

or transient or free inductive decay (FID).[54,56] Ions of different m/z with different cyclotron frequencies are simultaneously analyzed and collected in time-domain as raw data. The next step is to extract data about the different ion packets by a mathematical procedure known as a "Fourier transform" (FT). The signal from the time domain can be converted into the frequency domain by Fourier transform to give the mass to charge ratio and also intensity of the ions. Figure 1-6 shows the function of Fourier transform in FT-ICR-MS. The time domain raw data is converted to the frequency domain by a Fourier transform. The cyclotron frequencies of ions and their *m/z* are in inverse proportion so the frequency domain data can be readily converted into *m/z* of ions with mass calibration.

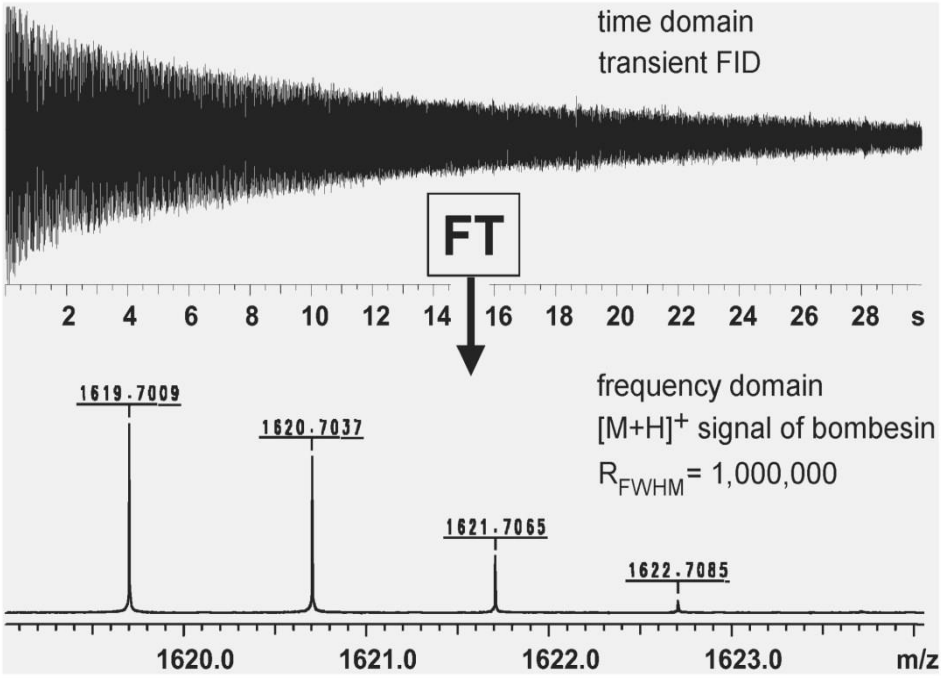


Figure 1-6 Function of Fourier transform in FT-ICR-MS. By courtesy of Bruker Daltonick, Bremen.

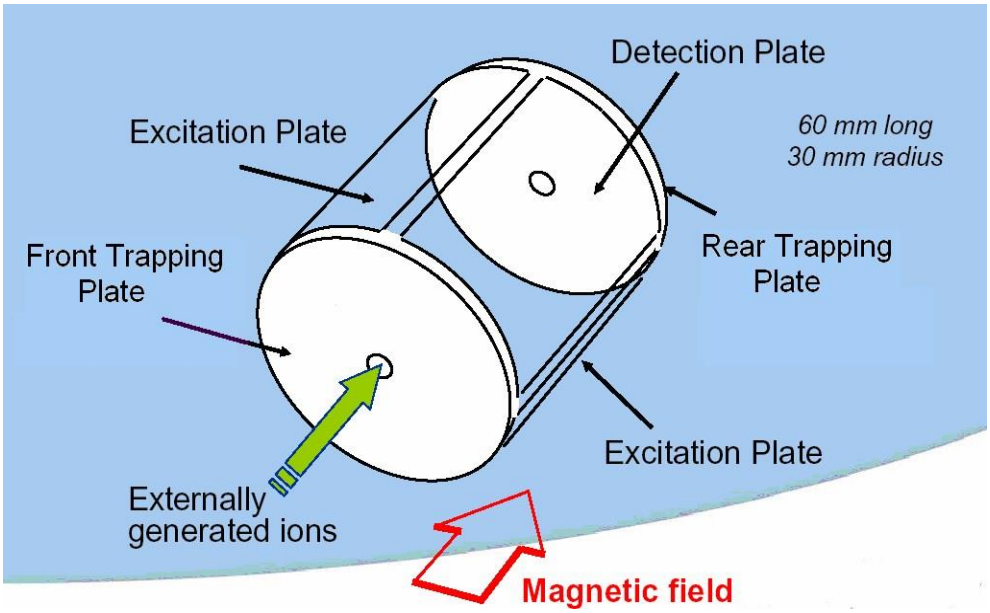


Figure 1-7 Schematic diagram of a typical ICR cell. By courtesy of Bruker Daltonick, Bremen.

The ICR cell, the mass analyzer of the FT-ICR-MS, located inside a homogeneous superconducting magnetic field, is responsible for trapping, exciting, and detecting ions. The most common shape of ICR cell is a closed cylinder constructed by three pairs of electric plates (see Figure 1-7). The first pair consists of two trapping plates placed at the ends of the cylinder with a hole in the center for ion transmission. These trapping plates are perpendicularly aligned to the magnetic field. At the beginning of analysis, ions are introduced along the central axis into the cell, and then a small DC voltage of the same polarity as the ions is applied on the trapping plates to store and confine the ions in the middle of the cell. At the same time another two pairs of electric plates are held at ground potential. After analysis, a large potential is applied to the trapping plates to quickly remove all remaining ions in the cell prior to the next acquisition. The second pair consists of excitation plates oppositely situated on the ICR cell wall and connected to an RF oscillator. The function of the excitation plates is to use a frequency sweep to quickly scan the range of frequencies corresponding to the *m/z* range of interest during ICR operation. For interested

ions with known mass range, an efficient approach named stored waveform inverse Fourier transform (SWIFT) was developed by Marshall *et al*. in the mid-1980s.[57,58] This approach uses an inverse Fourier transform (IFT) to calculate and produce an optimized excitation waveform. The stages of SWIFT include elaboration of the desired excitation spectrum, calculation of the excitation waveform by IFT, and application of the waveform to produce RF excitation frequency. The SWIFT is used not only for the isolation of interested ions but also the ejection of unwanted ions so it is a good method for tandem mass spectrometry experiments. The third pair of detector plates are responsible for collecting and converting the frequency signal to that of *m/z* values obtained in the corresponding mass spectrum.

The signal detected by FT-ICR-MS is proportional to the total charge and the proximity of the ions to the detector plates in the ICR cell. The FTICR-MS is extremely sensitive for multiply charged ions especially when ions are excited to a radius which is close to the detector plates in the ICR cell.[59] However, there are several disadvantages of the ICR cell. Firstly, the ICR cell is a type of trapping cell, so the space charge effect possibly occurs when too many different ion packets interact with each other through Coulombic repulsion and this effect has the influence on mass resolution and accuracy.[60] Space charge effects can be decreased by exciting the ions to larger cyclotron radii, controlling the number of ions injected into the ICR cell, or using higher magnetic field.[4,61] Another drawback of FT-ICR-MS is that it always needs to maintain the ultrahigh vacuum ($< 10^{-9}$ mbar) in the ICR cell in order to minimize the possibility of unwanted collisions between ions and residual gas molecules. Other disadvantages of FT-ICR-MS include high costs and large space requirements.[62]

The Bruker 9.4 T APEX FT-ICR-MS was the major instrument used in my research and is composed of four major sections: the ion source including ESI and MALDI, the Qh-Interface, the ion transfer optics, and the detector (see Figure 1-8). The ultrahigh vacuum in the ICR cell is

generated by six differential pumping stages consisting of two roughing pumps and four turbo molecular pumps. Three software programs have been used to manage the instruments and to process data. Bruker apexControl is used to control FT-ICR-MS.Bruker Hystar is used to control HPLC systems, which interface in diverting the solvent flow into FT-ICR-MS. Bruker DataAnalysis is used to analyze and process the data.

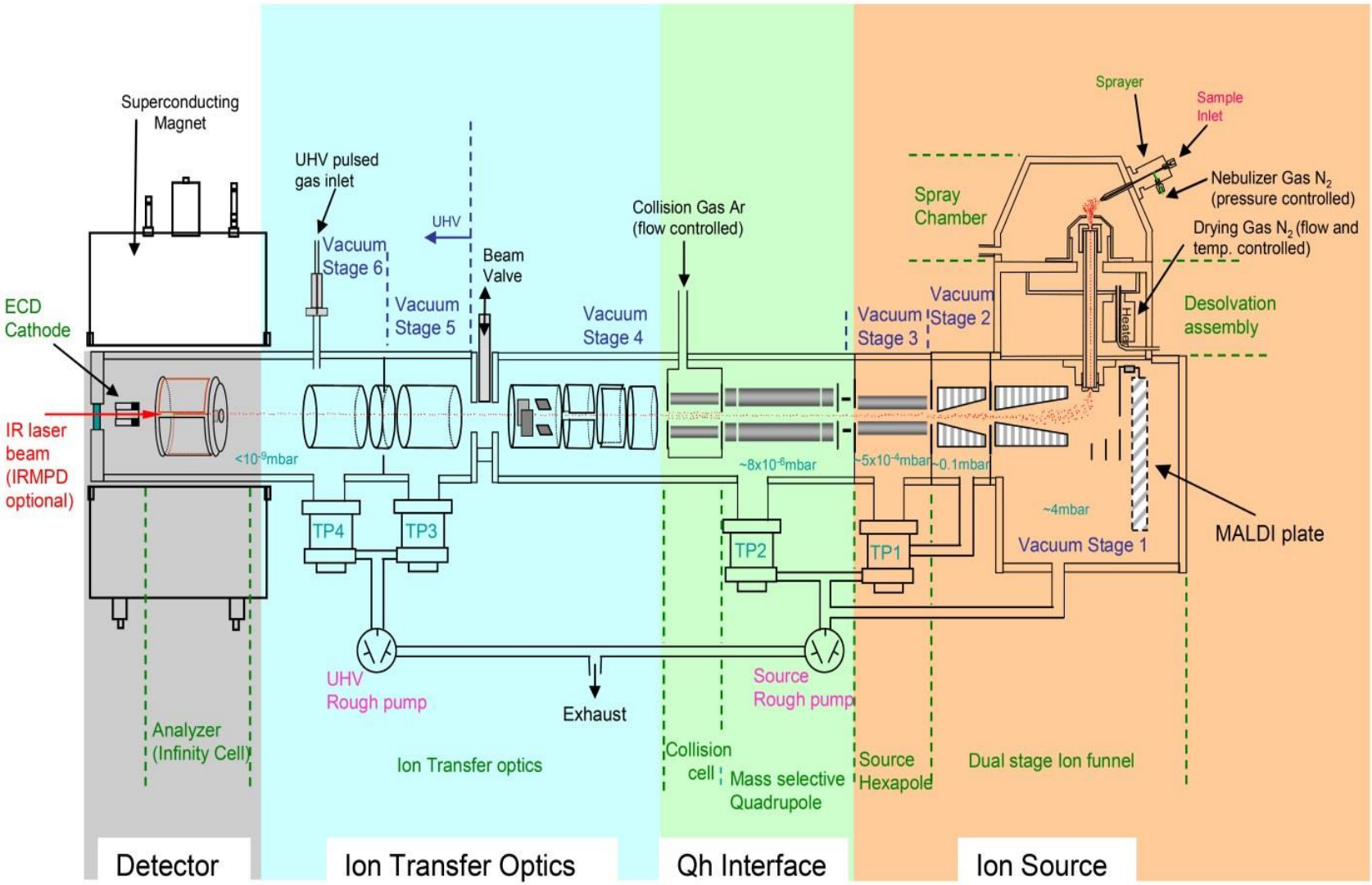


Figure 1-8 Schematic diagram of Bruker 9.4 T Apex FT-ICR-MS. By courtesy of Bruker Daltonick, Bremen.

1.2.4 Quadrupole Time-of-Flight Mass Spectrometry (Q-TOF-MS)

Q-TOF-MS is one of tandem mass instruments and consists of three quadrupoles, Q0, Q1 and Q2, followed by a reflecting Time-of-Flight (TOF) mass analyzer with orthogonal injection of ions (see Figure 1-9).[63] In single MS analysis, all of the three quadrupoles are operated in the RF-only mode to provide radial confinement of ions and also to transmit ions, while the TOF analyzer is used to record spectra. In MS/MS analysis, Q0 is operated in the RF-only mode to focus and guide ions, and Q1 is operated in the mass filter mode to isolate and transmit the parent ion of interest to the collision cell Q2. Before entering the collision cell Q2, the parent ion is accelerated by some electric potential to high kinetic energy. In Q2 where it undergoes collision-induced dissociation (CID), the parent ion collides with neutral gas molecules and generates fragment ions. After leaving the collision cell, fragment ions are accelerated again to the required energy and transmitted by ion optics into TOF section. Because the modulator region is the field-free of TOF, ions continue to fly in their original direction. Then, a pulsed electric field orthogonal to the ion trajectory pushes the ions into the accelerating column where ions are further accelerated to several keV per charge prior to arriving in the field-free flight tube, where TOF mass separation occurs. When ions fly in the flight tube, ions are separated according to their velocities. To enhance the resolution, an ion mirror is used to compensate for the initial energy distribution and to focus ions with the same *m/z* value. Finally, the microchannel plate (MCP) detector collects ions and generates signals.

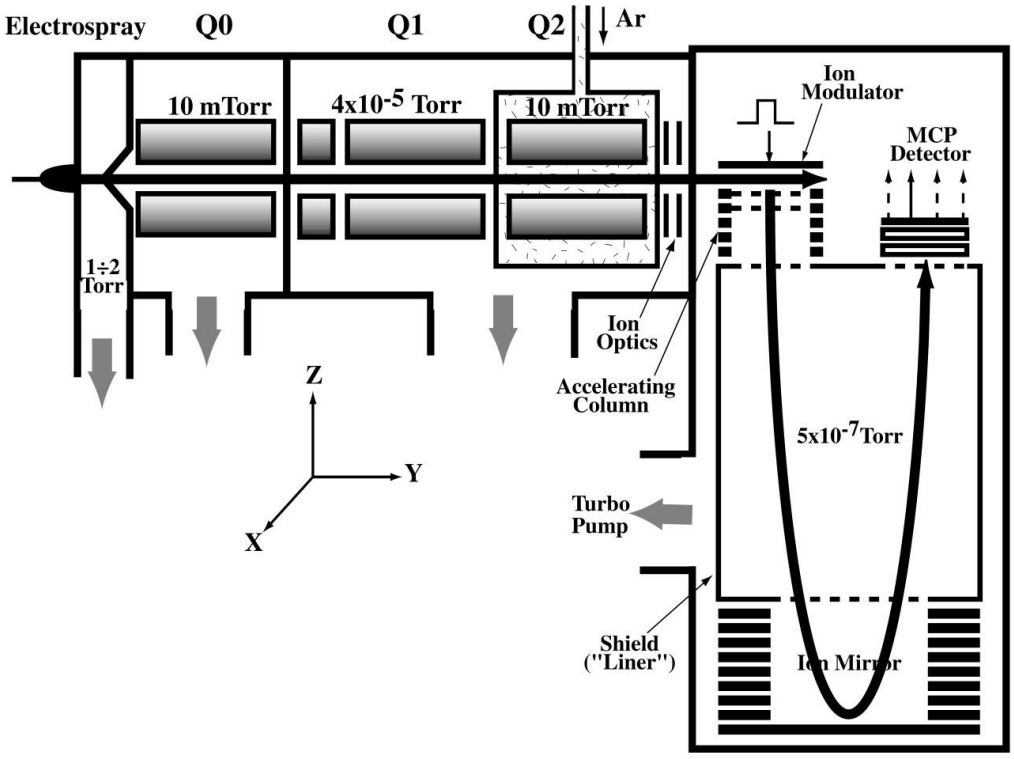


Figure 1-9 Schematic diagram of Q-TOF-MS.[63]

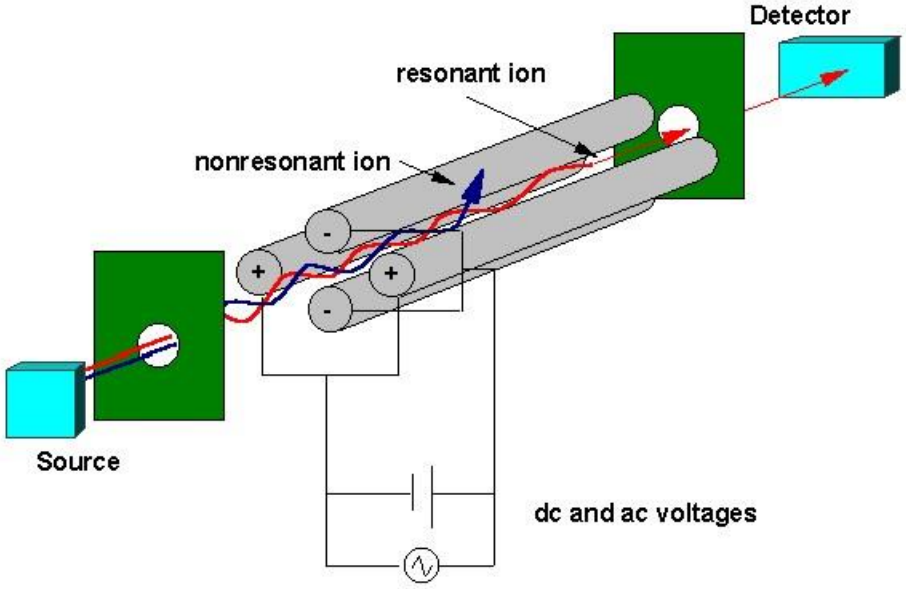


Figure 1-10 Schematic diagram of a linear quadrupoles.[64]

In Q-TOF-MS (Figure 1-9), linear quadrupoles function as mass filters or ion guides. A quadrupole mass analyzer consists of two pairs of cylindrical rods which are used to create a hyperbolic field to allow a certain *m/z* range to pass through for a given combination of potentials. A potential superimposed by radio frequency (RF) and direct current (DC) is applied to cylindrical rods. The form of the potential applied on the rods is $\Phi_0 = U + V\cos(\omega t)$, where U is the DC voltage and $V\cos(\omega t)$ is the RF potential with frequency $\omega/2\pi$. Although the potentials of the two pairs of rods have the same amplitude, the polarities of them are opposite (see Figure 1-10). As an ion travels down the quadrupole along the rods, it experiences an attractive force exerted by one of the rods which has the opposite polarity. The potential polarities on the rods switch periodically so that ions with stable trajectory will not hit the rods before they reach the detector.The stability of ion trajectory inside the quadrupole directly rely on the applied RF component (*V*) and DC component (*U*). The trajectory of the ion can be defined by Mathieu equations and Mathieu parameters a_u and q_u used to characterize the amplitudes of both components are described by Eq. 1.3.

$$a_u = \frac{8zeU}{m\omega^2 {r_0}^2} \qquad q_u = \frac{4zeV}{m\omega^2 {r_0}^2} \tag{1.3}$$

Where *e* is the charge of an electron, z is the number of charges, ω is the angular frequency of the RF component, and r_0 is the half distance because two opposite rods. Under normal operational mass analysis, only ions within a narrow m/z window and stable trajectories are transmitted corresponding to a_u = 0.706 and q_u = 0.237.[63] In the RF-only mode, the DC component is zero (a_u =0) and the quadrupole simultaneously rejects ions with m/z values lower than a certain cut-off value. Generally speaking, a quadrupole is a simple and low-cost instrument. Related devices are hexapole and octapole which are also used as ion guides and collision cells. The principles of the hexapole and the octapole are similar to the quadrupole, but the efficiencies of the hexapoles

and the octapoles are higher in ion transmission within the extended mass range at the expense of focusing power.

The other mass analyzer in Q-TOF-MS is a reflecting TOF mass spectrometer that measures ions based on the flight time that ions spend moving through the TOF tube. The essential principle of TOF is when a high voltage pulse (V_s) is applied to an ion with charge z, the electric potential energy $E_p = zeV_s$ is converted into the kinetic energy $E_k = ½mv^2$ which accelerates the ion to move forward to the detector. If the length of the TOF tube (L) is equal to the distance that the ion has to travel, the two equations can be arranged into Eq. 1.4 where the fight time (t) of an ion is proportional to the square root of the ion m/z.

$$(m/z)^{1/2} = \left(\frac{\sqrt{2eV_s}}{L}\right) t \qquad (1.4)$$

TOF-MS is direct and simply compatible with MALDI and other pulsed ion sources because the pulsed character of these ion sources provides a precisely defined ionization time and ionization region. However, for the continuous ion sources (e.g. ESI), it is a challenge to couple TOF-MS with them.[63] In general, the resolution of TOF-MS can be affected by initial spatial distribution, and initial ion kinetic energy distribution due to the direction of flight tube. Nowadays, these two limitations are avoided by the technique of orthogonal acceleration. Orthogonal acceleration not only changes the flight direction of ions which is orthogonal to the initial direction in which ions enter the TOF tube, but also is used as a pulsed ion sources; therefore, the flight time is independent of initial kinetic energy and spatial distribution of ions. Thus, this technique has been used to couple continuous ion sources with TOF-MS. Besides, an ion reflector, also called an ion mirror, is an ion optic device in which ions pass through a reflector and their flight is reversed (see Figure 1-11). The use of a reflector is to compensate for the initial energy distribution and to focus ions having the same m/z prior to the detector. Normally, the polarity of a potential applied on the

reflector is the same as that of the detected ions. As ions with the same m/z fly into the reflector, ions with greater kinetic energies penetrate deeper into the field and take a longer time to return than ions with smaller kinetic energies. Accordingly, the ions with high kinetic energies and smaller kinetic energies reach the detector at the same time and improve the spectrum resolution. For example, the resolution of Q-TOF-MS can easily achieve 50,000. Due to the high resolution of oaTOF-MS, the centroid of each peak can be accurately measured. In addition, calibration of TOF is fairly straightforward and accurate. According to Eq.1.4, the fight time of an ion is proportional to the square root of m/z of the ion. In principle, a mass-to-charge ratio window can be accurately calibrated by a simple two-point calibration.[63] Furthermore, Q-TOF-MS inherits properties from quadrupole and TOF-MS so other advantages of Q-TOF-MS include high ion transmission, MS/MS capacity, fast scan speed, and excellent mass accuracy as well.

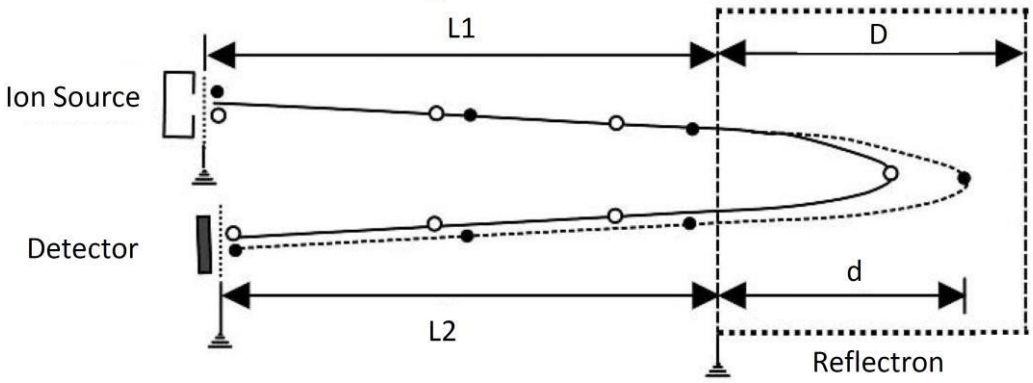


Figure 1-11 Schematic diagram of TOF-MS equipped with a reflectron. The black spots represent an ion with correct initial kinetic energy. The white spots are the ion of the same m/z but lower initial kinetic energy, so it spends shorter time in reflector than the black spot.

1.3 Microwave-assisted Extraction (MAE)

Microwave-assisted extraction is a method of using microwave dielectric heating to heat solvents in contact with a sample in order to partition analytes from the sample matrix into the solvent. Microwaves have great potential for heating material rapidly. Microwave ovens were invented after World War II, but they have only recently appeared in analytical laboratories. 1975 was the first time that microwave ovens were used in the laboratory by Abu-Samra *et al.* for performing trace analysis of metals from biological samples.[65] In the following years, microwave ovens were used for digestion methods of different sample types. The first publication of using microwave ovens on extractions appeared in 1986. Ganzler *et al.* presented the microwave-assisted extraction of organic compounds from biological samples and soil samples.[66] In the beginning of the 1990s, more and more research groups started to use microwave ovens for the purpose of extraction. Paré et al. patented a process for extracting natural products by microwave ovens in 1991.[67] In 1994, Lopez-Avila *et al.* presented the procedures of microwave-assisted extractions of organic compounds from standard reference soils and sediments.[68] This work was evaluated as a part of new sample preparation techniques by the US Environmental Protection Agency (EPA). Since then, numerous laboratories have investigated the analytical possibilities of microwave-assisted extractions in various applications. Nowadays, microwave-assisted extraction has become a relatively mature technique and standard protocols have been published.

Microwaves are a form of electromagnetic radiation located between the radio-frequency range at the lower frequency and infrared at the higher frequency in the electromagnetic spectrum, ranging from 300 MHz to 300 GHz.[69,70] Owing to the electromagnetic nature, microwaves are made up of two oscillating perpendicular fields, i.e. electric field and magnetic field. The electric field is responsible for heating which is a process that materials or solvents convert a part of the

absorbed electromagnetic energy to heat energy. The principle of heating using microwave is based on the direct effect of microwaves on polar materials or solvents via two phenomena named ionic conduction and dipole rotation.[71-74] The electrophoretic migration of ions under the influence of the changing electric field is referred to as ionic conduction.[75] The resistance of the solution to this flow of ions generates friction and then heats the solution. Dipolar rotation is due to realignment of dipoles of the molecules with the changing electric field.[76] This oscillation causes collisions with other molecules and then generates heat. With a frequency of 2450 MHz, used in commercial systems, the electric field component changes at the rate of 4.9×10^4 times per second.[77] There is a generation of heat through the friction force when the solvent molecules try to align themselves to keep in the same phase at a high speed changing electric field, but the molecules fail to realign themselves and start to vibrate. No heating is generated when the frequency is greater or less than 2450 MHz, and only dielectric materials or solvents with permanent dipoles would be heated up under microwave.[78] Therefore, the choice of solvents is an important factor of microwave-assisted extraction.

Microwave-assisted extraction has been developed and applied for extraction of different sample types. In this thesis, microwave-assisted extraction was mainly used to extract metabolites from different plants. In most cases, the target for microwave heating in dried plant materials is the minute microscopic trace of moisture occurring in plant cells. After plant cells heat up by microwave heating, evaporation of the moisture results in the generation of tremendous pressure on the cell wall. The cell wall is pushed from inside and ultimately ruptured due to the pressure. The active constituents from the ruptured cells release to the surrounding solvent thus improving the yield of phytoconstituents.[70] This yield can be further enhanced if the plant matrix is soaked in solvents with higher heating efficiency. The ether linkages of cellulose, which are the main constituent of plant cell wall, are hydrolyzed and converted into soluble fractions due to microwave

heating. Higher yield can also be obtained at higher temperature because solvents can access more easily into the cell.[79] The microwave-assisted extraction is a process that chemical substances within cells rapidly crack the cells and flow into the surrounding due to the sudden temperature rise and internal pressure increase.[80,81] Furthermore, the migration of dissolved ions not only increases solvent penetration into the plant matrix but also helps the release of chemical substances.[82] Most of the time, the sample is soaked in a single solvent or a mixture of solvents that absorb microwave energy strongly because the effect of microwave energy depends on the dielectric susceptibility of solvents and solid plant matrix.[83]

1.4 Differential Stable Isotopic Labeling and Derivatiztion for Metabolomics

The principle of the quantitation method used in LC-MS is not different from that used in many other analytical techniques. In analytical chemistry, comparison between the instrument response of a given sample and that of authentic standards can reflect the concentration of an analyte in a sample. There are three kinds of calibration methods often used in LC-MS based quantification. The first is the external calibration in which the signal response of an analyte is compared to a calibration curve developed in advance. Unfortunately, for complex biological samples, accuracy and precision of the quantification as well as detection capability for the analytes of interest could be severely affected by matrix effect and ion suppression.[84,85] Another method involves using standard additions in order to eliminate the matrix effect and a calibration curve is established by adding a series of known amounts of standards into the sample matrix. However, standard addition is very time-consuming due to the fact that a calibration curve for each analyte needs to be prepared. Thus, standard addition is commonly used for metabolic target analysis rather than metabolic profiling. Internal standard calibration is the third method and widely used in LC-MS-based quantitation. The theory is that the concentration of analytes in a

sample is determined based on the signal ratios between the analytes and the internal standards added to the sample. The analytical behavior of an internal standard is similar to that of the analyte of interest, but the signal of the internal standard can be distinguished from that of the analyte. Isotope analogues are the best choices because they almost show identical behaviors to the targeted analytes in sample preparation, RPLC, and ESI but different mass peaks in the spectra.[86,87] This method is referred to as stable isotopic labeling (SIL) dilution quantification which usually uses ^{2}H, ^{13}C, ^{15}N or ^{18}O to replace ^{1}H, ^{12}C, ^{14}N or ^{16}O in the analytes of interest, respectively. However, commercial SIL internal standards are expensive and limited so it is not practical to use SIL to study all the metabolites in a complex biological sample. To this end, a differential isotopic labeling (DIL) method has been developed and used in metabolomics.

A DIL method uses a chemical reaction to add an isotope tag to the analytes in one sample and another mass-different isotope tag to the same analytes in another comparative sample (or standards). Differentially labeled samples are mixed prior to mass spectrometric analysis, and mass shifts are observed in the mass spectrum. The peak abundance ratio of the isotope labeled analyte pair can reflect the relative quantification of the analytes in two comparative samples.[88] It is also possible to obtain an absolute quantification of analytes in a sample if the concentrations of analytes in the other sample are known. The accuracy and precision of quantification in DIL method is close to that in the isotope dilution quantification method, if the derivatization reaction yields in parallel light/heavy-isotope labeling experiments are similar. Compared to SIL, DIL has an important advantage in that a stable isotope labeled internal standard can be generated for every analyte. Therefore, DIL is a plausible approach for quantitative analysis in metabolomics. Compared to *in vivo* isotopic labeling approaches in living cells or organisms, DIL is more widely used in many different types of biological samples. More than 80% of metabolites contain at least one of the three functional groups: amine, organic acids, and carbonyl compounds.[89] Several

isotopic labeling reagents have been developed and used to label metabolites containing one of these three functional groups.[88,90,91] An early metabolomics study showed that the DIL method can be used to quantitatively analyze amino acids in urine and blood samples, by the use of isobaric tag for relative and absolute quantitation (iTRAQ) reagent, originally developed to label peptides for quantitative proteomics.[92] Yang et al. developed a labeling method involving derivatization with an N-hydroxysuccinimide ester of N-alkylnicotinic acid where deuterium replaced hydrogen in the alkyl chain to give a differential isotope tag.[93] The use of ^{13}C-/^{12}C-methylation for relative quantification of flavonoids was described by Fukusaki et al.[94] The use of ^{13}C-/^{12}C-based tags for differential isotope labeling of amine-containing metabolites was reported by Shortreed et al.[90] Guo et al. from our research group have developed a series of ^{13}C-/^{12}C-based differential isotope labeling methods for the analyses of amine-, phenol- and carboxyl-containing metabolites.[88,95]

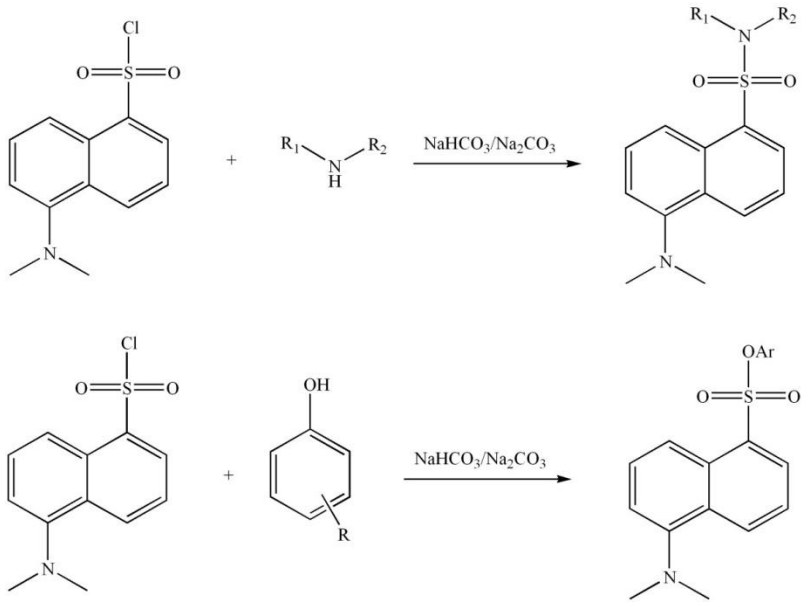


Figure 1-12 Reaction schemes for dansylation of primary/secondary amines and phenolic compounds.

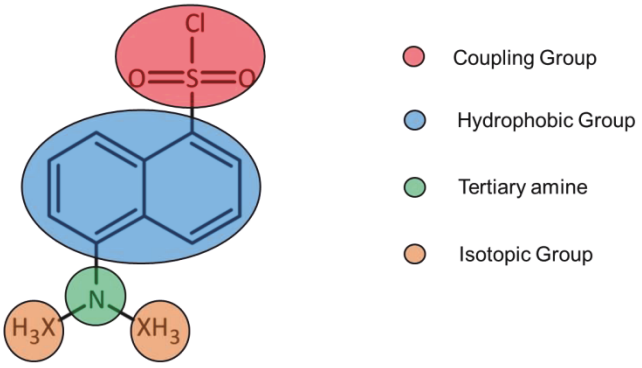


Figure 1-13 Dansyl molecule with highlighted chemical features.

In the approach used by Guo, et al., ^{13}C and ^{12}C dansyl chloride were used as derivatization reagents for differential isotopic labeling. Dansyl chloride is commonly used to label compounds containing amine and phenol functional groups (Figure 1-12). Dansylation, a simple and robust reaction, has been used for many years as pre-column LC derivatization for quantification of amino acids, biogenic amines and phenolic hydroxyls, as well as for signal enhancement in UV and fluorescence detection.[96-98] The dansyl label contains a hydrophobic naphthalene group (Figure 1-13) which allows relatively polar analytes to be more retained in RPLC. The elution of dansylated analytes is accomplished at a higher proportion of organic mobile phase. The higher percentage of organic modifier in the eluent helps to increase the ionization desolvation efficiency and the electrospray stability, thereby improving analyte response in ESI. In addition, the droplet surface affinity of the dansylated analyte is increased due to the hydrophobic naphthalene group, which in general provide higher ESI response.[99] Besides, the tertiary amine group also plays an important role in improvement of ESI reponse. It assists in the chargeability of the dansylated analyte and makes the dansylated analytes more favorable in the competition of the limited amount of charges within the ESI spray droplet. The improvement of ESI response is also expected because the naphthalene conjugation structure stabilizes the protonated charge at the tertiary amine group.

Furthermore, the mass of a dansylated analyte shifts from a low-mass region into a high-mass region that has less significant background noise from solvent clusters and common contaminants. As a result, these enhancements improve the coverage of phenol and amine containing metabolites and the confidence of metabolomics analysis.

1.5 Overview of Thesis

This thesis focuses on the development of LC-MS methods for studying amine- and phenol-containing metabolome with differential isotope labeling, as well as the related data processing software to quantify and qualify metabolites in biological systems. In the chemical isotope labeling or isotope coded derivatization LC-MS metabolomics platform, metabolites are typically detected as peak pairs with a defined mass difference governed by the mass difference of heavy- and light-isotope tags. How to deal with a large and unique dataset is a major issue for data processing. Chapter 2 describes the principle and preliminary results of software, IsoMS, which can rapidly handle our DIL method datasets and provide the information for quantification as well as identification. Chapter 3 demonstrates that dansylation of crude plant extracts could be applied to the rapid and sensitive analysis of plant metabolites as part of an LC/MS metabolomic profiling workflow. Flax bast (phloem) fibers were choosen because they are a readily available source of plant tissues of both scientific and economic relevance. The genome of flax was recently sequenced. Research on the structure and composition of flax fibers is accelerating due to their high tensile strength and potential for use in advanced materials. In addition to flax fibers, ginseng roots have great economic value as a source of medical components used in traditional Chinese medicines (TCMs). Chapter 4 presents the development of a robust metabolite extraction method from ginseng roots tailored to the analysis of amine- and phenol-containing metabolites, the application of the dansylation labeling method to gauge the detectability of the LC-MS technique

for the ginseng metabolome, and the study of spatial distribution of metabolites in ginseng roots. Besides studying the metabolome of plants with high economic value, studying the metabolic information to connect genotypes and phenotypes is important as well. Chapter 5 focuses on the development of metabolite extraction and LC-MS protocols for studying the metabolome of *Arabidopsis thaliana* by applying the $^{12}C/^{13}C$ dansylation labeling. These methods were used to study metabolic changing trends of different *TIFY* gene expression plants under methyl jasmonate treatment. In Chapter 6, the sensitive dansylation LC-MS method has been applied not only to study the metabolome profile of Goji but also to analyze the urine metabolomes of healthy individuals with and without drinking Goji tea to determine whether drinking Goji tea would have any effects on their urine metabolome profiles. Finally, Chapter 7 outlines the conclusions of my thesis and directions for future work.

Chapter 2

IsoMS: Automated Processing of LC-MS Data Generated by a Chemical Isotope Labeling Metabolomics Platform*

2.1 Introduction

Using an isotopic analog of a compound of interest as an internal standard is widely adopted for LC-MS-based chemical quantification, because of its high accuracy and precision. However, due to limited availability of isotopic analogs of metabolites, this approach is not feasible for metabolomic profiling. Metabolic or *in vivo* stable isotope labeling can be used to generate isotopic analogs of metabolites, which has been combined with LC-MS for studying the dynamics of metabolic fluxes.[100,101] However, it requires cell culturing and thus can't be readily applied for analyzing non-culturable samples (e.g., human biofluids). Alternatively, chemical derivatization can be used to introduce an isotope-coded mass tag to the metabolites that share a common functional group, such as amine, acid, phenol, etc. In this approach, a group-based sub-metabolome is labeled with a proper isotope labeling reagent for LC-MS analysis and the combined results of various sub-metabolomes represent a comprehensive profile of the whole metabolome. There are a number of isotope labeling reagents reported for labeling metabolites.[88,90,93,95,102-114] Using rational design of an isotope labeling reagent structure, simultaneous improvement in LC separation and MS detection, in addition to providing a mass tag, can be achieved to generate a more comprehensive profile of the sub-metabolome of a biological system, compared to conventional LC-MS methods.[88,95,115,116] Although several robust software tools have been developed for processing LC-MS data for metabolomics,[117-120] to our knowledge, there is no software available to handle the unique datasets generated by the chemical isotope labeling or

isotope coded derivatization (ICD) LC-MS metabolomics platform. We also found that software tools used for quantitative shotgun proteomics involving differential isotope labeling of peptides are generally not useful for processing the metabolomics data. Adapting these tools for chemical isotope labeling metabolomics likely requires modifying the codes to accommodate the special needs of dealing with low mass and unidentifiable analyte ions.

We have developed a data processing method and a software tool, IsoMS, for processing LC-MS data by exploring a mass spectral feature unique to the chemical labeling approach, i.e., metabolites are detected as peak pairs with a fixed mass difference in a mass spectrum. In a typical workflow of the chemical isotope labeling metabolomics platform, an individual sample is labeled with a light-chain reagent and a pooled sample produced by mixing aliquots of individual samples is labeled with a heavy-chain reagent. The mixture of the light-chain labeled individual sample and the heavy-chain labeled pooled sample is analyzed by LC-MS. A true metabolite should be detected as a pair of peaks with a defined mass difference governed by the mass difference of the heavy- and light-chain tags, while all the noises and background peaks are detected as singlet peaks. The intensity ratio of a given peak pair is reflective of the relative concentration of the metabolite in an individual sample to that in the pooled sample. Thus, the peak ratio values, along with the metabolite information (retention time and peak mass), can be directly exported into a data or statistical analysis program for further analysis.

In this chapter, I describe how the data processing method was implemented in IsoMS, the procedure for processing chemical isotope labeled LC-MS data, and the performance of the method for analyzing dansyl labeled metabolites and human urine samples. All the scripts and instructions including a user manual can be downloaded from www.mycompoundid.org/IsoMS.

2.2 Experimental

2.2.1 Reagents and Chemicals

$^{12}C_2$-Dansyl chloride, amine standards including 7-hydroxy-coumarin, tryptamine, phenol, D/L-tryptophan and N-methyl-D-phenylalanine, as well as formic acid were all purchased from Sigma-Aldrich (Oakville, ON). ^{13}C-dansyl chloride was synthesized in our lab and the chemicals used to synthesize this isotope reagent were also purchased from Sigma-Aldrich. Stock solutions of amino acid standards (10 mM each) were prepared in H_2O: ACN (50:50 (*v/v*)) and stored at 4°C. LC/MS-grade water, and acetonitrile (ACN) were purchased from Thermo Fisher Scientific (Edmonton, AB, Canada).

2.2.2 Urine Sample Collection

The second morning urine sample was collected each day from a healthy volunteer over five consecutive days. In each day, one urine sample before drinking coffee and two urine samples 1 or 2 h after drinking coffee were collected. The urine sample was centrifuged at 4,000 rpm for 10 min. The supernatant was filtered twice by 0.22 μM-pore-size Millipore filter (Millipore Corp., MA) and aliquoted into 0.5 mL vials. The urine samples were stored at -80 °C. A pooled urine sample was prepared by combining the aliquots of daily urine samples. The individual urine sample was derivatized by $^{12}C_2$-dansyl chloride (DnsCl) and the pooled urine sample was derivatized by $^{13}C_2$-DnsCl, according to the reported protocol.[116] An aliquot of the $^{12}C_2$-labeled individual sample was mixed with an aliquot of the $^{13}C_2$-labeled pooled sample and the mixture was then analyzed by LC-MS in triplicate.

2.2.3 LC-MS Analysis

For the urine metabolome profiling work, the LC-MS analysis was performed using an Agilent 1100 series binary system (Palo Alto, CA) connected with a Bruker Apex-Qe 9.4-T Fourier transform ion cyclotron resonance (FT-ICR) MS (Bruker, Billerica, MA) equipped with electrospray ionization (ESI). LC separation was performed on an Agilent Zorbax Eclipse Plus C18 column (2.1 mm × 100 mm, 1.8 μm particle size). The experimental setup conditions were the same as those reported elsewhere.[116] This method was also tested on LC-MS data generated by Bruker Maxis quadrupole time-of-flight (QTOF).

2.3 Results and Discussion

2.3.1 Workflow of IsoMS

Figure 2.1 shows the overall workflow for processing LC-MS data generated by a chemical isotope labeling metabolomics platform. IsoMS can be used to process data generated from different MS instruments, as long as they provide adequate resolving power to resolve the peaks within an isotope peak pair. As Figure 2.1 shows, there are generally three steps involved in data processing. The first step is to convert the raw LC-MS data into a centroid peak list file that contains information on retention time, m/z and peak intensity. An example of the peak list file for Bruker FT-ICR-MS data is shown in Table 2.1. All mass peaks with a user defined intensity threshold are retained in the peak list file. The peak list file also contains information on signal-to-noise ratio and peak width of a mass peak, which is useful for removing noisy spectra. This conversion can be done using the data analysis software of the manufacturer. The instruction (see Appendix A) and a script for Bruker data for converting the Bruker data can be downloaded from the MyCompoundID website. While there are software tools available for extracting and aligning peaks, IsoMS uses the raw peak list data as the starting point for further data processing. This

should provide a more uniform approach in dealing with chemical isotope labeling metabolome dataset, independent of any 2nd party software tools which may extract or align peaks differently from one software to another.

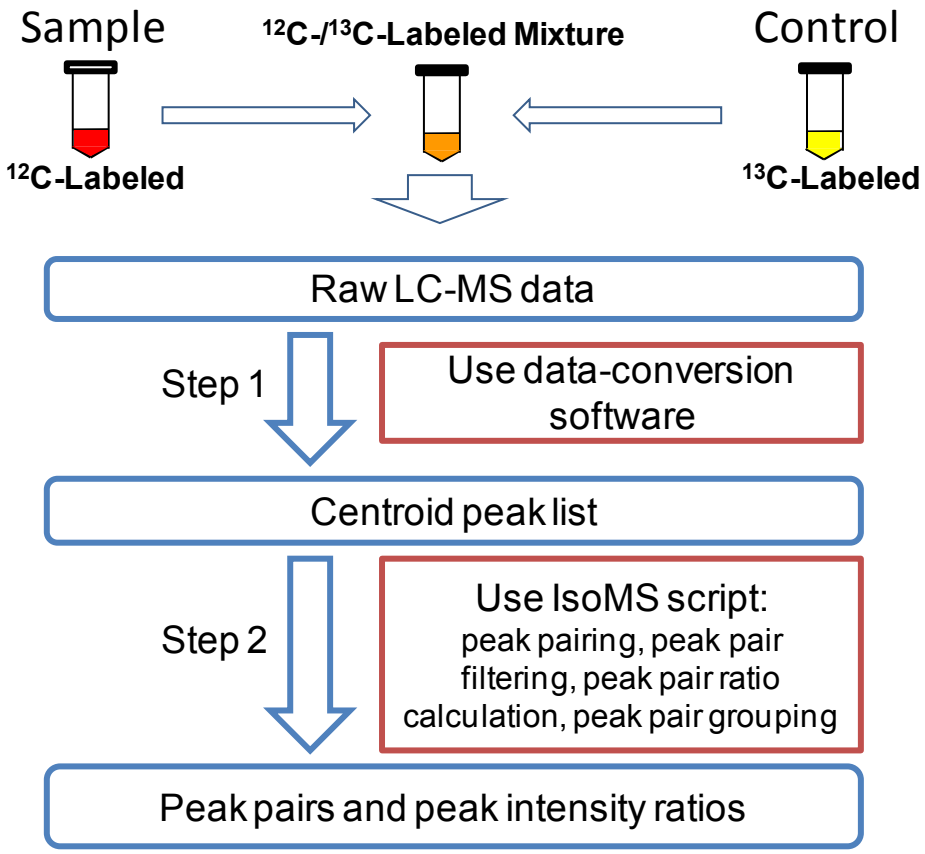


Figure 2.1 Workflow for IsoMS data processing.

Table 2.1 Partial list of a peak list file in CSV converted from the raw LC-MS data generated from Bruker FT-ICR-MS analysis of a $^{12}C_2$-/$^{13}C_2$-dansyl labeled human urine sample.*

sample	No	rt	mz	maxo	sn	Width
21	15	51.359	254.0756	28983.38	12.62992	3.74E-03
21	15	51.359	261.1306	33546.16	14.84752	4.70E-03
21	15	51.359	276.0572	31296.74	13.52674	5.39E-03
21	15	51.359	305.1571	26383.92	10.7661	0.005794
21	15	51.359	328.1002	40392.86	17.29376	6.57E-03
21	15	51.359	330.107	42637.58	18.3649	6.66E-03
21	15	51.359	338.3416	65853.51	29.31196	7.42E-03
21	15	51.359	371.3154	64699.14	27.61914	8.97E-03
21	15	51.359	391.2842	47467.85	19.3515	8.89E-03
21	15	51.359	393.8634	37584.48	14.84753	9.40E-03

*The column headings are: sample - sample identifier; No - scan number; rt - retention time (s); mz - m/z ratio; maxo - peak intensity counts; sn - signal-to-noise ratio; width - mass spectral peak width.

After the file conversion, all the peak list files from multiple LC-MS runs are placed in one folder, which are then automatically processed by IsoMS (i.e., Step 2 in Figure 2.1) to generate the CSV files containing information on sample identifier, retention time and m/z's of individual peak pairs, and peak ratio. An example is shown in Table 2.2 for the urine metabolome profiling work. The IsoMS algorithm processes the peak list data in the following manner.

Table 2.2 Partial list of a representative CSV file (displayed in Excel) generated by IsoMS processing of one LC-MS dataset.*

sample	Scan_ Num	rt	mz_light	mz_heavy	distance	int_light	int_heavy	ratio_light _to_heavy	sn_light	sn_heavy	nCharge	nTag	pairLev	isSaturated
1	35	122.33	444.0902	446.0961	2.005897	1.34E+05	1.80E+05	0.74	53.0542	71.9743	1	1	2	0
1	35	122.33	493.1053	495.1121	2.006834	3.99E+05	3.54E+05	1.13	156.3919	138.3566	1	1	1	0
1	36	125.93	405.1233	407.1300	2.006693	6.70E+05	6.40E+05	1.05	275.1115	262.507	1	1	1	0
1	38	133.13	478.1285	480.1356	2.007161	2.28E+05	3.05E+05	0.75	85.7736	115.3121	1	1	1	0
1	38	133.13	521.1709	523.1781	2.007157	4.24E+05	1.31E+05	3.24	162.9613	49.005	1	1	1	0
1	40	140.34	387.0685	389.0750	2.006579	6.43E+06	1.09E+07	0.59	2569.811	4374.875	1	1	1	0
1	40	140.34	501.1156	503.1248	2.009203	1.90E+05	8.64E+04	2.20	73.585	32.4479	1	1	2	0
1	41	143.93	418.1183	420.1252	2.006932	8.12E+05	7.29E+05	1.11	328.8194	295.2056	1	1	1	0
1	41	143.93	449.1150	451.1218	2.006834	3.89E+05	3.63E+05	1.07	156.0207	145.1485	1	1	1	0

*Scan_Num - scan number; rt - retention time (s); mz_light - m/z value of the light-labeled metabolite; mz_heavy - m/z value of the heavy-labeled metabolite; distance - mass difference between the heavy- and light-labeled metabolites within a pair; int_light - peak intensity of the light-labeled metabolite; int_heavy - peak intensity of the heavy-labeled metabolite; ratio_light_to_heavy - peak intensity ratio of the light-labeled metabolite and the heavy-labeled metabolite; sn_light - signal-to-noise ratio of the light-labeled metabolite; sn_heavy - signal-to-noise ratio of the heavy-labeled metabolite; nCharge - number of charge; nTag - number of tag attached to the metabolite; pairLev - level of the peak pair detected; isSaturated - whether one of the peaks within a pair is saturated or not (0=no; 1=yes)

2.3.2 Peak Pairing

IsoMS determines the charge state and isotope distribution of each ion in a mass spectrum. Two ions having a user-defined mass difference (e.g., 2.0067 for $^{12}C_2/^{13}C_2$-dansyated metabolites) within a tolerance window (< 5 ppm) are then paired to each other, if they have the same charge and isotopic pattern (Figure 2.2A). They are grouped and defined as a Level 1 peak pair. However, for low intensity peaks, one of the natural isotope peaks in the peak pair may be missing (Figure 2.2B). IsoMS detects and classifies them as Level 2 peak pairs. For very low abundance peaks, neither the natural isotope peaks within a peak pair is detected (Figure 2.2C). These ions are paired by IsoMS as Level 3 peak pairs. They are not as reliable as Levels 1 and 2 pairs, as the chance of random matches of two adjacent peaks increases significantly; in some cases, as high as 30% of these pairs were found to be false. In our current work, we only use Levels 1 and 2 peak pairs for further analysis, which usually gives less than 5% false positive rate (FPR) as determined by manual inspection of the peak pairs found.

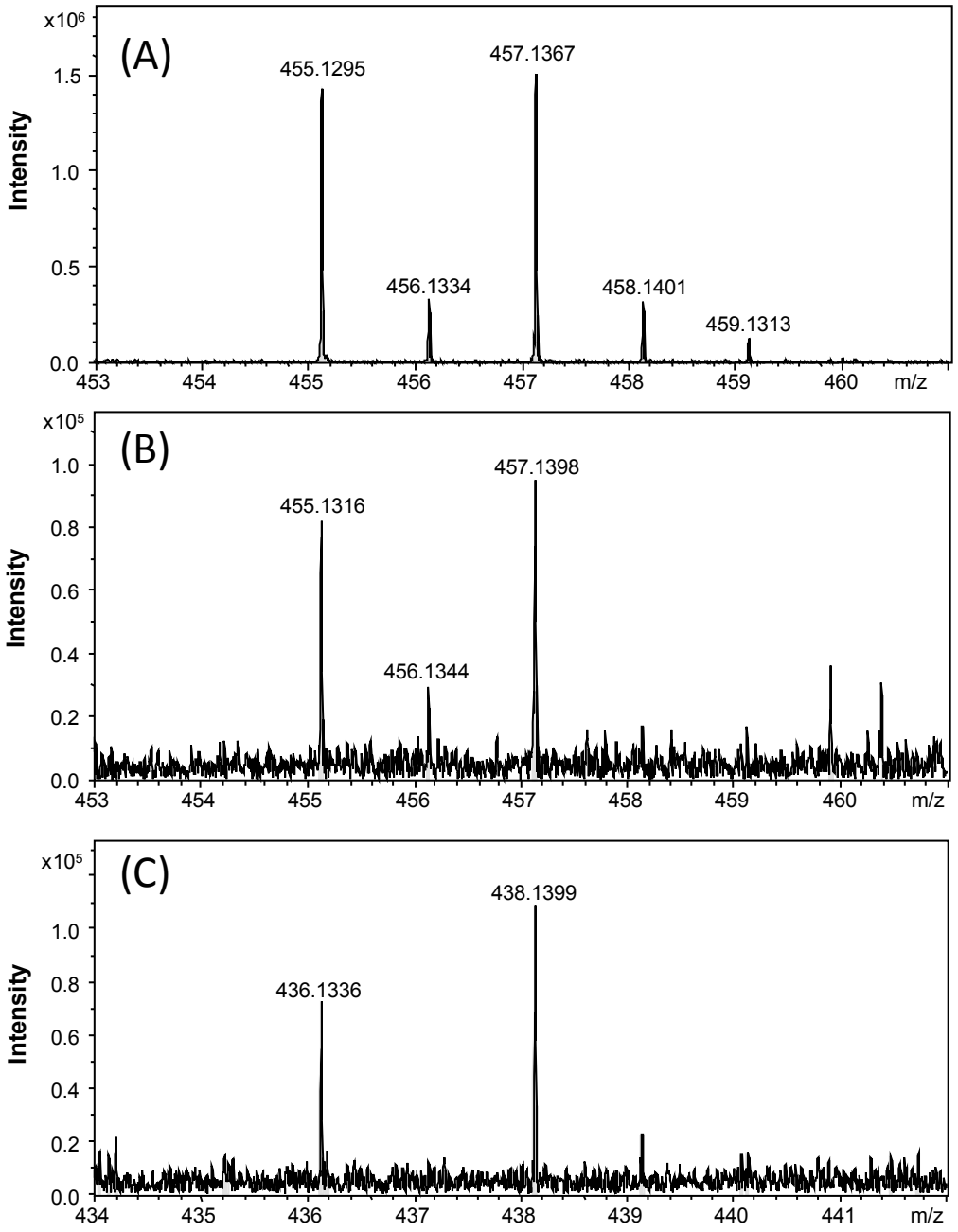


Figure 2.2 Mass spectra showing (A) Level 1 peak pair, (B) Level 2 peak pair, and (C) Level 3 peak pair.

2.3.3 Peak Pair Filtering

A labeled metabolite with a purposely designed tag, such as a dansyl group, is often detected in LC-MS in the form of MH^+. However, other forms, including adduct ions from Na^+, K^+ or NH_4^+, dimers and other multimers, and in-source fragment ions (usually after CO_2 loss), can sometimes be detected, particularly when the labeled metabolite concentration is high. Fortunately, these ions are detected within the same mass spectrum as that of MH^+ and thus can be readily filtered out after peak pairing. It should be noted that, if a chemical labeling reaction produces a labeled metabolite that forms MH^+ plus relatively high abundant adduct ions in a mass spectrum, filtering out the adduct ions may affect the quantification accuracy. In this case, MH^+ can still be used for the initial quantification. However, in the final result, one may use the summed area of the MH^+ and adduct ion peaks for quantifying the significant metabolites in comparative samples. IsoMS also allows a user to remove any peak pairs deemed to be from the background or blank. This can be done by entering the peak masses to be removed in a CSV file, isMZBackground. For FT-ICR-MS data, some noisy spectra with much reduced resolution are occasionally acquired (up to 15 or ~2.5% spectra in one LC-MS run), likely due to space charge issue related to ion detection (see an example in Figure 2.3). IsoMS has a filter to remove these noisy spectra based on the comparison of peak widths (resolutions) of a noisy spectrum to the neighboring spectra.

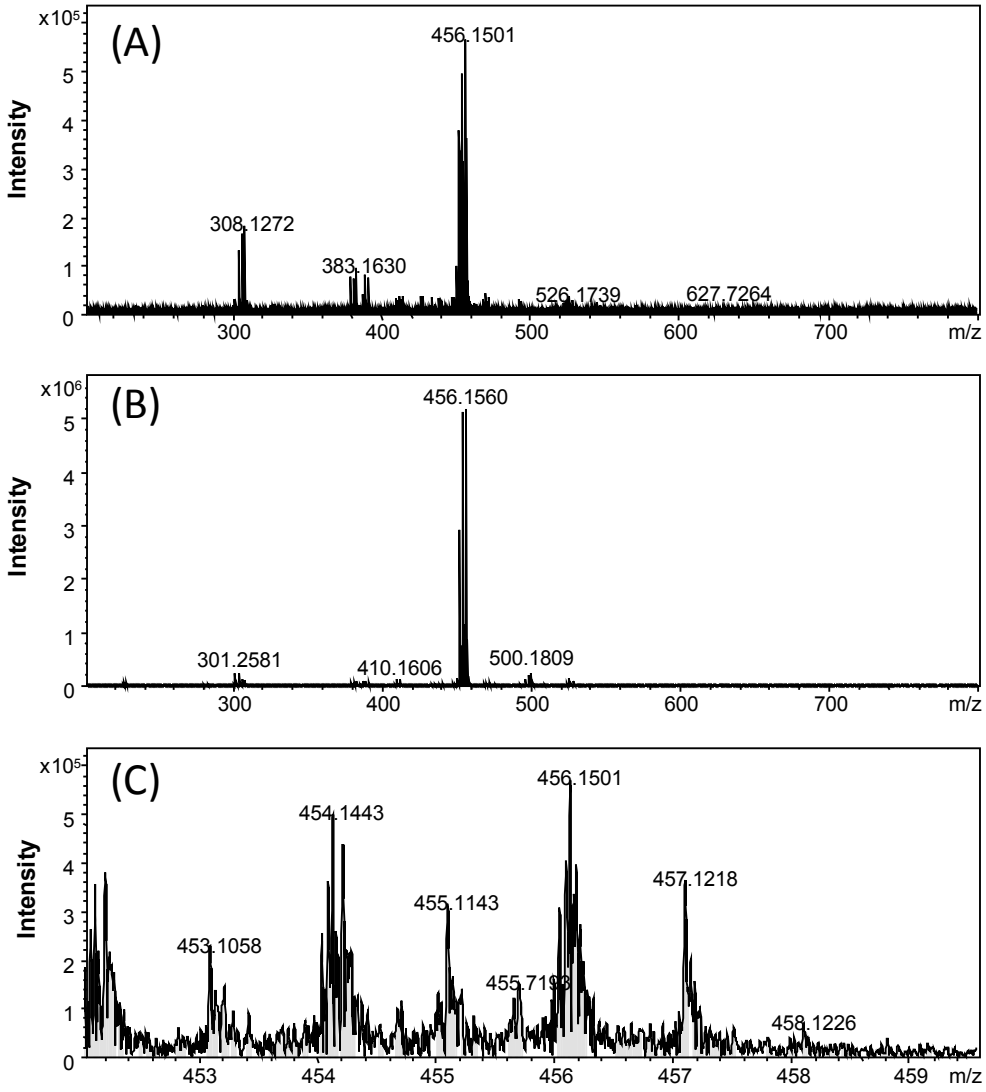


Figure 2.3 (A) noisy mass spectrum, (B) neighboring normal mass spectrum, (C) expanded region of peaks in the noisy mass spectrum. (A) and (C) show an example of noisy spectra found in LC-FT-ICR-MS analysis of 44etabolomics samples. This type of noise is likely caused by space charge problems in FT-ICR ion detection. They are usually detected in between two normal spectra, such as that shown in (B). Since a noisy spectrum contains many broad low intensity peaks (see the expanded spectrum in (C)), it can be readily filtered out based on their peak widths.

2.3.4 Peak Pair Grouping

After peak pair filtering to retain the only MH^+ peak pair from a metabolite, the intensity ratio of the light- and heavy-chain labeled peaks can be calculated in each mass spectrum. Thus, IsoMS can be used to determine the relative concentration of a labeled metabolite even when it only shows up in one spectrum, which can happen for low abundance analytes detected at a limited spectral acquisition speed, during the LC-MS run. However, many metabolites are repeatedly detected in several spectra. IsoMS groups the peak pairs found in adjacent spectra according to the light-chain labeled metabolite mass with a user-defined m/z tolerance window (e.g. 10 ppm). The peak ratio and retention time of the most intense peak pair with a group are retained. Note that IsoMS does not correct for natural isotope abundance in calculating the peak ratio. In general, for low mass metabolites (e.g., <800), the peak intensity contribution of the natural abundance $^{13}C_2$ peak of the $^{12}C_2$-labeled metabolite to that of the $^{13}C_2$-labeled metabolite peak is usually less than 10%, which is much less than the threshold value used to call for a significant metabolite concentration change (e.g., 1.5- or 2-fold change).

2.3.5 Validation

For each LC-MS run, IsoMS generates a list of m/z values of peak pairs, retention times and peak ratios of individual pairs as a CSV file (an example is shown in Table 2.1). In using IsoMS for processing LC-MS data, several parameters in the scripts can be fine-tuned, depending on the experimental conditions used. This is useful for optimizing the sensitivity and specificity for peak pair detection and peak ratio calculation. In our experience, after the initial adjustment of the parameters, there is no need to perform any further adjustment. The performance of the IsoMS method can be examined using two test samples. In one test, five standards, 7-hydroxy-coumarin, tryptamine, phenol, tryptophan and N-methyl-D-phenylalanine, labeled by $^{12}C_2$-/$^{13}C_2$-dansyl

chloride were analyzed by Bruker LC-FT-ICR-MS. Figures 2.4, Figure 2.5, and Table 2.3 show the ion chromatogram, a peak pair, and the final result obtained by IsoMS, respectively. IsoMS was able to pick the expected peak pairs correctly.

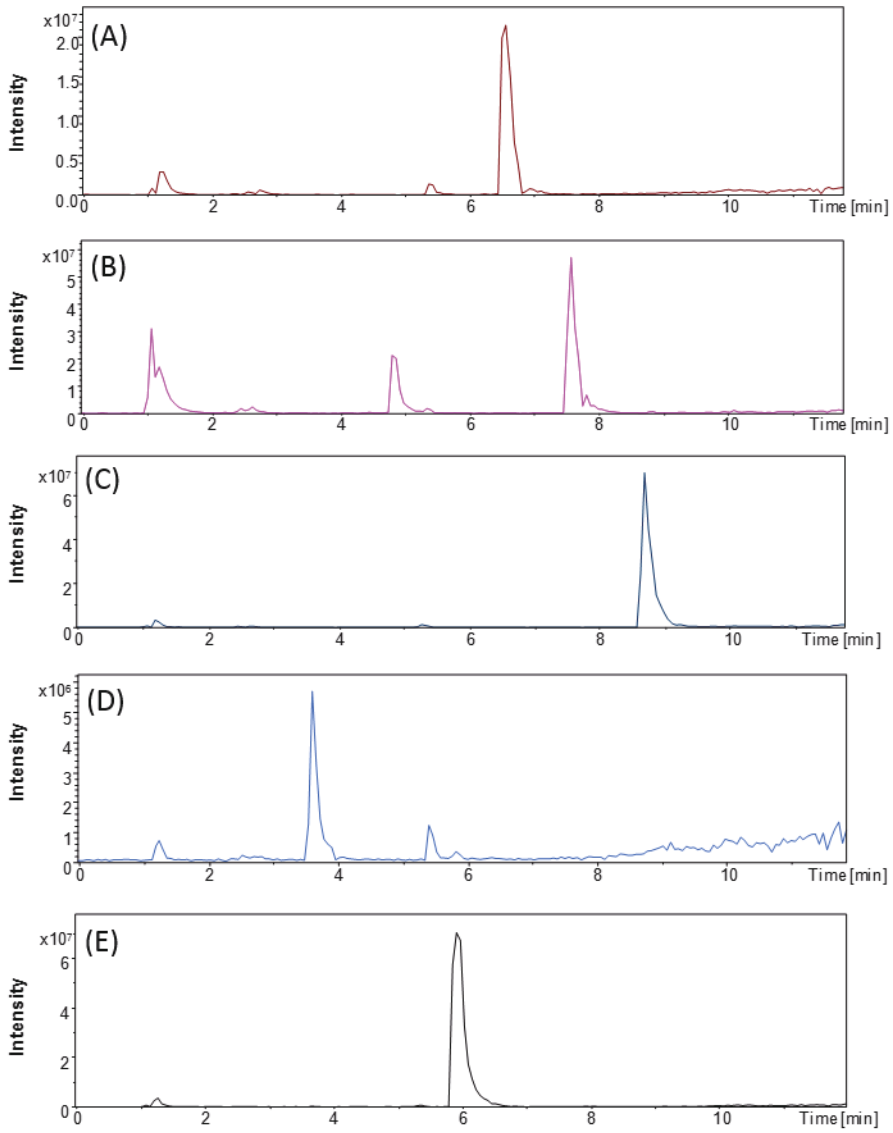


Figure 2.4 Base peak ion chromatograms obtained by LC-FT-ICR-MS analysis of individual $^{12}C_2$-/$^{13}C_2$-dansylated standards: (A) 7-hydroxycoumarine and its hydrolyzed by-product, (B) tryptamine, (C) phenol, (D) tryptophan and € *N*-methyl-*D*-phenylalanine.

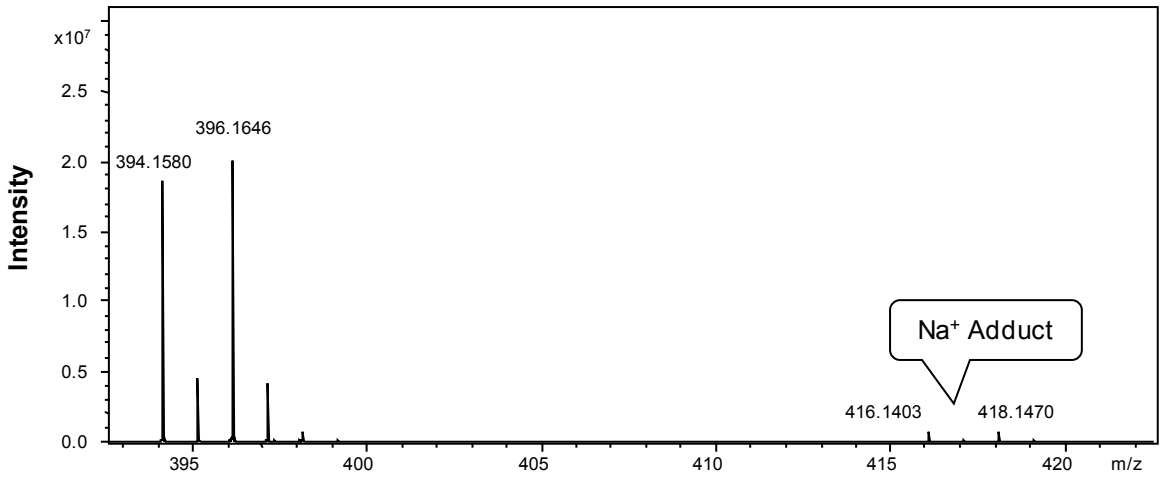


Figure 2.5 Mass spectrum of tryptamine showing a peak pair from MH^+ and a peak pair from the Na^+ adduct ions. IsoMS removes the adduct ion peak pairs and only retains one peak pair (MH^+) for one metabolite.

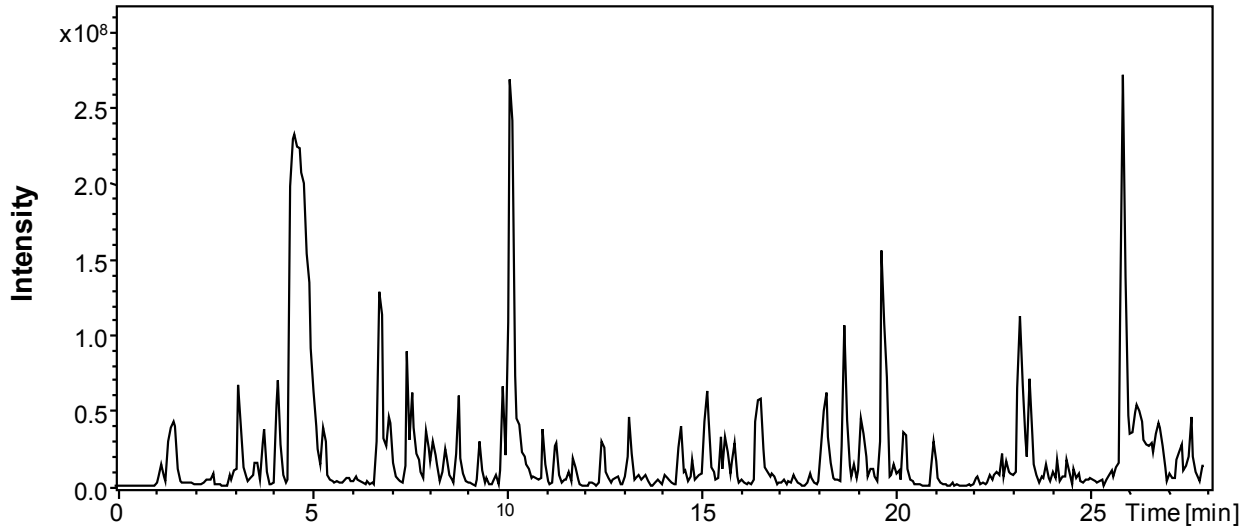


Figure 2.6 Base peak ion chromatogram of dansyl labeled human urine

Table 2.3 Final result from IsoMS processing of the LC-MS runs of five $^{12}C_2$-/$^{13}C_2$-danylated standards.*

sample	Scan Num	rt	mz_light	mz_heavy	distance	int_light	int_heavy	ratio_light to_heavy	sn_light	sn_heavy	nCharge	nTag	pairLev	isSaturated
Hydrolyzed 7-Hydroxycoumarin	81	294	414.1002	416.1069	2.0067	2.03E+07	2.13E+07	0.95	6753	7106	1	1	1	0
7-Hydroxycoumarin	127	459	396.0897	398.0963	2.0066	8.07E+07	8.29E+07	0.97	22012	22605	1	1	1	0
Tryptamine	110	397	394.1581	396.1647	2.0067	2.20E+07	2.39E+07	0.92	8514	9239	1	1	1	0
Phenol	146	530	328.1002	330.1067	2.0065	6.12E+07	7.08E+07	0.86	19854	22981	1	1	1	0
DL-Tryptophan	61	218	438.1479	440.1545	2.0067	4.34E+06	5.73E+06	0.76	1854	2449	1	1	1	0
N-methyl-D-Phenyalanine	99	357	413.1520	415.1585	2.0065	1.16E+08	1.19E+08	0.98	16410	16711	1	1	1	0

*Scan_Num – scan number; rt – retention time (s); mz_light – m/z value of the light-labeled metabolite; mz_heavy – m/z value of the heavy-labeled metabolite; distance – mass difference between the heavy- and light-labeled metabolites within a pair; int_light – peak intensity of the light-labeled metabolite; int_heavy – peak intensity of the heavy-labeled metabolite; ratio_light_to_heavy – peak intensity ratio of the light-labeled metabolite and the heavy-labeled metabolite; sn_light – signal-to-noise ratio of the light-labeled metabolite; sn_heavy – signal-to-noise ratio of the heavy-labeled metabolite; nCharge – number of charge; nTag – number of tag attached to the metabolite; pairLev – level of the peak pair detected; isSaturated – whether one of the peaks within a pair is saturated or not (0=no; 1=yes)

In another test, a differentially dansyl labeled human urine was analyzed and the ion chromatogram is shown in Figure 2.6. IsoMS detected 2044 peak pairs, and manual inspection of these peak pairs found 90 false peak pairs. Thus, the false positive rate was 4.4%, indicating that good specificity (<5%) could be obtained by IsoMS. Most of the false pairs had low signal intensities. Positive identification of the detected peak pairs is beyond the scope of this work.

2.4 Conclusions

We have developed a method based on peak pair picking and filtering in mass spectra to process LC-MS data generated by a chemical isotope labeling or isotope coded derivatization (ICD) metabolomics platform. This method has been implemented in a software tool, IsoMS, which can be freely downloaded from www.mycompoundid.org/IsoMS. IsoMS can be used to process the raw LC-MS datasets of multiple samples into a CSV file that can be readily exported for further data and statistical analysis. From the analysis of $^{12}C_2$-/$^{13}C_2$-dansyl labeled metabolite standards and human urine, we demonstrated that IsoMS is a reliable tool for processing chemical isotope labeling LC-MS data. We will continue to develop IsoMS by implementing other functionalities, such as automated statistical analysis and metabolome database search for compound identification.

Chapter 3

A Rapid and Sensitive Method for Plant Metabolomics Using Stable Isotope Dansylation: Demonstration in Plant Fibers

3.1 Introduction

Dansylation was described more than five decades ago as a derivatization method to improve the chromatographic separation of certain polar molecules, such as amino acids.[121] This procedure has been applied to the analysis of plant extracts since 1978.[122] Dansyl chloride reacts readily with primary amines, secondary amines and phenolic hydroxyls, all of which are found in many of the important metabolites within plant cells. In 2009, our group described the use of dansylation in combination with stable isotope labeling of urinary metabolites.[88] Differential isotope labeling with stable isotopes produces a mass difference that can be detected by mass spectrometry. Stable isotope labeling (SIL) and differential isotope labeling (DIL) techniques are applied to overcome matrix effect and ion suppression.[123] In our approach, crude extracts of metabolites are reacted with both ^{12}C or ^{13}C dansyl chloride and then analyzed by liquid chromatography mass spectrometry (LC-MS). The theoretical mass difference between a ^{12}C and ^{13}C dansyl label is 2.0067. Because the nonreactive metabolites or background ions do not produce $^{12}C/^{13}C$ peak pairs and thus do not have the specific mass difference between the peak pairs, dansylated metabolites can be readily filtered from background using IsoMS[124] and various in-house scripts.

Here I demonstrate, for the first time, that dansylation with stable isotopes can be applied to rapid and sensitive analysis of plant metabolites as part of a LC/MS metabolomic profiling workflow. We chose flax phloem (bast) fibers as a subject for analysis because they are a readily available source of plant tissues that is of both scientific and economic relevance: the genome of

flax was recently sequenced,[125] and research on the structure and composition of flax fibers is accelerating, owing to their high tensile strength and potential for use in advanced materials.[126] To date, most studies of flax fiber composition have focused on polysaccharides and lignin.[127,128] The procedure described here for optimization of extraction parameters and subsequent stable isotope labeling and analysis should be applicable to small amounts of starting tissue from a wide variety of tissues and species of plants.

3.2 Experimental

3.2.1 Reagents and Chemicals

^{12}C-dansyl chloride and amino acid standard solution were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). ^{13}C-dansyl chloride was synthesized in our lab[88] and the chemicals used to synthesize this isotope reagent were also purchased from Sigma-Aldrich. LC/MS grade water, acetonitrile (ACN), methanol (CH_3OH), acetone and formic acid were purchased from Thermo Fisher Scientific (Edmonton, AB, Canada).

3.2.2 Plant Material

Scutched, hackled, long-line flax fibers (i.e. phloem fibers) of field grown flax (*L. usitatissimum*) were obtained from an industrial scutching line operated by Terre de Lin (St Pierre Le Viger, France). Fibers have been stored at room temperature for over two years prior to metabolite extraction.

3.2.3 Microwave-assisted Extraction

Microwave-assisted extraction (MAE) was performed with a 1200 W domestic microwave oven (Panasonic, Toronto, ON, Canada). All microwave extractions were performed at 240 W power. Samples were transferred into polypropylene (PP) microcentrifuge tubes with o-ring seal

screw caps (Rose Scientific Ltd., Edmonton, AB, Canada) which were then placed in a PP plastic beaker with 200 mL water. The pressure gauge is absent for this type of domestic microwave oven. To determine the optimum solvent for extraction, 30 mg of flax fiber and 1 mL of solvent were mixed in a sample vial and microwave irradiated for 30 min. Four solvents, 80% ACN; 80% acetone; 80% CH_3OH and 100% H_2O, were tested and compared. For optimization of the sample-to-solvent ratio, different masses of fiber (20, 30, 40, 50, and 60 mg) in 1 mL of solvent were irradiated for 30 min. For optimization of extraction time, 50 mg of fiber and 1 mL of solvent were extracted for 15, 30, 45, and 60 min. In the above set of experiments, after the extraction, the extract was centrifuged at 14,000 rpm for 10 min to obtain the supernatant for futher analysis.

3.2.4 Ultrasonic-assisted Extraction

Ultrasonic-assisted extraction (UAE) was performed with a Branson ultrasonic cleaner 1510-MT (Branson Ultrasonics Corporation, Danbury, CT, USA). Ultrasonic extraction was conducted by mixing 30 mg of flax fiber and 1 mL of 80% methanol in a PP microcentrifuge tube with an o-ring seal screw cap, which was then placed in an ultrasonic bath at room temperature for 15 or 30 min. When extraction was completed, the extract was centrifuged at 14,000 rpm for 10 min to obtain supernatant.

3.2.5 Dansylation Labeling Reaction[88]

A 50 μL aliquot of supernatant from either the MAE or the UAE extraction was mixed with 25 μL sodium carbonate/sodium bicarbonate buffer (500 mM) and 25 μL ACN in a reaction vial. The samples were vortexed, spun down, and then mixed with 50 μL of freshly prepared ^{12}C-dansyl chloride solution (18 mg/mL) for light labeling or ^{13}C-dansyl chloride (18 mg/mL) for heavy labeling. The dansylation reaction was incubated for 60 min at 60 °C. After 60 min, mixtures were vortexed and then centrifuged, and NaOH (10 μL, 250 mM) was added to the reaction mixture to

consume the excess dansyl chloride and quench the dansylation reaction. After an additional 10 min of 60°C incubation, formic acid (50 μL, 425 mM in 50% acetonitrile) was added to consume the excess sodium hydroxide and to adjust pH value. The ^{13}C-labeled mixtures were combined with an equal volume of ^{12}C-labeled counterparts for LC-FT-ICR-MS analysis. The combined mixtures were centrifuged for 10 min at 14,000 rpm and then injected onto a RPLC column.

3.2.6 LC-MS Analysis and Data Processing

Analytes were separated on a HPLC system consisting of an Agilent 1100 liquid chromatograph (Palo Alto, CA, USA) equipped with a binary pump. The analytes were separated on a reversed-phase Eclipse plus C18 column (2.1 mm×100 mm, 1.8 μm particle size, 95 Å pore size) purchased from Agilent. Solvent A was 0.1% (v/v) formic acid in 5% (v/v) ACN, and solvent B was 0.1% (v/v) formic acid in ACN. The gradient elution profile was as follows: t = 0 min, 20% B; t = 1.0 min, 20% B; t = 16.5 min, 65% B; t = 19 min, 95% B; t = 28 min, 95% B; t = 30 min, 20% B. The flow rate was 180 μL/min, and the sample injection volume was 2.0 μL. The eluted solution from RPLC was split at a ratio of 1:3 before loaded to the electrospray ionization (ESI) source of a Bruker 9.4 Tesla Apex-Qe Fourier transform ion-cyclotron resonance (FT-ICR) mass spectrometer (Bruker, Billerica, MA, USA). All MS spectra were obtained in the positive ion mode.

The resulting mass spectral peaks were picked by using the Bruker DataAnalysis software 4.0 and an in-house visual basic (VB) script, and then processed by using an in-house peak-pair picking software, IsoMS[124], which was written in R language, for finding ^{12}C/^{13}C peak pairs with an accurate mass difference of 2.0067 Da. The IsoMS software also removes the redundant peaks of the same metabolite, such as natural isotopic peaks, common adduct ion peaks, dimer peaks, and multiply charged peaks. Only the protonated ion pair from a metabolite was exported for further analysis. For the comparsion of peak intensities, the extracted peak pair data were aligned

by retention time and accurate mass and analyzed by heatmap comparsion using Metaboanalyst[129] (www.metaboanalyst.ca). The data were then mean-centered and auto-scaled (unit variance) prior to analysis. The values of higher intensities were represented by red squares and the values of lower intensities by green squares. For further validation of compound identities, 40 different chemical standards were randomly chosen from HMDB[130] database search results, and they were derivatized by dansylation and then analyzed by LC-FTICR-MS in sequence after the flax samples. Retention time of the chromatographic peak and mass accuracy of the genuine standards were compared with those from flax samples.

3.3 Results and Discussion

3.3.1 Microwave vs. Ultrasonic Extraction

My objective was to demonstrate that dansylation of crude plant extracts could be used in a rapid and sensitive plant metabolomics workflow. For this purpose, we chose flax bast (phloem) fibers as a subject. These fibers were obtained from an industrial scutching line that separates bast fibers from surrounding stem tissues as an preliminary step in the production of linen. Dry, long-line flax fibers that had been stored at room temperature were used as input material for all metabolite profiling experiments described here. To optimize the metabolite extraction protocol for fibers, I varied solvent, incubation time, and sample:solvent ratio. Both microwave-assisted extraction (MAE) to ultrasonic-assisted extraction (UAE) were used to extract metabolites. All extracts were assayed by RPLC-FT-MS and the number of ^{12}C /^{13}C peak-pairs identified were used as a measure of the extraction yield.

I compared the efficacy of MAE and UAE of a 30 mg fiber sample in 80% methanol (Figure 3.1A). When the extraction incubation was limited to 15 min, MAE and UAE produced a similar number of metabolites. However, when the extraction time was increased to 30 min, the yield of

peak pairs increased for only the MAE method, and the yield was significantly larger for MAE (376±10) than for UAE (346±10). Because MAE appeared to be superior over UAE in these trials, I used only MAE in subsequent experiments.

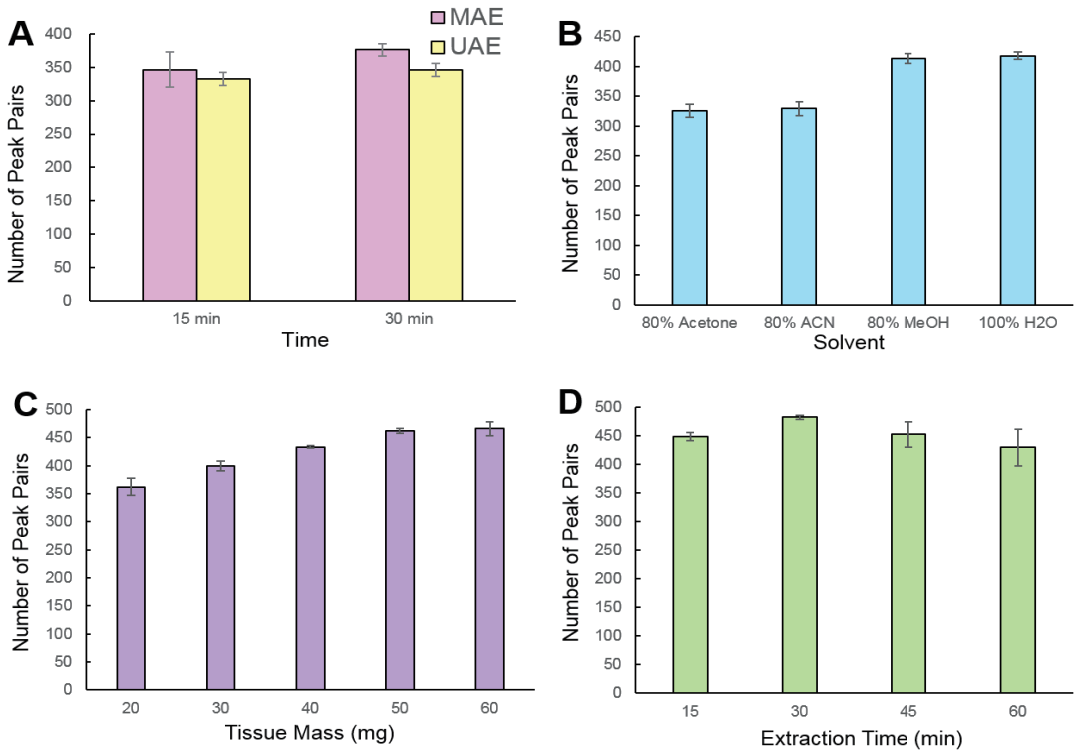


Figure 3.1 Impact of protocol optimization on analyte detection. The number of RPLC-FTMS peak-pairs identified in stable isotope dansylated extracts of fibers is shown for each of several different variations in the extraction protocol. Each experiment was repeated three times, and error bars show standard deviation. (A) Comparison of microwave assisted extraction (MAE) and ultrasonic-assisted extraction (UAE) at 15 and 30 min incubation time. (B) Comparison of four different extraction solvents. (C) Comparison of different sample:solvent ratios; each sample was extracted in 1 mL of solvent. (D) Comparison of different extraction times.

3.3.2 Comparison of Extraction Solvents

The choice of extraction solvent is an important consideration in metabolomics because different solvents can extract different analytes, and the solubility of analytes can affect extraction efficiency. I compared four solvents including 80% acetonitrile/water, 80% acetone/water, 80% CH_3OH/water and 100% H_2O. The result (Figure 3.1B) illustrated that the number of ^{12}C and ^{13}C labeled peak-pairs ranged from 325 to 417 pairs while the relative standard deviation (RSD) ranged from 1.44% to 3.70%. The maximum number of peak-pairs was obtained with either 80% CH_3OH or 100% H_2O. The relative intensities of same peak pairs among four solvents were presented in heatmap (Figure 3.2). In the heatmap, more red colored features indicate higher signal intensities of peak pairs and more green colored features indicate lower signal intensities of peak pairs. The heatmap showed that the intensities of most of the peak pairs extracted by 80% CH_3OH or 100% H_2O were higher than those by 80% acetonitrile or 80% acetone. Thus, 80% CH_3OH or 100% H_2O were chosen as the optimum extraction solvents.

To test the reproducibility of the extraction with different solvents, I conducted extractions in triplicate with either 80% CH_3OH or 100% H_2O. Each extraction started with 30 mg of tissue in 1 mL of solvent, with an incubation time of 30 min, using MAE. Each extraction was repeated three times and the list of ^{12}C and ^{13}C dansyl labeled peak pairs was produced by processing MS data with IsoMS. When 80% CH_3OH was used as an extraction solvent (Figure 3.3A), the total number of peak pairs was 443 and the number of peak pairs common between triplicate runs was 382 (86%). When 100% H_2O was used as an extraction solvent (Figure 3.3B), the number of total peak pairs was 474 and the number of common peak pairs between triplicate runs was 347 (73%).

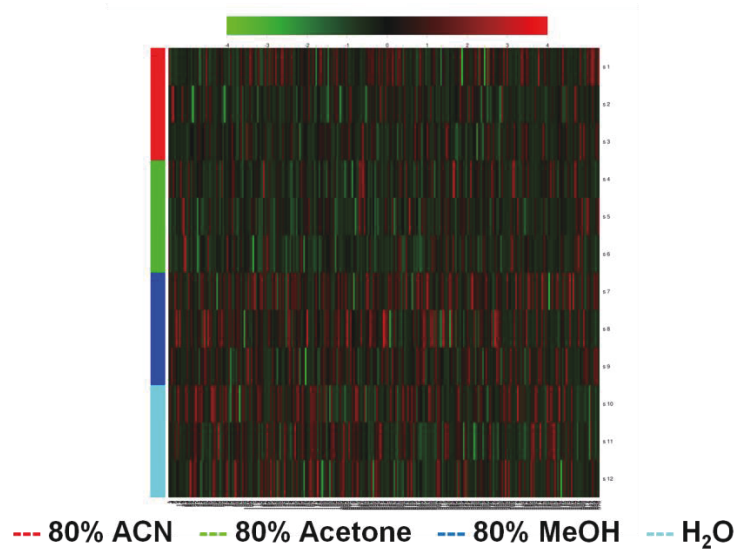


Figure 3.2 Comparsion of intensities of dansyl peak pairs. Heatmaps were shown comparison of the intensities of ^{12}C and ^{13}C labeled dansyl peak pairs in four solvent extraction methods. More red colored features indicate higher signal intensities of peak pairs.

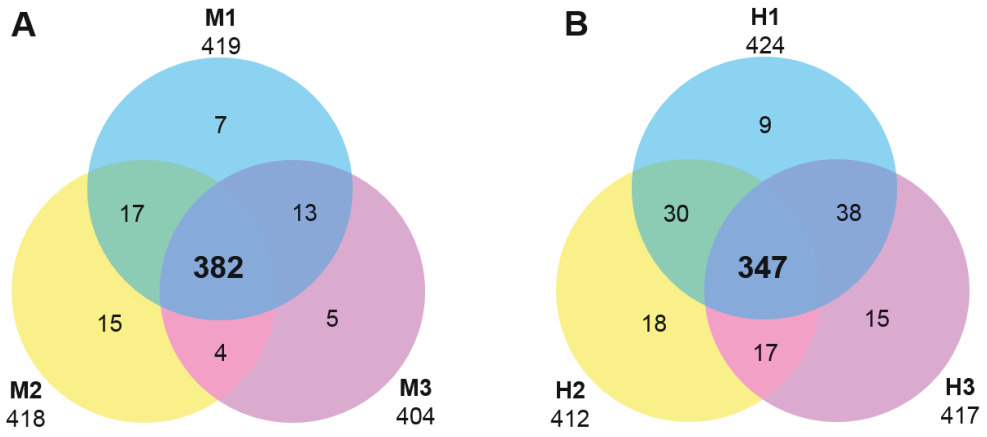


Figure 3.3 Reproducibility and overlap of analyte detection. The number of ^{12}C and ^{13}C labeled dansyl peak pairs is shown for triplicate experiments using either (A) 80% methanol (replicates M1, M2, M3) or (B) 100% water (replicates H1, H2, H3) as the extraction solvent. When using 80% CH_3OH as extraction solvent, the mean number of peak pairs was 413 ± 8, and the number of total peak pairs was 443. When using 100% H_2O as extraction solvent, the mean number of peak pairs was 417 ± 6, and the number of total peak pairs was 474.

I also compared the overlap between the peak-pairs identified by each solvent. Only 234 peak-pairs could be commonly identified from both extracts (Table 3.1). This represents 67% of the 347 peak pairs extracted with H_2O and 61% of the 382 peak pairs with 80% CH_3OH. 74 peak-pairs could be found using 100% H_2O, and 54 peak pairs could be found when using 80% CH_3OH. Thus, parallel extractions with two different solvents increased the number of detected analytes by ~1.4 fold, as compared with extractions with one solvent alone.

Table 3.1 Number of peak pairs common to different extraction solvents in one or more replicate analyses.

100% H_2O (number of replicates)	80% CH_3OH (number of replicates)	Number of peak pairs identified	Number of peak pairs identified in HMDB
3	3	234	145
3	2	23	8
3	1	16	4
3	0	74	49
3	3	234	145
2	3	60	23
1	3	34	13
0	3	54	21

Because extracts must be concentrated (e.g. by evaporation of the solvent) during the protocol, the volatility of solvents was also a consideration. In direct comparisons, I observed that extracts made with 80% CH_3OH could be concentrated twice as quickly as extracts made with 100% H_2O. Therefore, 80% CH_3OH offers extra benefit in minimizing sample processing time, so optimization of the remaining parameters was conducted with 80% CH_3OH as extracton solvent.

3.3.3 Sample-to-Solvent Ratio Optimization

The ratio between the starting mass of tissue and the solvent volume was expected to affect peak-pair yield. I compared the results from five different masses of flax fiber, ranging from 20 mg to 60 mg, extracted in 1 mL of 80% CH_3OH using MAE. As shown in Figure 3.1C, the number of peak-pairs ranged from 362 pairs to 466 pairs while the RSD ranged from 0.48% to 4.21%. The number of peak pairs increased as the mass of fiber increased within the fixed solvent volume. At lower sample:solvent ratios, many peak-pairs could not be distinguished from background signals, and were lost during analysis. The highest number of ^{12}C and ^{13}C labeled peak pairs were extracted from 60 mg fiber in 1 mL solvent; however, the yield at 60 mg/mL was not significantly different from 50 mg/mL. Therefore, 50 mg/mL was selected as the optimum sample:solvent ratio for these experiments

3.3.4 Extraction Time

During the extraction process, the tissue sample was incubated in solvent for a prescribed duration. I compared four different incubation times (15, 30, 45, 60 min) according to their peak-pair yield (Figure 3.1D). In each case, 50 mg of fiber was mixed with 1mL of 80% CH_3OH for extraction by MAE. The results showed that the number of peak pairs ranged from 429 pairs to 482 pairs while the RSD ranged from 0.62% to 7.36%. In general, longer extraction times resulted in the higher number of peak pairs. It was observed that increasing extraction time from 15 min to 30 min raised the number of peak pairs from 448 to 482. Although extension of extraction time may cause the decomposition of analytes, the benitfits of extracting more metabolites outweigh the possible risk. Thus 30 min was chosen as the optimum extraction time for this protocol.

3.3.5 Temperature Effect

During MAE, samples have the ability to convert microwave energy to heat, as such temperature could contribute to one of the effects on extraction efficiency. In this work, 50 mg of fiber mixed with 1mL of 80% CH_3OH was extracted for 30 min by using a microwave oven and a regular oven, respectively. After extracted peak pairs were aligned by retention time and accurate mass, total 1087 peak pairs were compared and illustrated with heatmap (Figure 3.4). The results show that the intensities of most of the peak pairs extracted by microwave oven were higher than those by regular oven. This suggests that heat was not the only factor on extraction efficiency of microwave-assisted extraction method.

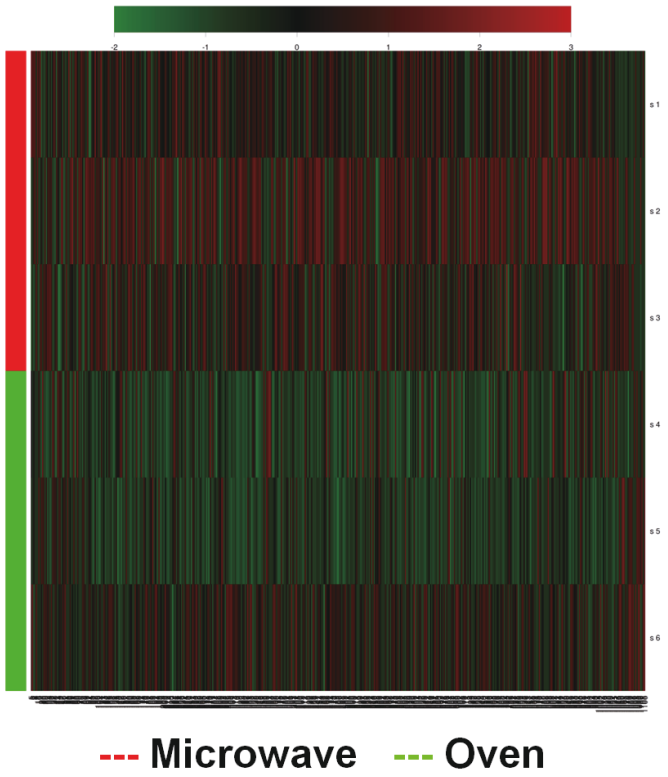


Figure 3.4 Temperature effect between microwave and oven. Heatmap was shown comparison of intensities of ^{12}C and ^{13}C labeled dansyl peak pairs in two different extraction methods. More red colored features indicate higher signal intensities of peak pairs.

3.3.6 Reproducibility of Optimized Extraction Condition

Reproducibility plays an important role in the precision of any new method. In this work, optimal results were obtained with a simple, 30 minute microwave-assisted extraction of 50 mg tissue in 1mL water or 80% CH_3OH/water. The list of ^{12}C and ^{13}C dansyl labeled peak pairs were checked manually after processing by IsoMS and other in-house programming. The results presented in Figure 3.5 showed that the number of total peak pairs was 620 and the number of common peak pairs between triplicate runs was 511 (82%), demonstrating reasonable reproducibility.

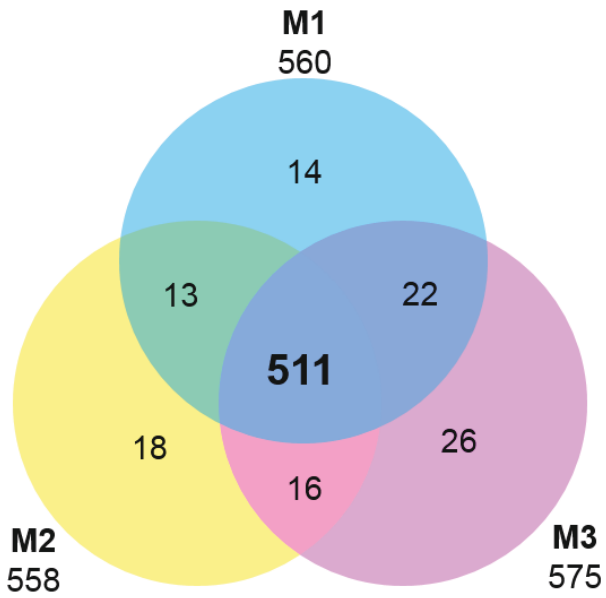


Figure 3.5 Reproducibility of optimized extraction condition. The number of ^{12}C and ^{13}C labeled dansyl peak pairs is shown for triplicate experiments in which 50 mg of flax fiber was extracted in 1 mL 80% MeOH under 30 min microwave-assisted extraction. The mean number of peak pairs was 564 ± 9, and the number of total peak pairs was 620.

3.3.7 Identification of Putative Metabolites

To assign identities to the peak-pairs obtained from flax fiber extracts of either 80% CH_3OH or 100% H_2O, I first searched the Human Metabolome Database (HMDB), containing mass spectral data of over 7900 small molecules.[130] I used a database of human metabolites because an equivalent resource is not yet available for plant metabolites. The search criterion for accurate mass tolerance was ±10 ppm. I was able to assign identities to 263 peak pairs, or 53% of the total non-redundant peak-pairs identified. Of the 263 peak pairs, 193 peak pairs were detected in both 100% H_2O and 80% CH_3OH extractions (Appendix B, Table B.1 and B.2), 49 metabolites were found in extracts with 100% H_2O (Appendix B, Table B.3 and B.4) and 21 metabolites were found in extracts with 80% CH_3OH (Appendix B, Table B.5 and B.6). The names of the putative metabolites from Table B.1 to B.6 (and their synonyms) were also searched in the PlantCyc database of plant metabolic pathways.[131] The presence or absence of each metabolite in the PlantCyc database is marked on the tables, to provide some context about whether a given compound is expected to be found in plants. However, it should be noted that PlantCyc is not a comprehensive list of all plant metabolites.

I selected a subset of 40 metabolites for which authentic standards were readily available, and compared the accurate mass and retention time of the dansylated authentic standards with the observations made on the flax metabolites. In 24/40 cases, the identity of the authentic standard was the same as one of the possible identities assigned to a compound by the HMDB search. Although the data for 16/40 of the HMDB-based identitifications did not match authentic standards, the majority (15/40) had alternative HMDB-based identitifications that could be matched to authentic standards. Only one of the authentic standards (DL-2-aminooctanoic acid) did not match corresponding entries in HMDB. Thus, the majority of identities assigned were consistent with

authentic standards, and these could also be used to distinguish between isomers in assigning identity to peak pairs.

The putative metabolites identified in the database search included most (16/20) of the common amino acids that are genetically encoded in proteins. This was increased to 19/20 when observed m/z and retention time were compared with authetic amino acid standards. Products of the breakdown of some amino acids and polyamines were also detected, including 2-aminomuconic semialdehyde and 4-guanidinobutanoic acid (which can be produced from Trp and Arg, respectively) and two 5-aminopentanal and 4-aminobutyraldehyde (which can be produced from putrescine). Given that the flax stems are typically in early stages of senescence at the time of harvest, it is not surpirsing to identify catabolites of amino acids in the dry fibers.

Many phenolic compounds were also identified among the putative metabolites. These include several well-established components of cell walls: ferulic acid, coumaric acid, and coniferyl alcohol, as well as phenol, catechol, and vanillins. In addition, I identified for the first time in flax fibers the polyphenol resveratrol.

Other compounds that were identified for the first time in flax included histamine, taurine, methionine sulfoxide, and a series of nitrophenolic comounds. Histamine has been previously described as a component of cotton fibers, although its funciton in plants is unknown. I note that despite the fact that fibers of this field crop are harvested and processed mechanically, the possibility remains that some of these compounds are anthropogenic in origin or are otherwise contaminants. Therefore, follow-up experiments are required to verify the natural existence in plants of these previously unidentified metabolites. However, for purposes of this work, the exact biological origin of the compounds is less important than an evaluation of our ability to rapidly identify metabolites.

3.4 Conclusions

The optimization process succesfully defined a simple extraction protocol that enabled identification of potentially hundreds of metabolites from a very small (~50mg) mass of starting tissue in a few hours. The extraction and downstream steps were succesful even when using dry flax fibers, which are very difficult to disrupt mechanically. This suggests that my method could be applicable to a variety of plant tissues (although the extraction parameters may need to be reoptimized in each case). The application of this metabolomics workflow to flax fibers, specifically, led to identification of many known plant metabolites (including amino acids, phenolics) and many metabolites whose presence in flax fibers has not been previously described. This will facilitate future research aimed to understand how flax fiber cells develop some of their unique properties (e.g. intrusive elongation, high tensile strength, hypolignification). Given that flax fibers have been used as textiles for at least 7,000 years, this simple and sensitive metabolite profiling procotol may also be useful in diverse fields such as anthropology and forensics.

Chapter 4

Effective Extraction Method and Stable-Isotope Dansylation Labeling Combined with RPLC-FTMS for the Analysis of Ginseng Root Metabolome

4.1 Introduction

Plant metabolomics has become an important area of research for understanding plant biology and developing new or improved plant-related products. One potential application is to study traditional Chinese medicines (TCMs) which have been developed in China over thousands of years. There are more than 100,000 recipes recorded in ancient literature and plants play a major role in TCMs.[132,133] Nowadays, TCMs have been widely used and accepted as complementary medicine in many countries. Many people use TCMs as their primary healthcare due to low toxicity and good therapeutic performance.[134,135] For example, ginseng, found in North America and in eastern Asia, is one of best-selling TCMs. Research has shown that ~80,000 tons of ginseng roots were produced from South Korea, China, Canada, and the US in 2010, estimated to be worth $2,084 million.[136] Many studies have indicated that ginseng has the abilities to fight against various diseases, boost immune systems, inhibit tumor growth, promote memory, act as an anti-oxidant, and etc.[137-142] Studies have shown that ginseng roots contains ginsenosides, polyacetylenes, polyphenolic compounds and acidic polysaccharides.[143-146] It is thought that the major pharmaceutically-active ingredients are ginsenosides which are a class of tetracyclic triterpenoid saponins.[143-146] As such, most of the studies of ginseng roots were focused on the development of techniques to analyze ginsenosides or on the effects of ginsenosides. Much less information is available about other metabolites in ginseng roots, so my research objective is to develop an analytical method for the metabolome of ginseng roots.

A number of techniques have been employed for studying plant metabolomics. However, comprehensive and quantitative non-targeted profiling of plant metabolome is an analytical challenge, due to a large diversity of chemical and physical properties of metabolites. Reversed phase liquid chromatography-mass spectrometry (RPLC−MS) is one of widely used and powerful methods to profile low molecular weight compounds owing to its sensitivity and resolution. However, RPLC-MS is not suitable for detecting polar and ionic metabolites. Recently, a dansylation labeling combined with stable isotope labeling technique has been developed by our group for absolute and relative quantification of the metabolome.[88] In our work, metabolites containing primary amine, secondary amine, or phenolic hydroxyl groups were labeled with ^{12}C or ^{13}C dansyl chloride prior to LC-MS analysis. This labeling strategy allows separation of polar or ionic metabolites on the RP column, as well as signal enhancement of 10 to 1000 fold. In the present study, I have developed a robust metabolite extraction method from ginseng roots tailored to the analysis of amine- and phenol-containing metabolites, applied the dansylation labeling method to evaluate the detectability of the LC-MS technique for the ginseng metabolome, and used these methods to study spatial distribution of metabolites in ginseng roots.

4.2 Experimental

4.2.1 Reagents and Chemicals

^{12}C-dansyl chloride and amino acid standard solution were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). ^{13}C-dansyl chloride was synthesized in our lab and the chemicals used to synthesize the isotope reagent were also purchased from Sigma-Aldrich. LC/MS grade water, acetonitrile (ACN), methanol (MeOH), acetone and formic acid were purchased from Thermo Fisher Scientific (Edmonton, AB, Canada).

4.2.2 Ultrasonic-assisted Extraction (UAE)

Ultrasonic-assisted extraction was performed with a Branson ultrasonic cleaner 1510-MT (Branson Ultrasonics Corporation, Danbury, CT, USA). Ginseng roots (*Panax quinquefolius)* were ground by using a pestle and mortar. Ultrasonic extraction was conducted by mixing 20 mg of ginseng roots and 1 mL of 80% MeOH or 80% ACN in a sample vial, which was then placed in an ultrasonic bath at room temperature for 5 min. When the extraction was completed, the extract was centrifuged at 14,000 rpm for 10 min to obtain a supernatant.

4.2.3 Microwave-assisted Extraction (MAE)

The procedure of MAE was described in Chapter 3.2.3, but with some minor changes tailored to the ginseng samples. All the microwave extractions were performed under 120 watts microwave irradiation power. Ginseng roots were ground by using a pestle and mortar. 20 mg of ginseng roots were mixed with 1 mL of 80% MeOH or 80% ACN in a sample vial, which was then placed in a microwave oven for 5 min. For optimizing extraction time, 20 mg of ginseng roots in 1 mL of 80% MeOH was extracted for 15, 30, 45, and 60 min. For choosing extraction solvent, the four different percentage of methanol studied were 40%, 60%, 80%, and 100%. For optimizing sample-to-solvent ratio, different amount of ginseng roots (5, 10, 20, 30, and 40 mg) in the same volume solvent were tested.

4.2.4 Dansylation Labeling Reaction

The dansylation labeling reaction has been decribed in Chapter 3.2.5. Briefly, a 50 μL sample was labeled with 50 μL of ^{12}C-dansyl chloride solution (18 mg/mL) for light labeling or ^{13}C-dansyl chloride (18 mg/mL) for heavy labeling. The ^{13}C-labeled mixtures were combined with an equal volume of their ^{12}C-labeled counterparts for LC-FT-ICR-MS analysis.

4.2.5 LC/MS Analysis

The HPLC system consisted of an Agilent 1100 liquid chromatograph (Palo Alto, CA, USA) equipped with a binary pump system. The analytes were separated by a reversed-phase Eclipse plus C18 column (2.1 mm×100 mm, 1.8 μm particle size, 95 Å pore size) purchased from Agilent. Solvent A was 0.1% (v/v) formic acid in 5% (v/v) ACN, and solvent B was 0.1% (v/v) formic acid in ACN. The gradient started at 20% B and stepwise increased to 99% B within 21.5 min, and hold at 99% for 2 min, then stepwise decreased to 20% B within 0.5 min, and finally hold at 20% B for 1 min. The flow rate was 180 μL/min, and the sample injection volume was 2.0 μL. The flow from RPLC was split at a ratio of 1:3 before a 45 mL/min flow was loaded to the electrospray ionization (ESI) source of a Bruker 9.4 Tesla Apex-Qe Fourier transform ion-cyclotron resonance (FT-ICR) mass spectrometer (Bruker, Billerica, MA, USA). The MS conditions used for FT-ICR-MS were as follows: nebulizer gas 2.3 L/min, drying gas temperature 190 ℃, drying gas flow 7.0 L/min, capillary voltage 4200 V, and scan range 200-1000. All MS spectra were obtained in the positive ion mode.

4.2.6 Data Processing and Metabolite Database Search

The steps of data processing were the same as those in Chapter 3.2.6. Putative metabolite identification was based on accurate mass matches of underivatized metabolites in HMDB (www.hmdb.ca)[89] and MycompoundID (www.mycompoundid.org),[147] using a mass accuracy tolerance of 10 ppm.

4.3 Results and Discussion

4.3.1 LC/MS Analysis

In this work, analytes were separated and detected by LC-FT/ICR-MS in 25 min. Figure 4.1A illustrates a representative base-peak ion chromatogram of ^{12}C and ^{13}C dansyl labeled metabolites extracted by MAE in 45 min from ginseng roots. A representative mass spectrum showing a pair of ^{12}C and ^{13}C peaks is presented in Figure 4.1B. Based on differential isotope labeling and dansylation reaction, metabolites containing primary amine, secondary amine, or phenolic hydroxyl groups could be detected and analyzed. As noted in Chapter 4, the theoretical mass difference between a ^{12}C and a ^{13}C dansyl tag is 2.0067 Da. Since nonreactive metabolites or background ions would not produce $^{12}C/^{13}C$ peak pairs with this specific mass difference, a list of specific metabolites from ginseng roots can be generated by software such as IsoMS.

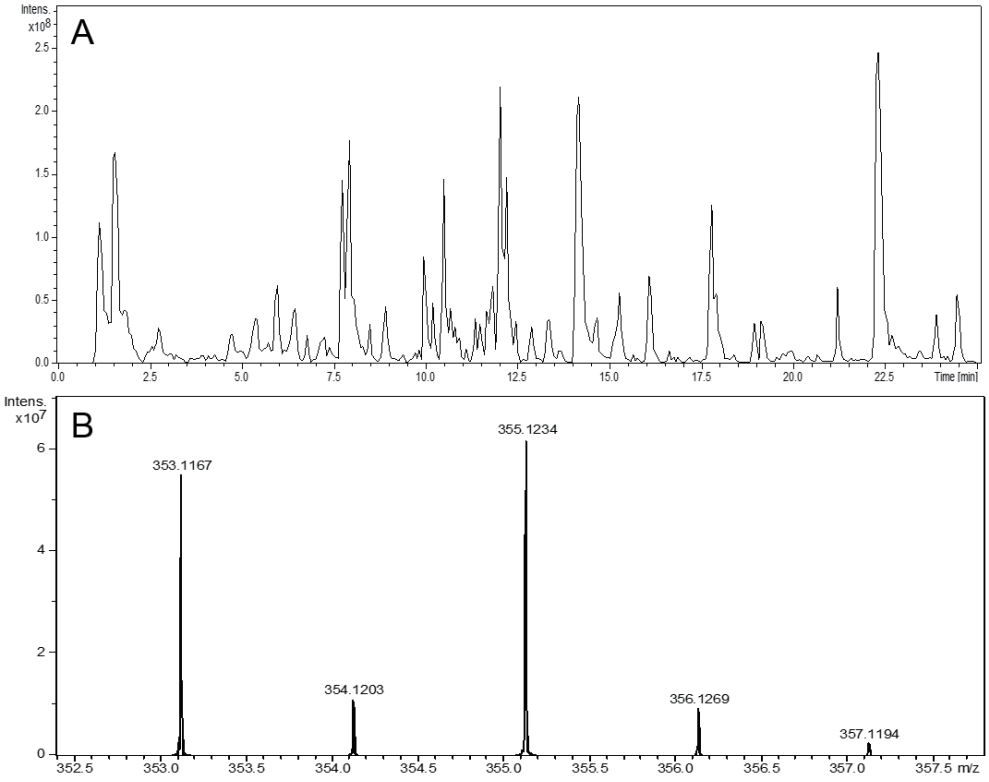


Figure 4.1 (A) Base peak ion chromatogram obtained by LC/FT-ICR-MS of ^{12}C -/^{13}C-dansylated ginseng root. (B) Expanded mass spectra of ^{12}C -/^{13}C-dansylated ion pairs from the chromatogram. Accurate mass match against the Human Metabolome Database indicates that the ion pairs were likely from threonine or other isomers.

4.3.2 Microwave vs. Ultrasonic Extraction

Extracting metabolites is an important step for plant metabolome analysis. The conventional solvent extraction with vortex, such as Soxhlet extraction, is time-consuming, requires large volumes of solvents, and may not be efficient.[148] In this work, I have studied and compared microwave-assisted extraction (MAE) vs ultrasonic-assisted extraction (UAE) for extracting metabolites from ginseng roots immersed in two different solvents. Since the sample for LC-FT-MS assay was prepared by mixing equal volumes of ^{12}C- and ^{13}C-dansyl labeled aliquots of the

same extract, the expected peak intensity ratio for each labeled metabolite should be 1.0. The IsoMS software was used to process the LC-MS data and calculate the intensity ratio for each peak pair. Peak pairs (<10% of the overall number of peak pairs detected) with ratios of smaller than 0.67 and larger than 1.5 (i.e., the outliers due to experimental variations) were excluded. The number of ^{12}C /^{13}C peak pairs identified was then used as a measure of the extraction yield.

I compared the efficacy of MAE and UAE of a 20 mg ginseng sample in 80% MeOH or 80% ACN (Figure 4.2A) for 5 min. The results showed both MAE and UAE produced a similar number of metabolites when 80% ACN was used as extraction solvent, and in both cases, the yield of peak pairs increased when 80% CH_3OH was used as extraction solvent. Figure 4.2B showes a clustered heatmap illustrating the relative intensities of the same peak pairs detected in different extraction solvents. In this map, the values of higher intensity were represented by red squares and the values of lower intensity by green squares. For each solvent there were prominent peak pairs apparent, but most of peak pairs show higher intensities in 80% CH_3OH. Therefore, the extraction efficiency of using 80% CH_3OH is better than that of 80% ACN. As the yield was larger for MAE (355±6) than for UAE (296±5), MAE appeared to be a superior method than UAE in these trials, so I focused on the development and application of MAE in the subsequent experiments.

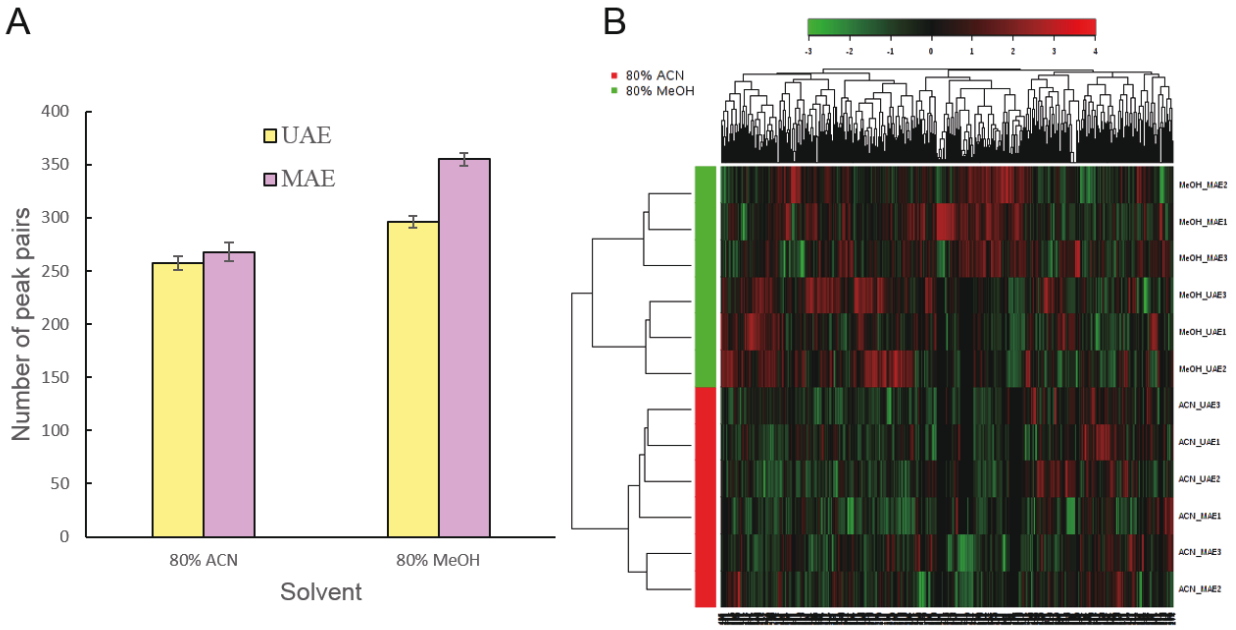


Figure 4.2 (A) Comparison of microwave assisted extraction (MAE) and ultrasonic-assisted extraction (UAE) with 80% MeOH or 80% ACN as extraction solvent. Each sample was done in experimental triplicates and error bars represent standard deviations. (B) Clustered heatmap was shown comparison of intensities of ^{12}C and ^{13}C labeled dansyl peak pairs in two different extraction solvents. More red colored features indicate higher signal intensities of peak pairs.

4.3.3 Optimization of Microwave-assisted Extraction Method

The microwave-assisted extraction technique has high and fast extraction performance ability with less solvent consumption.[148] In order to optimize the microwave-assisted extraction protocol for ginseng roots, I varied the extraction time, the concentration of MeOH, and sample-to-solvent ratio. All extracts were assayed by LC-FT-MS and the number of ^{12}C /^{13}C peak pairs identified was used as a measure of the extraction yield. Each sample was analyzed in experimental triplicates.

First, extraction time is a major factor needed to be taken into account for influencing the efficiency of MAE method. Four different extraction times, 15, 30, 45 and 60 min, were studied by extracting 20 mg of ginseng roots in 1 mL of solvent. The number of peak pairs obtained was plotted against time as shown in Figure 4.3A, which shows that the number of ^{12}C and ^{13}C labeled peak pairs ranged from 694 to 783 while the RSD ranged from 2.5% to 14.3%. In general, longer extraction results in the higher number of peak pairs. It was observed that increasing extraction time from 15 min to 45 min raised the number of peak pairs from 700 to 783. It seems that extension of extraction time to 60 min exhibits no further increases in peak pairs, this may also cause thermal degradation of analytes. Thus, 45 min were chosen as the optimum extraction time for following studies.

Next, the concentration of CH_3OH in H_2O is another important parameter of the MAE method, directly related to extraction efficiency. In this work, four different concentrations of CH_3OH were studied, 40%, 60%, 80%, and 100%. The result (Figure 4.3B) illustrated that the number of $^{12}C/^{13}C$ peak pairs ranged from 703 pairs to 792 pairs while the relative standard deviation (RSD) ranged from 0.7% to 4.2%. The result showed that the number of peak pairs increased with increasing the percentage of CH_3OH is increased from 40% to 80%. However, the number of peak pairs decreased with 100% CH_3OH. This suggests that the penetration and heating effect of extraction solvent with a small amount of H_2O in ginseng roots may be better than that without H_2O.[148,149] Thus, the extraction efficiency was optimum when using 80% MeOH as extraction solvent.

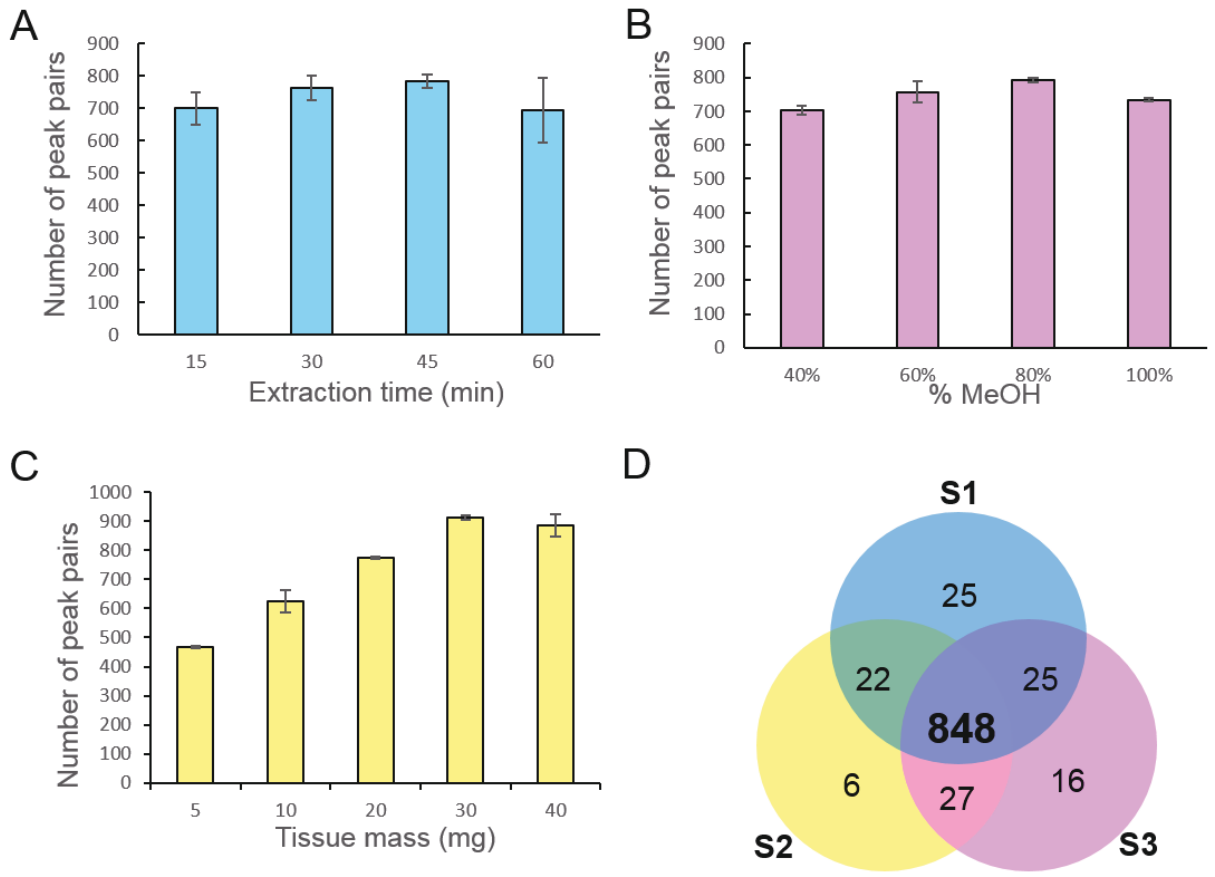


Figure 4.3 (A) Comparison of different extraction, (B) Comparison of different concentration of MeOH, (C) Comparison of different sample-to-solvent ratio, (D) A Venn diagram showes reproducibility at optimal extraction conditions. Each dataset was done in triplicates and error bars represent standard deviations in (A) to (C).

Last, I studied sample-to-solvent ratio as one of the influencing factors because the extraction yield can be affected by the sample amount in the fixed volume of solvent. Five different amounts, from 5 mg to 40 mg of ginseng roots, were studied under the 45 min microwave-assisted extraction in 1 mL 80% CH_3OH. As shown in Figure 4.3C, the number of ^{12}C and ^{13}C labeled peak pairs ranged from 466 to 913 while the RSD ranged from 0.8% to 6.1%. It was obvious that the number of peak pairs increased with increasing the amount of ginseng roots. Although low abundance peak

pairs could be produced with low amount of ginseng roots, they would not be detected because they are buried in the background; therefore, they were undetectable or were filtered out during the data processing. When extracting 30 mg of ginseng roots in 1 mL 80% MeOH, the number of labeled peak pairs was 913 and the value of RSD was 1.0%. However, increasing tissue amount had no further effect so 30 mg of ginseng roots is used for all following studies.

In this work, optimal results were obtained with a simple, 45 minute microwave-assisted extraction of 30 mg ginseng roots tissue in 1 mL water or 80% methanol. The list of ^{12}C and ^{13}C dansyl labeled peak pairs were generated by IsoM and other in-house programming (such as zero-filling[150]). The results presented in Figure 4.3D showed that the number of total peak pairs was 969 and the number of common peak pairs between triplicate runs was 848 (88%). The information of putative metabolites of ginseng roots was showed in Table C.1 (see Appendix C). Putative metabolite identification was based on accurate mass matches of underivatized metabolites in HMDB and MycompoundID, using a mass accuracy tolerance of 10 ppm. I used a database of human metabolites because an equivalent source is not yet available for plant metabolites.

4.3.4 Spatial Distribution of Metabolites in Ginseng Roots

In order to study the spatial distribution, a ginseng root was cut into small pieces with approximately 0.5 cm length from top to bottom. Each individual segment sample was labeled with ^{12}C-dansyl chloride, and a pooled sample was produced by mixing aliquots of individual ginseng segment samples and labeled with ^{13}C-dansyl chloride. A ^{12}C-dansyl labeled individual sample was mixed with an equal volume of the ^{13}C-dansyl labeled pooled sample, followed by running the misture in LC-MS to identify the isotope labeled paek pairs. The peak intensity ratios of the isotope labeled peak pairs were calculated. Since the same pooled sample was used, the peak ratios of each individual peak pair from different samples represented the different concentrations

of the same putative metabolite in these samples. Each ginseng segment sample was analyzed in experimental triplicates. Three ginseng roots were analyzed by the steps described as above.

I selected a subset of 16 amino acids for which authentic standards were readily available, and compared the accurate mass and retention time of the dansylated authentic standards with the observations made on the ginseng metabolites. Figure 4.4 shows the spatial distribution of 16 amino acids in three different ginseng roots. For each ginseng root, the average $^{12}C/^{13}C$ ratios of same amino acid fluctate in different segments. The RSD of the average $^{12}C/^{13}C$ ratio of Ginseng 1 ranged from 0.5% to 22.8%. For Ginseng 2, the RSD ranged from 0.9% to 22.8% and for Ginseng 3, the range was from 1.9% to 30.0%. The range of %RSD of these three ginseng roots was generally lower than or equal to 30%, indicating relatively good reproducibility. In this case, I did not find any trends showing that same amino acids were more abundant in different ginseng roots, but I demonstrated that this technique can successfully provide the spatial distribution of metabolites in the plant.

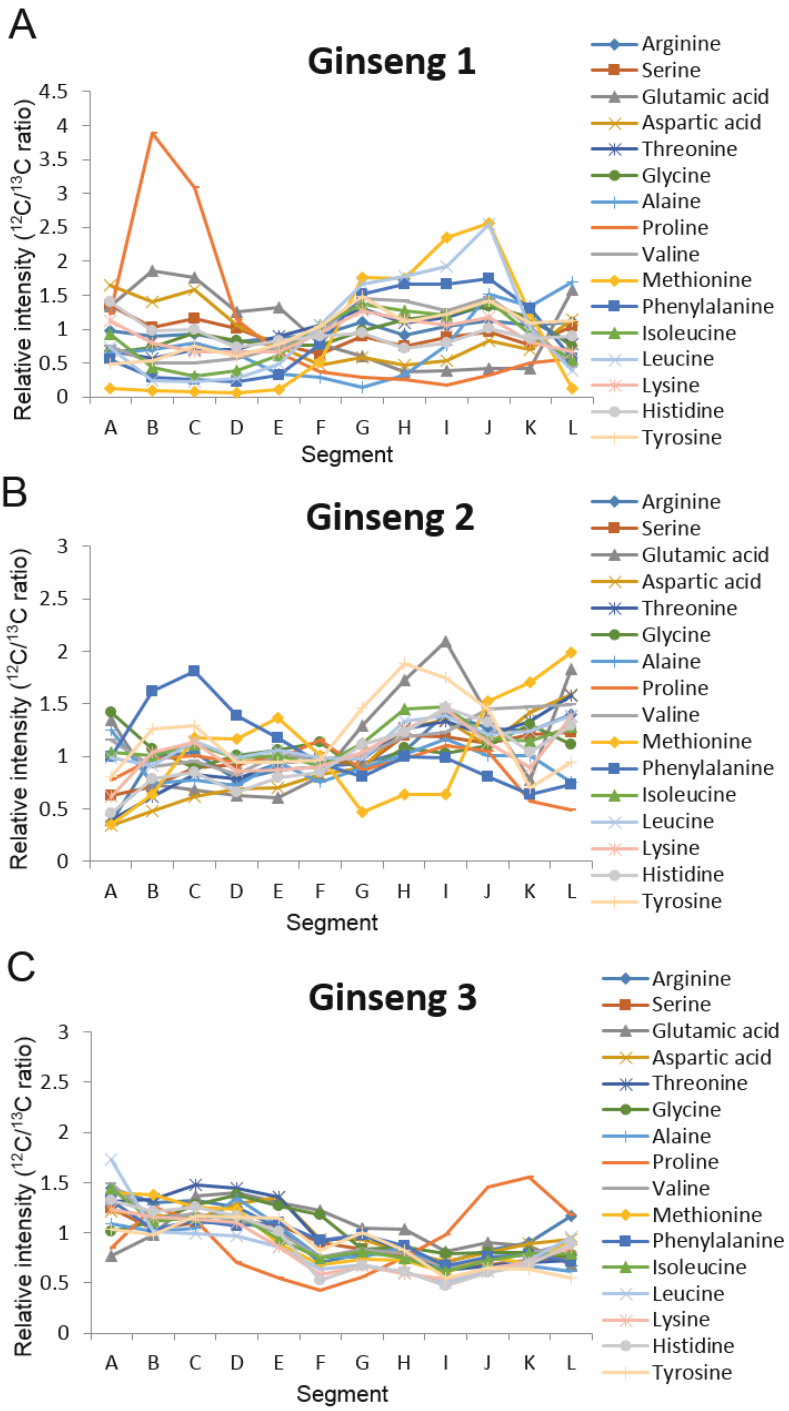


Figure 4.4 Spatial distribution of 16 amino acids (A) Ginseng 1, (B) Ginseng 2, (C) Ginseng 3. Each dataset was done in triplicates. Segment A represents the top of a ginseng root and segment L was the bottom of a ginseng root.

I also tried to find metabolites with high correlation in the same ginseng root. For each ginseng root, I analyzed changing trends of 12 segments and clustered the metabolites with similar changing trends. For Ginseng 1, I could find 15 groups and the result is shown in Table C.2 (see Appendix C). For Ginseng 2, the metabolites with similar changing trends could be summarized into 20 groups (Appendix C, Table C.3). Table C.4 (see Appendix C) illustrates there were 18 groups found in Ginseng 3. Figure 4.5 shows two examples of spatial distribution of metabolites in the same group of Ginseng 1. For measuring the same metabolites from different segments, although these segments came from the same ginseng root, the growth conditions at different periods should cause variation in the distribution of these metabolites. Besides, metabolites had similar changing trends so it might show some evidence that these metabolites have high correlation in plant growth. Therefore, this method would have potential to offer a preliminary understanding of plant metabolism.

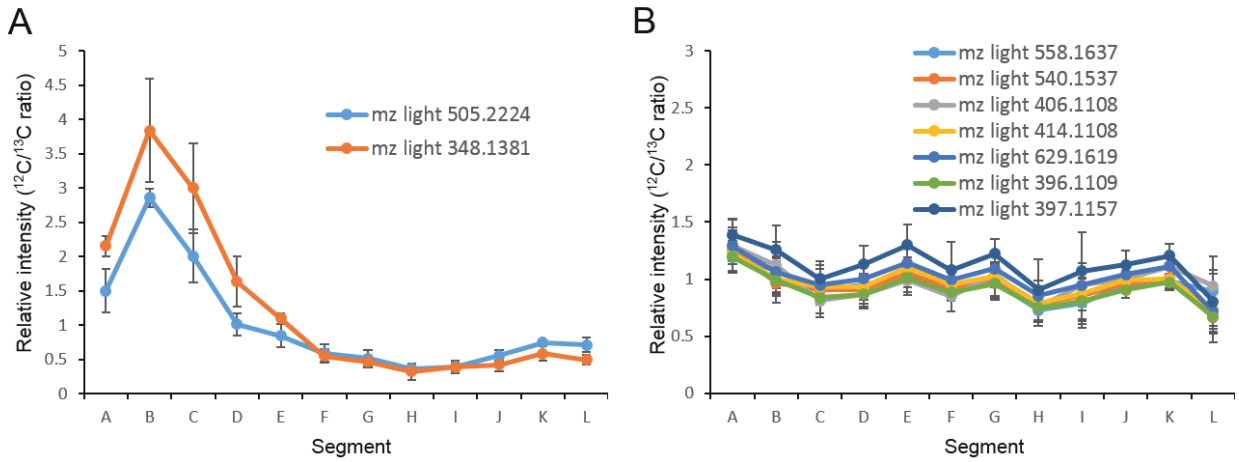


Figure 4.5 Spatial distribution of metabolites in Ginseng 1 (A) Group 2, (B) Group 7. Each dataset was done in triplicates and error bars represent standard deviations. Segment A represent the top of a ginseng root and segment L was the bottom of a ginseng root.

4.4 Conclusion

The optimization process successfully defined a simple extraction protocol that enabled identification of potentially hundreds of metabolites from a very small (~30 mg) mass of starting tissue in a few hours. Moreover, a reproducible and high-quality separation method has been developed for analyzing the isotope-tagged metabolites from ginseng roots. After labeling by the stable isotope dansylation reagents, the very polar metabolites changed their chromatographic retention behavior. As a result, the reversed-phase chromatographic separation can be easily done in 25 min and the sample throughput can be improved due to the use of one mode of separation, i.e., RPLC, instead of multiple modes of separation to handle metabolites of different ionic strength and polarity. There were more than 800 peak pairs detected by this optimal method with good reproducibility (88%) in triplicate runs. Application of this method was successfully demonstrated on the determination of spatial distribution of metabolites in ginseng root as well as correlation between different metabolites.

Chapter 5

Effective Extraction Method and Stable-Isotope Dansylation Labeling Combined with RPLC-FTMS for the Analysis of *Arabidopsis Thaliana* Metabolome

5.1 Introduction

Plant metabolomics has become an important area of research for understanding plant biology because it can provide information in connecting genotypes with phenotypes and understanding how biological systems behave at functional levels.[151,152] One of important models used in plant biology is the flowering plant *Arabidopsis thaliana*. Because this plant has many advantages, including a small size of genome, a rapid life cycle and large number of offspring,[153] it is widely used for understanding the genetic, cellular, and molecular biology of flowering plants.

The lipid-derived hormone jasmonates (JA), including jasmonic acid, methyl jasmonate (MeJA) and other derivatives, play important roles as signaling molecules in plant growth and stress responses caused by biotic and abiotic factors.[154,155] MeJA is involved in fruit ripening, root growth, flowering and etc.[156] MeJA can affect flowering time and flower morphology,[157] induce ethylene-forming enzyme activity for fruit maturation, and inhibit growth of plant roots.[158,159] In addition, when plants resist pathogens and insects, MeJA can be used to signal original plant's defense system which is called JA signaling, or it can be spread in the air to warn other unharmed plants to generate their defense systems. Therefore, MeJA plays a role in intra- and inter-plant signaling. Jasmonate ZIM domain (JAZ) family has been identified as a transcriptional repressor of JA signaling.[160-163] The defining feature of the JAZ family members is the highly conserved *TIFY* motif located within a larger conserved region in the ZIM domain.[160-163] Some studies have shown that the expressions of several *TIFY* genes are affected by wounding[164] or high salt

concentration[165] and indicated that these results are relative to JA signaling since JA signaling is highly connected to stress response.[166-168] In this study, I investigated the metabolic changes of different *TIFY* expression plants under MeJA treatment.

Comprehensive and quantitative non-targeted profiling for plant metabolome is an analytical challenge, due to a large diversity of chemical and physical properties of metabolites. One of the important tools chosen for plant metabolomics is liquid chromatography–mass spectrometry (LC-MS); but LC-MS does not directly generate quantitative information for non-targeted metabolites. Stable isotope labeling (SIL) combined with differential isotope labeling (DIL) techniques are not only the most reliable quantitative methods, but are also derived to overcome matrix effect and ion suppression.[123,169,170] Samples and internal standards are produced as isotopomers by these techniques, as well as analyzed by MS according to the mass difference. Recently, a dansylation labeling technique combined with stable isotope labeling technique has been reported by our group for absolute and relative quantification of the metabolome.[88] In our work, metabolites containing primary amine, secondary amine, or phenolic hydroxyl groups have been labeled with ^{12}C or ^{13}C dansyl chloride prior to LC-MS analysis. This labeling strategy allows separation of polar or ionic metabolites on the RP column, as well as MS signal enhancement of 10 to 1000 fold.

This chapter describes development of a LC-MS method for studying the metabolome of *Arabidopsis thaliana*. I have developed a robust metabolite extraction method from *Arabidopsis thaliana*, applied the dansylation labeling method to improve the detectability of the *Arabidopsis thaliana* metabolome by LC-MS, and used these methods to study metabolic changing trends of different *TIFY* expression plants under MeJA treatment.

5.2 Experimental

5.2.1 Chemicals and Reagents.

^{12}C-dansyl chloride and amino acid standard solution were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). ^{13}C-dansyl chloride was synthesized in our lab and the chemicals used to synthesize this isotope reagent were also purchased from Sigma-Aldrich. LC/MS grade water, acetonitrile (ACN), methanol (CH_3OH), acetone and formic acid were purchased from Thermo Fisher Scientific (Edmonton, AB, Canada).

5.2.2 Plant Growth Conditions

Arabidopsis thaliana ecotype Columbia (Col-0) was obtained from Dr. M. Deyholos (Department of Biological Sciences, University of Alberta, Canada). For soil-grown plants, *Arabidopsis thaliana* was grown in a controlled environment chamber at 21–23 °C and 60% relative humidity under a photoperiod of 16 h of light/8 h dark cycle with 100 μmol photons m^{-2} s^{-1}. The soil used in this work was Professional Growing Mix from Sunshine® containing Canadian sphagnum peat moss, perlite, a starter fertilizer charge, a controlled-release fertilizer and dolomite lime.

For cultured plants, wild-type, *AtTIFY10b* overxpression and knock-out *Arabidopsis* seeds were surface-sterilized by 50% (v/v) commercial bleach (Javex, Clorox), cold incubated at 4°C for two days in the dark, and then sown in petri-dishes (100 x15 mm) containing Murashige and Skoog medium (MS medium, pH 5.7 by KOH) and 0.7% phytablend (MS-agar). For methyl jasmonate (MeJA) treatment, 7 days old of 0.5× MS growing seedlings, were treated with 50 μM MeJA, Samples were taken as three biological replicates at 0 h, 0.5 h, and 24 h time points after treatment, followed by lyophilization for 36 hours. All of plants were stored in -80 °C freezer until further use.

5.2.3 Microwave-assisted Extraction (MAE)

The procedure of MAE was described in Chapter 3.2.3, but with some minor changes tailored to *Arabidopsis thaliana*. All the microwave extractions were performed under 120 watts microwave irradiation power. For the purpose of choosing a good extraction solvent, three solvents, 80% ACN, 80% acetone, and 80% CH_3OH, were tested. For optimizing extraction time, the mixture consisted of 20 mg of *Arabidopsis thaliana* and 1 mL of solvent was extracted for 15, 30, 45, and 60 min. For optimizing sample-to-solvent ratio, different amount of *Arabidopsis thaliana* (10, 20, 30, 40, and 50 mg) in the same volume solvent were irradiated.

5.2.4 Dansylation Labeling Reaction[88]

The dansylation labeling reaction has been decribed in Chapter 3.2.5. Briefly speaking, a 50 μL aliquot of supernatant from MAE extraction was labeled with 50 μL of freshly prepared ^{12}C-dansyl chloride solution (18 mg/mL) for light labeling or ^{13}C-dansyl chloride (18 mg/mL) for heavy labeling. The ^{13}C-labeled mixtures were combined with an equal volume of the ^{12}C-labeled counterparts for LC-FT-ICR-MS analysis.

5.2.5 LC/MS Analysis

The LC conditions used for Agilent HPLC and the MS conditions used for Bruker FT-ICR-MS were described in Chapter 4.2.5, but with some minor changes. The gradient started at 20% B and stepwise increased to 35% B within 3.5 min, followed by 14.5-min ramp to 65% B, then raised to 95% B within 3 min and hold at 95% for 5 min, followed by 0.5-min ramp to 99% B and hold at 99% for 2 min, then stepwise decreased to 20% B within 1 min, and finally hold at 20% B for 1.5 min.

5.2.6 Data Processing and Metabolite Database Search

The steps of data processing were as same as those in Chapter 3.2.6. Putative metabolite identification was based on accurate mass matches of underivatized metabolites in HMDB[89] and MycompoundID[147], using a mass accuracy tolerance of 10 ppm.

5.3 Results and Discussion

5.3.1 LC/MS Analysis

In this work, analytes were separated and detected by LC-FT-ICR-MS in 31 min. Figure 5.1A shows a representative base-peak ion chromatogram of ^{12}C and ^{13}C dansyl labeled metabolites extracted by the MAE method for 45 min from *Arabidopsis thaliana*. A representative mass spectrum showing a pair of ^{12}C and ^{13}C peaks is presented in Figure 5.1B. Based on differential isotope labeling technique and dansylation reaction, metabolites containing primary amine, secondary amine, or phenolic hydroxyl groups could be detected and analyzed. The theoretical mass difference between a ^{12}C and a ^{13}C dansyl tag is 2.0067 Da. Since nonreactive metabolites or background ions would not produce $^{12}C/^{13}C$ peak pairs with this specific mass difference, a list of ^{12}C and ^{13}C dansyl labeled metabolites can be generated by softwares based on IsoMS.

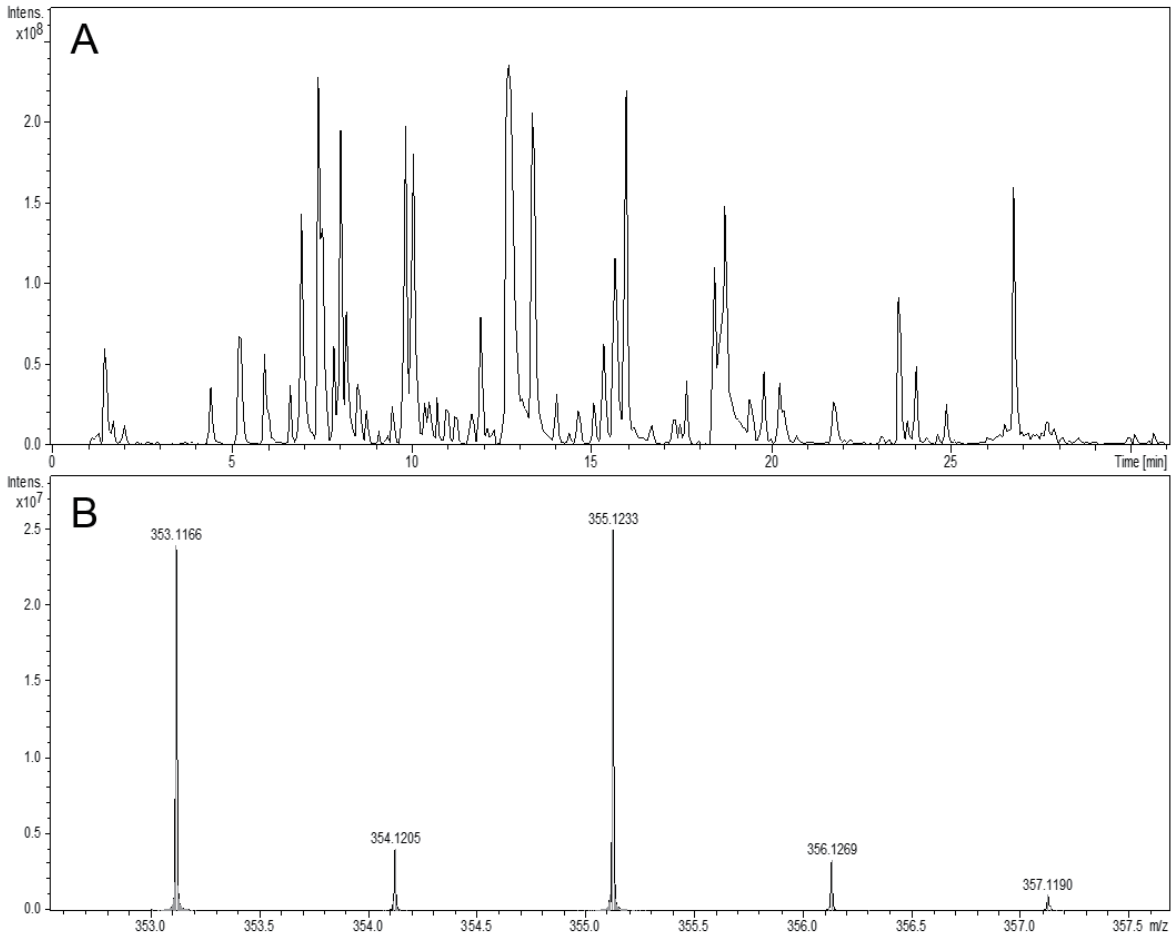


Figure 5.1 (A) Base peak ion chromatogram obtained by LC/FT-ICR-MS of ^{12}C-/^{13}C-dansylated arabidopsis thaliana. (B) Expanded mass spectra of ^{12}C-/^{13}C-dansylated ion pairs from the chromatogram. Accurate mass match against the Human Metabolome Database indicates that the ion pairs were likely from threonine or other isomers.

5.3.2 Extraction Conditions of Microwave-assisted Extraction Method

Extracting metabolites is an important step for plant metabolome analysis. The conventional solvent extraction with vortex was time-consuming and not very efficient. I focused on the development and application of the microwave-assisted extraction method. The microwave-assisted extraction technique has high and fast extraction performace ability with less solvent

consumption.[148] To optimize the microwave-assisted extraction protocol for *Arabidopsis thaliana*, I varied the solvent, extraction time, and sample-to-solvent ratio. All extracts were assayed by LC-FT-MS and the number of ^{12}C /^{13}C peak pairs identified was used as a measure of the extraction yield.

First of all, the choice of extraction solvent is an important parameter of the MAE method because different solvents can extract different analytes, and also the solubility of analytes in the extraction solvent has the effect on extraction efficiency. In this work, three different solvents were studied, including 80% ACN, 80% acetone, and 80% MeOH. The result (Figure 5.2A) shows that the number of $^{12}C/^{13}C$ peak pairs ranged from 425 to 632 while the relative standard deviation (RSD) ranged from 7.6% to 14.9%. The extraction efficiency was optimum when using 80% CH_3OH as extraction solvent.

Next, extraction time is another condition needed to be taken into account for influencing the efficiency of MAE method. Four different extraction times, from 15 min to 60 min, were studied by extracting 20 mg of *Arabidopsis thaliana* in 1 mL of solvent. The number of peak pairs obtained from the software was plotted against the extraction time as shown in Figure 5.2B. The result showed that the number of ^{12}C and ^{13}C labeled peak pairs ranged from 608 to 729 while the RSD ranged from 2.5% to 6.3%. In general, longer extraction results in the higher number of peak pairs. It was observed that increasing extraction time from 15 min to 45 min raised the number of peak pairs from 608 to 716. However, the extension of extraction time to 60 min not only had no further effect, but also caused the decomposition of analytes. Based on this finding, 45 min was chosen as the optimum extraction time for the following studies.

Last, I studied sample-to-solvent ratio as one of the influencing factors because the extracted amount of analytes can be affected by the different amount of sample in a fixed volume of solvent. There were five different amounts, from 10 mg to 50 mg of *Arabidopsis thaliana*, studied under

the 45 min microwave-assisted extraction in 1 mL 80% methanol. As shown in Figure 5.2C, the number of ^{12}C and ^{13}C labeled peak pairs ranged from 588 to 781 while the RSD ranged from 1.4% to 9.8%. It was obvious that the number of peak pairs increased with increasing the amount of *Arabidopsis thaliana*. Although the low abundance peak pairs could be produced at the low amount of *Arabidopsis thaliana*, they could not be detected because they are buried in the background; therefore, they were undetectable or were filtered out during the data processing. When extracting 40 mg of *Arabidopsis thaliana* in 1 mL methanol, the number of labeled peak pairs was 781 and the value of RSD was 1.9%. Since further increasing tissue amount resulted no additional benefits, 40 mg of *Arabidopsis thaliana* is used for the following studies.

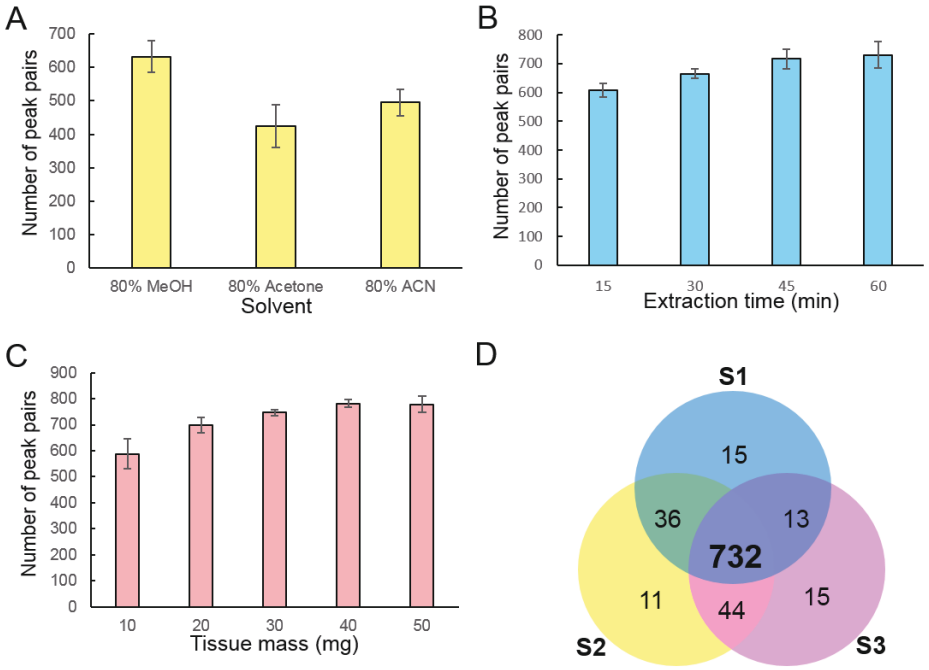


Figure 5.2 (A) Comparison of different extraction solvents, (B) Comparison of different extraction time, (C) Comparison of different sample-to-solvent ratio, (D) A Venn diagram shows reproducibility at optimal extraction conditions. Each dataset was done in triplicates and error bars represent standard deviations in (A) to (C).

Reproducibility plays an important role in the precision of a new method. In this work, optimal results were obtained with a simple, 45 min microwave-assisted extraction of 40 mg tissue in 1 mL water or 80% methanol. The list of ^{12}C and ^{13}C dansyl labeled peak pairs were generated by IsoM and other in-house programming. The results are presented in Figure 5.2D, showing that the number of total peak pairs was 866 and the number of common peak pairs between triplicate runs was 732 (84%).

5.3.3 Metabolites Affected by TIFY Expression Levels

In this work, the microwave assisted extraction method was applied to study putative metabolite of *Arabidopsis thaliana* with three different *TIFY* gene expression levels, including overexpression *TIFY* (OXT), wild type (WT), and knock-out *TIFY* (KOT). To compare the plant metabolome, a pooled sample was produced by mixing aliquots of individual plant samples and labeled with ^{13}C-dansyl chloride. An equal volume of the ^{12}C-dansyl labeled individual sample was mixed with the ^{13}C-dansyl labeled pooled sample, followed by running the mixture in LC-MS. The peak intensity ratios of the isotope labeled peak pairs detected in the mass spectra were calculated. Because the same pooled sample was used, the peak ratios of each individual peak pair from different samples represented the different concentrations of the same putative metabolite in these samples. The peak ratio values found in different samples were then exported into a statistics tool. Each plant sample was analyzed in biological triplicates and experimental duplicates.

The results of the analysis are shown in Figure 5.3. The unsupervised principal component analysis (PCA) was applied to generate an overview on how the data were scattered (see Figure 5.3A). The PCA score plot clearly demonstrated that the three plants with different *TIFY* gene levels were well separated. The first principal component resolved WT from OXT and KOT samples, whereas the second principal component resolved OXT from KOT samples. Although

WT, OXT, and KOT all belong to *Arabidopsis thaliana*, there were substantial differences in their metabolic profiles because of different TIFY levels. Figure 5.3B shows a clustered heatmap which illustrates the relative intensities of the same peak pairs detected among all three plant types. In the clustered heatmap, the values of higher intensity were represented by red squares and the values of lower intensity by green squares. Although most of peak pairs were shown to have higher concentrations in WT samples, it could be seen that for each plant type there were prominent peak pairs specific to this plant type. The results from PCA and heatmap analysis indicated the potential of this method for differentiating different gene levels from the same plant.

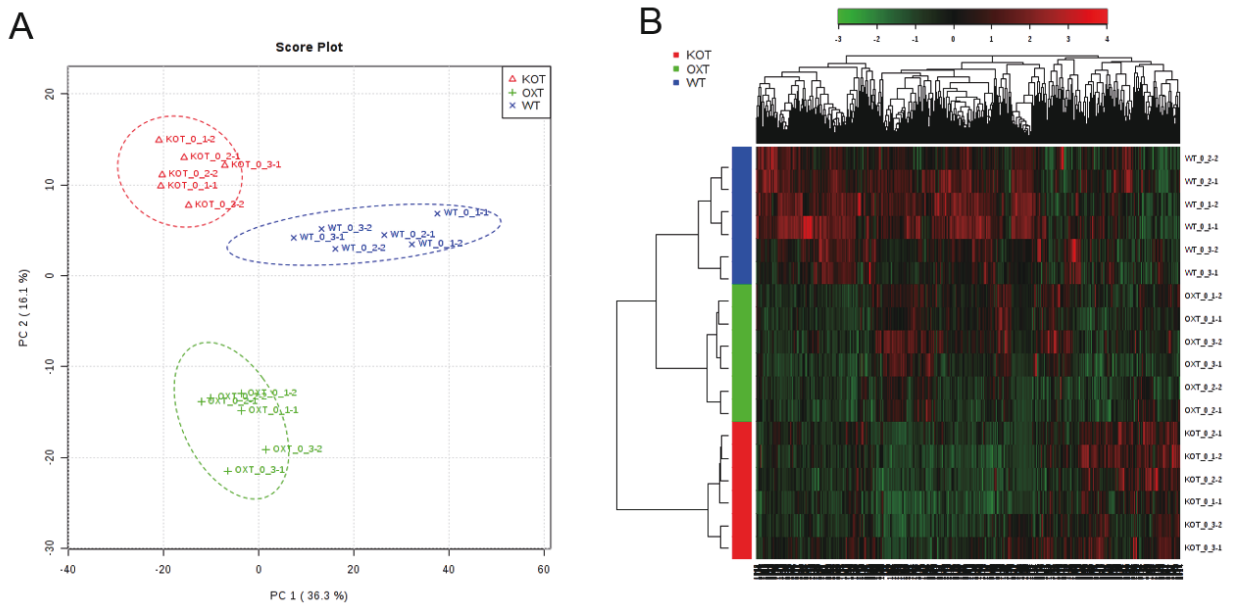


Figure 5.3 (A) PCA score plot (B) Clustered heatmap. Both showed comparison between three plant types with different gene expression levels, including wild type (WT), knock-out (KOT) and overexpression (OXT) TIFY. Each plant type was done in biological triplicates and experimental duplicates.

In order to gain information of putative metabolites, I did a comparison between two plant types. The data was analyzed by using a Volcano plot which is another statistical tool that combines the information of fold change and t-test to determine the significant metabolites that can be used to separate the two groups. Using the 1.5-fold change and p-value 0.01 as the threshold cutoff values, the Volcano plots are shown in Figure 5.4. According to Figure 5.4A and 5.4B, most of peak pairs were found to be significant in the upper left. Negative fold changes of these peak pairs indicated that the concentrations of the putative metabolites were down regulated in OXT and KOT samples compared to WT samples. Besides Volcano plots, the variable importance in projection (VIP) that contribute to the differentiation were determined by the supervised partial least squares discriminant analysis (PLS-DA). The threshold of VIP scores was set larger than 1.5, which showed significant differences for each discriminant metabolite between different plant types. By combining the results from the PLS-DA and Volcano plot analyses, Table D.1 (see Appendix D) illustrates that 121 peak pairs had significant changes in different TIFY levels. Table D.1 shows the average $^{12}C/^{13}C$ ratio of these peak pairs in each plant type, with the % RSD included in parentheses. I found that the range of % RSD is generally from 2% to 30%, indicating that a good reproducibility was obtained. However, I noted that for a few peak pairs, the % RSD was larger than 40% and a detailed examination of the data variations revealed that the large %RSD was mainly attributed to the biological variations. For putative metabolite identification based on accurate mass of each underivatized metabolite matches with metabolites in HMDB and MycompoundID, a mass accuracy tolerance of 10 ppm was used. I used a database of human metabolites because an equivalent resource is not yet available for plant metabolites.

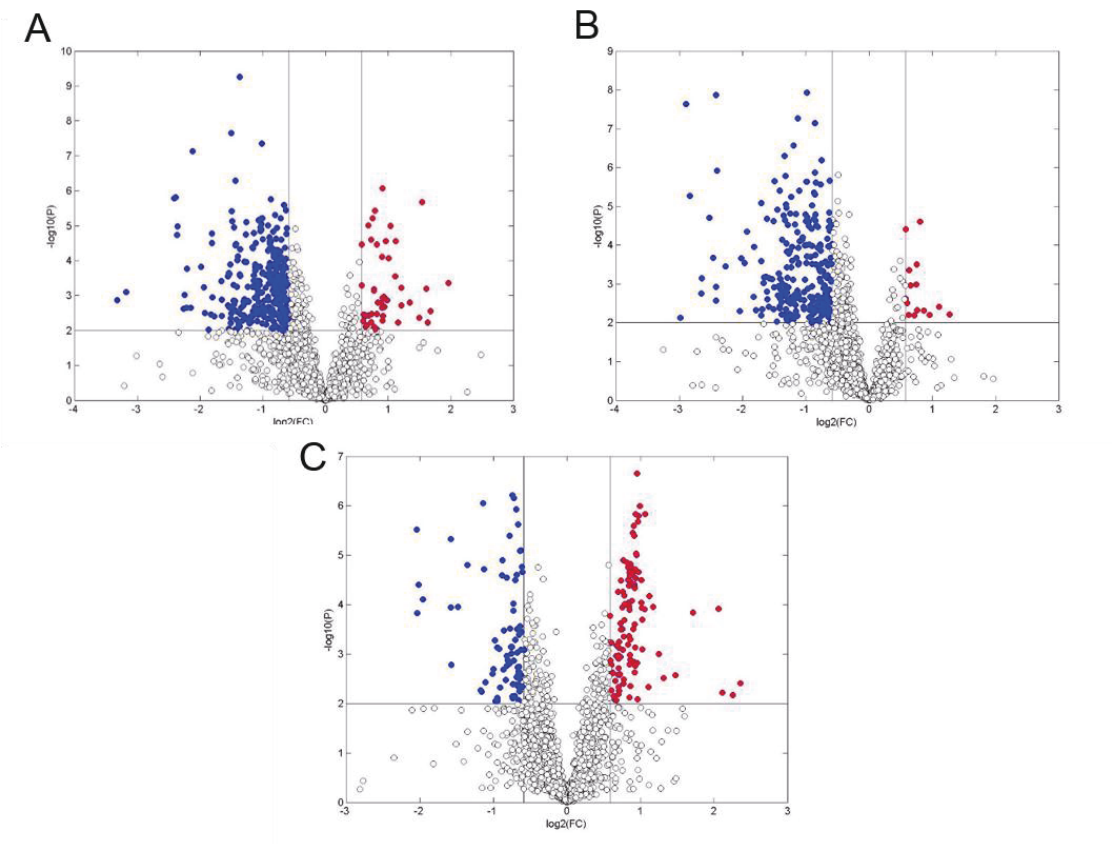


Figure 5.4 Volcano plot. (A) KOT verse WT (B) OXT verse WT (C) OXT verse KOT. The threshold cutoff values were 1.5-fold change and p-value 0.01.

5.3.4 Metabolites Affected by Methyl Jasmonate (MeJA) Treatment

In order to study natural metabolic changes of *Arabidopsis thaliana* under MeJA treatment, the wild type plants were treated with MeJA for 0.5 h and 24 h, respectively. I focused on the trends of metabolites that were always up-regulated or down-regulated during the 24-h MeJA treatment and used the data from 0.5 h treatment to confirm the trends. The microwave assisted extraction method was applied to study metabolic changes of wild type plants under MeJA treatment, including 0.5 h and 24 h. Untreated wild type plants (0 h) were also planted and considered as a control group. A reference sample was generated by mixing aliquots of individual plant samples

and labeled with ^{13}C-dansyl chloride. An equal volume of the ^{12}C-dansyl labeled individual sample was mixed with the ^{13}C-dansyl labeled pooled sample, followed by running the mixture in LC-MS. Each sample was analyzed in biological triplicates and experimental duplicates.

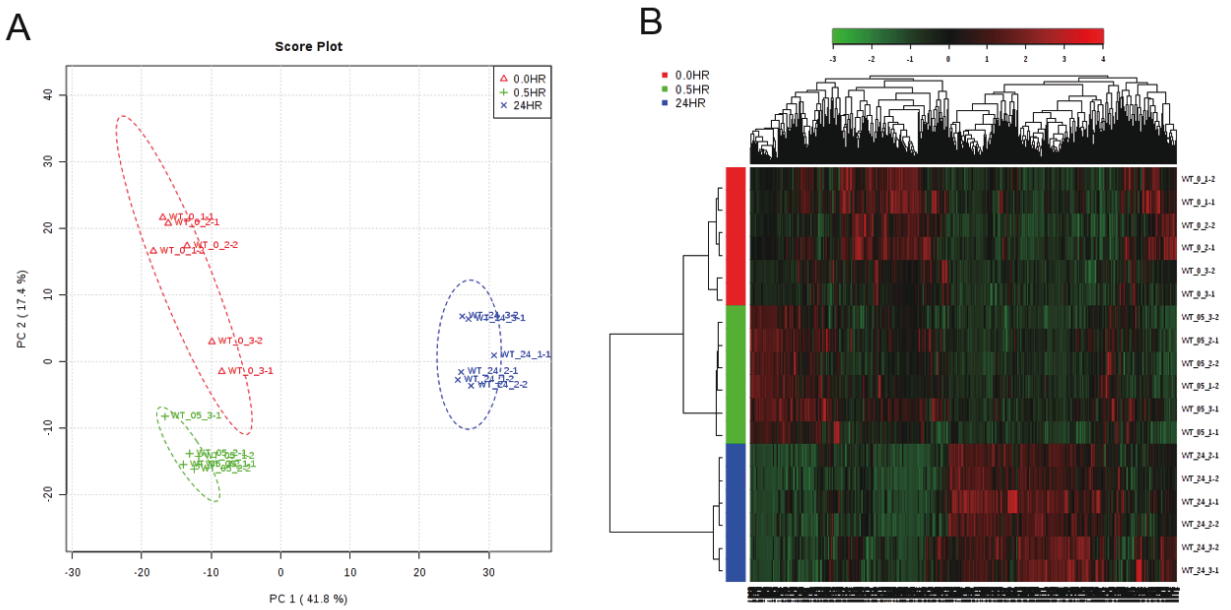


Figure 5.5 (A) PCA score plot (B) Clustered heatmap. Both show comparison between three different treatment time points, including 0.0 h (without treatment), 0.5 h and 24 h. Each time point was done in biological triplicates and experimental duplicates.

Figure 5.5A shows the PCA score plot which generated an overview on how the data were scattered and the first two principal components captured 69% of the total variation. It clearly demonstrated that plants with different MeJA treatment time points can be well separated. Especially after 24 h MeJA treatment, samples were significantly distinguished from samples without treatment and with 0.5 h treatment. This suggests that the metabolome of *Arabidopsis thaliana* has undergone significant variations after 24 h treatment. In addition, a clustered heatmap (Figure 5.5B) illustrates the relative intensities of same peak pairs detected among these different

treatment time points. I found that most of peak pairs were affected after 24 h MeJA treatment. The results of the PCA and the heatmap also showed the potential of this method for discriminating plants with different treatment time points. In order to gain information of metabolic trends for up-regulation and down-regulation, I used Volcano plot and VIP of PLS-DA to do data comparison between samples without treatment and with 24h treatment, followed by checking the trends with those treated with 0.5 h. Table D.2 (see Appendix D) shows the list of peak pairs with up-regulated or down-regulated trends caused by MeJA treatment. I found that after 24 hr MeJA treatment, 30 of the peak pairs picked were up-regulated, while 23 of those were down-regulated. Table D.2 contains the average $^{12}C/^{13}C$ ratio of these peak pairs in each treatment time point, with the % RSD included in parentheses. Most of peak pairs had %RSD smaller than 35%; however, there were a few peak pairs with higher %RSD because of biological variations.

5.3.5 Metabolites Affected by MeJA Treatment and TIFY Expression Levels

Table D.2 summarizes 53 peak pairs with up-regulated or down-regulated trends induced by MeJA treatment. In order to recognize whether these peak pairs were in response to different TIFY expression levels, KOT and OXT plants were also analyzed without or with MeJA treatment. The same pooled sample was used to determine the peak ratios of each individual peak pair from different samples with different concentrations of the same putative metabolite.

Figure 5.6 shows a heatmap illustrating concentration changing trends of 53 peak pairs induced by MeJA treatment in different TIFY expression level plants. They were divided into 4 patterns based on the concentration changing trends. Group 1 indicated that the concentration changing trends of 14 peak pairs in KOT, WT, and OXT plants were the same, Group 2 showed the concentration changing trends in KOT plants were similar with those in OXT plants, and Group

3 showed the trends in KOT plants were similar with WT plants. In these three groups, I didn't find significant evidence to show metabolic changes were related to TIFY expression.

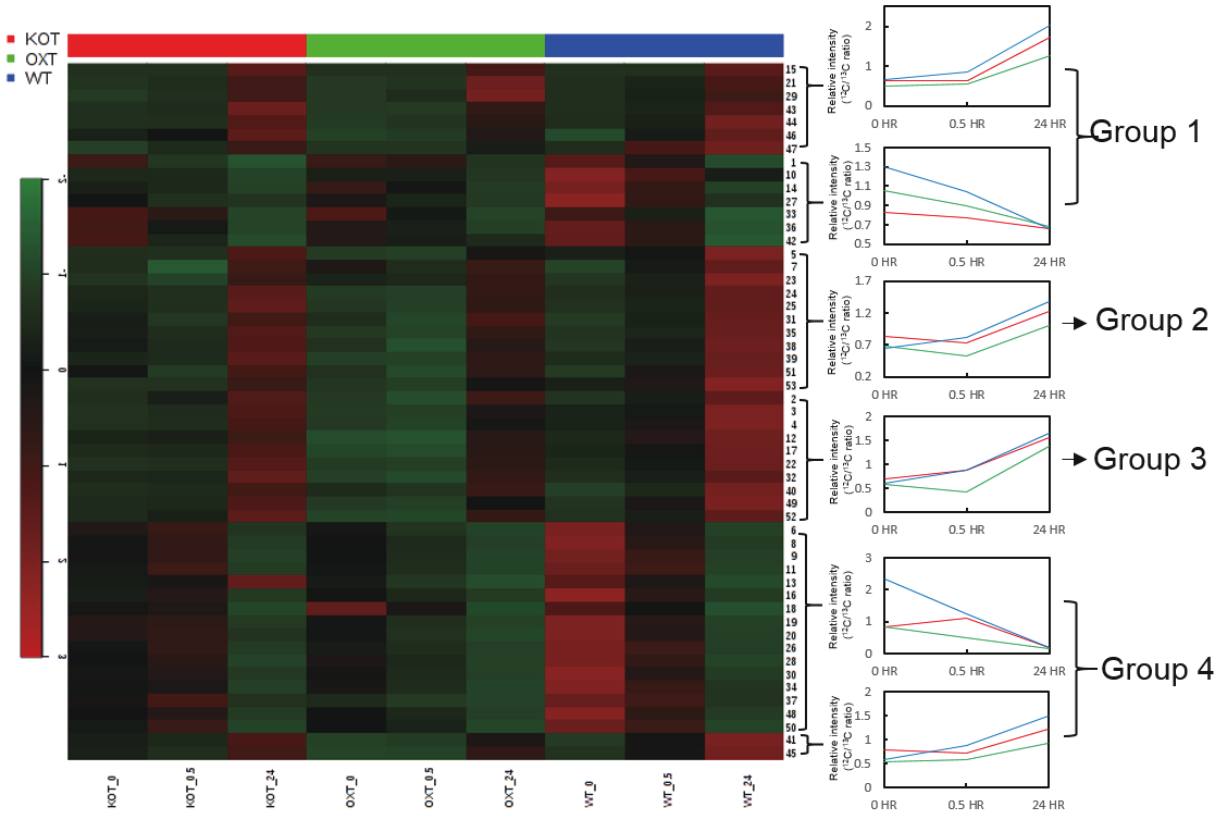


Figure 5.6 A heatmap shows concentration changing trends of 53 peak pairs induced by MeJA treatment in different TIFY expression level plants.

I was more interested in Group 4 because this group showed that the concentration changing trends of 18 peak pairs (italic font in Appendix D, Table D.2) in OXT plants were similar with those in WT plants, but the slopes of concentration change in OXT plants were smaller than those in WT plants. It showed that the rate of production or consumption of the same metabolites in OXT plants was slower than those in WT plants. Besides, at the beginning half hour treatment, the trends of the same metabolites in KOT plants were always opposite to those in WT and OXT plants. After induced by MeJA treatment, plants would produce or consume certain compounds to help activate

the *TIFY* gene. However, KOT plants do not contain *TIFY* gene so those compounds would not be used. The trends in KOT plants were different from those in WT and OXT plants. Therefore, the putative metabolites in Group 4 were relative not only to MeJA treatment, but also to *TIFY* gene expression levels.

5.4 Conclusion

The optimization process successfully defined a simple extraction protocol that enabled identification of potentially hundreds of metabolites from a very small (~40 mg) mass of starting tissue in a few hours. In addition, a reproducible and high-quality separation method has been developed for analyzing the isotope-tagged metabolites from *Arabidopsis thaliana* for 31 min. More than 700 peak pairs were detected by this optimal method with good reproducibility (84%) in triplicate runs.

Although absolute concentrations of metabolites could not be measured due to the lack of proper standards, relative concentrations could be determined by using a labeled pooled sample. This way of relative quantification allowed me to compare the same metabolites in different individual samples. In different gene expression level plants, there were 121 peak pairs with significant metabolic changes due to different TIFY levels and the relative concentrations of most of them were higher in WT plants. In the study of MeJA treatment, most of peak pairs were affected after 24 hr MeJA treatment, and 53 peak pairs always showed up-regulated or down-regulated trends. 18 out of the peak pairs picked were also affected by *TIFY* gene expression levels. In the present work, putative metabolite identification was based on accurate mass of each underivatized metabolite matches with metabolites in HMDB and MycompoundID. In the future, the work on identification of metabolites with authentic standards will be performed.

Chapter 6

High-Performance Isotope-Labeling Liquid Chromatography Mass Spectrometry for Investigating the Effect of Drinking Goji Tea on Urine Metabolome Profiling*

6.1 Introduction

Urine metabolomics has been widely used to search for potential metabolite biomarkers for disease diagnosis or prognosis.[171-177] However, urine metabolome can be very complex because, along with human endogenous metabolites, exogenous metabolites (such as those from food, drinks, nutritional supplements, and pharmaceuticals) can also be present in urine. The presence of these exogenous metabolites in significant amounts or varieties in urine can complicate biomarker discovery, as any chemical residues from food, drinks, etc. can affect the overall urine metabolome composition.[178] Another potential complication in urine metabolomics is related to the effect of diet or other factors on metabolome profiles.[179-183] For example, diet-induced metabolome changes may mask the differences of the metabolomes between a diseased group and a control group. Thus, it is critical to minimize the influence of these external factors in the profiling of endogenous metabolites in urine samples for discovering specific disease biomarkers.

The effect of external factors on the urine metabolome may not be detectable if an analytical technique is not sufficiently sensitive or quantitative. Technological advances in recent years have made it possible to detect and quantify many metabolites present at low concentrations in urine.[184,185] Although sensitive techniques are very useful for detecting low-abundance metabolites as potential biomarkers, they also increase the likelihood of detecting exogenous metabolites in an untargeted metabolome profiling work, as well as of revealing subtle changes due to external factors such as diet. Thus, it is necessary to pay special attention to the potential interferences of external factors for urine metabolomics. We have recently developed a highly sensitive isotope

labeling liquid chromatography mass spectrometry (LC-MS) for quantitative metabolomics based on rational design of chemical labeling reagents for simultaneous improvement in metabolite separation, detection, quantification and identification.[88,95] For example, urine metabolites that contain primary amine, secondary amine, or phenolic hydroxyl groups can be labeled with ^{12}C- or ^{13}C-dansyl chloride prior to LC-MS analysis.[88] This labeling strategy allows the separation of polar or ionic metabolites on a reversed phase (RP) column, as well as signal enhancement of tenfold to a thousandfold, compared to the unlabeled counterparts. In a human urine sample, we can detect more than 1000 putative metabolites from more than 10000 features typically found in a 30-min LC-MS run.

In this chapter, I report my investigation of using the dansylation isotope labeling LC-MS method to study the effect of drinking Goji tea on urine metabolome profile. The objectives of this study were: (1) to present the metabolome profile of Goji, which has never been reported; and (2) to examine how the intake of Goji tea affects the urine metabolome of a small number of healthy individuals. The ultimate goals were to determine the external factors that can cause significant alternation of the urine metabolome, as well as to make recommendations to avoid or properly control these external factors during sample collection, in order to eliminate or reduce their influence in profiling the human endogenous metabolites in urine. This work was not intended to study the effect of drinking Goji tea on human health, which would require a much larger sample size from diverse populations.

Goji tea is considered a nutritional supplement in some regions of the world. Goji, the red and sweet fruit of *Lycium barbarum*, is native to southeastern Europe and Asia,[186] and commonly used in meals, drinks, and traditional Chinese medicine (TCM).[187] Like many fruits, Goji contains many nutrients including vitamins, trace minerals, amino acids, and essential oils,[188-190] as well as active ingredients such as *Lycium Barbarum* polysaccharides (LBP), scopoletin, zeaxanthin

dipalmitate, and β-carotene.[186,190-193] Dried Goji berries and Goji tea made from the berries are being sold as nutritional supplements or health food products. In this chpater, I applied the very sensitive dansylation LC-MS method of targeting amine- and phenol- containing metabolites to analyze the urine metabolomes of healthy individuals with and without drinking Goji tea to determine whether drinking Goji tea would have any effects on their urine metabolome profiles.

6.2 Experimental

6.2.1 Reagents and Chemicals

^{12}C-dansyl chloride was purchased from Sigma-Aldrich Canada (Markham, ON, Canada). ^{13}C-dansyl chloride was synthesized in our lab[88] and the chemicals used to synthesize this isotope reagent were also purchased from Sigma-Aldrich. LC/MS-grade water, acetonitrile (ACN), and formic acid were purchased from Thermo Fisher Scientific (Edmonton, AB, Canada).

6.2.2 Goji Tea

Goji fruits purchased from a local Chinese herbal medicine store were imported from Qinghai, China. Goji tea was made by mixing 15 g Goji in 500 mL water, heating the water to its boiling point, and then leaving it to stand for 1 h at room temperature. For Goji metabolome analysis, two aliquots of 50 μL of Goji tea were labeled with ^{12}C- and ^{13}C-dansyl chloride, respectively. A dansylated Goji tea sample was prepared by mixing equal volumes of ^{12}C- and ^{13}C-dansyl labeled solutions. The mixture sample was centrifuged for 10 min at 14000 rpm, after which it was ready for analysis.

6.2.3 Urine Sample Collection

Urine samples were collected from seven healthy individuals, males and females, with ethics approval from the University of Alberta. The samples were stored at 4 °C immediately after

collection. The urine samples were centrifuged at 4000 rpm for 10 min, after which the supernatants were filtered twice through a 0.2 μm filter. The individually filtered urine was aliquoted and stored at -80 °C until further use.

6.2.4 Dansylation Labeling Reaction

The frozen urine samples were thawed in an ice bath and then diluted twofold, followed by dansylation labeling.[88,116] The 50 μL sample was mixed with 25 μL sodium carbonate/sodium bicarbonate buffer (500 mM) and 25 μL ACN in a reaction vial. The solutions were vortexed, spun down, and mixed with either 50 μL of freshly prepared ^{12}C-dansyl chloride solution (18 mg/mL) for light labeling or ^{13}C-dansyl chloride (18 mg/mL) for heavy labeling. The dansylation reaction was allowed to proceed for 60 min at 40 °C. After 60 min, the mixtures were vortexed and spun down, and then 10 μL of sodium hydroxide (250 mM) was added to the reaction mixture to consume the excess dansyl chloride and quench the dansylation reaction. After an additional 10 min of 40 °C incubation, a 50 μL formic acid (425 mM in 50% acetonitrile) was added to consume the excess sodium hydroxide and to adjudge pH value. The ^{12}C-labeled or ^{13}C-labeled mixtures were centrifuged for 10 min at 14000 rpm, after which they were ready to be injected onto an ultrahigh performance liquid chromatography (UPLC) column for UV quantification.[194] According to the UV quantification results, an appropriate volume of ^{13}C-labeled sample was combined with a ^{12}C-labeled sample in a fixed amount ratio. The mixtures were centrifuged for 10 min at 14000 rpm and were then ready to be injected into LC-MS for analysis.

6.2.5 LC-UV Quantification

A Waters ACQUITY UPLC system (Milford, MA, USA) with a binary solvent manager, a sampler manager, and a photodiode array (PDA) detector was used for the quantification step. First, 2 μL of the labeled urine solutions were injected onto a Waters ACQUITY BEH C18 column (2.1

mm × 5 cm, 1.7 μm particle size, 130 Å pore size) for a fast step-gradient run. Solvent A was 0.1% (*v*/*v*) formic acid in 5% (*v*/*v*) acetonitrile; Solvent B was 0.1% (*v*/*v*) formic acid in acetonitrile. The gradient started with 0% B for 1 min and was stepwise increased to 95% within 0.01 min and held at 95% B for 1 min to ensure complete elution of all labeled metabolites. The gradient was restored to 0% B in 0.5 min and held in this condition for 3.5 min to re-equilibrate the column. The flow rate was 0.45 mL/min. The detection wavelength was set at 338 nm.

6.2.6 LC-MS Analysis

The HPLC system consisted of an Agilent 1100 liquid chromatograph (Palo Alto, CA, USA) equipped with a binary pump system. The labeled urine samples were separated with a reversed-phase Eclipse plus C18 column purchased from Agilent (2.1 mm×100 mm, 1.8 μm particle size, 95 Å pore size) in an optimal injection amount. Solvent A was 0.1% (*v*/*v*) formic acid in 5% (*v*/*v*) ACN; Solvent B was 0.1% (*v*/*v*) formic acid in ACN. The gradient started at 20% B and stepwise increased to 35% B within 3.5 min, followed by 14.5-min ramp to 65% B, then raised to 99% B within 6 min and finally held at 99% B for 4 min to ensure complete elution of all labeled metabolites. The flow rate was 180 μL/min. The flow from HPLC was loaded to the electrospray ionization (ESI) source of a Bruker maXis Impact Ultra-High Resolution Quadrupole Time-of-Flight (QTOF) mass spectrometer (Bruker, Billerica, MA, USA). The MS conditions were as follows: nebulizer gas pressure 1.8 bar; drying gas temperature 230°C; drying gas flow 8 L/min; capillary voltage 4500 V; and scan range 150–800. All MS spectra were obtained in the positive-ion mode.

6.2.7 Data Processing and Metabolite Database Search

The steps of data processing were the same as those in Chapter 3.2.6. Putative metabolite identification was based on accurate mass matches of underivatized metabolites in HMDB[89] and MycompoundID[147], using a mass accuracy tolerance of 10 ppm.

6.3 Results and Discussion

6.3.1 Study Design

Urine typically contains many endogenous metabolites as well as exogenous metabolites; the latter are metabolic breakdown compounds from a wide range of foods, drinks, nutritional supplements, and other sources. The presence of exogenous metabolites in high abundance may complicate the discovery of potential biomarker metabolites from urine. To study the effect of taking nutritional supplements (in this case, Goji tea) on urine metabolome, I designed two experiments: one to study the short-term effect (<3 h) and the other to study a longer-term effect (12 h). Figure 6.1 shows the workflows of urine sample collections for both experiments. To study the short-term effect (Figure 6.1(a)), after overnight fasting, five urine samples were collected from one volunteer each day for three continuous days. The first and second urine samples for each day were used as the control group. After the second urine sample was collected, the volunteer drank 500 mL of Goji tea. Three urine samples were then collected at 1-h intervals and treated as an experimental group. This workflow yielded a total of 15 urine samples from one individual. Each urine sample was analyzed in experimental triplicates.

In studying the longer-term effect (Figure 6.1(b)), urine samples were collected from six volunteers (three males and three females) over six continuous days. For the first three days, volunteers were asked to drink 500 mL of water in the evenings, 12 h before their urine samples were to be collected. These urine samples were used as the control group. For the following three

days, they were asked to drink 500 mL of Goji tea in the evenings, 12 h before their urine samples were to be collected. These urine samples were treated as an experimental group. This workflow generated 36 urine samples in total. For each sample, experimental duplicates were performed.

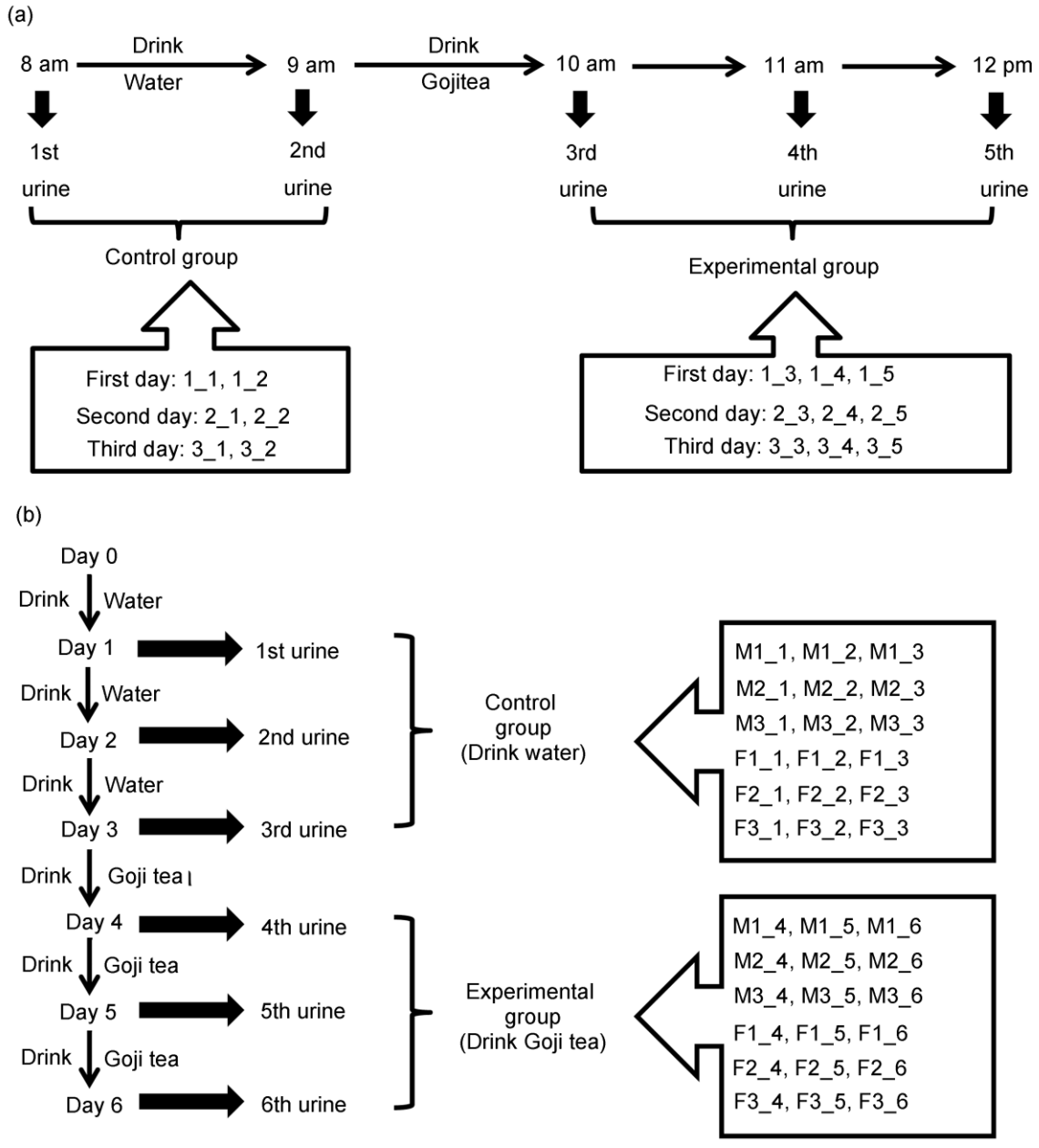


Figure 6.1 Experimental design for urine sample collection: (a) studying the short term effect of drinking Goji tea on urine metabolome and (b) studying the longer term effect of drinking Goji tea on urine metabolome (drinking water was used as a control).

6.3.2 Goji Metabolome Profiling

Before analyzing the human urine samples, I profiled the Goji metabolome to determine the detectability of the dansyl-labeling LC-MS method on the Goji metabolites. Figure 6.2 shows the base-peak ion chromatograms of two separate LC-MS runs from the same injection amount of plain Goji tea and dansylated Goji tea. Many peaks with high intensities were detectable in the dansylated Goji tea, whereas only a few low-intensity peaks were detected in the plain Goji tea at the same injection amount.

Because the dansylated Goji tea sample was prepared by mixing equal volumes of ^{12}C- and ^{13}C-dansyl-labeled aliquots of the same tea sample, the expected peak intensity ratio for each labeled metabolite should have been closed to 1.0. The IsoMS software was used to process the LC-MS data and calculate the peak ratio for each peak pair picked by the software. I excluded <10% of the overall number of peak pairs detected that had ratios of <0.67 and >1.5 (i.e., the outliers due to experimental variations). The final results show that the average number of peak pairs or putative metabolites detected in the Goji tea was 1629±15 (n=3) and that 1475 (91%) metabolites were commonly found in the triplicate experiments.

To putatively identify these peak pairs, the accurate mass of each metabolite from an individual peak pair was fed into a search using MyCompoundID against the HMBD and EML databases. Of the 1475 commonly found metabolites, 361 peak pairs matched the HMDB metabolites and 575 matched the predicted metabolites in EML, for a total of 936 matches (63%) (Appendix E, Tables E.1 and E.2). Thus, many of the detected peak pairs could be putatively identified based on this accurate mass match. Of course, authentic standards are needed for positive identification, which is beyond the scope of this work. It should be noted that I used a database of human metabolites for matches because an equivalent resource is not available for plant

metabolites. One would expect that plants and humans share many similar metabolic processes and that they should share many common metabolites. I reasoned that the unique metabolites found in Goji tea, which are different from humans either in identity or quantity, could serve as the tracers to examine the effect of drinking Goji tea on the human urine metabolome profile.

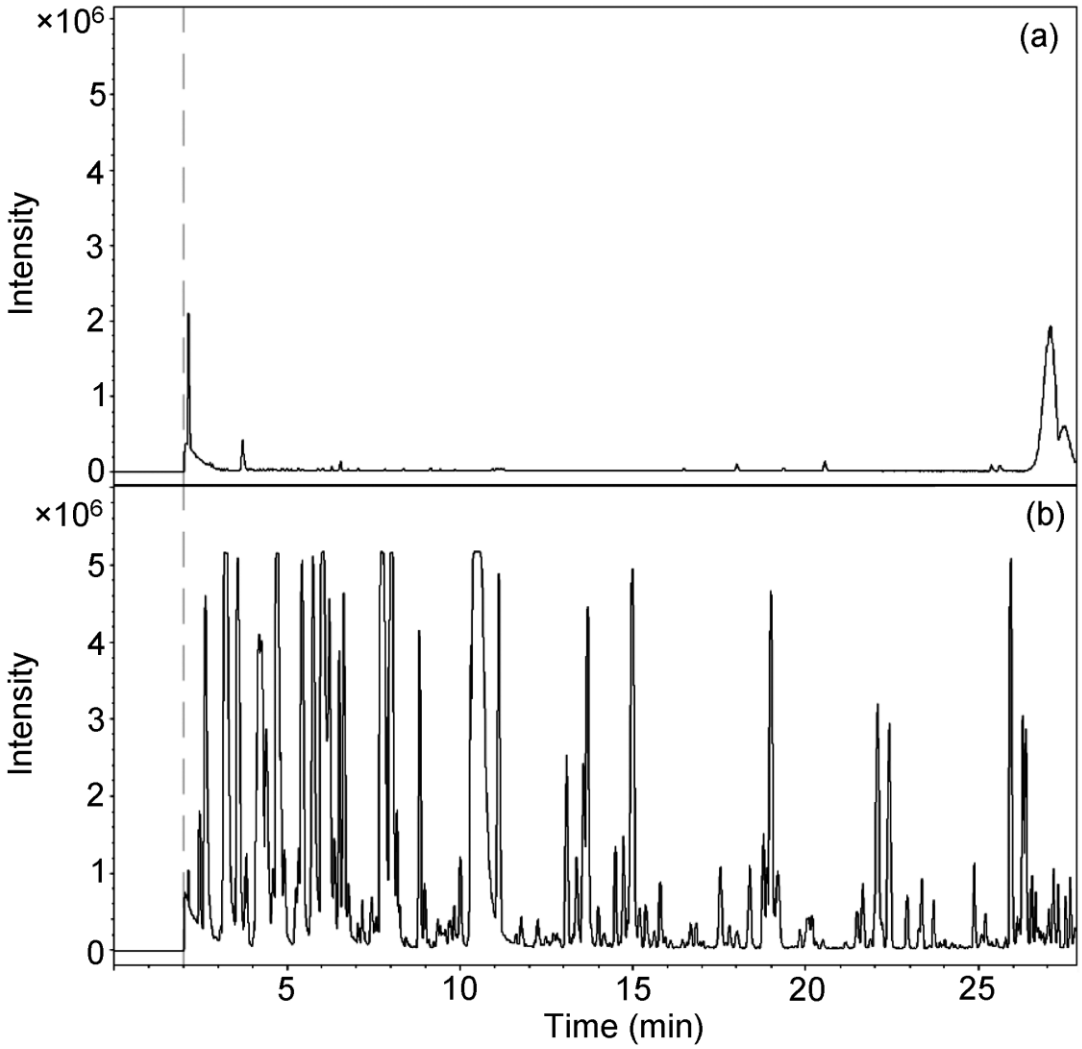


Figure 6.2 Base peak ion chromatograms obtained by LC-MS from (a) plain Goji tea and (b) dansylated Goji tea.

6.3.3 Short Term Effect on Urine Metabolome

To generate a reference sample for comparison with the individual samples, aliquots of individual urine samples were taken and mixed to produce a pooled sample, which was then labeled by ^{13}C-dansyl chloride. An equal amount of ^{12}C-dansyl labeled individual sample was mixed with the ^{13}C-dansyl-labeled pooled sample, followed by running the mixture in LC-MS. The peak-intensity ratios of the isotope-labeled peak pairs detected in the mass spectra were calculated. Because the same pooled sample was used, the peak ratios of each individual peak pair from different samples represented the different concentrations of the same putative metabolite in those samples. The peak ratio values found in different samples were then exported into a statistics tool for multivariate analysis. The results of the analysis are shown in Figure 6.3; each point represents an average of triplicate runs. Principal component analysis (PCA) was applied to generate an overview on how the data were scattered (Figure 6.3(a)). The PCA plot shows that there is no apparent separation between the control group and the experimental group. The data points appear to be clustered by different days.

In order to determine if the control group and the experimental group could be separated, partial least squares discriminant analysis (PLS-DA) was applied; the resulting score plot is shown in Figure 6.3(b). This plot shows that the two groups are separated from each other. Next, I applied the permutation test (n=100) to the PLS-DA model to determine whether this separation is valid. Figure 6.3(c) shows that the observed statistic p-value is 0.83 which is much greater than the validation threshold of 0.01. Thus, the PLS-DA model is not valid, which indicates that the two groups are not really separate from each other.

I also analyzed the data using a Volcano plot, which is another statistical tool that combines the fold-change and t-test information to determine the significant metabolites that can be used to

separate the two groups. Using the twofold change and a *p*-value of 0.05 as the threshold cutoff values, the Volcano plot is shown in Figure 6.3(d). Only six peak pairs are found to be significant; they appear in the upper right. Positive fold changes of these six peak pairs indicate that the concentrations of these putative metabolites are higher in the control group. Three out of the six peak pairs could match with some metabolites in HMDB[89] or METLIN[195] (Appendix E, Table E.3). The *p*-value of these metabolites for group separation ranges from 0.015 to 0.049. If a cutoff value of 0.05 is used, these metabolites would be considered significant.

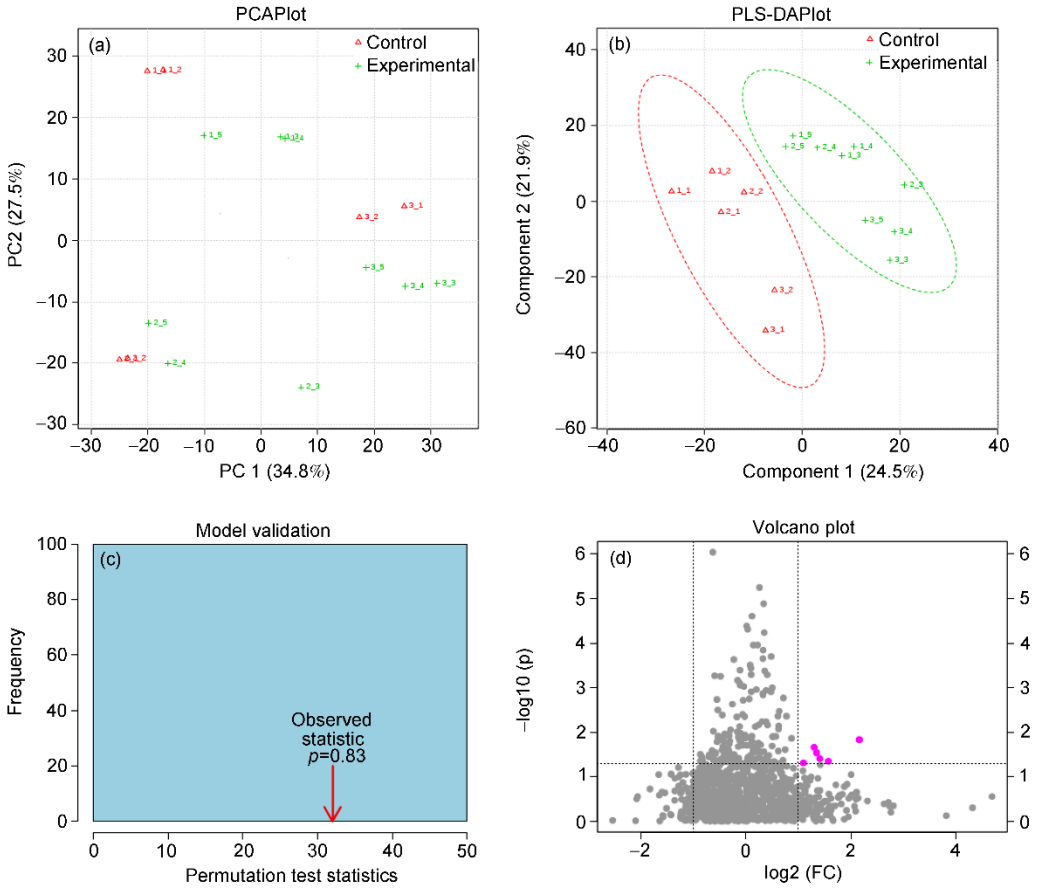


Figure 6.3 Multivariate analysis of the metabolome data obtained from the urine samples collected in the short term effect study: (a) PCA score plot, (b) PLS-DA score plot (c) permutation test of the PLS-DA model, and (d) Volcano plot. Each point in PCA and PLS-DA score plots represents the average of three replicate experiments.

In my work, a moderate amount of Goji tea from a typical-sized drink was consumed by an individual, since it would be a better representation of the participants in a biomarker discovery study in which some participants may ingest nutritional supplements (such as Goji tea) in moderate amounts. Results from the PLS-DA and Volcano plot analyses indicate that there is little or no separation between the control and experimental groups. In other words, drinking a moderate amount of Goji tea (500 mL) did not have a significant effect on urine metabolome even after the urine samples were collected within 1 to 3 h.

To compare the urine metabolomes at the individual metabolite level, I examined the metabolite distributions in the Goji tea and urine samples collected before and after subjects drank the Goji tea. Figure 6.4(a) shows a comparison of the number of peak pairs found in urine before drinking tea (control group) and in the Goji tea itself. The total number of peak pairs detected in the six control-urine samples is 2168. Of these, 622 (18%) peak pairs match the total number of peak pairs found in the Goji tea (i.e., 1917 from three runs). Figure 6.4(b) shows a comparison of the peak pair numbers found in urine before and after subjects drank the Goji tea (i.e., control group vs. experimental group). The total number of peak pairs detected in the experimental group is 2187. Of these, 2143 (98%) peak pairs are found in both groups. Although there are 25 unique peak pairs found in the control group and 44 unique peak pairs in the experimental group, most of these peak pairs are only found in an individual sample or were found in one day. After removing these pairs, only two unique peak pairs are consistently found in the experimental group. Comparing these two pairs with those found in the Goji tea, one with m/z 285.0644 and a retention time of 1283 s was also observed in the Goji tea. As shown in Figure 6.4(a), the number of the unique peak pairs found in the Goji tea itself is 1295; however, only 1 could be observed in the urine samples after subjects drank the Goji tea. It is clear that the concentrations of these compounds from the Goji tea were so diluted that they could not be detected in the urine samples.

Alternatively, these compounds might have been absorbed or converted into other metabolites by the human body. In any case, these results are consistent with the global classification results generated by the PLS-DA and Volcano plot models: the urine metabolome was not significantly affected by drinking a moderate amount of the Goji tea.

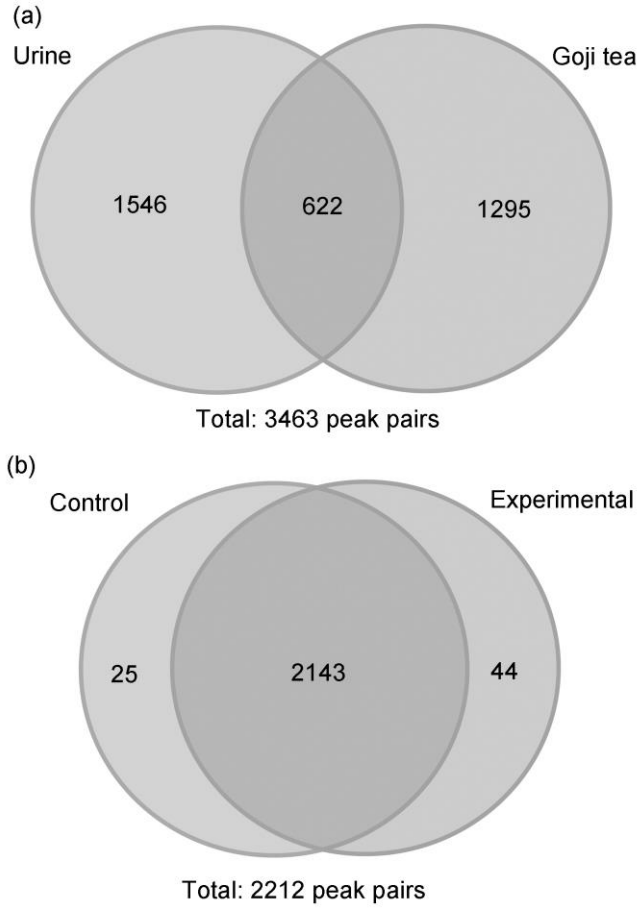


Figure 6.4 Venn diagrams of the distributions of the numbers of peak pairs or putative metabolites detected: (a) comparison between the number detected in urine before drinking Goji tea (control group) and the number detected in the Goji tea itself, and (b) comparison between the number detected in urine before drinjing Goji tea (control group) and the number detected in urine after drinking Goji tea (experimental group).

6.3.4 Longer Term Effect on Urine Metabolome

Although the results described above illustrate that there is no short-term effect on urine metabolome, I cannot rule out the possibility of some metabolism occurring over an extended period, after the intake of food or drink that would result in the late appearance of these metabolites in urine. Fasting for 12 h is a common practice before urine- sample collection. I examined the effect on urine metabolome of drinking Goji tea 12 h before sample collection. To compare the urine metabolomes from six individuals and six days, a pooled sample was produced by mixing aliquots of individual urine samples and labeling them with ^{13}C-dansyl chloride. An equal amount of ^{12}C-dansyl-labeled individual sample was mixed with ^{13}C-dansyl-labeled pooled sample, followed by running the mixture in LC-MS. The whole dataset of 36 samples was analyzed by statistical tools; results are shown in Figure 5. Each point represents an average of duplicate experiments.

Figure 6.5(a) depicts the PCA score plot, which does not show any clear separation between drinking water or Goji tea 12 h before urine samples were collected; moreover, most of the data points are clustered within an individual. The PLS-DA score plot is shown in Figure 6.5(b) and the permutation test (n =100) result of the PLS-DA model is shown in Figure 6.5(c). The p-value from the test is 0.98 which indicates that the model is not valid (i.e., there is no separation between the control and experimental groups). This conclusion is supported by the Volcano plot shown in Figure 6.5(d), where there is no significant metabolite found that would differentiate the two groups. These results indicate that 12 h after taking the Goji tea, there was no effect on the urine metabolome profiles.

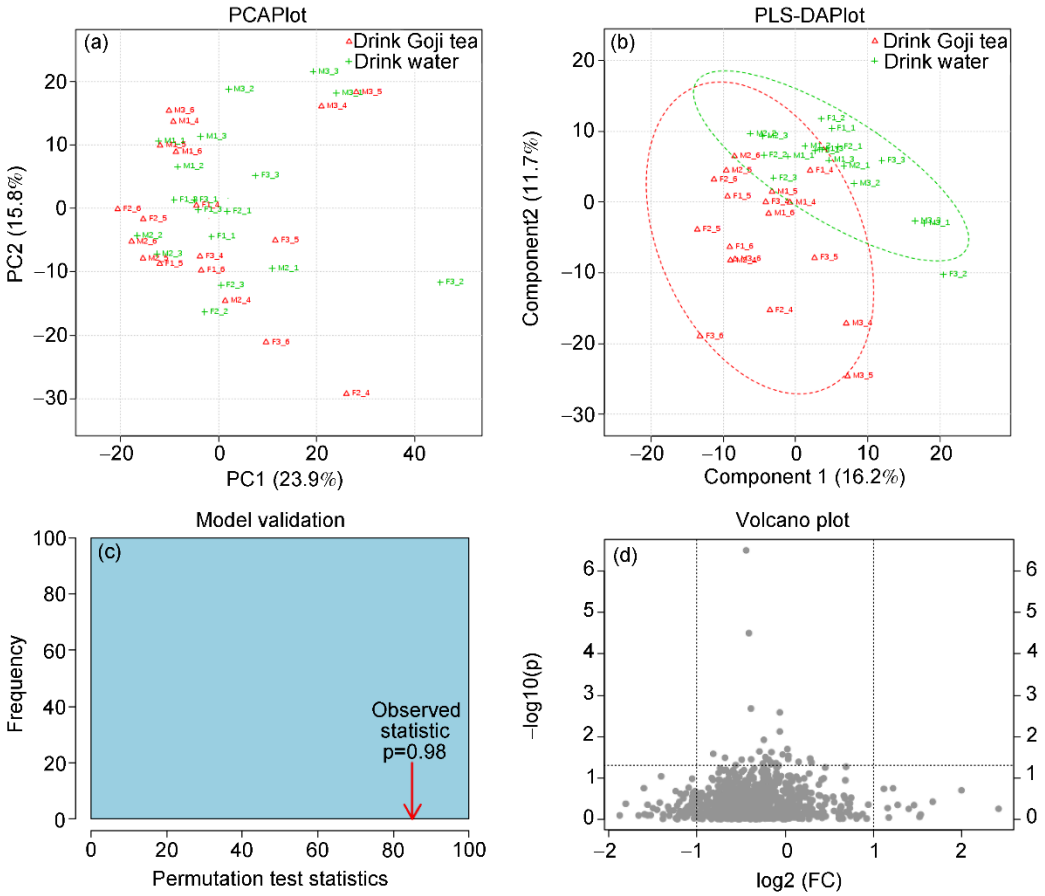


Figure 6.5 Multivariate analysis of the metabolome data obtained from the urine samples collected in the longer term effect study: (a) PCA score plot, (b) PLS-DA score plot (c) permutation test of the PLS-DA model, and (d) Volcano plot. Each point in PCA and PLS-DA score plots represents the average of two replicate experiments.

6.4 Conclusions

In this work, a dansylation-isotope-labeling LC-MS method was used to study the effect of drinking Goji tea on human urine metabolome profiles. This sensitive method allows the detection of an average of 1629±15 (n=3) peak pairs or putative metabolites in the Goji tea itself, compared to very few metabolites detectable without labeling using the same amount of injection into LC-MS. From the studies of the short-term (<3 h) and longer-term (12 h) effects of drinking Goji tea,

it is clear that the consumption of a moderate amount of Goji tea does not affect the urine metabolome significantly. Fasting for 12 h should be sufficient to remove any potential interference of the Goji metabolites in human urine metabolome profiling. Although this work focused only on the profiling of amine- and phenol- containing metabolites in urine, the high sensitivity of the technique allows the examination of even the smallest changes in urine metabolome. I have shown that a conventional LC-MS method, without dansylation labeling of the metabolites, could detect only very few metabolites in Goji tea that was prepared at a typical concentration used for consumption. Therefore, I did not expect that a conventional LC-MS could detect any Goji metabolites in urine after consumption. In the future, I will apply this sensitive analytical approach and the same study design to examine the effects of other types of food or drinks on urine metabolome, with the goal of developing guidelines about avoiding certain types of commonly consumed food and drinks for urine metabolomics-based biomarker discovery studies.

Chapter 7

Conclusions and Future Work

The purpose of metabolic profiling is to identify or quantify a number of metabolites from biological samples; however, it is a challenge to achieve comprehensive metabolomic profiling owing to the complexities of the metabolites with diverse chemical and physical properties as well as a vast dynamic range of concentrations. Reverse phase liquid chromatography coupled with mass spectrometry (RPLC-MS) is one of the important analytical techniques in metabolomics and has a great potential to be the mainstream technique for metabolomic profiling. RPLC not only provides premium resolution to most of medium polar and small non-polar analytes, but also deals with ion suppression effect and complexity of the mass spectra in the time dimension. Furthermore, it supplies the information of retention time of analytes that can be used as a criterion for metabolite identification. MS measures the mass-to-charge ratio (*m/z*) of charged analytes that is used for mass determination, and the information of *m/z* can be used as another criterion for identification. However, not all of metabolites can be separated and analyzed by RPLC-MS because a large proportion of metabolites belong to highly polar compounds. In addition, datasets generated by RPLC-MS based metabolomic experiments are normally huge and complicated due to chemical noise, instrument noise or contamination. In order to overcome the limitations of RPLC-MS for metabolome analysis, our lab has been involved in the development of differential isotopic labeling (DIL) methods over the past several years. Compared to stable isotopic labeling (SIL) method, a DIL method can effectively compensate the matrix effect and instrument drift without purchasing expensive and limited isotopic internal standards for analyzing metabolites in a complex biological sample. In our group, a differential ^{13}C-/^{12}C-isotope dansylation labeling approach has been developed for studying amine- and phenol-containing metabolomic profiling. The hydrophobicity

of polar and ionic amine- and phenol-containing metabolites is increased by dansylation; therefore, most of amine- and phenol-containing metabolites can be retained and separated by RPLC-MS. Thus, the metabolome coverage for comprehensive metabolomics is also improved.

The aims of this thesis work focused on the establishment of a data processing method and software to quantify and identify metabolites, the development of extraction methods for different plant samples, and the application of a differential ^{13}C-/^{12}C-isotope dansylation labeling technique targeted at amine- and phenol-containing metabolites in plant and human urine samples. All of the samples were derivitized by ^{13}C-/^{12}C-isotope dansylation reaction and analyzed by RPLC-MS. The relative quantification, peak pair identification, and putative identification of metabolites can be achieved by this approach. An overview of metabolomics, chemical derivatization and several key instruments related to MS-based metabolomics studies are briefly introduced in Chapter 1.

A reliable software tool for processing the DIL method data is an urgent requirement of high throughput RPLC-MS analysis in metabolomics. In Chapter 2, a software tool, named IsoMS, was developed and described for processing raw LC-MS data by exploring a mass spectral feature unique to the chemical labeling approach. The software can automatically pick up peak pairs with a defined mass difference governed by the mass difference of the heavy- and light-chain tag, and generate their relative quantitative information by calculating peak ratio values of given peak pairs, as well as filter various noise caused by solvents or contamination. The pair list exported by IsoMS can be used for further data and statistical analysis. The performance of IsoMS has been tested by various samples labeled with ^{13}C-/^{12}C-dansyl chloride in our group. In the current work, we only use Levels 1 and 2 peak pairs for further analysis. The differentially dansyl labeled human urine samples was analyzed and detected by IsoMS and the result showed 2044 peak pairs detected by IsoMS. After manual inspection of these peak pairs, only 90 false peak pairs were found and the false positive rate was 4.4%, indicating that good specificity (<5%) could be obtained by IsoMS.

The relative quantitative information has been used on multivariate analysis to discover meaningful metabolites that can be putative biomarkers in some research. Therefore, IsoMS is a reliable tool for processing chemical isotope labeling LC-MS data.

In Chapter 3, starting from dry flax fibers, optimal results were obtained with a simple, 30 minute microwave-assisted extraction of 50 mg tissue in 1mL water or 80% methanol, followed by stable isotope dansylation, and analyzed by reverse phase liquid chromatography fourier transform mass spectrometry (RPLC-FT-MS). The extraction and downstream steps were succesful even when using dry flax fibers, which are very difficult to disrupt mechanically. We found that approximately 400 chromatographic peaks could be reproducibly distinguished when 80% methanol was used for extraction. This protocol was sufficient for the identification of more than 200 distinct polar compounds when RPLC-FT-MS data were searched against the HMDB spectral database. The entire extraction, derivatization, and chromatography procotol for up to 8 samples can typically be completed under 13 hours.

In Chapter 4, a reproducible and high-quality separation method has been developed for analyzing the isotope-tagged metabolites from ginseng roots. After labeling by stable isotope dansylation reagents, metabolites were easily separated and detected by LC-FT/ICR-MS in 25 min. There were more than 800 peak pairs detected with good reproducibility (88%) in triplicate runs by an optimal extraction method (i.e. 45 minute microwave-assisted extraction of 30 mg tissue in 1mL water or 80% methanol). For spatial distribution of 16 amino acids, although we did not find a changing trend of same amino acids in different ginseng roots, this technique successfully provided information of a spatial distribution in each ginseng root. For each ginseng root, metabolites tend to have similar changing trends, which could be observed according to the information of spatial distribution. This suggests that metabolites in one ginseng root have high

correlation. Therefore, this method has a potential to give a preliminary understanding of plant metabolism.

In Chapter 5, a differential ^{13}C-/^{12}C-isotope dansylation labeling technique with LC-FT-MS for the analysis of amine- and phenol-containing metabolites has been applied to characterize the *Arabidopsis thaliana* metabolome. Optimal conditions for a robust metabolite extraction method from *Arabidopsis thaliana* were achieved by heating 40 mg tissue in 1mL 80% methanol under 45 min microwave-assisted extraction. We optimized a process that can successfully analyze the isotope-tagged metabolites from *Arabidopsis thaliana* in 31 min. More than 700 ^{13}C-/^{12}C-isotope dansylated peak pairs were detected by the optimal method with good reproducibility (84%) in triplicate runs. In different *TIFY* gene expression level plants, there were 121 peak pairs with significant metabolic changes and the relative concentrations of most of them were higher in WT plants. In the study of methyl jasmonate (MeJA) treatment, most of peak pairs were affected after 24 hr MeJA treatment, and 30 of the peak pairs picked were always up-regulated, while 23 of them was always down-regulated. Therefore, there were total 53 peak pairs induced by MeJA treatment even after 24 hr. In addition, 18 out of the 53 peak pairs were also affected by *TIFY* gene expression levels because the slopes of concentration changing trends in OXT plants were smaller than those in WT plants, and at the beginning half hour treatment, the trends in KOT plants were always opposite to those in WT and OXT plants.

In Chapter 6, an investigation of using the dansylation isotope labeling LC-MS method to study the effect of drinking Goji tea on urine metabolome profile was reported. From the studies of the short-term (<3 h) and longer-term (12 h) effects of drinking a moderate amount of Goji tea, it clearly showed that the consumption of Goji tea did not affect the urine metabolome significantly. This sensitive LC-MS method allowed the detection of an average of 1629±15 (n=3) peak pairs or putative metabolites in the Goji tea itself, compared to very few metabolites detectable without

labeling using the same amount of injection into LC-MS. Therefore, I did not have any expectation that any Goji metabolites could be detected in urine by a traditional LC-MS after consumption without the use of stable isotope dansylation labelling.

Future research for data processing will involve further development of IsoMS by implementing other functionalities, such as automated statistical analysis and metabolome database search for compound identification. A differential ^{13}C-/^{12}C-isotope dansylation labeling technique targeted at amine- and phenol-containing metabolites has been successfully applied on metabolomic profiling of different type samples including plants and urine. A number of other different labeling chemistries are currently being tested or developed in our group. In the future, combination of all metabolomic profiles generated by different differential isotopic labeling techniques that derive different functional groups of metabolites will increase the metabolome coverage for comprehensive metabolomics, which would enlarge the understanding of plant or urine metabolome. In addition, another way to improve the metabolome coverage of plant or urine in the future is to use two-dimensional LC methods as separation techniques prior to mass spectrometric analysis to separate the co-eluted metabolites that occur in one-dimensional LC.

References

(1) Oliver, S. G.; Winson, M. K.; Kell, D. B.; Baganz, F. *Trends in Biotechnology* **1998**, *16*, 373-378.

(2) Fiehn, O. *Comparative and Functional Genomics* **2001**, *2*, 155-168.

(3) German, J. B.; Watkins, S. M.; Fay, L. B. *Journal of the American Dietetic Association* **2005**, *105*, 1425-1432.

(4) Brown, S. C.; Kruppa, G.; Dasseux, J. L. *Mass Spectrometry Reviews* **2005**, *24*, 223-231.

(5) Becker, S.; Kortz, L.; Helmschrodt, C.; Thiery, J.; Ceglarek, U. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* **2012**, *883-884*, 68-75.

(6) Dettmer, K.; Aronov, P. A.; Hammock, B. D. *Mass Spectrometry Reviews* **2007**, *26*, 51-78.

(7) Woo, S. S.; Song, J. S.; Lee, J. Y.; In, D. S.; Chung, H. J.; Liu, J. R.; Choi, D. W. *Phytochemistry* **2004**, *65*, 2751-2761.

(8) Kato, H.; Izumi, Y.; Hasunuma, T.; Matsuda, F.; Kondo, A. *Journal of Bioscience and Bioengineering* **2012**, *113*, 665-673.

(9) Ellis, D. I.; Goodacre, R. *The Analyst* **2006**, *131*, 875-885.

(10) Johnson, H. E.; Broadhurst, D.; Goodacre, R.; Smith, A. R. *Phytochemistry* **2003**, *62*, 919-928.

(11) Vikram, A.; Lui, L.; Hossain, A.; Kushalappa, A. *Annals of Applied Biology* **2006**, *148*, 17-26.

(12) Schuhmacher, R.; Krska, R.; Weckwerth, W.; Goodacre, R. *Analytical and Bioanalytical Chemistry* **2013**, *405*, 5003-5004.

(13) Lewis, I. A.; Schommer, S. C.; Hodis, B.; Robb, K. A.; Tonelli, M.; Westler, W. M.; Sussman, M. R.; Markley, J. L. *Analytical Chemistry* **2007**, *79*, 9385-9390.

(14) Slupsky, C. M. *Biomarkers in Medicine* **2010**, *4*, 195-197.

(15) Wishart, D. S. *TrAC Trends in Analytical Chemistry* **2008**, *27*, 228-237.

(16) Kuhara, T.; Shinka, T.; Inoue, Y.; Ohse, M.; Zhen-wei, X.; Yoshida, I.; Inokuchi, T.; Yamaguchi, S.; Takayanagi, M.; Matsumoto, I. *Journal of Chromatography. B, Biomedical Sciences and Applications* **1999**, *731*, 141-147.

(17) Chalcraft, K. R.; Lee, R.; Mills, C.; Britz-McKibbin, P. *Analytical Chemistry* **2009**, *81*, 2506-2515.

(18) Maxwell, E. J.; Chen, D. D. *Analytica Chimica Acta* **2008**, *627*, 25-33.

(19) Cai, J.; Henion, J. *Journal of Chromatography A* **1995**, *703*, 667-692.

(20) Hammarqvist, F.; Andersson, K.; Luo, J. L.; Wernerman, J. *Clinical Nutrition* **2005**, *24*, 236-243.

(21) Wang, M.; Lamers, R. J.; Korthout, H. A.; van Nesselrooij, J. H.; Witkamp, R. F.; van der Heijden, R.; Voshol, P. J.; Havekes, L. M.; Verpoorte, R.; van der Greef, J. *Phytotherapy Research : PTR* **2005**, *19*, 173-182.

(22) De Vos, R. C.; Moco, S.; Lommen, A.; Keurentjes, J. J.; Bino, R. J.; Hall, R. D. *Nature Protocols* **2007**, *2*, 778-791.

(23) Wishart, D. S. *American Journal of Transplantation : Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons* **2005**, *5*, 2814-2820.

(24) Antignac, J.-P.; Courant, F.; Pinel, G.; Bichon, E.; Monteau, F.; Elliott, C.; Le Bizec, B. *TrAC Trends in Analytical Chemistry* **2011**, *30*, 292-301.

(25) Lewis, G. D.; Asnani, A.; Gerszten, R. E. *Journal of the American College of Cardiology* **2008**, *52*, 117-123.

(26) Kaddurah-Daouk, R.; Krishnan, K. R. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology* **2009**, *34*, 173-186.

(27) Fiehn, O. *Plant Molecular Biology* **2002**, *48*, 155-171.

(28) Soga, T.; Baran, R.; Suematsu, M.; Ueno, Y.; Ikeda, S.; Sakurakawa, T.; Kakazu, Y.; Ishikawa, T.; Robert, M.; Nishioka, T.; Tomita, M. *The Journal of Biological Chemistry* **2006**, *281*, 16768-16776.

(29) Arlt, W.; Biehl, M.; Taylor, A. E.; Hahner, S.; Libe, R.; Hughes, B. A.; Schneider, P.; Smith, D. J.; Stiekema, H.; Krone, N.; Porfiri, E.; Opocher, G.; Bertherat, J.; Mantero, F.; Allolio, B.; Terzolo, M.; Nightingale, P.; Shackleton, C. H.; Bertagna, X.; Fassnacht, M.; Stewart, P. M. *The Journal of Clinical Endocrinology and Metabolism* **2011**, *96*, 3775-3784.

(30) Wishart, D. S.; Knox, C.; Guo, A. C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. *Nucleic Acids Research* **2008**, *36*, D901-906.

(31) Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J.; Liu, P.; Yallou, F.; Bjorndahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A. *Nucleic Acids Research* **2013**, *41*, D801-807.

(32) Ryan, D.; Robards, K. *Analytical Chemistry* **2006**, *78*, 7954-7958.

(33) Dunn, W. B.; Ellis, D. I. *TrAC Trends in Analytical Chemistry* **2005**, *24*, 285-294.

(34) Metz, T. O.; Page, J. S.; Baker, E. S.; Tang, K.; Ding, J.; Shen, Y.; Smith, R. D. *TrAC Trends in Analytical Chemistry* **2008**, *27*, 205-214.

(35) Theodoridis, G.; Gika, H. G.; Wilson, I. D. *TrAC Trends in Analytical Chemistry* **2008**, *27*, 251-260.

(36) Fekete, S.; Olah, E.; Fekete, J. *Journal of Chromatography. A* **2012**, *1228*, 57-71.

(37) Gritti, F.; Guiochon, G. *Analytical Chemistry* **2008**, *80*, 5009-5020.

(38) Gritti, F.; Martin, M.; Guiochon, G. *Analytical Chemistry* **2009**, *81*, 3365-3384.

(39) Gritti, F.; Guiochon, G. *Analytical Chemistry* **2008**, *80*, 6488-6499.

(40) Zeleny, J. *Physical Review* **1914**, *3*, 69.

(41) Dole, M.; Mack, L.; Hines, R.; Mobley, R.; Ferguson, L.; Alice, M. d. *The Journal of Chemical Physics* **1968**, *49*, 2240-2249.

(42) Yamashita, M.; Fenn, J. B. *The Journal of Physical Chemistry* **1984**, *88*, 4451-4459.

(43) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64-71.

(44) Kebarle, P. *Journal of Mass Spectrometry : JMS* **2000**, *35*, 804-817.

(45) Kebarle, P.; Verkerk, U. H. *Mass Spectrometry Reviews* **2009**, *28*, 898-917.

(46) Kebarle, P.; Tang, L. *Analytical Chemistry* **1993**, *65*, 972A-986A.

(47) Mack, L. L.; Kralik, P.; Rheude, A.; Dole, M. *The Journal of Chemical Physics* **1970**, *52*, 4977-4986.

(48) Iribarne, J.; Thomson, B. *The Journal of Chemical Physics* **1976**, *64*, 2287-2294.

(49) Fernandez De la Mora, J. *Analytica Chimica Acta* **2000**, *406*, 93-104.

(50) Mora, J. F. d. l.; Van Berkel, G. J.; Enke, C. G.; Cole, R. B.; Martinez-Sanchez, M.; Fenn, J. B. *Journal of Mass Spectrometry* **2000**, *35*, 939-952.

(51) Gamero-Castano, M.; De La Mora, J. F. *The Journal of Chemical Physics* **2000**, *113*, 815-832.

(52) Lawrence, E. O.; Livingston, M. S. *Physical Review* **1932**, *40*, 19.

(53) MB, C.; Marshall, A. *Chemical Physics Letters* **1974**, *25*, 282-283.

(54) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrometry Reviews* **1998**, *17*, 1-35.

(55) Marshall, A. G.; Hendrickson, C. L. *International Journal of Mass Spectrometry* **2002**, *215*, 59-75.

(56) Barrow, M. P.; Burkitt, W. I.; Derrick, P. J. *The Analyst* **2005**, *130*, 18-28.

(57) Cody, R.; Hein, R.; Goodman, S.; Marshall, A. G. *Rapid Communications in Mass Spectrometry* **1987**, *1*, 99-102.

(58) Guan, S.; Marshall, A. G. *International Journal of Mass Spectrometry and Ion Processes* **1996**, *157*, 5-37.

(59) Zhang, L. K.; Rempel, D.; Pramanik, B. N.; Gross, M. L. *Mass Spectrometry Reviews* **2005**, *24*, 286-309.

(60) Ledford, E. B., Jr.; Rempel, D. L.; Gross, M. L. *Analytical Chemistry* **1984**, *56*, 2744-2748.

(61) Bristow, A. W. *Mass Spectrometry Reviews* **2006**, *25*, 99-111.

(62) Cottingham, K. *Analytical Chemistry* **2006**, *78*, 655-657.

(63) Chernushevich, I. V.; Loboda, A. V.; Thomson, B. A. *Journal of Mass Spectrometry : JMS* **2001**, *36*, 849-865.

(64) Stafford, G., Jr. *Journal of the American Society for Mass Spectrometry* **2002**, *13*, 589-596.

(65) Abu-Samra, A.; Morris, J. S.; Koirtyohann, S. *Analytical Chemistry* **1975**, *47*, 1475-1477.

(66) Ganzler, K.; Salgo, A.; Valko, K. *Journal of Chromatography* **1986**, *371*, 299-306.

(67) Pare, J. R. J.; Sigouin, M.; Lapointe, J.; Google Patents, 1991.

(68) Lopez-Avila, V.; Young, R.; Beckert, W. F. *Analytical Chemistry* **1994**, *66*, 1097-1106.

(69) Letellier, M.; Budzinski, H. *Analusis* **1999**, *27*, 259-271.

(70) Hemalatha, S.; Mandal, V.; Mohan, Y. *Pharmacognosy Reviews* 2007; Vol. 1, p 7-18.

(71) Bethe, H. A. *Physical Review* **1947**, *72*, 339-341.

(72) ThueÂry, J. *Microwaves: Industrial, Scientific and Medical Applications, Part* **1992**, *1*, 83-125.

(73) Demesmay, C.; Olle, M. *Spectra 2000 Analyse* **1993**, *22*, 27-32.

(74) Sinquin, A.; Gorner, T.; Dellacherie, E. *Analusis* **1993**, *21*, 1-10.

(75) Anastassiades, J.; Kaminsky, D.; Perea, E.; Poezevara, A. *Solid-state Microwave Generation*; Chapman & Hall, 1992

(76) Burfoot, J. C. *Ferroelectrics: an Introduction to the Physical Principles*; Van Nostrand New York, 1967.

(77) Burkert, R.; Helberg, H.; Von Schütz, J. *Synthetic Metals* **1993**, *56*, 2519-2524.

(78) Carroll, J. E. *Hot Electron Microwave generators*; Edward Arnold London, 1970.

(79) Latha, C. *Biotechnology Letters***2007**, *29*, 319-322.

(80) Zhou, H.-Y.; Liu, C.-Z. *Journal of Chromatography A* **2006**, *1129*, 135-139.

(81) Kratchanova, M.; Pavlova, E.; Panchev, I. *Carbohydrate Polymers* **2004**, *56*, 181-185.

(82) Luque de Castro, M.; Jimenez-Carmona, M.; Fernandez-Perez, V. *TrAC Trends in Analytical Chemistry* **1999**, *18*, 708-716.

(83) Zuloaga, O.; Etxebarria, N.; Fernández, L. A.; Madariaga, J. M. *Talanta* **1999**, *50*, 345-357.

(84) Stokvis, E.; Rosing, H.; Beijnen, J. H. *Rapid Communications in Mass Spectrometry* **2005**, *19*, 401-407.

(85) Van Eeckhaut, A.; Lanckmans, K.; Sarre, S.; Smolders, I.; Michotte, Y. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* **2009**, *877*, 2198-2207.

(86) Wu, L.; Mashego, M. R.; van Dam, J. C.; Proell, A. M.; Vinke, J. L.; Ras, C.; van Winden, W. A.; van Gulik, W. M.; Heijnen, J. J. *Analytical Biochemistry* **2005**, *336*, 164-171.

(87) Goodlett, D. R.; Keller, A.; Watts, J. D.; Newitt, R.; Yi, E. C.; Purvine, S.; Eng, J. K.; von Haller, P.; Aebersold, R.; Kolker, E. *Rapid Communications in Mass Spectrometry : RCM* **2001**, *15*, 1214-1221.

(88) Guo, K.; Li, L. *Analytical Chemistry* **2009**, *81*, 3919-3932.

(89) 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. *Nucleic Acids Research* **2007**, *35*, D521-526.

(90) Shortreed, M. R.; Lamos, S. M.; Frey, B. L.; Phillips, M. F.; Patel, M.; Belshaw, P. J.; Smith, L. M. *Analytical Chemistry* **2006**, *78*, 6398-6403.

(91) Yang, W. C.; Regnier, F. E.; Jiang, Q.; Adamec, J. *Journal of Chromatography. A* **2010**, *1217*, 667-675.

(92) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nature Biotechnology* **1999**, *17*, 994-999.

(93) Yang, W. C.; Regnier, F. E.; Sliva, D.; Adamec, J. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* **2008**, *870*, 233-240.

(94) Fukusaki, E.; Harada, K.; Bamba, T.; Kobayashi, A. *Journal of Bioscience and Bioengineering* **2005**, *99*, 75-77.

(95) Guo, K.; Li, L. *Analytical Chemistry* **2010**, *82*, 8789-8793.

(96) Seiler, N.; Knodgen, B. *Journal of Chromatography* **1978**, *145*, 29-39.

(97) Zotou, A.; Loukou, Z.; Soufleros, E.; Stratis, I. *Chromatographia* **2003**, *57*, 429-439.

(98) Minocha, R.; Long, S. *Journal of Chromatography. A* **2004**, *1035*, 63-73.

(99) Keller, B. O.; Sui, J.; Young, A. B.; Whittal, R. M. *Analytica Chimica Acta* **2008**, *627*, 71-81.

(100) Creek, D. J. *Bioanalysis* **2013**, *5*, 1807-1810.

(101) Bennett, B. D.; Yuan, J.; Kimball, E. H.; Rabinowitz, J. D. *Nature Protocols* **2008**, *3*, 1299-1311.

(102) Bruheim, P.; Kvitvang, H. F.; Villas-Boas, S. G. *Journal of Chromatography. A* **2013**, *1296*, 196-203.

(103) Guo, K.; Ji, C.; Li, L. *Analytical Chemistry* **2007**, *79*, 8631-8638.

(104) O'Maille, G.; Go, E. P.; Hoang, L.; Want, E. J.; Smith, C.; O'Maille, P.; NordstrÖm, A.; Morita, H.; Qin, C.; Uritboonthai, W. *Journal of Spectroscopy* **2008**, *22*, 327-343.

(105) Shanaiah, N.; Desilva, M. A.; Nagana Gowda, G. A.; Raftery, M. A.; Hainline, B. E.; Raftery, D. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 11540-11544.

(106) Abello, N.; Geurink, P. P.; van der Toorn, M.; van Oosterhout, A. J.; Lugtenburg, J.; van der Marel, G. A.; Kerstjens, H. A.; Postma, D. S.; Overkleeft, H. S.; Bischoff, R. *Analytical Chemistry* **2008**, *80*, 9171-9180.

(107) Giavalisco, P.; Kohl, K.; Hummel, J.; Seiwert, B.; Willmitzer, L. *Analytical Chemistry* **2009**, *81*, 6546-6551.

(108) Armenta, J. M.; Cortes, D. F.; Pisciotta, J. M.; Shuman, J. L.; Blakeslee, K.; Rasoloson, D.; Ogunbiyi, O.; Sullivan, D. J., Jr.; Shulaev, V. *Analytical Chemistry* **2010**, *82*, 548-558.

(109) Tang, Z.; Guengerich, F. P. *Analytical Chemistry* **2010**, *82*, 7706-7712.

(110) Huang, Y. Q.; Liu, J. Q.; Gong, H.; Yang, J.; Li, Y.; Feng, Y. Q. *The Analyst* **2011**, *136*, 1515-1522.

(111) Wang, H.; Wang, H.; Zhang, L.; Zhang, J.; Guo, Y. *Analytica Chimica Acta* **2011**, *690*, 1-9.

(112) Dai, W.; Huang, Q.; Yin, P.; Li, J.; Zhou, J.; Kong, H.; Zhao, C.; Lu, X.; Xu, G. *Analytical Chemistry* **2012**, *84*, 10245-10251.

(113) Zhang, S.; You, J.; Ning, S.; Song, C.; Suo, Y. R. *Journal of Chromatography. A* **2013**, *1280*, 84-91.

(114) Tayyari, F.; Gowda, G. A.; Gu, H.; Raftery, D. *Analytical Chemistry* **2013**, *85*, 8715-8721.

(115) Peng, J.; Li, L. *Analytica Chimica Acta* **2013**, *803*, 97-105.

(116) Zhou, R.; Guo, K.; Li, L. *Analytical Chemistry* **2013**, *85*, 11532-11539.

(117) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Analytical Chemistry* **2006**, *78*, 779-787.

(118) Melamud, E.; Vastag, L.; Rabinowitz, J. D. *Analytical Chemistry* **2010**, *82*, 9818-9826.

(119) Bueschl, C.; Kluger, B.; Berthiller, F.; Lirk, G.; Winkler, S.; Krska, R.; Schuhmacher, R. *Bioinformatics* **2012**, *28*, 736-738.

(120) Chokkathukalam, A.; Jankevics, A.; Creek, D. J.; Achcar, F.; Barrett, M. P.; Breitling, R. *Bioinformatics* **2013**, *29*, 281-283.

(121) Gros, C.; Labouesse, B. *European Journal of Biochemistry / FEBS* **1969**, *7*, 463-470.

(122) Macnicol, P. K. *Analytical Biochemistry* **1978**, *85*, 71-78.

(123) Niessen, W. M.; Manini, P.; Andreoli, R. *Mass Spectrometry Reviews* **2006**, *25*, 881-899.

(124) Zhou, R.; Tseng, C. L.; Huan, T.; Li, L. *Analytical Chemistry* **2014**, *86*, 4675-4679.

(125) Wang, Z.; Hobson, N.; Galindo, L.; Zhu, S.; Shi, D.; McDill, J.; Yang, L.; Hawkins, S.; Neutelings, G.; Datla, R.; Lambert, G.; Galbraith, D. W.; Grassa, C. J.; Geraldes, A.; Cronk, Q. C.; Cullis, C.; Dash, P. K.; Kumar, P. A.; Cloutier, S.; Sharpe, A. G.; Wong, G. K.; Wang, J.; Deyholos, M. K. *The Plant Journal : for Cell and Molecular Biology* **2012**, *72*, 461-473.

(126) Deyholos, M. K. *Israel Journal of Plant Sciences* **2006**, *54*, 273-280.

(127) Goubet, F.; Bourlard, T.; Girault, R.; Alexandre, C.; Vandevelde, M.-C.; Morvan, C. *Carbohydrate Polymers* **1995**, *27*, 221-227.

(128) Kim, J.-W.; Mazza, G. *Journal of Agricultural and Food Chemistry* **2009**, *57*, 1805-1813.

(129) Xia, J.; Psychogios, N.; Young, N.; Wishart, D. S. *Nucleic Acids Research* **2009**, *37*, W652-660.

(130) 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. *Nucleic Acids Research* **2009**, *37*, D603-610.

(131) Zhang, P.; Dreher, K.; Karthikeyan, A.; Chi, A.; Pujar, A.; Caspi, R.; Karp, P.; Kirkup, V.; Latendresse, M.; Lee, C.; Mueller, L. A.; Muller, R.; Rhee, S. Y. *Plant Physiology* **2010**, *153*, 1479-1491.

(132) Li, W. F.; Jiang, J. G.; Chen, J. *Archives of Medical Research* **2008**, *39*, 246-251.

(133) Chen, K.; Yu, B. *Chinese Medical Journal* **1999**, *112*, 934-937.

(134) Gray, M. J.; Chang, D.; Zhang, Y.; Liu, J.; Bensoussan, A. *Biomedical Chromatography : BMC* **2010**, *24*, 91-103.

(135) Li, P.; Qi, L.-W.; Liu, E.-H.; Zhou, J.-L.; Wen, X.-D. *TrAC Trends in Analytical Chemistry* **2008**, *27*, 66-77.

(136) Baeg, I. H.; So, S. H. *Journal of Ginseng Research* **2013**, *37*, 1-7.

(137) Lee, T.-f.; Shiao, Y.-J.; Chen, C.-F.; Wang, L. C. *Planta Medica* **2001**, *67*, 634-637.

(138) Hu, C.; Kitts, D. D. *Journal of the American Oil Chemists' Society* **2001**, *78*, 249-255.

(139) Duda, R. B.; Zhong, Y.; Navas, V.; Li, M. Z.; Toy, B. R.; Alavarez, J. G. *Journal of Surgical Oncology* **1999**, *72*, 230-239.

(140) Vuksan, V.; Sievenpiper, J. L.; Koo, V. Y.; Francis, T.; Beljan-Zdravkovic, U.; Xu, Z.; Vidgen, E. *Archives of Internal Medicine* **2000**, *160*, 1009-1013.

(141) Kim, J. H. *J Ginseng Res* **2012**, *36*, 16-26.

(142) Biondo, P. D.; Robbins, S. J.; Walsh, J. D.; McCargar, L. J.; Harber, V. J.; Field, C. J. *Applied Physiology, Nutrition, and Metabolism* **2008**, *33*, 966-975.

(143) Attele, A. S.; Wu, J. A.; Yuan, C. S. *Biochemical Pharmacology* **1999**, *58*, 1685-1693.

(144) Kitts, D.; Hu, C. *Public Health Nutrition* **2000**, *3*, 473-485.

(145) Lee, E. J.; Shaykhutdinov, R.; Weljie, A. M.; Vogel, H. J.; Facchini, P. J.; Park, S. U.; Kim, Y. K.; Yang, T. J. *Journal of Agricultural and Food Chemistry* **2009**, *57*, 7513-7522.

(146) MacCrehan, W. A.; White, C. M. *Analytical and Bioanalytical Chemistry* **2013**, *405*, 4511-4522.

(147) Li, L.; Li, R.; Zhou, J.; Zuniga, A.; Stanislaus, A. E.; Wu, Y.; Huan, T.; Zheng, J.; Shi, Y.; Wishart, D. S.; Lin, G. *Analytical Chemistry* **2013**, *85*, 3401-3408.

(148) Mandal, V.; Mohan, Y.; Hemalatha, S. *Pharmacognosy Reviews* **2007**, *1*, 7.

(149) Li, H.; Chen, B.; Nie, L.; Yao, S. *Phytochemical analysis : PCA* **2004**, *15*, 306-312.

(150) Huan, T.; Li, L. *Analytical Chemistry* **2014**, *87*, 1306-1313.

(151) Sumner, L. W.; Mendes, P.; Dixon, R. A. *Phytochemistry* **2003**, *62*, 817-836.

(152) Arbona, V.; Iglesias, D. J.; Talon, M.; Gomez-Cadenas, A. *Journal of Agricultural and Food Chemistry* **2009**, *57*, 7338-7347.

(153) Meinke, D. W.; Cherry, J. M.; Dean, C.; Rounsley, S. D.; Koornneef, M. *Science* **1998**, *282*, 662, 679-682.

(154) Creelman, R. A.; Tierney, M. L.; Mullet, J. E. *Proceedings of the National Academy of Sciences of the United States of America* **1992**, *89*, 4938-4941.

(155) Seo, H. S.; Song, J. T.; Cheong, J. J.; Lee, Y. H.; Lee, Y. W.; Hwang, I.; Lee, J. S.; Choi, Y. D. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 4788-4793.

(156) Cheong, J. J.; Choi, Y. D. *Trends in Genetics : TIG* **2003**, *19*, 409-413.

(157) Radhika, V.; Kost, C.; Boland, W.; Heil, M. *Plos One* **2010**, *5*, e9265.

(158) Wasternack, C. *Annals of Botany* **2007**, *100*, 681-697.

(159) Wasternack, C.; Hause, B. *Annals of Botany* **2013**, *111*, 1021-1058.

(160) Chini, A.; Fonseca, S.; Fernandez, G.; Adie, B.; Chico, J. M.; Lorenzo, O.; Garcia-Casado, G.; Lopez-Vidriero, I.; Lozano, F. M.; Ponce, M. R.; Micol, J. L.; Solano, R. *Nature* **2007**, *448*, 666-671.

(161) Yan, Y.; Stolz, S.; Chetelat, A.; Reymond, P.; Pagni, M.; Dubugnon, L.; Farmer, E. E. *Plant Cell* **2007**, *19*, 2470-2483.

(162) Thines, B.; Katsir, L.; Melotto, M.; Niu, Y.; Mandaokar, A.; Liu, G.; Nomura, K.; He, S. Y.; Howe, G. A.; Browse, J. *Nature* **2007**, *448*, 661-665.

(163) Chung, H. S.; Howe, G. A. *Plant Cell* **2009**, *21*, 131-145.

(164) Major, I. T.; Constabel, C. P. *The New Phytologist* **2006**, *172*, 617-635.

(165) Jiang, Y.; Deyholos, M. K. *Bmc Plant Biology* **2006**, *6*, 25.

(166) Creelman, R. A.; Mullet, J. E. *Proceedings of the National Academy of Sciences of the United States of America* **1995**, *92*, 4114-4119.

(167) Halim, V. A.; Vess, A.; Scheel, D.; Rosahl, S. *Plant Biology* **2006**, *8*, 307-313.

(168) Vanholme, B.; Grunewald, W.; Bateman, A.; Kohchi, T.; Gheysen, G. *Trends Plant Science* **2007**, *12*, 239-244.

(169) Ong, S. E.; Foster, L. J.; Mann, M. *Methods* **2003**, *29*, 124-130.

(170) Leitner, A.; Lindner, W. *Proteomics* **2006**, *6*, 5418-5434.

(171) Truong, M.; Yang, B.; Jarrard, D. F. *The Journal of Urology* **2013**, *189*, 422-429.

(172) Rosser, C. J.; Urquidi, V.; Goodison, S. *Biomarkers in Medicine* **2013**, *7*, 779-790.

(173) Storr, M.; Vogel, H. J.; Schicho, R. *Current Opinion in Gastroenterology* **2013**, *29*, 378-383.

(174) Weiss, R. H.; Kim, K. *Nature Neviews. Nephrology* **2012**, *8*, 22-33.

(175) Zhang, T.; Wu, X.; Ke, C.; Yin, M.; Li, Z.; Fan, L.; Zhang, W.; Zhang, H.; Zhao, F.; Zhou, X.; Lou, G.; Li, K. *Journal of Proteome Research* **2013**, *12*, 505-512.

(176) Zhang, A.; Sun, H.; Wu, X.; Wang, X. *Clinica Chimica Acta; International Journal of Clinical Chemistry* **2012**, *414*, 65-69.

(177) Ye, G.; Zhu, B.; Yao, Z.; Yin, P.; Lu, X.; Kong, H.; Fan, F.; Jiao, B.; Xu, G. *Journal of Proteome Research* **2012**, *11*, 4361-4372.

(178) Emwas, A.-H. M.; Salek, R. M.; Griffin, J. L.; Merzaban, J. *Metabolomics* **2013**, *9*, 1048-1072.

(179) Beckmann, M.; Lloyd, A. J.; Haldar, S.; Fave, G.; Seal, C. J.; Brandt, K.; Mathers, J. C.; Draper, J. *The Proceedings of the Nutrition Society* **2013**, *72*, 352-361.

(180) Pujos-Guillot, E.; Hubert, J.; Martin, J. F.; Lyan, B.; Quintana, M.; Claude, S.; Chabanas, B.; Rothwell, J. A.; Bennetau-Pelissero, C.; Scalbert, A.; Comte, B.; Hercberg, S.; Morand, C.; Galan, P.; Manach, C. *Journal of Proteome Research* **2013**, *12*, 1645-1659.

(181) Lampe, J. W.; Navarro, S. L.; Hullar, M. A.; Shojaie, A. *The Proceedings of the Nutrition Society* **2013**, *72*, 207-218.

(182) Hu, C.; Xu, G. *TrAC Trends in Analytical Chemistry* **2013**, *52*, 36-46.

(183) Medina, S.; Domínguez-Perles, R.; Ferreres, F.; Tomás-Barberán, F. A.; Gil-Izquierdo, Á. *TrAC Trends in Analytical Chemistry* **2013**, *52*, 88-99.

(184) Putri, S. P.; Yamamoto, S.; Tsugawa, H.; Fukusaki, E. *Journal of Bioscience and Bioengineering* **2013**, *116*, 9-16.

(185) Wu, Z.; Huang, Z.; Lehmann, R.; Zhao, C.; Xu, G. *Chromatographia* **2009**, *69*, 23-32.

(186) Potterat, O. *Planta Medica* **2010**, *76*, 7-19.

(187) Burke, D.; Smidt, C.; Vuong, L. *Current Topics in Nutraceutical Research* **2005**, *3*, 259.

(188) Altintas, A.; Kosar, M.; Kirimer, N.; Baser, K.; Demirci, B. *Chemistry of Natural Compounds* **2006**, *42*, 24-25.

(189) Yin, G.; Dang, Y. *Carbohydrate Polymers* **2008**, *74*, 603-610.

(190) Tang, W. M.; Chan, E.; Kwok, C. Y.; Lee, Y. K.; Wu, J. H.; Wan, C. W.; Chan, R. Y.; Yu, P. H.; Chan, S. W. *Inflammopharmacology* **2012**, *20*, 307-314.

(191) Weller, P.; Breithaupt, D. E. *Journal of Agricultural and Food Chemistry* **2003**, *51*, 7044-7049.

(192) Inbaraj, B. S.; Lu, H.; Hung, C. F.; Wu, W. B.; Lin, C. L.; Chen, B. H. *Journal of Pharmaceutical and Biomedical Analysis* **2008**, *47*, 812-818.

(193) Liu, X. L.; Sun, J. Y.; Li, H. Y.; Zhang, L.; Qian, B. C. *China journal of Chinese Materia Medica* **2000**, *25*, 481-483.

(194) Wu, Y. M.; Li, L. *Analytical Chemistry* **2012**, *84*, 10723-10731.

(195) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. *Therapeutic Drug Monitoring* **2005**, *27*, 747-751.

Appendix A

IsoMS User Manual For FT-ICR-MS (version 1.0; March 9, 2014)

- IsoMS is a program written in R for processing data generated by differential chemical isotopic labeling LC-MS. It consists of two scripts, IsoMS and IsoMS-align. IsoMS is for peak picking, peak pairing and peak ratio calculation from an LC-MS dataset. IsoMS-align is for alignment of peak pairs and their peak ratios from multiple datasets. These two scripts are freely available for non-commercial use from www.mycompoundid.org/IsoMS.
- The overall workflow of IsoMS for processing multiple LC-MS datasets produced from running multiple samples in a typical metabolomics study includes three steps: 1) converting the raw LC-MS data into a centroid peak list file (CSV format) and storing all the converted files in a folder, 2) using IsoMS to process all the files automatically, and 3) using IsoMS-align to align the peak pairs and their ratios from all the datasets to produce the final CSV file that can be exported for data and statistics analysis.
- The instruction for data conversion in Step 1 and the script (BD-convert) used to convert the data file into a format compatible with IsoMS can be downloaded from the MyCompoundID website.
- The instructions for installing and using the IsoMS (i.e., IsoMSFT for this instruction) and IsoMS-align scripts are given below.

1. Installation of IsoMS

1) Two versions of IsoMS, IsoMSFT and IsoMSTOF, have been developed for processing the FT-ICR-MS data and TOF- or QTOF-MS data, respectively. For processing the FT-ICR-MS data, download the IsoMSFT version (zipped file).
2) R graphic user interface (RGui) should be installed in a computer with Windows XP or Windows 7. RGui can be downloaded from the following URL: http://www.r-project.org/. Since IsoMS has been packaged with RGui 3.0, it is recommended to install the same version to run IsoMS.
3) If a different version of RGui is installed, R-tools is required for packaging the IsoMS source files which can be found in the src folder within the IsoMS unzipped folder (Figure 1). R-tools can be downloaded from the URL: http://cran.rstudio.com/.

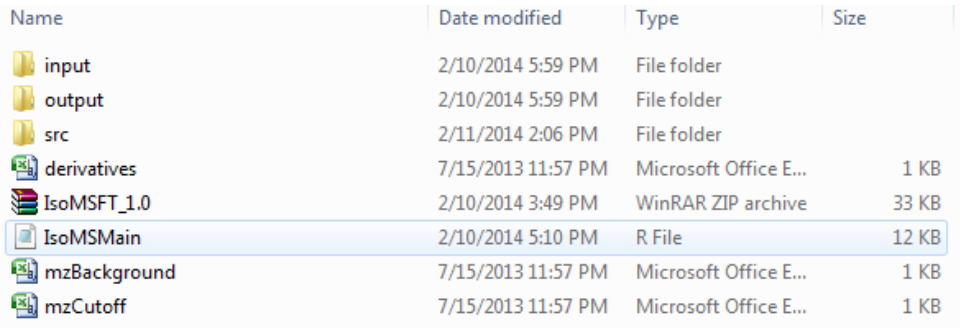


Figure 1. Unzipped IsoMS folder (for FT-MS version).

4) During the installation of R-tools, the environmental variables should be checked (see Figure 2). Otherwise a user needs to edit the R environmental variables after installation by right clicking Computer → Properties → Advance system properties → Advance → Environmental variables.

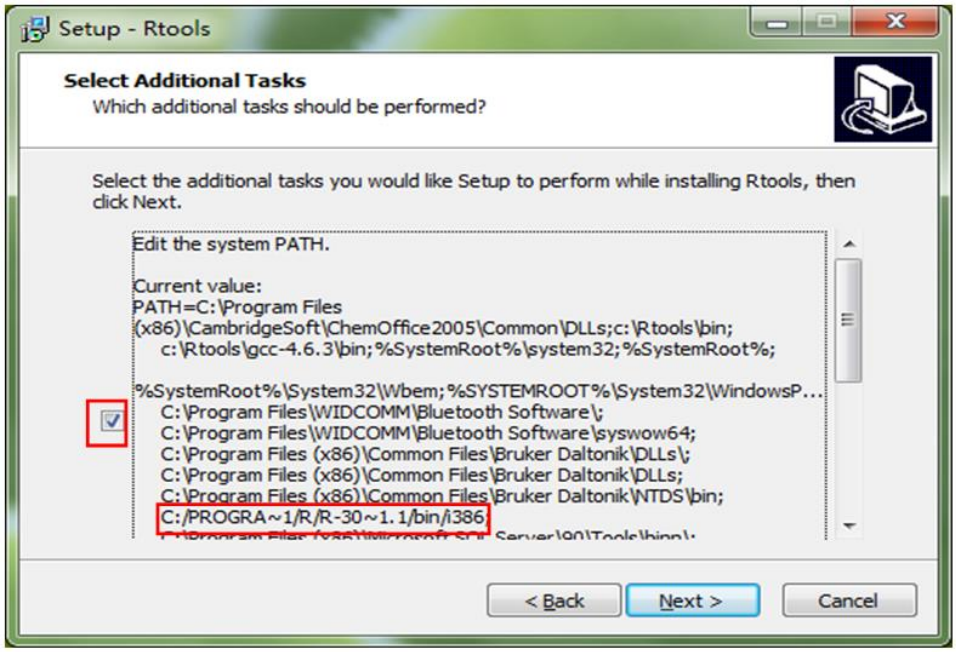


Figure 2. Select/check the environmental variables during the installation of R-tools.

5) Windows 7 users must change "user account control" to the lowest level in order to package the IsoMS scripts by opening Control panel → User account → User account settings.
6) After downloading IsoMSFT.zip, unzip it. In the unzipped file folder (Figure 1), there is another zipped file, IsoMSFT_1.0; DO NOT unzip it. Install it in RGui by clicking Packages → Install packages(s) from local zip files (Figure 3), and then select the zipped IsoMS package named as "IsoMSFT_1.0".

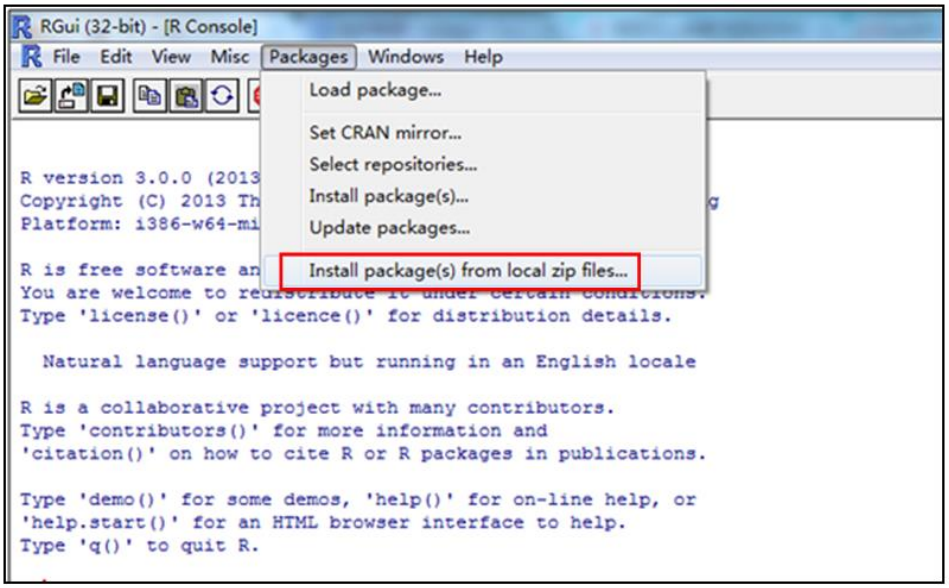


Figure 3. Install the IsoMS package in RGui.

7) Assign the unzipped folder of IsoMS (Figure 1) as the working folder of RGui by clicking: File → Change dir..... (Figure 4).

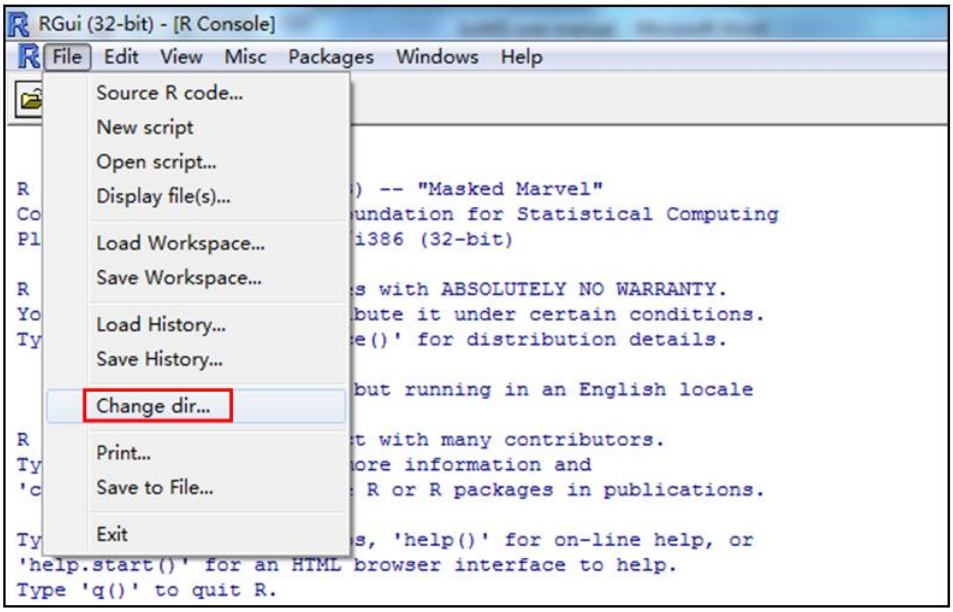


Figure 4. Change the working folder of RGui.

2. Running IsoMS

1) Run IsoMS by clicking: File → Source R code... to open the working folder (Figure 5), and then click "IsoMSMain".

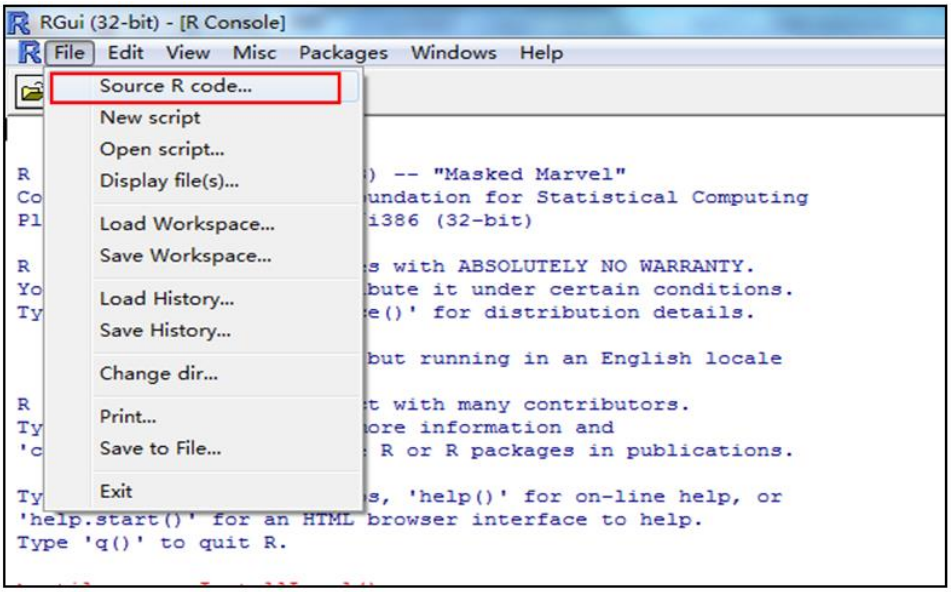
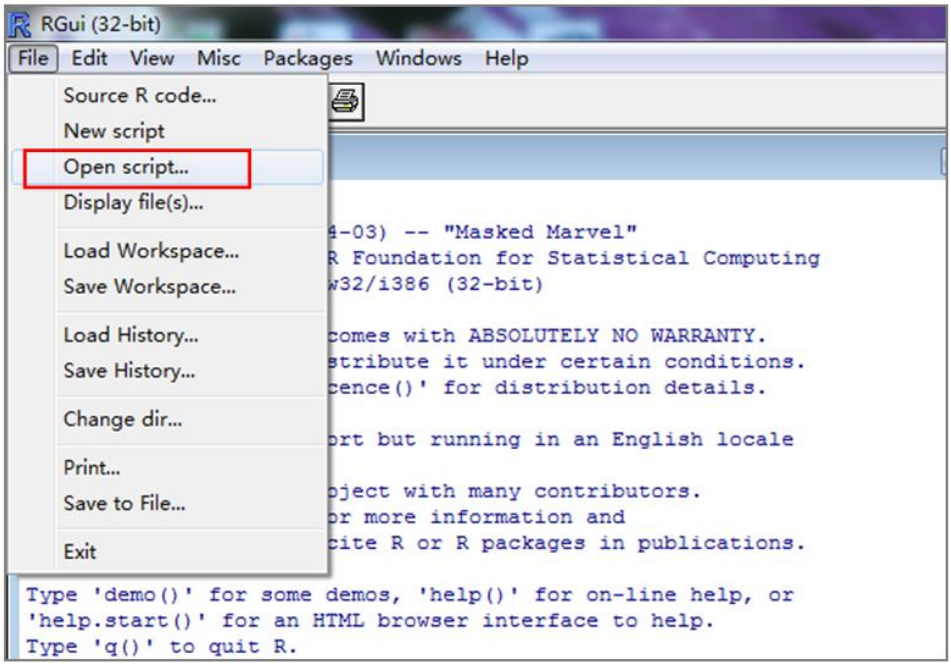


Figure 5. Run IsoMSMain in RGui.

2) A user has the option to change the IsoMS parameters based on the LC-MS data acquired and the desired output. To open the parameter page in RGui, click File → Open scripts → IsoMSMain (Figure 6A). A partial list of the parameter page is shown in Figure 6B.

(A)



(B)

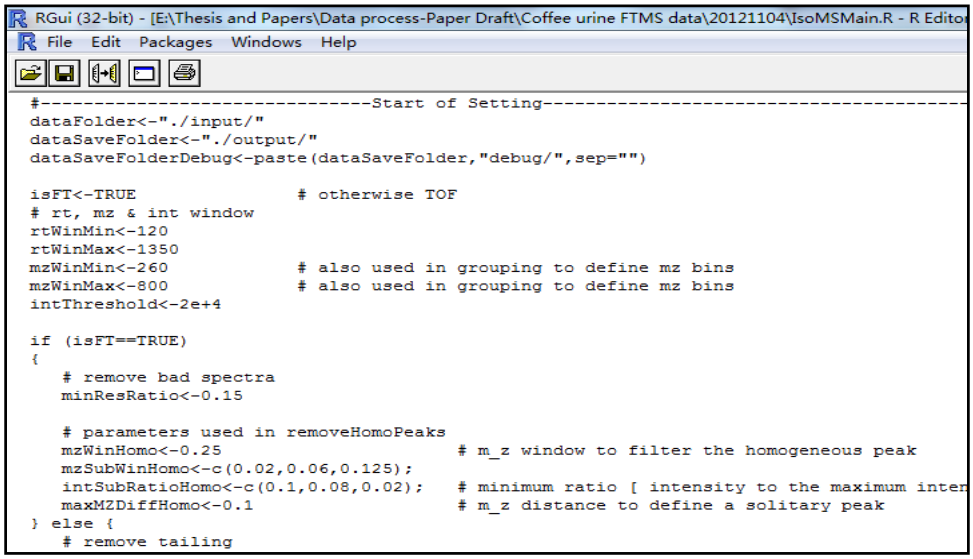


Figure 6. (A) Open IsoMSMain script. (B) Partial list of the parameter page of IsoMSMain.

3) Change the parameters. Table 1 lists the parameters and their functions. These parameters should be adjusted according to the LC-MS data obtained.

Table 1. IsoMS parameters that need to be changed according to the LC-MS data obtained.

Parameter	Function
dataFolder	Location of source data files. Default folder is "input" in the working folder (see Figure 1). Note: if a user opts to use a folder containing all the peak list files outside this working folder, replace "input" with the directory of the folder containing the peak list files.
dataSaveFolder	Location of resultant data files. Default folder is "output" in the working folder (see Figure 1). Note: if a user opts to use a folder outside this working folder, replace "output" with the directory of the folder to contain all the resultant data files.
rtWinMin	Chromatogram cutoff window: Spectra before this retention time are not processed. Note: use this parameter to exclude any mass spectra showing only background peaks at the beginning of the chromatogram where no metabolite peak pairs are expected.
rtWinMax	Chromatogram cutoff window: Spectra after this retention time are not processed.

	Note: use this parameter to exclude any mass spectra collected after metabolite elution where no metabolite peak pairs are expected.	
mzWinMin	Mass spectrum cutoff window: Spectral peaks with lower m/z than the cutoff value are not processed. Note: this window can be useful to exclude mass spectral region where no peak pairs are expected (e.g., below the m/z value of a labeling reagent).	
mzWinMax	Mass spectrum cutoff window: Spectral peaks with higher m/z than the cutoff value are not processed.	
intThreshold	Intensity threshold. Spectral peaks with lower intensity are not processed.	
intSaturated	Upper limit of intensity or ion count to define the peak saturation. Note: the value of this limit is dependent on the dynamic range of an instrument. It can be determined by inspecting the mass spectra obtained under the experimental conditions; the upper limit is the max intensity or ion count where the peak saturation occurs.	
isMZBackground	No change is needed. Always select =TRUE to use a user-defined mzBackground csv file to filter out the background peaks. This file should be placed in the working folder (see Figure 1). Note 1: The mzBackground.csv file contains a list of masses of m/z_light peaks of known background peak pairs to be excluded in the final results. An example of the mzBackground csv file is given in the working folder. Users should change the mass values in this file according to the background peak pairs found in their own work. Note 2: Any other peaks a user wants to remove can be added to this mzBackground file. For example, if a blank sample is run and the blank dataset has been processed by IsoMS, the m/z_light masses of the peak pairs found in the blank can be added to the mass column in the mzBackground file. Only the m/z_light masses need to be added. The program automatically removes the corresponding m/z_heavy peaks. Note 3: If there is no background peak to be removed, the mzBackground csv file should still contain "m/z" in row 1, but leave all the other cells blank.	
mzTolGroup	Mass tolerance to group the same peak pairs from the neighboring spectra.	Note: the mass tolerance value used is depending on the mass measurement accuracy of the instrument used. There is no need to
mzTolIso	Mass tolerance to search ^{13}C isotopologues within a pair.	
mz_tol_dt	Mass tolerance to filter derivative pairs.	

mzTolChargeVec	Mass tolerance to calculate the charge number of an ion.	adjust the individual mass tolerance values (i.e., use the default values already given), unless a user wants to fine-tune these values to see if any additional improvement on the specificity and sensitivity of IsoMS can be made.
mzTolFixLev3	Mass tolerance to pair the peaks for level 3 pairs.	
mzTolDimer	Mass tolerance to filter dimer or other multimer pairs.	
mzTolBackground	Mass tolerance to filter background noises.	
mz_tol_ctfilter	Mass tolerance to filter multiply charged ions.	
minIntensity	Lowest intensity of exported peak pairs.	
minRatio	The peak intensity ratio range of exported peak pairs.	
MaxRatio		
minSN	Only for FTMS data: the lowest S/N of exported peak pairs.	

3. Installing and running IsoMS-align

1) Download the IsoMS-align script from the MyCompoundID website.
2) Assign the folder of IsoMS-align as the working folder of RGui by cliking: File → Change dir (Figure 7).

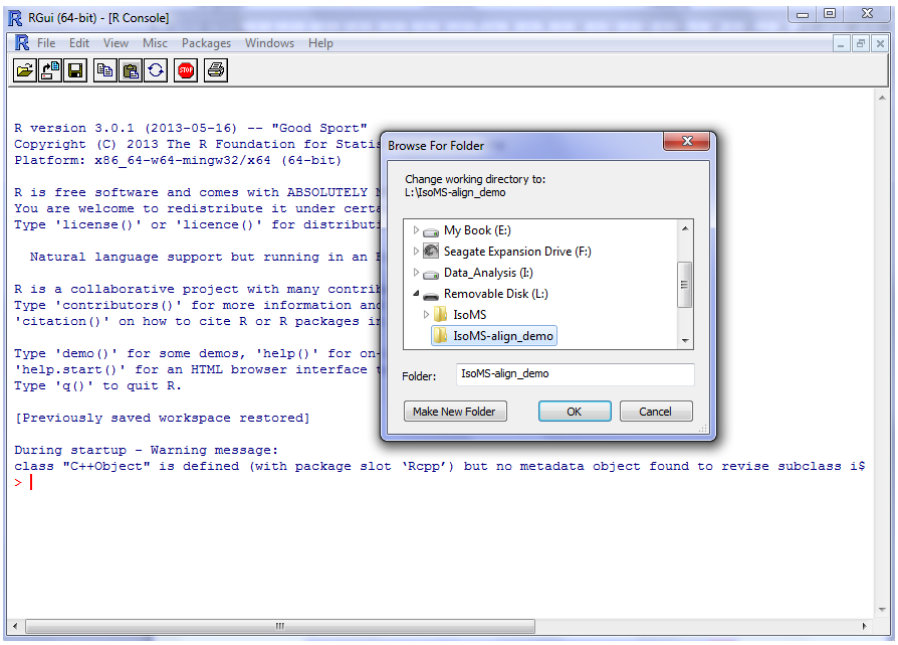


Figure 7. Select the alignment script.

3) Open the IsoMS-align script (Figure 8) and change the parameters therein (see Table 2 for the explanation of these parameters).

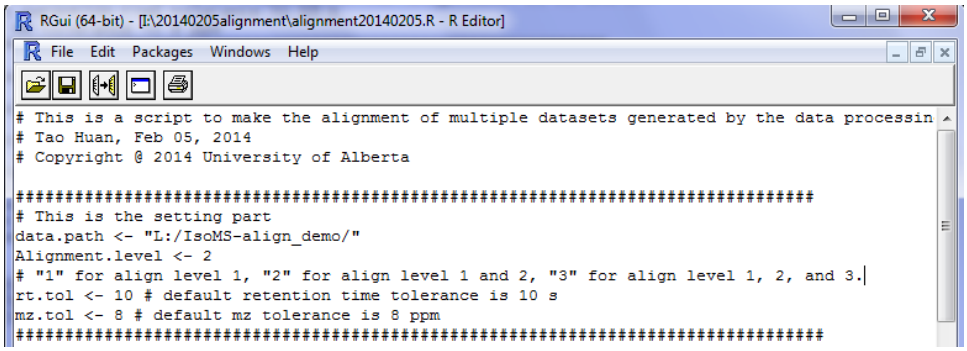


Figure 8. Change the parameters in IsoMS-align.

Table 2. IsoMS-align parameters that need to be changed.

Parameter	Function
data.path	Set the data path to the folder that contains all the peak pair csv files.
Alignment.level	Select the levels of peak pairs to perform the alignment.
rt.tol	Set the retention time tolerance during the alignment process.
mz.tol	Set the mz tolerance during the alignment process.

4) Save the parameter changes to the script. Type in the command code in RGui as shown in red in Figure 9 and press enter to run the script.

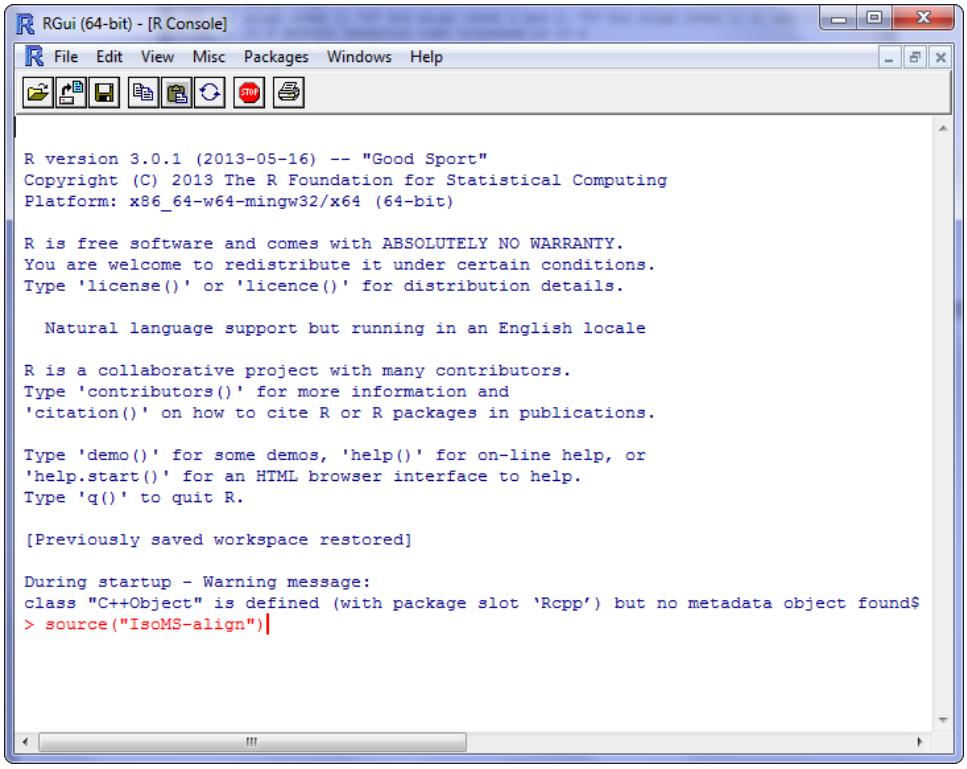


Figure 9. Type in the command code in RGui and then run the script.

5) After running the script, a new csv file named "Alignment_result.csv" will be created. This csv file contains the alignment result.

Appendix B

Table B.1 Putative metabolites identified by HMDB database search using the molecular masses of peak-pairs commonly detected in both 100% H_2O and 80% CH_3OH extraction methods**.**

Retention Time (sec)	m/z light	HMDB ID	Common Name	In PlantCyc ?	Identified by standard*
960	293.1319	HMDB34006	1-Propylamine	N	
435	295.1112	HMDB00149	Ethanolamine	Y	
1089	305.1320	HMDB31641	Pyrrolidine	Y	
1079	307.1113	HMDB12114	(3S)-3,6-Diaminohexanoate	N	
		HMDB12115	3,5-diaminohexanoic	N	
		HMDB00182	L-lysine	Y	
		HMDB03405	D-Lysine	N	
1170	307.1476	HMDB34198	2-Methyl-1-propylamine	N	
		HMDB32179	sec-Butylamine	N	
		HMDB31321	1-Butylamine	Y	
		HMDB41878	Diethylamine	N	
478	309.0903	HMDB00123	Glycine	Y	Y
664	309.0905	HMDB00123	Glycine	Y	N
603	309.1267	HMDB12136	1-Amino-propan-2-ol	N	
530	309.1267	HMDB12136	1-Amino-propan-2-ol	N	
1267	310.0750	HMDB00292	Xanthine	Y	
1181	317.5933	HMDB29873	(S)-Spinacine	N	
		HMDB60362	2,4-Diamino-6-nitrotoluene	N	
939	321.1267	HMDB01080	4-Aminobutyraldehyde	Y	
		HMDB31215	*N*-Ethylacetamide	N	
		HMDB31581	Morpholine	N	
		HMDB33870	Butyramide	N	
698	321.1268	HMDB01080	4-Aminobutyraldehyde	Y	
		HMDB31215	*N*-Ethylacetamide	N	
		HMDB31581	Morpholine	N	
		HMDB33870	Butyramide	N	

1221	321.1633	HMDB31659	3-Methyl-1-butylamine	N	
		HMDB32392	2-Methylbutylamine	N	
1366	323.0718	HMDB32951	4,7-Dihydroxy-2H-1-benzopyran-2-one	N	
		HMDB32950	5,7-Dihydroxy-4H-1-benzopyran-4-one	N	
		HMDB30819	Aesculetin	Y	
560	323.1060	HMDB00056	Beta-Alanine	Y	N
		HMDB00161	L-Alanine	Y	Y
		HMDB00271	Sarcosine	N	
		HMDB01310	D-Alanine	Y	
1294	324.0792	HMDB03501	3,4-Dihydroxy-trans-cinnamate	Y	
		HMDB32057	(E)-3-(2,3-Dihydroxyphenyl)-2-propenoic acid	N	
		HMDB01964	Caffeic acid	Y	
		HMDB32131	3-(3,5-Dihydroxyphenyl)-2-propenoic acid	N	
1249	324.5957	HMDB02184	L-Threo-3-Phenylserine	N	
		HMDB06050	*o*-Tyrosine	N	
		HMDB00158	Tyrosine	Y	
		HMDB59720	Meta-Tyrosine	N	
		HMDB60280	4,6,7-Trihydroxy-1,2,3,4-tetrahydroisoquinoline	N	
		HMDB03831	Beta-Tyrosine	N	
1343	326.0769	HMDB01866	3,4-Dihydroxymandelic acid	N	
		HMDB13198	4-O-Methyl gallic acid	N	
		HMDB29650	Methyl 2,4,6-trihydroxybenzoate	N	
1296	326.0770	HMDB01866	3,4-Dihydroxymandelic acid	N	
		HMDB13198	4-O-Methyl gallic acid	N	
		HMDB29650	Methyl 2,4,6-trihydroxybenzoate	N	
1271	331.5851	HMDB06116	3-Hydroxyhippuric acid	N	
		HMDB04067	Leucodopachrome	Y	

		HMDB39119	L-Dopaquinone	Y	
		HMDB00840	Salicyluric acid	N	
		HMDB02016	4-Carboxyphenylglycine	N	
1297	331.6032	HMDB14903	Metyrosine	Y	
		HMDB60065	*N*-methyl-4,6,7-trihydroxy-1,2,3,4-tetrahydroisoquinoline	N	
		HMDB41943	*N*-acetyldopamine	N	
		HMDB29217	Tyrosine methylester	N	
907	335.1425	HMDB12815	5-aminopentanal	N	
562	337.1216	HMDB00112	Gamma-Aminobutyric acid	Y	Y
		HMDB00452	L-Alpha-aminobutyric acid	Y	
		HMDB00650	D-Alpha-aminobutyric acid	N	
		HMDB01906	2-Aminoisobutyric acid	N	
		HMDB02166	(S)-β -aminoisobutyric acid	Y	
		HMDB02299	(R)-β-aminoisobutyric acid	Y	
		HMDB03911	3-Aminoisobutanoic acid	Y	
655	337.1217	HMDB00112	Gamma-Aminobutyric acid	Y	N
		HMDB00452	L-Alpha-aminobutyric acid	Y	
		HMDB00650	D-Alpha-aminobutyric acid	N	
		HMDB01906	2-Aminoisobutyric acid	N	
		HMDB02166	(S)-β-aminoisobutyric acid	Y	
		HMDB02299	(R)-β-aminoisobutyric acid	Y	
		HMDB03911	3-Aminoisobutanoic acid	Y	
313	339.1010	HMDB00187	L-Serine	Y	Y
		HMDB03406	D-Serine	Y	
1101	343.1111	HMDB35882	2-Acetylpyrrole	N	
		HMDB01169	4-aminophenol	N	
900	345.0916	HMDB28785	Cysteinyl-Threonine	N	
		HMDB00099	Cystathionine	Y	
		HMDB29058	Threoninyl-Cysteine	N	
		HMDB00455	Allocystathionine	N	
799	346.0856	HMDB60760	4-Carboxypyrazole	N	

		HMDB00300	Uracil	Y	
1126	347.0878	HMDB01904	3-Nitrotyrosine	N	
566	348.1375	HMDB00323	3-Amino-2-piperidone	N	
		HMDB61162	N-Mononitrosopiperazine	N	
725	349.1216	HMDB00162	L-Proline	Y	Y
		HMDB03411	D-Proline	N	
		HMDB12880	Acetamidopropanal	N	
		HMDB30409	4-Amino-2-methylenebutanoic acid	N	
		HMDB34208	Pterolactam	N	
767	351.1373	HMDB00883	L-Valine	Y	Y
		HMDB02141	*N*-Methyl-a-aminoisobutyric acid	N	
		HMDB03355	5-Aminopentanoic acid	N	
		HMDB13716	Norvaline	Y	
849	351.1373	HMDB00883	L-Valine	Y	N
		HMDB02141	*N*-Methyl-a-aminoisobutyric acid	N	
		HMDB03355	5-Aminopentanoic acid	N	
		HMDB13716	Norvaline	Y	
643	351.1373	HMDB00883	L-Valine	Y	N
		HMDB02141	*N*-Methyl-a-aminoisobutyric acid	N	
		HMDB03355	5-Aminopentanoic acid	N	
		HMDB13716	Norvaline	Y	
463	353.1166	HMDB00167	L-Threonine	Y	Y
		HMDB00719	L-Homoserine	Y	N
		HMDB04041	L-Allothreonine	Y	
		HMDB61148	Hydroxyethyl glycine	N	
1042	358.1108	HMDB59709	2-Hydroxybenzyl alcohol	N	
		HMDB59712	3-Hydroxybenzyl alcohol	N	
		HMDB00873	4-methylcatechol	N	N
		HMDB11724	4-hydroxybenzyl	N	
		HMDB01398	Guaiacol	N	
127	359.0732	HMDB00251	Taurine	N	Y

781	360.1014	HMDB02024	Imidazoleacetic acid	N	
898	360.1014	HMDB02024	Imidazoleacetic acid	N	
758	363.1007	HMDB01843	*N*-Acryloylglycine	N	
		HMDB00805	2-Pyrrolidone-5-carboxylic acid	N	
686	363.1010	HMDB01843	*N*-Acryloylglycine	N	
		HMDB00805	2-Pyrrolidone-5-carboxylic acid	N	
848	363.1374	HMDB00070	Pipecolic acid	N	
		HMDB00716	L-Pipecolic acid	N	
		HMDB04226	N4-Acetylaminobutanal	N	
		HMDB05960	D-Pipecolic acid	N	
		HMDB29435	L-trans-4-Methyl-2-pyrrolidinecarboxylic acid	N	
896	365.1529	HMDB00172	L-Isoleucine	Y	N
		HMDB00557	L-Alloisoleucine	N	
		HMDB00687	L-Leucine	Y	Y
		HMDB01645	L-Norleucine	Y	
		HMDB01901	Aminocaproic acid	Y	N
		HMDB03640	Beta-Leucine	N	
216	366.1482	HMDB03374	D-Ornithine	N	
		HMDB00214	Ornithine	Y	
419	367.0958	HMDB00191	L-Aspartic acid	Y	N
		HMDB06483	D-Aspartic acid	N	
		HMDB11753	Iminodiacetate	N	
375	367.0959	HMDB00191	L-Aspartic acid	Y	Y
		HMDB06483	D-Aspartic acid	N	
		HMDB11753	Iminodiacetate	N	
1182	367.1111	HMDB59805	5-Hydroxyindole	N	
		HMDB29757	4-Hydroxybenzeneacetonitrile	Y	
586	367.1321	HMDB39499	1,4-Dideoxy-1,4-imino-D-ribitol	N	
		HMDB31658	L-2-Amino-5-hydroxypentanoic acid	N	
		HMDB32356	*N*-Lactoyl ethanolamine	N	

1064	367.6067	HMDB00085	Deoxyguanosine	N	
		HMDB00050	Adenosine	N	
		HMDB14340	Vidarabine	N	
754	369.1126	HMDB00034	Adenine	Y	Y
757	370.0967	HMDB00157	Hypoxanthine	Y	
		HMDB14581	Allopurinol	Y	
695	370.0968	HMDB00157	Hypoxanthine	Y	
		HMDB14581	Allopurinol	Y	
1237	370.1109	HMDB03767	4-Hydroxyphenylacetaldehyde	Y	
1051	371.1061	HMDB01123	2-Aminobenzoic acid	Y	
		HMDB01891	m-Aminobenzoic acid	N	N
		HMDB01392	p-Aminobenzoic acid	Y	N
		HMDB15687	Salicylamide	N	
1093	372.0900	HMDB00500	4-Hydroxybenzoic acid	Y	Y
		HMDB01895	Salicylic acid	Y	N
		HMDB02466	3-Hydroxybenzoic acid	Y	
		HMDB04062	Gentisate aldehyde	N	
		HMDB59965	3,4-Dihydroxybenzaldehyde	Y	
900	372.1012	HMDB00301	Urocanic acid	N	
		HMDB34174	(E)-Urocanic acid	N	
		HMDB02730	Nicotinamide N-oxide	N	
1173	372.1264	HMDB04284	Tyrosol	Y	
		HMDB40177	3-Ethyl-1,2-benzenediol	N	
		HMDB32036	4-Ethoxyphenol	N	
		HMDB32136	2-Methoxy-4-methylphenol	N	
		HMDB40174	4-Ethyl-1,2-benzenediol	N	
		HMDB32137	2,6-Dimethyl-1,4-benzenediol	N	
921	373.0852	HMDB01232	4-Nitrophenol	N	
1012	373.0852	HMDB01232	4-Nitrophenol	N	
963	373.0853	HMDB01232	4-Nitrophenol	N	
1273	373.0855	HMDB01232	4-Nitrophenol	N	

1292	375.1011	HMDB01280	2-aminomuconic semialdehyde	N	
1196	379.1114	HMDB29737	1H-Indole-3-carboxaldehyde	Y	
1268	379.1114	HMDB29737	1H-Indole-3-carboxaldehyde	Y	
242	379.1437	HMDB03464	4-Guanidinobutanoic acid	N	
219	380.1275	HMDB00641	L-Glutamine	Y	Y
		HMDB02031	Ureidoisobutyric acid	Y	
		HMDB03423	D-Glutamine	Y	
		HMDB06899	Alanylglycine	N	
370	381.1114	HMDB00148	L-Glutamic acid	Y	Y
		HMDB02393	*N*-Methyl-D-aspartic acid	Y	
		HMDB02931	*N*-Acetylserine	N	
		HMDB03011	O-Acetylserine	Y	
		HMDB03339	D-Glutamic acid	Y	
		HMDB06556	L-4-Hydroxyglutamate semialdehyde	N	
933	383.1054	HMDB04058	5,6-Dihydroxyindole	Y	
		HMDB60318	(S)-4-Hydroxymandelonitrile	Y	
1321	385.0679	HMDB40578	4-Thiocyanatophenol	Y	
621	385.1077	HMDB00132	Guanine	Y	
		HMDB00403	2-Hydroxyadenine	N	
		HMDB00542	8-Hydroxyadenine	N	
894	385.1222	HMDB12992	Leukoaminochrome	N	
		HMDB29703	Methyl 2-aminobenzoate	N	
		HMDB60680	2-Amino-3-methylbenzoate	N	
		HMDB01859	Acetaminophen	N	
		HMDB32609	2-(Methylamino)benzoic acid	N	
		HMDB12219	Dopamine-quinone	Y	
657	386.0918	HMDB00786	Oxipurinol	N	
		HMDB01182	6,8-Dihydroxypurine	N	
		HMDB00292	Xanthine	Y	
1224	386.1057	HMDB00020	p-Hydroxyphenylacetic acid	N	
		HMDB00440	3-Hydroxyphenylacetic acid	N	

		HMDB00669	Ortho-Hydroxyphenylacetic acid	N	
		HMDB02390	3-Cresotinic acid	N	
		HMDB03791	3,4-Dihydroxyphenylacetaldehyde	Y	
		HMDB04815	4-Hydroxy-3-methylbenzoic acid	N	
		HMDB12308	Vanillin	Y	
1276	386.1058	HMDB00020	*p*-Hydroxyphenylacetic acid	N	
		HMDB00440	3-Hydroxyphenylacetic acid	N	
		HMDB00669	Ortho-Hydroxyphenylacetic acid	N	
		HMDB02390	3-Cresotinic acid	N	
		HMDB03791	3,4-Dihydroxyphenylacetaldehyde	Y	
		HMDB04815	4-Hydroxy-3-methylbenzoic acid	N	
		HMDB12308	Vanillin	Y	
1038	386.1059	HMDB00020	*p*-Hydroxyphenylacetic acid	N	
		HMDB00440	3-Hydroxyphenylacetic acid	N	
		HMDB00669	Ortho-Hydroxyphenylacetic acid	N	
		HMDB02390	3-Cresotinic acid	N	
		HMDB03791	3,4-Dihydroxyphenylacetaldehyde	Y	
		HMDB04815	4-Hydroxy-3-methylbenzoic acid	N	
		HMDB12308	Vanillin	Y	
1286	389.0804	HMDB02916	4-Nitrocatechol	N	
		HMDB06200	2,4-Dihydroxy-nitrophenol	N	
1100	389.1278	HMDB00177	L-Histidine	Y	Y
		HMDB34267	L-2-Amino-3-(1-pyrazolyl)propanoic acid	Y	
876	390.6002	HMDB33662	(S)-5'-Deoxy-5'-(methylsulfinyl)adenosine	N	
1224	393.0957	HMDB37501	Dihydroisorhamnetin	N	
		HMDB02937	Tamarixetin	N	
934	393.1843	HMDB00991	DL-2-Aminooctanoic acid	N	N
1307	394.1215	HMDB37321	Oleacein	N	
610	395.1272	HMDB00510	Aminoadipic acid	Y	N

		HMDB29423	Acetylhomoserine	Y	
548	395.1273	HMDB00510	Aminoadipic acid	Y	N
		HMDB29423	Acetylhomoserine	Y	
1130	398.1056	HMDB01713	*m*-Coumaric acid	Y	
		HMDB02035	4-Hydroxycinnamic acid	Y	
		HMDB02641	2-Hydroxycinnamic acid	Y	
		HMDB41592	Coumarinic acid	Y	
		HMDB30677	cis-p-Coumaric acid	N	
902	399.1373	HMDB00159	L-Phenylalanine	Y	Y
		HMDB04992	Benzocaine	N	
		HMDB06044	Norsalsolinol	N	
1230	400.1212	HMDB00375	3-(3-Hydroxyphenyl)propanoic acid	Y	
		HMDB02199	Desaminotyrosine	N	
		HMDB05175	Homovanillin	N	
1326	400.1214	HMDB00375	3-(3-Hydroxyphenyl)propanoic acid	Y	
		HMDB02199	Desaminotyrosine	N	
		HMDB05175	Homovanillin	N	
1036	400.1217	HMDB00375	3-(3-Hydroxyphenyl)propanoic acid	Y	
		HMDB02199	Desaminotyrosine	N	
		HMDB05175	Homovanillin	N	
916	401.1165	HMDB60374	3-Methoxyanthranilate	N	
1184	401.1284	HMDB29873	(S)-Spinacine	N	
		HMDB60362	2,4-Diamino-6-nitrotoluene	N	
1080	402.1006	HMDB00130	Homogentisic acid	Y	N
		HMDB00484	Vanillic acid	N	Y
		HMDB00750	3-Hydroxymandelic acid	N	
		HMDB00822	p-Hydroxymandelic acid	N	
		HMDB01336	3,4-Dihydroxybenzeneacetic acid	N	
		HMDB01868	5-Methoxysalicylic acid	N	N
		HMDB06242	3,4-Dihydroxymandelaldehyde	N	
885	402.1006	HMDB00130	Homogentisic acid	Y	N

		HMDB00484	Vanillic acid	N	N
		HMDB00750	3-Hydroxymandelic acid	N	
		HMDB00822	p-Hydroxymandelic acid	N	
		HMDB01336	3,4-Dihydroxybenzeneacetic acid	N	
		HMDB01868	5-Methoxysalicylic acid	N	N
		HMDB06242	3,4-Dihydroxymandelaldehyde	N	
1274	403.0960	HMDB00439	2-Furoylglycine	N	
1214	412.1211	HMDB39471	3-(4-Hydroxy-2-methoxyphenyl)-2-propena	N	
1225	414.1006	HMDB03501	3,4-Dihydroxy-trans-cinnamate	N	
		HMDB00707	4-Hydroxyphenylpyruvic acid	Y	
		HMDB32057	(E)-3-(2,3-Dihydroxyphenyl)-2-propenoic acid	Y	
		HMDB32131	3-(3,5-Dihydroxyphenyl)-2-propenoic acid	N	
		HMDB11663	3-Hydroxyphenylpyruvic acid	N	
		HMDB06915	2-Hydroxy-3-(4-hydroxyphenyl)propenoic acid	Y	
1036	416.1161	HMDB00118	Homovanillic acid	N	N
		HMDB00333	Isohomovanillic acid	N	
		HMDB00423	3,4-Dihydroxyhydrocinnamic acid	Y	
		HMDB00755	Hydroxyphenyllactic acid	Y	N
		HMDB02643	3-(3-hydroxyphenyl)-3-hydroxypropanoic acid	N	
		HMDB04061	3-Methoxy-4-hydroxyphenylglycolaldehyde	N	
1201	416.1163	HMDB00118	Homovanillic acid	N	N
		HMDB00333	Isohomovanillic acid	N	
		HMDB00423	3,4-Dihydroxyhydrocinnamic acid	Y	
		HMDB00755	Hydroxyphenyllactic acid	Y	N
		HMDB02643	3-(3-hydroxyphenyl)-3-hydroxypropanoic acid	N	

		HMDB04061	3-Methoxy-4-hydroxyphenylglycolaldehyde	N	
921	416.1165	HMDB00118	Homovanillic acid	N	N
		HMDB00333	Isohomovanillic acid	N	
		HMDB00423	3,4-Dihydroxyhydrocinnamic acid	Y	
		HMDB00755	Hydroxyphenyllactic acid	Y	N
		HMDB02643	3-(3-hydroxyphenyl)-3-hydroxypropanoic acid	N	
		HMDB04061	3-Methoxy-4-hydroxyphenylglycolaldehyde	N	
849	418.1318	HMDB35056	1-(3-Hydroxy-4-methoxyphenyl)-1,2-ethanediol	N	
		HMDB01490	Vanylglycol	N	
		HMDB40603	(±)-erthyro-1-(4-Hydroxyphenyl)-1,2,3-propanetriol	N	
1334	421.1269	HMDB38111	Kievitone hydrate	N	
		HMDB38112	Kievitol	N	
		HMDB33279	8-Hydroxypinoresinol	N	
1044	421.1790	HMDB37790	Polyethylene glycol	Y	
		HMDB13010	N-Heptanoylglycine	N	
1125	428.1163	HMDB00954	trans-Ferulic acid	Y	Y
		HMDB00955	Isoferulic acid	N	N
792	430.0955	HMDB59809	4-Hydroxy-benzenepropanedioate	N	
1096	430.0956	HMDB59809	4-Hydroxy-benzenepropanedioate	N	
1032	430.1322	HMDB60733	3-(3-Hydroxyphenyl)-2-methyllactic acid	N	
		HMDB60398	5-Hydroxyconiferyl alcohol	N	
		HMDB29645	Xanthoxylin	N	
430	431.1750	HMDB29422	L-Histidine trimethylbetaine	N	
867	432.1113	HMDB00291	Vanillylmandelic acid	N	
		HMDB03503	3-(3,4-Dihydroxyphenyl)lactic acid	Y	
		HMDB33836	Ethyl gallate	N	

1049	432.1114	HMDB00291	Vanillylmandelic acid	N	
		HMDB03503	3-(3,4-Dihydroxyphenyl)lactic acid	Y	
		HMDB33836	Ethyl gallate	N	
806	438.1481	HMDB00929	L-Tryptophan	Y	Y
		HMDB13609	D-Tryptophan	Y	
1207	440.1163	HMDB29775	Vinyl caffeate	N	
		HMDB36627	Eugenin	N	
		HMDB29472	Isoeugenitol	N	
1064	450.1481	HMDB35665	L-1,2,3,4-Tetrahydro-beta-carboline-3-carboxylic acid	Y	
1072	458.1264	HMDB41691	5-(3',4',5'-Trihydroxyphenyl)-gamma-valerolactone	N	
846	459.1220	HMDB60026	Vanilloylglycine	N	
776	460.1058	HMDB12283	Prephenic acid	Y	
		HMDB12199	Chorismic acid	Y	
572	478.1275	HMDB00767	Pseudouridine	N	
665	494.1951	HMDB11171	L-gamma-glutamyl-l-leucine	N	
		HMDB11170	L-gamma-glutamyl-l-isoleucine	N	
1428	541.2732	HMDB38457	Dihydrocapsaicin	N	

* Y: identified with authentic standards based on accurate mass and retention time.

N: not matched with authentic standards

Table B.2 List of putative metabolites identified by the EML predicted human metabolite library using the molecular masses of peak-pairs commonly detected in both 100% H_2O and 80% CH_3OH extraction methods. *Click on the link of the matches to see the matched metabolites in EML at the website www.mycompoundid.org.

	Retention Time (sec)	m/z light	Number of matched entries
1	1345	305.0717	6
2	1287	305.0718	6
3	993	308.0951	40
4	707	310.0744	15
5	1383	310.0819	31
6	1273	311.5693	6
7	1374	318.0796	24
8	1137	318.5884	10
9	898	319.1112	16
10	959	322.0744	30
11	732	323.1425	13
12	1264	325.0694	13
13	1082	329.0954	5
14	399	336.1377	10
15	429	337.1580	5
16	506	337.1581	5
17	952	338.1420	11
18	1282	340.0747	10
19	781	340.1212	19
20	710	340.1214	19
21	474	345.0904	8
22	876	346.0744	8
23	1018	350.1059	58
24	1003	356.0359	1
25	1228	356.0953	22

26	1193	357.0905	9
27	885	359.1060	15
28	934	360.0901	15
29	1162	362.1056	32
30	784	362.1532	8
31	422	363.1486	4
32	536	363.1733	10
33	628	368.1161	64
34	795	368.1526	11
35	1116	368.1527	11
36	687	376.0848	6
37	748	376.0850	6
38	1318	376.1578	28
39	828	379.1686	19
40	794	380.1163	60
41	835	380.1165	60
42	1209	384.0901	25
43	721	384.1475	17
44	1271	386.1533	8
45	153	388.1075	12
46	1021	388.1214	7
47	1067	394.0510	2
48	534	398.1268	74
49	1040	399.1010	17
50	808	400.0849	18
51	747	405.1162	50
52	848	405.1163	50
53	1022	407.1996	8
54	1323	408.1189	6
55	124	411.9502	2
56	809	416.0796	13

57	836	417.1112	23
58	122	421.9792	1
59	1077	423.1008	8
60	914	423.1948	5
61	833	432.1478	6
62	716	444.0747	3
63	736	448.1423	6
64	422	452.0733	1
65	453	454.0911	3
66	1383	454.2413	3
67	837	458.0722	3
68	633	460.1171	19
69	1006	460.1425	12
70	122	499.9952	1
71	1240	522.2880	7
72	123	525.0838	3
73	120	571.0611	1

Table B.3 Putative metabolites unique to 100% H_2O extracts identified by HMDB database search.

Retention Time (sec)	m/z light	HMDB ID	Common Name	In PlantCyc ?	Identified by standard*
1199	278.1083	HMDB01414	Putrescine	Y	
1395	289.0768	HMDB00957	Pyrocatechol	Y	
		HMDB02434	Hydroquinone	N	
1214	292.1056	HMDB12176	5-Aminopentanamide	N	
1317	302.6004	HMDB00306	Tyramine	Y	
		HMDB04989	*m*-Tyramine	N	
		HMDB60681	4-Hydroxy-2,6-dimethylaniline	N	
984	319.1111	HMDB02039	2-Pyrrolidinone	N	
872	321.1266	HMDB31215	*N*-Ethylacetamide	N	
		HMDB31581	Morpholine	N	
		HMDB33870	Butyramide	N	
		HMDB01080	4-Aminobutyraldehyde	Y	
1369	331.0873	HMDB30765	trans-4-Hydroxymellein	N	
915	335.1423	HMDB12815	5-aminopentanal	Y	
499	348.1012	HMDB60245	L-3-Cyanoalanine	Y	
		HMDB00076	Dihydrouracil	Y	
557	362.1168	HMDB29874	Squamolone	N	
		HMDB31547	L-Cyclo(alanylglycyl)	N	
		HMDB00079	Dihydrothymine	Y	
783	363.1007	HMDB00805	2-Pyrrolidone-5-carboxylic acid	N	
		HMDB01843	*N*-Acryloylglycine	N	
		HMDB00267	Pyroglutamic acid	N	
217	366.1117	HMDB00026	Ureidopropionic acid	Y	
		HMDB00168	L-Asparagine	Y	Y
		HMDB11733	Glycyl-glycine	N	
498	367.0957	HMDB00191	L-Aspartic acid	Y	N
		HMDB06483	D-Aspartic acid	N	
		HMDB11753	Iminodiacetate	N	

358	367.1322	HMDB39499	1,4-Dideoxy-1,4-imino-D-ribitol	N	
		HMDB31658	L-2-Amino-5-hydroxypentanoic acid	N	
		HMDB32356	N-Lactoyl ethanolamin	N	
1291	375.1009	HMDB01280	2-Aminomuconic acid semialdehyde	N	
880	375.6041	HMDB03333	8-hydroxy-deoxyguanosine	Y	
		HMDB00133	Guanosine	Y	
460	379.1321	HMDB29426	L-cis-5-Hydroxy-2-piperidinecarboxylic acid	N	
		HMDB59595	(*S*)-2-amino-6-oxohexanoate	Y	
		HMDB12151	2-Keto-6-aminocaproate	N	
		HMDB29425	(2*R*,4*S*)-4-(Hydroxymethyl)-2-pyrrolidinecarboxylic acid	N	
		HMDB03681	4-Acetamidobutanoic acid	N	
		HMDB00730	Isobutyrylglycine	N	
		HMDB00808	*N*-Butyrylglycine	N	
		HMDB01263	Allysine	N	
		HMDB12131	(*S*)-5-Amino-3-oxohexanoate	N	
679	381.1480	HMDB29449	L-Ribo-2-Amino-4-hydroxy-3-methylpentanoic acid	N	
		HMDB33453	4-Epifagomine	N	
1103	385.1077	HMDB00132	Guanine	Y	
		HMDB00403	2-Hydroxyadenine	N	
		HMDB00542	8-Hydroxyadenine	N	
525	393.1120	HMDB29433	L-2-Amino-4-methylenepentanedioic acid	N	
		HMDB34442	L-trans-alpha-Amino-2-carboxycyclopropaneacetic acid	N	
485	394.1429	HMDB28680	Alanyl-Alanine	Y	
		HMDB31411	4-Acetamido-2-aminobutanoic acid	Y	
250	399.1042	HMDB02005	Methionine sulfoxide	Y	Y

		HMDB37057	Ethiin	Y	
1284	407.1119	HMDB37250	Muscomin	N	
554	408.1591	HMDB03357	*N*-Acetylornithine	Y	
		HMDB34365	L-Theanine	N	
		HMDB28854	Glycyl-Valine	N	
		HMDB29127	Valyl-Glycine	N	
160	408.1701	HMDB00517	L-Arginine	Y	Y
		HMDB03416	D-Arginine	N	
300	409.1535	HMDB00904	Citrulline	Y	
1395	414.1370	HMDB12915	Coniferyl alcohol	Y	
785	422.1743	HMDB00206	N6-Acetyl-L-lysine	N	N
		HMDB00446	N-Alpha-acetyllysine	N	
		HMDB00759	Glycyl-L-leucine	N	
		HMDB28844	Glycyl-Isoleucine	N	
		HMDB28907	Isoleucyl-Glycine	N	
		HMDB28929	Leucyl-Glycine	N	
		HMDB28700	Alanyl-Valine	N	
		HMDB29120	Valyl-Alanine	N	
242	424.1533	HMDB29054	Threoninyl-Alanine	N	
		HMDB01370	Diaminopimelic acid	Y	
		HMDB31412	2-Amino-4-(lactoylamino)butyric acid	N	
		HMDB28697	Alanyl-Threonine	Y	
		HMDB39222	N-(gamma-Glutamyl)ethanolamine	N	
799	436.1899	HMDB28691	Alanyl-Leucine	Y	
		HMDB28690	Alanyl-Isoleucine	N	
		HMDB28900	Isoleucyl-Alanine	N	
		HMDB28922	Leucyl-Alanine	N	
772	448.1906	HMDB03581	Dethiobiotin	Y	
		HMDB29030	Prolyl-Valine	N	
		HMDB29135	Valyl-Proline	N	

394	452.1485	HMDB06248	5-L-Glutamyl-L-alanin	N
		HMDB29040	Serinyl-Hydroxyproline	N
		HMDB28872	Hydroxyprolyl-Serine	N
		HMDB03764	Glutamylalanine	Y
		HMDB36301	*N*-gamma-L-Glutamyl-D-alanine	N
876	462.2056	HMDB11174	L-isoleucyl-L-proline	Y
		HMDB11175	L-leucyl-L-proline	N
999	478.2369	HMDB28911	Isoleucyl-Leucine	N
		HMDB28910	Isoleucyl-Isoleucine	N
		HMDB28932	Leucyl-Isoleucine	N
		HMDB28933	Leucyl-Leucine	Y
342	501.1544	HMDB00085	Deoxyguanosine	Y
		HMDB00050	Adenosine	Y
		HMDB14340	Vidarabine	N
196	509.1696	HMDB05766	Norophthalmic acid	N
		HMDB11738	gamma-Glutamylglutamine	N
294	510.1541	HMDB11737	Gamma Glutamylglutamic acid	N
138	517.1497	HMDB00133	Guanosine	Y
		HMDB03333	8-hydroxy-deoxyguanosine	Y

* Y: identified with authentic standards based on accurate mass and retention time.

N: not matched with authentic standards

Table B.4 Putative metabolites unique to 100% H_2O extracts identified by the EML predicted human metabolite library.

	Retention Time (sec)	m/z light	Number of matched entries
1	336	292.0302	1
2	1289	298.0641	8
3	426	339.1010	2
4	353	398.1268	74
5	342	426.6910	6
6	1154	433.1789	9
7	951	434.1741	11
8	1318	451.1372	14
9	1166	476.2216	4
10	323	576.1739	48

Table B.5 Putative metabolites unique to 80% CH_3OH extracts identified by HMDB database search..

Retention Time (sec)	m/z light	HMDB ID	Common Name	In PlantCyc?	Identified by standard*
1453	289.0767	HMDB02434	Hydroquinone	Y	
		HMDB00957	Pyrocatechol	Y	
		HMDB32037	1,3-Benzenediol	N	
1447	318.0797	HMDB29644	2',4',6'-Trihydroxyacetophenon	N	
		HMDB06242	3,4-Dihydroxymandelaldehyde	Y	
		HMDB00130	Homogentisic	Y	
		HMDB00822	*p*-Hydroxymandelic acid	N	
1275	328.1005	HMDB00228	Phenol	Y	Y
921	335.1425	HMDB12815	5-Aminopentanal	Y	
827	337.1216	HMDB38394	O-Acetylethanolamine	N	
		HMDB31654	3-Aminobutanoic acid	Y	
		HMDB02299	(*R*)-β-aminoisobutyric acid	Y	
		HMDB41945	*N*-ethylglycine	Y	
		HMDB03911	beta-aminoisobutyric acid	Y	
403	345.1381	HMDB00870	Histamine	N	
		HMDB60263	Histamium	Y	
191	366.1483	HMDB00214	Ornithine	Y	N
		HMDB03374	D-Ornithine	N	
304	367.1322	HMDB39499	1,4-Dideoxy-1,4-imino-D-ribitol	N	
		HMDB31658	L-2-Amino-5-hydroxypentanoic acid	Y	
		HMDB32356	*N*-Lactoyl ethanolamine	N	
1317	379.0980	HMDB30144	alpha-Cotonefuran	N	
		HMDB32322	*cis*-3 and *trans*-2-Hexenyl propionate	Y	
		HMDB01871	Epicatechin	Y	
		HMDB41293	6-Hydroxy-alpha-pyrufuran	N	
		HMDB29253	3-Hydroxyphloretin	N	

		HMDB37953	(-)-Catechin	N	
		HMDB37954	(+)-Epicatechin	N	
		HMDB41310	Luteoforol	N	
		HMDB02780	Catechin	Y	
1209	390.0824	HMDB29276	Cis-Caffeoyl tartaric acid	N	
		HMDB13680	Caftaric acid	N	
621	395.1272	HMDB00510	Aminoadipic acid	Y	N
		HMDB29423	Acetylhomoserine	Y	
1476	413.1295	HMDB35789	Matairesinol	N	
		HMDB35698	(-)-Matairesinol	Y	
		HMDB38113	5-Deoxykievitone hydrate	N	
		HMDB35467	Glicophenone	N	
1104	414.1370	HMDB12915	Coniferyl alcohol	Y	
1271	416.1165	HMDB00118	Homovanillic acid	N	Y
		HMDB00333	Isohomovanillic acid	N	
		HMDB00423	3,4-Dihydroxyhydrocinnamic acid	N	
		HMDB00755	Hydroxyphenyllactic acid	Y	N
		HMDB02643	3-(3-hydroxyphenyl)-3-hydroxypropanoic acid	N	
		HMDB04061	3-Methoxy-4-hydroxyphenylglycolaldehyde	N	
407	422.1746	HMDB00206	*N6*-Acetyl-L-lysine	Y	Y
		HMDB00446	*N*-Alpha-acetyllysine	Y	
		HMDB00759	Glycyl-L-leucine	N	
187	424.1537	HMDB01370	Diaminopimelic acid	Y	
		HMDB29054	Threoninyl-Alanine	N	
		HMDB31412	2-Amino-4-[(2-hydroxy-1-oxopropyl)amino]butanoic acid		
		HMDB28697	Alanyl-Threonin	N	
		HMDB39222	*N*-(gamma-Glutamyl)ethanolamine	N	
1297	462.1372	HMDB03747	Resveratrol	Y	
		HMDB34118	(Z)-Resveratrol	N	

1275	490.1318	HMDB41647	2-Dehydro-O-desmethylangolensin	N
		HMDB29519	(2*S*)-Liquiritigenin	Y
		HMDB37316	Isoliquiritigenin	Y
		HMDB29462	(*E*)-2,4,4'-Trihydroxychalcone	N
		HMDB30808	(*S*)-Pinocembrin	Y
		HMDB36457	Emodinanthranol	Y
271	501.1545	HMDB00085	Deoxyguanosine	Y
		HMDB00050	Adenosine	Y
		HMDB14340	Vidarabine	N
		HMDB37842	*N*-(1-Deoxy-1-fructosyl)serine	N
131	509.1704	HMDB05766	Norophthalmic acid	N
		HMDB11738	Gamma-Glutamylglutamine	N
983	524.1296	HMDB37533	Rustoside	N
		HMDB37928	Quercetin 3-[rhamnosyl-(1->2)-alpha-L-arabinopyranoside]	N
		HMDB37408	Neocarlinoside	N
		HMDB38469	Graveobioside A	N
		HMDB39507	Kaempferol 3-xylosylglucoside	N
		HMDB37430	Kaempferol 3-[apiosyl-(1->2)-galactoside]	N
		HMDB30721	Isocarlinoside	N

* Y: identified with authentic standards based on accurate mass and retention time.

N: not matched with authentic standards

Table B.6 Putative metabolites unique to 80% CH_3OH extracts identified by the EML predicted human metabolite library.

	Retention Time (sec)	m/z light	Number of matched entries
1	265	292.0302	1
2	1439	304.0822	18
3	1213	336.0901	52
4	1331	340.1367	8
5	1418	363.2098	1
6	1486	363.2102	1
7	935	390.1008	17
8	1292	428.1527	16
9	1425	432.1641	1
10	1284	442.1325	24
11	263	470.1208	5
12	1476	483.2678	1
13	421	499.1470	3
14	1252	632.1592	10

Appendix C

Table C.1 Putative metabolites of ginseng root were identified based on accurate mass match with metabolites in HMDB or with predicted metabolic products from one metabolic reaction in EML at the website of MyCompoundID. The criteria for accurate mass match is within a mass error of less than 10 ppm.

Retention Time (sec)	m/z light	HMDB	Name	# matched entries in EML
121	414.1219			87
121	381.1229			5
121	576.1748			50
122	408.1701	HMDB04370	*N*-Methyltryptamine	
		HMDB03416	D-Arginine	
		HMDB00517	L-Arginine	
123	410.1020	HMDB00828	Ureidosuccinic acid	
		HMDB38668	L-3-Amino-2-(oxalylamino)propanoic acid	
		HMDB29402	L-2-Amino-3-(oxalylamino)propanoic acid	
157	438.1697	HMDB33891	*N*6-Acetyl-5S-hydroxy-L-lysine	
		HMDB39948	3,4-Dihydroxy-2-hydroxymethyl-1-pyrrolidinepropanamide	
		HMDB29052	Serinyl-Valine	
		HMDB29136	Valyl-Serine	
171	366.1119	HMDB00026	Ureidopropionic acid	
		HMDB00168	L-Asparagine	
		HMDB33780	D-Asparagine	
		HMDB11733	Glycyl-glycine	
		HMDB12265	*N*-Carbamoylsarcosine	
172	414.1218			87
172	408.1703	HMDB04370	*N*-Methyltryptamine	
		HMDB03416	D-Arginine	
		HMDB00517	L-Arginine	
183	424.1174	HMDB11165	L-beta-aspartyl-L-glycine	

		HMDB28837	Glycyl-Aspartate	
		HMDB28753	Aspartyl-Glycine	
186	366.1270			2
200	409.1542	HMDB00904	Citrulline	
		HMDB03148	Argininic acid	
208	380.1275	HMDB00641	L-Glutamine	
		HMDB02031	Ureidoisobutyric acid	
		HMDB03423	D-Glutamine	
		HMDB06899	Alanylglycine	
220	379.1435	HMDB03464	4-Guanidinobutanoic acid	
221	501.1551	HMDB00050	Adenosine	
		HMDB00830	Neuraminic acid	
		HMDB00085	Deoxyguanosine	
222	414.1218			87
226	364.0851			40
231	576.1744			50
234	408.1702	HMDB04370	*N*-Methyltryptamine	
		HMDB03416	D-Arginine	
		HMDB00517	L-Arginine	
236	501.1550	HMDB00050	Adenosine	
		HMDB00830	Neuraminic acid	
		HMDB00085	Deoxyguanosine	
244	300.0689			1
256	426.1223			38
257	366.1119	HMDB00026	Ureidopropionic acid	
		HMDB00168	L-Asparagine	
		HMDB33780	D-Asparagine	
		HMDB11733	Glycyl-glycine	
		HMDB12265	*N*-Carbamoylsarcosine	
258	399.1045	HMDB04089	Formylanthranilic acid	
		HMDB02005	Methionine sulfoxide	

265	353.1165	HMDB61148	Hydroxyethyl glycine	
		HMDB04041	Allothreonine	
		HMDB00719	Homoserine	
		HMDB00167	Threonine	
279	558.1642			31
282	501.1562	HMDB14340	Vidarabine	
		HMDB00050	Adenosine	
		HMDB00085	Deoxyguanosine	
286	339.1010	HMDB03406	D-Serine	
		HMDB00187	Serine	
290	408.1702	HMDB04370	*N*-Methyltryptamine	
		HMDB03416	D-Arginine	
		HMDB00517	L-Arginine	
297	380.1275	HMDB00641	L-Glutamine	
		HMDB02031	Ureidoisobutyric acid	
		HMDB03423	D-Glutamine	
		HMDB06899	Alanylglycine	
302	422.1857	HMDB29416	L-Targinine	
		HMDB00670	Homoarginine	
304	414.1218			87
310	353.1166	HMDB61148	Hydroxyethyl glycine	
		HMDB04041	Allothreonine	
		HMDB00719	Homoserine	
		HMDB00167	Threonine	
315	576.1747			50
316	381.1115	HMDB00148	L-Glutamic acid	
		HMDB02393	*N*-Methyl-D-aspartic acid	
		HMDB02931	*N*-Acetylserine	
		HMDB03011	*O*-Acetylserine	
		HMDB03339	D-Glutamic acid	
		HMDB06556	L-4-Hydroxyglutamate semialdehyde	

317	367.1324	HMDB39499	1,4-Dideoxy-1,4-imino-D-ribitol	
		HMDB31658	L-2-Amino-5-hydroxypentanoic acid	
		HMDB32356	*N*-Lactoyl ethanolamine	
319	362.1170	HMDB29874	Squamolone	
		HMDB31547	L-Cyclo(alanylglycyl)	
		HMDB00079	Dihydrothymine	
323	367.0959	HMDB11753	Iminodiacetic acid	
		HMDB00191	Aspartic acid	
		HMDB06483	D-Aspartic acid	
324	408.1590	HMDB34365	L-Theanine	
		HMDB28854	Glycyl-Valine	
		HMDB29127	Valyl-Glycine	
		HMDB03357	*N*-Acetylornithine	
327	436.2268	HMDB03406	D-Serine	
		HMDB00187	Serine	
328	339.1012			21
330	363.1487			4
332	362.1646			2
334	363.2103			1
334	436.2016	HMDB01539	Asymmetric dimethylarginine	
		HMDB03334	Symmetric dimethylarginine	
335	426.1221			38
341	492.1802	HMDB38676	(2*S*,2'*S*)-Pyrosaccharopine	
		HMDB38614	gamma-L-Glutamyl-L-pipecolic acid	
342	292.1114			1
343	431.1753			3
345	378.1120			15
345	337.1581			5
345	363.1122			6
349	422.1747	HMDB28844	Glycyl-Isoleucine	
		HMDB28907	Isoleucyl-Glycine	

		HMDB28929	Leucyl-Glycine	
		HMDB00759	Glycyl-l-leucine	
		HMDB28700	Alanyl-Valine	
		HMDB00206	*N*-acetyl-lysine	
		HMDB00446	*N*-Alpha-acetyllysine	
		HMDB29120	Valyl-Alanine	
352	381.1116	HMDB00148	L-Glutamic acid	
		HMDB02393	*N*-Methyl-D-aspartic acid	
		HMDB02931	*N*-Acetylserine	
		HMDB03011	*O*-Acetylserine	
		HMDB03339	D-Glutamic acid	
		HMDB06556	L-4-Hydroxyglutamate semialdehyde	
356	507.1548			40
359	353.1165	HMDB61148	Hydroxyethyl glycine	
		HMDB04041	Allothreonine	
		HMDB00719	Homoserine	
		HMDB00167	Threonine	
360	367.1324	HMDB39499	1,4-Dideoxy-1,4-imino-D-ribitol	
		HMDB31658	L-2-Amino-5-hydroxypentanoic acid	
		HMDB32356	*N*-Lactoyl ethanolamine	
360	392.1277			17
361	558.1641			31
362	411.1577			14
363	422.1858	HMDB29416	L-Targinine	
		HMDB00670	Homoarginine	
364	414.1216			87
364	476.1969			3
365	429.1595			8
370	463.2374			2
373	379.1324	HMDB29426	L-cis-5-Hydroxy-2-piperidinecarboxylic acid	
		HMDB59595	(s)-2-amino-6-oxohexanoate	

		HMDB12151	2-Keto-6-aminocaproate	
		HMDB03681	4-Acetamidobutanoic acid	
		HMDB00730	Isobutyrylglycine	
		HMDB00808	*N*-Butyrylglycine	
		HMDB01263	Allysine	
		HMDB12131	(S)-5-Amino-3-oxohexanoate	
373	502.1769	HMDB28865	Hydroxyprolyl-Histidine	
		HMDB12881	*N*-Acetylcarnosine	
		HMDB28886	Histidinyl-Hydroxyproline	
373	641.3623			29
373	588.2111			19
375	415.1438			9
375	295.1111	HMDB00149	Ethanolamine	
377	337.1582			5
378	339.1374			14
378	464.1968			4
379	415.0995			9
384	492.1803	HMDB38676	(2*S*,2'*S*)-Pyrosaccharopine	
		HMDB38614	gamma-L-Glutamyl-L-pipecolic acid	
385	558.1645			31
386	483.1448	HMDB01526	S-Acetyldihydrolipoamide	
388	478.1284	HMDB00767	Pseudouridine	
388	367.0962	HMDB11753	Iminodiacetic acid	
		HMDB00191	Aspartic acid	
		HMDB06483	D-Aspartic acid	
388	337.1219	HMDB00112	Gamma-Aminobutyric acid	
		HMDB00452	L-Alpha-aminobutyric acid	
		HMDB00650	D-Alpha-aminobutyric acid	
		HMDB01906	2-Aminoisobutyric acid	
		HMDB02166	(*S*)-β-aminoisobutyric acid	
		HMDB02299	(*R*)-β-aminoisobutyric acid	

		HMDB03911	3-Aminoisobutanoic acid	
388	318.0796			13
390	472.1900			4
391	414.1219			87
394	408.1592	HMDB34365	L-Theanine	
		HMDB28854	Glycyl-Valine	
		HMDB29127	Valyl-Glycine	
		HMDB03357	*N*-Acetylornithine	
394	463.1648	HMDB28739	Asparaginyl-Proline	
		HMDB29012	Prolyl-Asparagine	
396	353.1166	HMDB61148	Hydroxyethyl glycine	
		HMDB04041	Allothreonine	
		HMDB00719	Homoserine	
		HMDB00167	Threonine	
397	480.1915			4
397	348.1013	HMDB60245	L-3-Cyanoalanine	
		HMDB00076	Dihydrouracil	
397	380.1275	HMDB00641	L-Glutamine	
		HMDB02031	Ureidoisobutyric acid	
		HMDB03423	D-Glutamine	
		HMDB06899	Alanylglycine	
401	423.1699	HMDB00679	Homocitrulline	
401	415.1328	HMDB02184	L-Threo-3-Phenylserine	
		HMDB06050	o-Tyrosine	
		HMDB00158	Tyrosine	
		HMDB59720	Meta-Tyrosine	
		HMDB60280	4,6,7-Trihydroxy-1,2,3,4-tetrahydroisoquinoline	
		HMDB03831	Beta-Tyrosine	
403	337.1581			5
403	391.1438			12
405	343.1112	HMDB01169	4-Aminophenol	

		HMDB35882	2-Acetylpyrrole	
		HMDB40582	1-Methyl-2-pyrrolecarboxaldehyde	
407	309.0903	HMDB14691	Acetohydroxamic Acid	
		HMDB00123	Glycine	
407	428.1377			33
410	363.1488			4
412	393.0577			1
414	431.1274	HMDB38336	2-Hydroxy-3-(3,4-dihydroxyphenyl)propanamide	
		HMDB00181	L-Dopa	
		HMDB00609	DL-Dopa	
		HMDB38750	N-Hydroxy-L-tyrosine	
415	344.1065	HMDB03905	Imidazole-4-acetaldehyde	
		HMDB60741	3-Hydroxy-4-aminopyridine	
417	467.1386			1
417	531.1655	HMDB05862	2-Methylguanosine	
		HMDB06038	3'-O-Methylguanosine	
419	486.1808	HMDB28893	Histidinyl-Proline	
		HMDB29019	Prolyl-Histidine	
423	558.1650			31
424	436.2014	HMDB01539	Asymmetric dimethylarginine	
		HMDB03334	Symmetric dimethylarginine	
424	403.1438			11
425	363.1737			10
425	378.1120			15
425	390.0645			7
428	446.1751	HMDB11180	L-prolyl-l-proline	
432	348.1378	HMDB00323	3-Amino-2-piperidone	
432	396.1114			81
433	393.1598			8
435	565.1862			8
435	403.1330	HMDB01537	6-Hydroxydopamine	

		HMDB04817	5-Hydroxydopamine	
		HMDB37685	xi-Norepinephrine	
		HMDB00216	Norepinephrine	
441	376.1802			2
442	463.1653	HMDB28739	Asparaginyl-Proline	
		HMDB29012	Prolyl-Asparagine	
442	365.0805	HMDB01131	Iminoaspartic acid	
444	348.1013	HMDB60245	L-3-Cyanoalanine	
		HMDB00076	Dihydrouracil	
445	357.1270	HMDB32578	4-Hydroxybenzylamine	
		HMDB29300	*p*-Anisidine	
446	366.1009			69
446	517.1506	HMDB00133	Guanosine	
		HMDB03333	8-hydroxy-deoxyguanosine	
446	377.1644			3
447	362.1170	HMDB29874	Squamolone	
		HMDB31547	L-Cyclo(alanylglycyl)	
		HMDB00079	Dihydrothymine	
447	414.1220			87
450	395.1272	HMDB33747	meso-2,2'-Iminobispropanoic acid	
		HMDB29423	Acetylhomoserine	
		HMDB00510	Aminoadipic acid	
456	378.1121			15
457	381.1117	HMDB00148	L-Glutamic acid	
		HMDB02393	N-Methyl-D-aspartic acid	
		HMDB02931	*N*-Acetylserine	
		HMDB03011	*O*-Acetylserine	
		HMDB03339	D-Glutamic acid	
		HMDB06556	L-4-Hydroxyglutamate semialdehyde	
458	428.1376			33
458	501.1554	HMDB00050	Adenosine	

		HMDB14340	Vidarabine
		HMDB00085	Deoxyguanosine
458	379.1325	HMDB29426	L-cis-5-Hydroxy-2-piperidinecarboxylic acid
		HMDB59595	(*S*)-2-amino-6-oxohexanoate
		HMDB12151	2-Keto-6-aminocaproate
		HMDB03681	4-Acetamidobutanoic acid
		HMDB00730	Isobutyrylglycine
		HMDB00808	*N*-Butyrylglycine
		HMDB01263	Allysine
		HMDB12131	(*S*)-5-Amino-3-oxohexanoate
461	415.1333	HMDB02184	L-Threo-3-Phenylserine
		HMDB06050	*o*-Tyrosine
		HMDB00158	Tyrosine
		HMDB59720	Meta-Tyrosine
		HMDB60280	4,6,7-Trihydroxy-1,2,3,4-tetrahydroisoquinoline
		HMDB03831	Beta-Tyrosine
461	353.1164	HMDB61148	Hydroxyethyl glycine
		HMDB04041	Allothreonine
		HMDB00719	Homoserine
		HMDB00167	Threonine
464	521.1711	HMDB02089	*N*-Ribosylhistidine
465	323.1059	HMDB31219	Ethyl carbamate
		HMDB00056	Beta-alanine
		HMDB01310	Alanine
		HMDB00271	Sarcosine
		HMDB00161	L-Alanine
468	389.0789		[6]
468	540.1539		[14]
468	550.1760		[8]
468	478.1286	HMDB00767	Pseudouridine
469	323.1167		[1]

471	396.1114			81
476	507.2644			1
477	348.1378	HMDB00323	3-Amino-2-piperidone	
486	367.1323	HMDB39499	1,4-Dideoxy-1,4-imino-D-ribitol	
		HMDB31658	L-2-Amino-5-hydroxypentanoic acid	
		HMDB32356	*N*-Lactoyl ethanolamine	
486	405.1170			50
490	458.1863			12
491	349.1217	HMDB03411	D-Proline	
		HMDB00162	L-Proline	
		HMDB12880	Acetamidopropanal	
492	394.1432			35
494	506.1961			8
494	453.1697	HMDB12204	Zeatin	
494	442.1534			26
495	416.1281			20
496	381.1115	HMDB00148	L-Glutamic acid	
		HMDB02393	*N*-Methyl-D-aspartic acid	
		HMDB02931	*N*-Acetylserine	
		HMDB03011	*O*-Acetylserine	
		HMDB03339	D-Glutamic acid	
		HMDB06556	L-4-Hydroxyglutamate semialdehyde	
497	434.1750			11
505	361.1220	HMDB29434	(±)-4-Methylene-2-pyrrolidinecarboxylic acid	
505	409.1432	HMDB38594	Calystegine B2	
		HMDB34252	2-Aminoheptanedioic acid	
		HMDB36383	Calystegine B5	
506	350.1535	HMDB12176	5-Aminopentanamide	
509	389.1282	HMDB00177	L-Histidine	
509	436.1911	HMDB28691	Alanyl-Leucine	
		HMDB28690	Alanyl-Isoleucine	

		HMDB28900	Isoleucyl-Alanine	
		HMDB28922	Leucyl-Alanine	
509	377.1642			3
510	345.1380	HMDB00870	Histamine	
		HMDB60263	Histamium	
510	309.1267	HMDB12136	1-Amino-2-propano	
512	391.0964			19
512	579.2017			4
513	616.2432			6
515	460.1176			19
518	395.1273	HMDB00510	Aminoadipic acid	
		HMDB33747	meso-2,2'-Iminobispropanoic acid	
		HMDB29423	Acetylhomoserine	
519	548.1584	HMDB34060	6-O-Acetylarbutin	
		HMDB40720	2-O-Acetylarbutin	
519	420.1591	HMDB29010	Prolyl-Alanine	
		HMDB28695	Alanyl-Proline	
522	378.0675			2
524	414.1221			87
527	546.1645			41
531	422.1748	HMDB28844	Glycyl-Isoleucine	
		HMDB28907	Isoleucyl-Glycine	
		HMDB28929	Leucyl-Glycine	
		HMDB00759	Glycyl-l-leucine	
		HMDB28700	Alanyl-Valine	
		HMDB00206	*N*-acetyl-lysine	
		HMDB00446	*N*-Alpha-acetyllysine	
		HMDB29120	Valyl-Alanine	
534	309.0906	HMDB14691	Acetohydroxamic Acid	
		HMDB00123	Glycine	
540	417.0396	HMDB31466	Selenomethyl selenocysteine	

		HMDB04119	Selenohomocysteine	
		HMDB04113	Se-Methylselenocysteine	
541	545.2435			1
542	334.1220			3
544	398.1270			74
545	560.1804			38
548	450.2062	HMDB29140	Valyl-Valine	
549	351.1371	HMDB13716	Norvaline	
		HMDB03355	5-aminovalerate	
		HMDB02141	*N*-Methyl-a-aminoisobutyric acid	
		HMDB00883	Valine	
551	443.1637			18
552	454.1435	HMDB00472	5-Hydroxy-L-tryptophan	
552	379.1324	HMDB29426	L-cis-5-Hydroxy-2-piperidinecarboxylic acid	
		HMDB59595	(s)-2-amino-6-oxohexanoate	
		HMDB12151	2-Keto-6-aminocaproate	
		HMDB03681	4-Acetamidobutanoic acid	
		HMDB00730	Isobutyrylglycine	
		HMDB00808	*N*-Butyrylglycine	
		HMDB01263	Allysine	
		HMDB12131	(*S*)-5-Amino-3-oxohexanoate	
554	376.0850			6
554	321.0904	HMDB03609	2-Aminoacrylic acid	
555	331.6031			29
556	437.1378			36
559	351.1011	HMDB06454	L-2-Amino-3-oxobutanoic acid	
		HMDB00532	Acetylglycine	
		HMDB12249	L-Aspartate-semialdehyde	
561	337.1218	HMDB00112	Gamma-Aminobutyric acid	
		HMDB00452	L-Alpha-aminobutyric acid	
		HMDB00650	D-Alpha-aminobutyric acid	

		HMDB01906	2-Aminoisobutyric acid	
		HMDB02166	(*S*)-β-aminoisobutyric acid	
		HMDB02299	(*R*)-β-aminoisobutyric acid	
		HMDB03911	3-Aminoisobutanoic acid	
562	363.1010	HMDB01843	*N*-Acryloylglycine	
562	393.1482	HMDB00339	2-Methylbutyrylglycine	
		HMDB00678	Isovalerylglycine	
		HMDB00927	Valerylglycine	
		HMDB11757	*N*-Acetylvaline	
		HMDB12175	5-Acetamidovalerate	
563	414.1220			87
563	318.0795			13
563	506.1482	HMDB29943	Arbutin	
563	488.0834			2
564	566.1699	HMDB31721	Leonuriside A	
564	536.1593			7
565	307.1112	HMDB01106	3-Aminopropionaldehyde	
		HMDB02134	Aminoacetone	
567	364.1690	HMDB60268	*N*-Acetylputrescinium	
567	350.1171	HMDB39111	L-Acetopine	
569	492.1435			30
570	369.0938	HMDB00742	Homocysteine	
		HMDB02108	Methylcysteine	
571	362.1169	HMDB29874	Squamolone	
		HMDB31547	L-Cyclo(alanylglycyl)	
		HMDB00079	Dihydrothymine	
574	428.1164	HMDB00955	Isoferulic acid	
		HMDB00954	trans-Ferulic acid	
577	436.1903	HMDB28691	Alanyl-Leucine	
		HMDB28690	Alanyl-Isoleucine	
		HMDB28900	Isoleucyl-Alanine	

		HMDB28922	Leucyl-Alanine	
581	478.1534			7
581	450.2062	HMDB29140	Valyl-Valine	
583	309.0904	HMDB14691	Acetohydroxamic Acid	
		HMDB00123	Glycine	
583	310.0745			15
584	500.1965			6
585	405.1166			50
585	409.1434	HMDB38594	Calystegine B2	
		HMDB34252	2-Aminoheptanedioic acid	
		HMDB36383	Calystegine B5	
586	321.1268	HMDB01080	4-Aminobutyraldehyde	
		HMDB31215	*N*-Ethylacetamide	
		HMDB31581	Morpholine	
		HMDB33870	Butyramide	
588	424.0998			2
591	362.1534			8
596	448.1426			6
596	463.1536			16
597	363.6406			2
599	349.1217	HMDB03411	D-Proline	
		HMDB00162	L-Proline	
		HMDB12880	Acetamidopropanal	
599	324.0901			87
601	414.1220			87
602	598.2934			5
603	632.1331			8
607	426.1194			15
607	445.1544	HMDB15078	Zalcitabine	
607	438.1488	HMDB00929	L-Tryptophan	
		HMDB13609	D-Tryptophan	

609	344.1063	HMDB03905	Imidazole-4-acetaldehyde	
		HMDB60741	3-Hydroxy-4-aminopyridine	
610	510.1912	HMDB00279	Saccharopine	
611	558.1643			31
611	454.1430	HMDB00472	5-Hydroxy-L-tryptophan	
611	337.1219	HMDB00112	Gamma-Aminobutyric acid	
		HMDB00452	L-Alpha-aminobutyric acid	
		HMDB00650	D-Alpha-aminobutyric acid	
		HMDB01906	2-Aminoisobutyric acid	
		HMDB02166	(*S*)-β-aminoisobutyric acid	
		HMDB02299	(*R*)-β-aminoisobutyric acid	
		HMDB03911	3-Aminoisobutanoic acid	
611	424.0899			11
612	405.1166			50
613	436.1908	HMDB28691	Alanyl-Leucine	
		HMDB28690	Alanyl-Isoleucine	
		HMDB28900	Isoleucyl-Alanine	
		HMDB28922	Leucyl-Alanine	
614	357.1251			4
614	502.1647	HMDB29051	Serinyl-Tyrosine	
		HMDB29114	Tyrosyl-Serine	
616	647.1741	HMDB32398	Methyl n-formylanthranilate	
618	431.1276	HMDB38336	2-Hydroxy-3-(3,4-dihydroxyphenyl)propanamide	
		HMDB00181	L-Dopa	
		HMDB00609	DL-Dopa	
		HMDB38750	*N*-Hydroxy-L-tyrosine	
621	420.1591	HMDB29010	Prolyl-Alanine	
		HMDB28695	Alanyl-Proline	
621	453.1096	HMDB39387	Methyl 2,6-dihydroxy-4-quinolinecarboxylate	
623	531.1438			25
625	335.1425	HMDB12815	5-Aminopentanal	

629	424.0856			4
629	351.1372	HMDB13716	Norvaline	
		HMDB03355	5-aminovalerate	
		HMDB02141	*N*-Methyl-a-aminoisobutyric acid	
		HMDB00883	Valine	
631	340.0880	HMDB30483	Questiomycin A	
631	321.1142	HMDB04370	*N*-Methyltryptamine	
		HMDB03416	D-Arginine	
		HMDB00517	L-Arginine	
632	383.1074			16
634	444.1592			23
637	447.1334			19
637	430.0957			31
639	555.1917	HMDB30382	2-[4-(3,4-Methylenedioxyphenyl)butyl]-4(1H)-quinolinone	
640	433.0951			7
640	365.1168	HMDB06272	5-Amino-2-oxopentanoic acid	
		HMDB02104	L-Glutamic-gamma-semialdehyde	
		HMDB01149	5-Aminolevulinic acid	
641	315.0848			81
642	414.1219			87
643	406.1208			29
645	395.1274	HMDB00510	Aminoadipic acid	
		HMDB33747	meso-2,2'-Iminobispropanoic acid	
		HMDB29423	Acetylhomoserine	
645	405.0969			7
645	324.0901			87
646	378.0676			2
646	409.1430	HMDB38594	Calystegine B2	
		HMDB34252	2-Aminoheptanedioic acid	
		HMDB36383	Calystegine B5	

649	396.6137			4
649	444.6024			7
651	371.1073	HMDB01123	2-Aminobenzoic acid	
		HMDB01891	m-Aminobenzoic acid	
		HMDB01392	p-Aminobenzoic acid	
		HMDB15687	Salicylamide	
655	474.1686			18
655	423.1591			28
655	438.1484	HMDB00929	L-Tryptophan	
		HMDB13609	D-Tryptophan	
655	417.1122			23
655	590.1329			15
655	400.0853			18
658	367.6378			3
659	450.2061	HMDB29140	Valyl-Valine	
659	448.1904	HMDB03581	Dethiobiotin	
		HMDB29030	Prolyl-Valin	
		HMDB29135	Valyl-Proline	
662	572.1588	HMDB41191	Hydrojuglone glucoside	
		HMDB29681	3-O-*p*-Coumaroylquinic acid	
		HMDB34242	alpha-Hydrojuglone 4-O-b-D-glucoside	
662	602.1691	HMDB39959	3-O-Caffeoyl-1-O-methylquinic acid	
		HMDB39960	3-O-Caffeoyl-4-O-methylquinic acid	
663	620.1431			3
664	351.1372	HMDB13716	Norvaline	
		HMDB03355	5-aminovalerate	
		HMDB02141	*N*-Methyl-a-aminoisobutyric acid	
		HMDB00883	Valine	
665	379.1326	HMDB29426	L-cis-5-Hydroxy-2-piperidinecarboxylic acid	
		HMDB59595	(*S*)-2-amino-6-oxohexanoate	
		HMDB12151	2-Keto-6-aminocaproate	

		HMDB29425	(2*R*,4*S*)-4-(Hydroxymethyl)-2-pyrrolidinecarboxylic acid	
		HMDB03681	4-Acetamidobutanoic acid	
		HMDB00730	Isobutyrylglycine	
		HMDB00808	*N*-Butyrylglycine	
		HMDB01263	Allysine	
		HMDB12131	(*S*)-5-Amino-3-oxohexanoate	
669	429.1757	HMDB60953	2-hydroxymexiletine	
		HMDB60954	*p*-hydroxymexiletine	
672	399.1376	HMDB00159	L-Phenylalanine	
		HMDB04992	Benzocaine	
		HMDB06044	Norsalsolinol	
676	445.1542	HMDB15078	Zalcitabine	
676	364.6249	HMDB28947	Lysyl-Aspartate	
		HMDB02248	Gamma glutamyl ornithine	
		HMDB04985	Aspartyllysine	
		HMDB04987	alpha-Aspartyl-lysine	
		HMDB28758	Aspartyl-Lysine	
676	357.1249			4
676	389.1171			24
677	466.1435			12
679	424.0899			11
681	459.1323			20
681	318.0799			13
683	362.1060			32
683	503.1601			15
685	414.1219			87
686	399.1310	HMDB33538	Fragransol B	
686	405.1167			50
687	417.0396	HMDB31466	Selenomethyl selenocysteine	
		HMDB04119	Selenohomocysteine	

		HMDB04113	Se-Methylselenocysteine	
687	384.0881			1
688	337.1217	HMDB00112	Gamma-Aminobutyric acid	
		HMDB00452	L-Alpha-aminobutyric acid	
		HMDB00650	D-Alpha-aminobutyric acid	
		HMDB01906	2-Aminoisobutyric acid	
		HMDB02166	(*S*)-β-aminoisobutyric acid	
		HMDB02299	(*R*)-β-aminoisobutyric acid	
		HMDB03911	3-Aminoisobutanoic acid	
689	365.1530	HMDB01901	Aminocaproic acid	
		HMDB01645	L-Norleucine	
		HMDB00687	L-Leucine	
		HMDB00557	L-Alloisoleucine	
		HMDB00172	L-Isoleucine	
		HMDB03640	Beta-Leucine	
		HMDB13773	D-Leucine	
690	458.1747			10
690	345.6427			1
691	402.6464	HMDB29099	Tyrosyl-Arginine	
		HMDB28721	Arginyl-Tyrosine	
		Kyotorphin	Arginyl-Tyrosine	
692	433.0955			8
692	378.1848	HMDB02284	*N*-Acetylcadaverine	
		HMDB12135	1-(3-Aminopropyl)-4-aminobutana	
694	546.1643			41
694	324.0902			87
699	321.1270	HMDB01080	4-Aminobutyraldehyde	
		HMDB31215	*N*-Ethylacetamide	
		HMDB31581	Morpholine	
		HMDB33870	Butyramide	
705	331.1115			8

705	494.1754			10
706	376.1692			5
707	395.1220			7
708	443.1266			43
708	399.1375	HMDB00159	L-Phenylalanine	
		HMDB04992	Benzocaine	
		HMDB06044	Norsalsolinol	
708	351.1373	HMDB13716	Norvaline	
		HMDB03355	5-aminovalerate	
		HMDB02141	*N*-Methyl-a-aminoisobutyric acid	
		HMDB00883	Valine	
709	389.1284	HMDB00177	L-Histidine	
709	457.1428	HMDB00866	Acetyltyrosine	
711	462.1584			3
714	346.0747			8
715	381.1116	HMDB00148	L-Glutamic acid	
		HMDB02393	*N*-Methyl-D-aspartic acid	
		HMDB02931	*N*-Acetylserine	
		HMDB03011	*O*-Acetylserine	
		HMDB03339	D-Glutamic acid	
		HMDB06556	L-4-Hydroxyglutamate semialdehyde	
717	321.0905	HMDB03609	2-Aminoacrylic acid	
719	375.6041	HMDB03333	8-Hydroxy-deoxyguanosine	
		HMDB00133	Guanosine	
719	328.1221	HMDB29416	L-Targinine	
		HMDB00670	Homoarginine	
720	682.2913			1
720	409.1428	HMDB38594	Calystegine B2	
		HMDB34252	2-Aminoheptanedioic acid	
		HMDB36383	Calystegine B5	
724	572.1589	HMDB41191	Hydrojuglone glucoside	

		HMDB29681	3-O-p-Coumaroylquinic acid	
		HMDB34242	alpha-Hydrojuglone 4-O-b-D-glucoside	
727	494.2369			1
727	467.1389			1
727	379.1323	HMDB29426	L-cis-5-Hydroxy-2-piperidinecarboxylic acid	
		HMDB59595	(*S*)-2-amino-6-oxohexanoate	
		HMDB12151	2-Keto-6-aminocaproate	
		HMDB03681	4-Acetamidobutanoic acid	
		HMDB00730	Isobutyrylglycine	
		HMDB00808	*N*-Butyrylglycine	
		HMDB01263	Allysine	
		HMDB12131	(*S*)-5-Amino-3-oxohexanoate	
729	602.1694	HMDB39959	3-O-Caffeoyl-1-O-methylquinic acid	
		HMDB39960	3-O-Caffeoyl-4-O-methylquinic acid	
731	365.1529	HMDB01901	Aminocaproic acid	
		HMDB01645	L-Norleucine	
		HMDB00687	L-Leucine	
		HMDB00557	L-Alloisoleucine	
		HMDB00172	L-Isoleucine	
		HMDB03640	Beta-Leucine	
		HMDB13773	D-Leucine	
735	629.1625			1
735	319.1111	HMDB02039	2-Pyrrolidinone	
735	462.2062	HMDB11174	L-isoleucyl-L-proline	
		HMDB11175	L-leucyl-L-proline	
		HMDB28915	Isoleucyl-Proline	
		HMDB28937	Leucyl-Proline	
737	405.1166			50
737	337.1221	HMDB00112	Gamma-Aminobutyric acid	
		HMDB00452	L-Alpha-aminobutyric acid	
		HMDB00650	D-Alpha-aminobutyric acid	

		HMDB01906	2-Aminoisobutyric acid	
		HMDB02166	(S)-β-aminoisobutyric acid	
		HMDB02299	(R)-β-aminoisobutyric acid	
		HMDB03911	3-Aminoisobutanoic acid	
737	446.1269			12
738	396.1113	HMDB29874	Squamolone	
		HMDB31547	L-Cyclo(alanylglycyl)	
		HMDB00079	Dihydrothymine	
739	464.2218	HMDB29131	Valyl-Leucine	
		HMDB28920	Isoleucyl-Valine	
		HMDB29130	Valyl-Isoleucine	
		HMDB28942	Leucyl-Valine	
741	362.1172	HMDB29874	Squamolone	
		HMDB31547	L-Cyclo(alanylglycyl)	
		HMDB00079	Dihydrothymine	
743	410.1268			37
744	423.1948			5
744	373.0854			9
746	373.0855			9
749	335.1424	HMDB12815	5-Aminopentanal	
749	432.1471	HMDB33093	3-(4-Hydroxy-3-methoxyphenyl)-1,2-propanediol	
749	546.1652			41
750	399.1377	HMDB00159	L-Phenylalanine	
		HMDB04992	Benzocaine	
		HMDB06044	Norsalsolinol	
757	416.1641			16
759	481.1431			5
759	558.1648			31
760	319.1111	HMDB02039	2-Pyrrolidinone	
763	416.1163	HMDB00423	3,4-Dihydroxyhydrocinnamic acid	
		HMDB00333	Isohomovanillic acid	

		HMDB00118	Homovanillic acid	
		HMDB00755	Hydroxyphenyllactic acid	
		HMDB02643	3-(3-hydroxyphenyl)-3-hydroxypropanoic acid	
		HMDB04061	3-Methoxy-4-hydroxyphenylglycolaldehyde	
764	423.1587			28
764	360.0902	HMDB13675	1,3,5-trihydroxybenzene	
764	412.0827			1
767	390.1008			17
768	353.1530			10
771	321.1267	HMDB01080	4-Aminobutyraldehyde	
		HMDB31215	*N*-Ethylacetamide	
		HMDB31581	Morpholine	
		HMDB33870	Butyramide	
772	405.1168			50
773	414.1219			87
775	540.1534			14
776	402.1009	HMDB00130	Homogentisic acid	
		HMDB00484	Vanillic acid	
		HMDB00750	3-Hydroxymandelic acid	
		HMDB00822	*p*-Hydroxymandelic acid	
		HMDB01336	3,4-Dihydroxybenzeneacetic acid	
		HMDB01868	5-Methoxysalicylic acid	
		HMDB06242	3,4-Dihydroxymandelaldehyde	
779	338.1421			11
779	528.1537			32
780	457.1429	HMDB00866	Acetyltyrosine	
780	427.1325			24
780	377.1168	HMDB00894	Vinylacetylglycine	
780	414.1004	HMDB03501	3,4-Dihydroxy-trans-cinnamate	
		HMDB01964	Caffeic acid	
780	617.3018			2

782	513.1692			6
783	449.1646			3
784	351.1373	HMDB13716	Norvaline	
		HMDB03355	5-aminovalerate	
		HMDB02141	*N*-Methyl-a-aminoisobutyric acid	
		HMDB00883	Valine	
785	396.1113			31
786	331.1112			8
787	483.1588			6
787	399.6228			8
788	459.1498			4
791	479.2213			16
792	335.1424	HMDB12815	5-Aminopentanal	
792	382.6124	HMDB05862	2-Methylguanosine	
		HMDB06038	3'-O-Methylguanosine	
793	363.1481			4
793	444.1479	HMDB41406	Bancroftinone	
		HMDB33798	3-Methyl-1-(2,4,6-trihydroxyphenyl)-1-butanone	
		HMDB29187	5-(3',5')-Dihydroxyphenyl-gamma-valerolactone	
		HMDB29233	3,4-Dihydroxyphenylvaleric acid	
793	385.1218	HMDB01859	Acetaminophen	
		HMDB02210	2-Phenylglycine	
		HMDB12219	Dopamine quinone	
		HMDB12992	Leukoaminochrome	
794	445.1429	HMDB12992	Leukoaminochrome	
		HMDB11754	Methyldopa	
		HMDB01434	3-Methoxytyrosine	
799	442.1323	HMDB38510	6-Methoxymellein	
801	392.2005			3
802	337.1222	HMDB00112	Gamma-Aminobutyric acid	
		HMDB00452	L-Alpha-aminobutyric acid	

		HMDB00650	D-Alpha-aminobutyric acid	
		HMDB01906	2-Aminoisobutyric acid	
		HMDB02166	(*S*)-β-aminoisobutyric acid	
		HMDB02299	(*R*)-β-aminoisobutyric acid	
		HMDB03911	3-Aminoisobutanoic acid	
803	576.1746			50
806	328.1220	HMDB29416	L-Targinine	
		HMDB00670	Homoarginine	
806	478.2373	HMDB28911	Isoleucyl-Leucine	
		HMDB28910	Isoleucyl-Isoleucine	
		HMDB28932	Leucyl-Isoleucine	
		HMDB28933	Leucyl-Leucine	
806	360.1016	HMDB00262	Thymine	
806	426.1849			8
806	389.1278	HMDB00177	L-Histidine	
807	387.1012	HMDB01972	3-Aminosalicylic acid	
		HMDB14378	Aminosalicylic Acid	
		HMDB01476	3-Hydroxyanthranilic acid	
		HMDB14389	Mesalazine	
810	319.1112	HMDB02039	2-Pyrrolidinone	
810	430.1800			9
810	386.1537			8
810	365.1531	HMDB01901	Aminocaproic acid	
		HMDB01645	L-Norleucine	
		HMDB00687	L-Leucine	
		HMDB00557	L-Alloisoleucine	
		HMDB00172	L-Isoleucine	
		HMDB03640	Beta-Leucine	
		HMDB13773	D-Leucine	
811	334.6141			3
811	472.1905			4

811	373.0855			9
811	617.3018			2
812	439.1438			2
812	321.1269	HMDB01080	4-Aminobutyraldehyde	
		HMDB31215	*N*-Ethylacetamide	
		HMDB31581	Morpholine	
		HMDB33870	Butyramide	
814	342.6298	HMDB60442	beta-Alanyl-L-lysine	
		HMDB28692	Alanyl-Lysine	
		HMDB12230	Gamma-glutamyl-L-putrescine	
		HMDB28944	Lysyl-Alanine	
815	473.1380			17
815	349.1582	HMDB31651	*N*-(3-Methylbutyl)acetamide	
		HMDB34203	*N*-(2-Methylpropyl)acetamide	
816	397.1162			15
816	315.0847			81
816	361.1219	HMDB29434	(±)-4-Methylene-2-pyrrolidinecarboxylic acid	
817	307.1113	HMDB01106	3-Aminopropionaldehyde	
		HMDB02134	Aminoacetone	
822	629.1625			1
822	415.0850			8
826	414.1218			87
829	393.1481	HMDB00339	2-Methylbutyrylglycine	
		HMDB00678	Isovalerylglycine	
		HMDB00927	Valerylglycine	
		HMDB11757	*N*-Acetylvaline	
		HMDB12175	5-Acetamidovalerate	
834	546.1645			41
835	432.1488	HMDB29706	2-Amino-3-methylimidazo[4,5-f]quinoline	
837	343.1220	HMDB28938	Leucyl-Serin	
		HMDB29042	Serinyl-Isoleucine	

		HMDB28916	Isoleucyl-Serine	
		HMDB29137	Valyl-Threonine	
		HMDB29074	Threoninyl-Valine	
		HMDB29043	Serinyl-Leucine	
		HMDB14515	Meprobamate	
839	300.1032	HMDB00214	Ornithine	
		HMDB03374	D-Ornithine	
840	425.6124			15
843	346.0746			8
844	479.1643			5
844	400.1696			5
844	410.1062			8
844	363.1380			30
848	445.1412	HMDB11754	Methyldopa	
		HMDB01434	3-Methoxytyrosine	
		HMDB60747	3-*O*-Methyl-a-methyldopa	
855	376.0854			6
860	386.1061	HMDB00440	3-Hydroxyphenylacetic acid	
		HMDB00020	*p*-Hydroxyphenylacetic acid	
		HMDB00669	Ortho-Hydroxyphenylacetic acid	
		HMDB02390	3-Cresotinic acid	
		HMDB03791	3,4-Dihydroxyphenylacetaldehyde	
		HMDB04815	4-Hydroxy-3-methylbenzoic acid	
860	603.4269			1
860	409.1431	HMDB38594	Calystegine B2	
		HMDB34252	2-Aminoheptanedioic acid	
		HMDB36383	Calystegine B5	
861	459.1701			28
863	656.3622			37
863	346.1111			16

868	450.1485	HMDB35665	L-1,2,3,4-Tetrahydro-beta-carboline-3-carboxylic acid	
869	379.1687			19
870	372.1270	HMDB04284	Tyrosol	
873	494.1748			8
873	365.1531	HMDB01901	Aminocaproic acid	
		HMDB01645	L-Norleucine	
		HMDB00687	L-Leucine	
		HMDB00557	L-Alloisoleucine	
		HMDB00172	L-Isoleucine	
		HMDB03640	Beta-Leucine	
		HMDB13773	D-Leucine	
874	402.1007	HMDB00130	Homogentisic acid	
		HMDB00484	Vanillic acid	
		HMDB00750	3-Hydroxymandelic acid	
		HMDB00822	p-Hydroxymandelic acid	
		HMDB01336	3,4-Dihydroxybenzeneacetic acid	
		HMDB01868	5-Methoxysalicylic acid	
		HMDB06242	3,4-Dihydroxymandelaldehyde	
874	460.1429	HMDB60736	3-(4-Hydroxy-3-methoxyphenyl)-2-methyllactic acid	
		HMDB41679	4-Hydroxy-(3',4'-dihydroxyphenyl)-valeric acid	
874	452.1644	HMDB14673	Mephenytoin	
875	348.6300			3
876	474.1464			28
877	307.1110	HMDB12114	(3*S*)-3,6-Diaminohexanoate	
		HMDB12115	3,5-diaminohexanoic	
		HMDB00182	L-lysine	
		HMDB03405	D-Lysine	
878	423.1011			8
879	362.1063			32

879	337.0676			2
880	576.1746			50
881	392.1166			30
881	365.1319			4
882	459.1591	HMDB39238	Cytokinin B	
883	351.1375	HMDB13716	Norvaline	
		HMDB03355	5-aminovalerate	
		HMDB02141	*N*-Methyl-a-aminoisobutyric acid	
		HMDB00883	Valine	
885	329.0956	HMDB13751	2-hydroxypyridine	
886	430.1323			25
888	432.1497	HMDB29706	2-Amino-3-methylimidazo[4,5-f]quinoline	
888	458.1273	HMDB41691	5-(3',4',5'-Trihydroxyphenyl)-gamma-valerolactone	
889	401.1168	HMDB60374	3-Methoxyanthranilate	
892	414.1219			87
895	356.6454	HMDB29132	Valyl-Lysine	
		HMDB28964	Lysyl-Valine	
895	472.1904			4
895	327.6427	HMDB02189	*N*8-Acetylspermidine	
		HMDB60288	*N*8-Acetylspermidinium	
		HMDB01276	*N*1-Acetylspermidine	
895	335.1425	HMDB12815	5-Aminopentanal	
895	441.1480	HMDB00860	Phenylpropionylglycine	
		HMDB00512	Acetylphenylalanine	
		HMDB60608	*N*-isopropylterephthalamic acid	
		HMDB02042	3-Phenylpropionylglycine	
896	414.1372	HMDB12915	Coniferyl alcohol	
899	409.1219	HMDB00197	Indoleacetic acid	
		HMDB03157	Guanidinosuccinic acid	
		HMDB04073	5-Hydroxyindoleacetaldehyde	

900	420.0927			5
901	320.6349			5
902	546.1645			41
905	407.1641	HMDB00747	Isovalerylalanine	
		HMDB00701	Hexanoylglycine	
		HMDB11756	*N*-Acetylleucine	
905	389.1280	HMDB00177	L-Histidine	
906	446.1638			3
906	343.1114	HMDB01169	4-Aminophenol	
		HMDB35882	2-Acetylpyrrole	
		HMDB40582	1-Methyl-2-pyrrolecarboxaldehyde	
906	375.6047	HMDB03333	8-Hydroxy-deoxyguanosine	
		HMDB00133	Guanosine	
906	366.1009			69
906	311.5931	HMDB00177	L-Histidine	
907	385.1079	HMDB00132	Guanine	
		HMDB00542	8-hydroxyadenine	
		HMDB00403	2-Hydroxyadenine	
910	428.1164	HMDB00955	Isoferulic acid	
		HMDB00954	trans-Ferulic acid	
914	377.1535			19
914	398.1059	HMDB01713	*m*-Coumaric acid	
		HMDB02641	2-Hydroxycinnamic acid	
917	425.1539			10
921	387.1013	HMDB01972	3-Aminosalicylic acid	
		HMDB14378	Aminosalicylic Acid	
		HMDB01476	3-Hydroxyanthranilic acid	
		HMDB14389	Mesalazine	
921	335.1299	HMDB01539	Asymmetric dimethylarginine	
		HMDB03334	Symmetric dimethylarginine	
925	307.1475	HMDB34198	2-Methyl-1-propylamine	

		HMDB32179	sec-Butylamine	
		HMDB31321	1-Butylamine	
		HMDB41878	Diethylamine	
925	363.2100			1
928	493.1432			1
929	449.2108	HMDB13279	*N*-Nonanoylglycine	
932	376.1327			13
934	540.1567			4
934	413.1531			27
935	378.1007			43
936	455.1639			16
936	400.1692			5
936	398.1058	HMDB01713	*m*-Coumaric acid	
		HMDB02641	2-Hydroxycinnamic acid	
937	307.1113	HMDB12114	(3*S*)-3,6-Diaminohexanoate	
		HMDB12115	3,5-diaminohexanoic	
		HMDB00182	L-lysine	
		HMDB03405	D-Lysine	
937	617.3020			2
942	365.1531			3
942	363.1122			6
946	395.1063			14
946	409.1219	HMDB00197	Indoleacetic acid	
		HMDB03157	Guanidinosuccinic acid	
		HMDB04073	5-Hydroxyindoleacetaldehyde	
950	389.1284	HMDB00177	L-Histidine	
951	343.1112	HMDB01169	4-Aminophenol	
		HMDB35882	2-Acetylpyrrole	
		HMDB40582	1-Methyl-2-pyrrolecarboxaldehyde	
953	414.1371	HMDB12915	Coniferyl alcohol	
954	372.1744			3

957	367.1220	HMDB29005	Phenylalanyl-Threonine	
		HMDB29068	Threoninyl-Phenylalanine	
		HMDB33598	*N*5-(4-Methoxybenzyl)glutamine	
961	370.1589			3
962	379.1687			19
963	416.1442			1
963	343.1220	HMDB28938	Leucyl-Serin	
		HMDB29042	Serinyl-Isoleucine	
		HMDB28916	Isoleucyl-Serine	
		HMDB29137	Valyl-Threonine	
		HMDB29074	Threoninyl-Valine	
		HMDB29043	Serinyl-Leucine	
		HMDB14515	Meprobamate	
964	393.1486			42
964	424.2268			1
965	608.2975			3
968	329.0956	HMDB13751	2-hydroxypyridine	
968	442.2048			2
971	348.1198			2
971	428.1164	HMDB00955	Isoferulic acid	
		HMDB00954	trans-Ferulic acid	
972	391.1690	HMDB14447	Tranexamic Acid	
973	421.1794	HMDB13010	*N*-Heptanoylglycine	
		HMDB37790	Polyethylene glycol	
974	401.1169	HMDB60374	3-Methoxyanthranilate	
975	360.1195	HMDB28893	Histidinyl-Proline	
		HMDB29019	Prolyl-Histidine	
976	401.1510	HMDB12162	4-Methoxytyramine	
		HMDB04826	*p*-Synephrine	
		HMDB02182	Phenylephrine	
		HMDB00022	3-Methoxytyramine	

976	429.1483	HMDB14903	Metyrosine	
980	387.1061			14
983	357.0905			9
987	378.0830			1
988	558.1644			31
988	363.2101			1
989	430.1799			9
991	335.1425	HMDB03406	D-Serine	
		HMDB00187	Serine	
993	314.1189	HMDB02038	*N*6-Methyllysine	
		HMDB06009	Isoputreanine	
995	360.1142	HMDB29046	Serinyl-Phenylalanine	
		HMDB29004	Phenylalanyl-Serine	
		HMDB29098	Tyrosyl-Alanine	
		HMDB28699	Alanyl-Tyrosine	
996	374.1114	HMDB28760	Aspartyl-Phenylalanine	
		HMDB00706	L-Aspartyl-L-phenylalanine	
		HMDB28991	Phenylalanyl-Aspartate	
		HMDB11167	L-beta-aspartyl-l-phenylalanine	
997	416.1166	HMDB00423	3,4-Dihydroxyhydrocinnamic acid	
		HMDB00333	Isohomovanillic acid	
		HMDB00118	Homovanillic acid	
		HMDB00755	Hydroxyphenyllactic acid	
		HMDB02643	3-(3-hydroxyphenyl)-3-hydroxypropanoic acid	
		HMDB04061	3-Methoxy-4-hydroxyphenylglycolaldehyde	
1002	315.0848			81
1002	414.1225			87
1002	466.1799	HMDB01389	Melatonin	
1004	401.1281	HMDB29873	(*S*)-Spinacine	
		HMDB60362	2,4-Diamino-6-nitrotoluene	
1005	479.1639			5

1005	353.5981			17
1006	465.1484			3
1006	503.1615			4
1009	336.0900			52
1010	333.1631	HMDB32416	2-Methylpiperidine	
		HMDB31404	Cyclohexylamine	
1011	307.1112	HMDB12114	(3*S*)-3,6-Diaminohexanoate	
		HMDB12115	3,5-diaminohexanoic	
		HMDB00182	L-lysine	
		HMDB03405	D-Lysine	
1011	494.2363			1
1012	321.1632	HMDB31659	3-Methyl-1-butylamine	
		HMDB32392	2-Methylbutylamine	
1013	415.1452			4
1013	637.4068			5
1013	576.1749			50
1016	363.1739			10
1021	357.1269	HMDB32578	4-Hydroxybenzylamine	
		HMDB29300	*p*-Anisidine	
1025	431.1096	HMDB33179	*N*-(4-hydroxyphenyl)ethoxycarbothioamide	
1027	355.1112	HMDB04461	Benzamide	
1028	386.1059	HMDB00440	3-Hydroxyphenylacetic acid	
		HMDB00020	*p*-Hydroxyphenylacetic acid	
		HMDB00669	Ortho-Hydroxyphenylacetic acid	
		HMDB02390	3-Cresotinic acid	
		HMDB03791	3,4-Dihydroxyphenylacetaldehyde	
		HMDB04815	4-Hydroxy-3-methylbenzoic acid	
1028	379.1687			19
1029	285.5900	HMDB00112	Gamma-Aminobutyric acid	
		HMDB00452	L-Alpha-aminobutyric acid	
		HMDB00650	D-Alpha-aminobutyric acid	

		HMDB01906	2-Aminoisobutyric acid	
		HMDB02166	(*S*)-β-aminoisobutyric acid	
		HMDB02299	(*R*)-β-aminoisobutyric acid	
		HMDB03911	3-Aminoisobutanoic acid	
1030	385.0678			1
1034	356.0952	HMDB11718	4-Hydroxybenzaldehyde	
		HMDB34170	2-Hydroxybenzaldehyde	
1036	319.1475	HMDB34301	Piperidine	
1039	412.1216	HMDB39471	3-(4-Hydroxy-2-methoxyphenyl)-2-propenal	
1042	339.6008	HMDB12992	Leukoaminochrome	
		HMDB11754	Methyldopa	
		HMDB01434	3-Methoxytyrosine	
1044	375.0833			1
1044	370.1110	HMDB03767	4-Hydroxyphenylacetaldehyde	
1045	421.2159			9
1048	444.1484			15
1052	384.1743			2
1052	381.1194	HMDB00594	Glutamylphenylalanine	
		HMDB29106	Tyrosyl-Hydroxyproline	
		HMDB01894	Aspartame	
		HMDB28875	Hydroxyprolyl-Tyrosine	
		HMDB38514	L-gamma-Glutamyl-beta-phenyl-beta-L-alanine	
1052	314.1189	HMDB02038	*N*(6)-Methyllysine	
		HMDB06009	Isoputreanine	
1052	349.1581	HMDB31651	*N*-(3-Methylbutyl)acetamide	
		HMDB34203	*N*-(2-Methylpropyl)acetamide	
1054	400.1217	HMDB00375	3-(3-Hydroxyphenyl)propanoic acid	
		HMDB02199	Desaminotyrosine	
		HMDB05175	Homovanillin	
1054	464.1641			3
1063	363.1009			30

1066	379.1689			19
1066	546.1643			41
1069	418.1144	HMDB39959	3-O-Caffeoyl-1-O-methylquinic acid	
1069	382.1113			27
1070	402.1375			20
1071	403.1092	HMDB41191	Hydrojuglone glucoside	
		HMDB34242	alpha-Hydrojuglone 4-O-b-D-glucoside	
1071	324.5953	HMDB02184	L-Threo-3-Phenylserine	
		HMDB06050	*o*-Tyrosine	
		HMDB00158	Tyrosine	
		HMDB59720	Meta-Tyrosine	
		HMDB60280	4,6,7-Trihydroxy-1,2,3,4-tetrahydroisoquinoline	
		HMDB03831	Beta-Tyrosine	
1072	364.0813			10
1074	414.1252			1
1074	474.1588			13
1079	393.1848			13
1079	315.0848			81
1082	407.2004	HMDB00991	alpha-Aminooctanoic acid	
		HMDB14375	Pregabalin	
1085	403.1437			11
1086	414.1372			5
1086	318.6009			11
1093	361.0709			2
1095	444.1479	HMDB41406	Bancroftinone	
		HMDB33798	3-Methyl-1-(2,4,6-trihydroxyphenyl)-1-butanone	
		HMDB29187	5-(3',5')-Dihydroxyphenyl-gamma-valerolactone	
		HMDB29233	3,4-Dihydroxyphenylvaleric acid	
1095	427.1687			13
1099	472.1427			4
1099	372.1266	HMDB04284	Tyrosol	

1100	428.1162	HMDB00955	Isoferulic acid	
		HMDB00954	trans-Ferulic acid	
1103	377.1894			2
1104	379.1113	HMDB29737	1H-Indole-3-carboxaldehyde	
1107	416.1165	HMDB00423	3,4-Dihydroxyhydrocinnamic acid	
		HMDB00333	Isohomovanillic acid	
		HMDB00118	Homovanillic acid	
		HMDB00755	Hydroxyphenyllactic acid	
		HMDB02643	3-(3-hydroxyphenyl)-3-hydroxypropanoic acid	
		HMDB04061	3-Methoxy-4-hydroxyphenylglycolaldehyde	
1109	388.1579			3
1116	314.1189	HMDB02038	N(6)-Methyllysine	
		HMDB06009	Isoputreanine	
1116	410.6378			7
1118	411.0334			1
1119	416.1528			11
1120	384.1265			31
1120	348.4316	HMDB36203	2-Ethoxy-4-(4-methyl-1,3-dioxolan-2-yl)phenol	
1121	458.1636			3
1121	313.0690	HMDB32323	Hexylamine	
1124	421.1792	HMDB13010	*N*-Heptanoylglycine	
		HMDB37790	Polyethylene glycol	
1125	386.1059	HMDB00440	3-Hydroxyphenylacetic acid	
		HMDB00020	*p*-Hydroxyphenylacetic acid	
		HMDB00669	Ortho-Hydroxyphenylacetic acid	
		HMDB02390	3-Cresotinic acid	
		HMDB03791	3,4-Dihydroxyphenylacetaldehyde	
		HMDB04815	4-Hydroxy-3-methylbenzoic acid	
1126	381.1192	HMDB00594	Glutamylphenylalanine	
		HMDB29106	Tyrosyl-Hydroxyproline	
		HMDB01894	Aspartame	

		HMDB28875	Hydroxyprolyl-Tyrosine	
		HMDB38514	L-gamma-Glutamyl-beta-phenyl-beta-L-alanine	
1128	335.1788			1
1128	414.1219			87
1133	442.1321	HMDB38510	6-Methoxymellein	
1135	350.1298	HMDB29064	Threoninyl-Isoleucine	
		HMDB29065	Threoninyl-Leucine	
		HMDB13075	Spermic acid 2	
		HMDB28939	Leucyl-Threonine	
		HMDB28917	Isoleucyl-Threonine	
1138	428.1530	HMDB32590	Zingerone	
		HMDB37271	2,6-Dimethoxy-4-(1-propenyl)phenol	
		HMDB31984	2,5-Dimethoxy-4-(2-propenyl)phenol	
		HMDB40730	Butyl salicylate	
		HMDB32575	Butylparaben	
		HMDB34206	Ethyl 4-methylphenoxyacetate	
1145	324.5954	HMDB02184	L-Threo-3-Phenylserine	
		HMDB06050	o-Tyrosine	
		HMDB00158	Tyrosine	
		HMDB59720	Meta-Tyrosine	
		HMDB60280	4,6,7-Trihydroxy-1,2,3,4-tetrahydroisoquinoline	
		HMDB03831	Beta-Tyrosine	
1146	389.0806			6
1148	608.2973			3
1151	393.1844			13
1155	385.0678			1
1159	338.1191	HMDB60813	Descarbonyl-lacosamid	
1160	363.1738			10
1162	363.2101			1
1163	396.1114			81
1166	331.6032			29

1167	416.1440			1
1174	576.1746			50
1175	321.1268	HMDB13287	Ne,Ne dimethyllysine	
1177	428.1529	HMDB32590	Zingerone	
		HMDB37271	2,6-Dimethoxy-4-(1-propenyl)phenol	
		HMDB31984	2,5-Dimethoxy-4-(2-propenyl)phenol	
		HMDB40730	Butyl salicylate	
		HMDB32575	Butylparaben	
		HMDB34206	Ethyl 4-methylphenoxyacetate	
1181	384.1265			31
1185	338.5929			43
1185	442.1320	HMDB38510	6-Methoxymellein	
		HMDB36328	Nordihydrocapsaicin	
1187	324.6011			9
1188	527.2579	HMDB29846	Nonivamide	
1192	345.6006	HMDB00866	Acetyltyrosine	
1195	372.1265	HMDB04284	Tyrosol	
1197	325.6087			8
1199	391.2052			1
1200	435.1951	HMDB41540	*N*-(5-Methyl-3-oxohexyl)alanine	
1203	345.6427			1
1208	529.3737			3
1214	400.1215	HMDB00375	3-(3-Hydroxyphenyl)propanoic acid	
		HMDB02199	Desaminotyrosine	
		HMDB05175	Homovanillin	
1214	337.1217	HMDB00112	Gamma-Aminobutyric acid	
		HMDB00452	L-Alpha-aminobutyric acid	
		HMDB00650	D-Alpha-aminobutyric acid	
		HMDB01906	2-Aminoisobutyric acid	
		HMDB02166	(*S*)-β-aminoisobutyric acid	
		HMDB02299	(*R*)-β-aminoisobutyric acid	

		HMDB03911	3-Aminoisobutanoic acid	
1217	555.2889	HMDB36330	Homodihydrocapsaicin	
1218	541.2738	HMDB38457	Dihydrocapsaicin	
1218	409.1218	HMDB00197	Indoleacetic acid	
		HMDB03157	Guanidinosuccinic acid	
		HMDB04073	5-Hydroxyindoleacetaldehyde	
1225	352.6085			9
1232	331.6031			29
1232	513.2785			6
1242	412.1578			16
1254	398.1423			12
1262	305.0716			6
1264	336.5772	HMDB00881	Xanthurenic acid	
		HMDB32963	Zeanic acid	
1271	510.2315			2
1280	388.0771			1
1281	344.1070	HMDB03905	Imidazole-4-acetaldehyde	
		HMDB60741	3-Hydroxy-4-aminopyridine	
1282	289.5743			8
1282	405.2207	HMDB36195	N,2,3-Trimethyl-2-(1-methylethyl)butanamide	
1287	440.2260			1
1293	496.2521			4
1302	414.1217			87
1306	454.2412			3
1335	363.2095			1
1342	421.2625			3
1349	432.1633			1
1354	468.2572	HMDB40179	2,6-Di-tert-butyl-4-ethylphenol	
1360	331.0873	HMDB00955	Isoferulic acid	
		HMDB00954	trans-Ferulic acid	
1365	496.2521			4

1387	351.1373	HMDB13716	Norvaline	
		HMDB03355	5-aminovalerate	
		HMDB02141	*N*-Methyl-a-aminoisobutyric acid	
		HMDB00883	Valine	
1387	569.3045			29
1397	456.2206			2
1397	414.1217			87
1400	483.2681			1
1405	302.0846	HMDB03767	4-Hydroxyphenylacetaldehyde	
1418	549.3364			8
1420	458.2002			3
1448	454.2412			3
1462	319.0652			2

Table C.2 Summary of the metabolites with similar changing trends in Ginseng 1. Putative metabolites were identified based on accurate mass match with with metabolites in HMDB or with predicted metabolic products from one metabolic reaction in EML at the website of MyCompoundID. The criteria for accurate mass match is within a mass error of less than 10 ppm.

Group no.	Retention Time (sec)	m/z light	HMDB	Name	# matched entries in EML
1	166	576.1748			50
	173	596.0667			1
	183	575.1660			4
	185	614.1326			7
	290	576.1744			50
	322	576.1980			15
	328	456.1330			53
	360	576.1741			48
	438	558.1637			31
	502	363.1701			0
	663	414.1215			87
	721	405.1053			3
	731	323.6035			1
	837	405.1164			50
2	167	366.1008			69
	599	460.1175	HMDB01904	3-Nitrotyrosine	
3	212	576.1742			50
	290	414.1216			87
	666	406.1212			8
	757	809.1552			1
	991	349.1580			21
4	219	462.1807			2
	794	445.1538	HMDB15078	Zalcitabine	
5	332	501.1551	HMDB00830	Neuraminic acid	
			HMDB00050	Adenosine	

			HMDB00085	Deoxyguanosine	
	1088	368.6114			6
6	364	505.2224	HMDB29011	Prolyl-Arginine	
			HMDB28717	Arginyl-Proline	
	535	348.1381	HMDB00323	3-Amino-2-piperidone	
			HMDB61162	*N*-Mononitrosopiperazine	
7	395	558.1637			31
	545	540.1537			14
	690	406.1108			4
	821	414.1214			87
	899	629.1619			3
	1042	396.1109			31
	1048	397.1157	HMDB40930	1,5-bis(4-hydroxy-3-methoxyphenyl)-1,4-pentadien-3-one	
			HMDB40315	Moracin P	
			HMDB41299	omega-Hydroxymoracin N	
			HMDB30606	1,7-Dihydroxy-3-methoxy-2-prenylxanthone	
			HMDB34015	1,6-Dihydroxy-3-methoxy-2-prenylxanthone	
			HMDB30605	1,5-Dihydroxy-3-methoxy-2-prenylxanthone	
			HMDB40313	Moracin O	
8	419	408.1589	HMDB03357	*N*-Acetylornithine	
			HMDB29127	Valyl-Glycine	
			HMDB28854	Glycyl-Valine	
			HMDB34365	L-Theanine	
	959	319.1098			0
9	511	517.1503	HMDB00133	Guanosine	
			HMDB03333	8-Hydroxy-deoxyguanosine	

	545	478.1281	HMDB00767	Pseudouridine	
10	624	546.1640			41
	1033	540.1536			14
11	694	349.1207	HMDB00162	L-Proline	
			HMDB03411	D-Proline	
			HMDB30409	4-Amino-2-methylenebutanoic acid	
			HMDB12880	Acetamidopropanal	
	1115	305.1318	HMDB31641	Pyrrolidine	
12	736	576.1748			50
	873	558.1637			31
13	781	809.3005			2
	928	396.1108			0
	1120	396.6125			2
14	805	602.1688	HMDB39959	3-O-Caffeoyl-1-O-methylquinic acid	
			HMDB39960	3-O-Caffeoyl-4-O-methylquinic acid	
	897	602.1687	HMDB39959	3-O-Caffeoyl-1-O-methylquinic acid	
			HMDB39960	3-O-Caffeoyl-4-O-methylquinic acid	
15	902	415.0849			8
	947	338.1419			11

Table C.3 Summary of the metabolites with similar changing trends in Ginseng 2. Putative metabolites were identified based on accurate mass match with with metabolites in HMDB or with predicted metabolic products from one metabolic reaction in EML at the website of MyCompoundID. The criteria for accurate mass match is within a mass error of less than 10 ppm.

Group no.	Retentio Time (sec)	m/z light	HMDB	Name	# matched entries in EML
1	128	449.1085			0
	818	590.1333	HMDB03841	5-Amino-6-(5'-phosphoribitylamino)uracil	
			HMDB41709	Caffeoyl C1-glucuronide	
			HMDB41705	Caffeic acid 3-O-glucuronide	
			HMDB41707	Caffeic acid 4-O-glucuronide	
2	149	496.1859			1
	1149	348.0899			28
	1268	348.1194			2
3	153	410.0969			0
	821	620.1437			2
4	183	575.1660			4
	721	324.0875			7
5	184	598.1565	HMDB33304	Gerberinol	
	290	414.1216			87
	744	406.1188			25
	458	348.1009	HMDB60245	L-3-Cyanoalanine	
			HMDB00076	Dihydrouracil	
	1231	298.5849			6
6	247	424.1533	HMDB01370	Diaminopimelic acid	
			HMDB29054	Threoninyl-Alanine	
			HMDB31412	2-Amino-4-[(2-hydroxy-1-oxopropyl)amino]butanoic acid	
			HMDB28697	Alanyl-Threonine	
			HMDB39222	*N*-(gamma-Glutamyl)ethanolamine	

	831	576.1743			50
7	264	560.1486			2
	875	647.1728			1
8	264	558.1638			31
	549	562.1351			19
	740	424.0903			11
9	267	380.1274	HMDB02031	Beta-ureidoisobutyrate	
			HMDB00641	L-Glutamine	
			HMDB06899	Alanylglycine	
			HMDB28687	Alanyl-Glycine	
			HMDB03423	D-Glutamine	
	379	362.1142			0
10	364	505.2224	HMDB29011	Prolyl-Arginine	
			HMDB28717	Arginyl-Proline	
	694	349.1207	HMDB03411	D-proline	
			HMDB30409	4-Amino-2-methylenebutanoic acid	
			HMDB34208	Pterolactam	
			HMDB12880	Acetamidopropanal	
			HMDB00162	L-Proline	
11	391	381.1116	HMDB06556	L-4-Hydroxyglutamate semialdehyde	
			HMDB03339	D-glutamic acid	
			HMDB02931	*N*-Acetylserine	
			HMDB00148	L-Glutamic acid	
			HMDB03011	*O*-Acetylserine	
			HMDB33550	3-(Carboxymethylamino)propanoic acid	
	1212	420.0928			5
12	480	350.1531	HMDB12176	5-Aminopentanamide	
	935	416.1639			16
13	545	540.1537			14
	781	809.3005			2

	803	629.1644			14
	814	810.2320			48
	899	629.1619			3
	902	415.0849			8
	1315	414.1219			87
14	624	546.1640			41
	867	462.1582			3
15	673	536.1585			7
	726	692.2379	HMDB37108	3-Hydroxychavicol 1-[rhamnosyl-(1->6)-glucoside]	
16	805	602.1688	HMDB39959	3-O-Caffeoyl-1-O-methylquinic acid	
			HMDB39960	3-O-Caffeoyl-4-O-methylquinic acid	
	1328	359.1059			16
17	806	572.1583	HMDB41191	Hydrojuglone glucoside	
			HMDB29681	3-O-p-Coumaroylquinic acid	
			HMDB34242	alpha-Hydrojuglone 4-O-b-D-glucoside	
	873	558.1637			31
	911	572.1583	HMDB41191	Hydrojuglone glucoside	
			HMDB29681	3-O-p-Coumaroylquinic acid	
			HMDB34242	alpha-Hydrojuglone 4-O-b-D-glucoside	
18	808	337.1216	HMDB00112	Gamma-Aminobutyric acid	
			HMDB00452	L-Alpha-aminobutyric acid	
			HMDB00650	D-Alpha-aminobutyric acid	
			HMDB01906	2-Aminoisobutyric acid	
			HMDB02166	(*S*)-β-aminoisobutyric acid	
			HMDB02299	(*R*)-β-aminoisobutyric acid	
			HMDB03911	3-Aminoisobutanoic acid	
	1095	365.1527	HMDB00172	L-Isoleucine	
			HMDB00557	L-Alloisoleucine	

			HMDB00687	L-Leucine	
			HMDB01645	L-Norleucine	
			HMDB01901	Aminocaproic acid	
			HMDB03640	Beta-Leucine	
19	857	351.1342			0
	1243	379.1684			19
20	1122	307.1109	HMDB12114	(3*S*)-3,6-Diaminohexanoate	
			HMDB12115	3,5-diaminohexanoic	
			HMDB00182	L-lysine	
			HMDB03405	D-Lysine	
	1154	311.5929	HMDB00177	Histidine	

Table C.4 Summary of the metabolites with similar changing trends in Ginseng 3. Putative metabolites were identified based on accurate mass match with with metabolites in HMDB or with predicted metabolic products from one metabolic reaction in EML at the website of MyCompoundID. The criteria for accurate mass match is within a mass error of less than 10 ppm.

Group no.	Retention Time (sec)	m/z light	HMDB	Name	# matched entries in EML
1	135	389.1237			0
	205	381.1189			0
	274	380.1236			0
	1162	307.1084			0
2	144	576.1640			22
	153	410.0969			0
	848	620.1310			6
3	166	576.1664	HMDB31992	Dulxanthone A	
			HMDB34182	1,5,8-Trihydroxy-3-methyl-2-prenylxanthone	
			HMDB32736	Dulxanthone D	
			HMDB31289	1,4,6-Trihydroxy-5-methoxy-7-prenylxanthone	
			HMDB30575	4',5,6,7-Tetramethoxyflavone	
	754	424.0858			0
	866	405.1119	HMDB31992	Dulxanthone A	
			HMDB34182	1,5,8-Trihydroxy-3-methyl-2-prenylxanthone	
			HMDB32736	Dulxanthone D	
			HMDB31289	1,4,6-Trihydroxy-5-methoxy-7-prenylxanthone	
			HMDB30575	4',5,6,7-Tetramethoxyflavone	
4	178	576.1888	HMDB41273	Myzodendrone	
			HMDB32767	Sphalleroside A	
			HMDB39334	Citrusin D	

	269	414.1173			0
5	221	509.1627			3
	379	362.1142			0
6	263	383.1231			0
	428	492.1731			1
7	301	414.1172			0
	330	394.1391			0
	871	399.1330			0
8	732	369.1092			0
	918	335.1394			0
	1391	349.1555			0
9	559	540.1448			17
	1085	397.1115			3
10	595	499.1938			2
	754	647.1599	HMDB30408 HMDB00704	2-Amino-3,4-dihydroxypentanedioic acid Isoxanthopterin	
11	606	420.1541			0
	821	421.1382			0
12	633	457.0592			0
	824	396.1071			0
13	832	602.1577			3
	933	602.1576			3
14	669	363.0973			0
	1358	336.0870			0
15	695	405.1122	HMDB31992 HMDB34182 HMDB32736 HMDB31289	Dulxanthone A 1,5,8-Trihydroxy-3-methyl-2-prenylxanthone Dulxanthone D 1,4,6-Trihydroxy-5-methoxy-7-prenylxanthone	

			HMDB30575	4',5,6,7-Tetramethoxyflavone	
	850	414.1170			0
16	696	406.1168			1
	799	809.1329			2
17	720	349.1162			0
	742	349.1326			3
	1337	326.6320			0
18	748	405.1116	HMDB31992	Dulxanthone A	
			HMDB34182	1,5,8-Trihydroxy-3-methyl-2-prenylxanthone	
			HMDB32736	Dulxanthone D	
			HMDB31289	1,4,6-Trihydroxy-5-methoxy-7-prenylxanthone	
			HMDB30575	4',5,6,7-Tetramethoxyflavone	
	752	810.2091			20
	852	590.1230			14

Appendix D

Table D.1 Summary of the discriminant metabolites determined by VIP score and Volcano plot for difference *TIFY* gene expressions. The discriminant metabolites were identified based on accurate mass match with metabolites in HMDB or with predicted metabolic products from one metabolic reaction in EML at the website of MyCompoundID. The criteria for accurate mass match is within a mass error of less than 10 ppm. (* RT means retention time. ** The number in parenthesis refers to %RSD)

	RT* (sec)	m/z light	$^{12}C/^{13}C$ OXT**	$^{12}C/^{13}C$ KOT**	$^{12}C/^{13}C$ WT**	Name	HMDB	# matched entries in EML
1	133	414.1222	1.11(10)	0.59(16)	1.26(24)	Paraxanthine	HMDB01860	
2	137	598.1768	1.00(13)	0.54(16)	0.94(26)	Hydroxytorsemide	HMDB60982	
3	148	414.1243	1.02(14)	0.60(13)	0.99(21)	Adrenochrome o-semiquinone	HMDB12883	
4	149	575.1791	1.07(22)	0.47(5)	0.97(30)			20
5	150	598.1691	1.06(12)	0.63(7)	1.02(19)	Bumetanide	HMDB15024	
6	155	598.1576	1.06(4)	0.54(30)	1.11(15)			6
7	167	509.1705	1.52(28)	0.46(25)	1.46(37)	Norophthalmic acid Gamma-Glutamyl Glutamine	HMDB05766 HMDB11738	
8	202	558.1659	1.07(11)	0.57(8)	0.99(17)			33
9	204	414.1225	1.16(12)	0.62(8)	1.04(20)	Paraxanthine	HMDB01860	
10	216	576.1812	0.67(27)	0.51(28)	1.49(27)	Clozapine N-oxide	HMDB60900	
11	235	414.1253	1.10(14)	0.62(10)	1.05(21)	Adrenochrome o-semiquinone	HMDB12883	
12	237	576.2004	1.02(12)	0.56(14)	0.94(21)			20

13	245	576.1776	1.03(14)	0.57(14)	0.98(18)			51
14	253	456.1340	1.17(20)	0.56(15)	1.15(14)			53
15	254	464.1973	0.48(24)	0.86(16)	0.88(25)			4
16	267	468.1442	0.71(22)	0.87(17)	1.78(12)	L-beta-aspartyl-L-threonine Threoninyl-Aspartate Aspartyl-Threonine	HMDB11169 HMDB29057 HMDB28763	
17	293	370.0953	0.57(15)	0.73(13)	1.28(16)			14
18	299	364.1809	0.45(12)	0.71(6)	0.69(10)	Agmatine Agmatinium	HMDB01432 HMDB60259	
19	304	452.1483	0.32 (8)	0.69(17)	0.73(11)	5-L-Glutamyl-L-alanine Serinyl-Hydroxyproline Hydroxyprolyl-Serine Glutamylalanine *N*-gamma-L-Glutamyl-D-alanine	HMDB06248 HMDB29040 HMDB28872 HMDB03764 HMDB36301	
20	306	305.1321	0.49(21)	0.75(9)	0.66(13)	Pyrrolidine	HMDB31641	
21	324	558.1765	1.15(12)	0.59(7)	1.02(19)			12
22	331	558.1671	1.04(11)	0.62(13)	1.01(16)			34
23	332	378.1975	0.66(8)	0.89(10)	1.19(13)			3
24	340	364.1808	0.46(9)	0.75(7)	0.71(7)	Agmatine Agmatinium	HMDB01432 HMDB60259	
25	367	624.1938	1.25(10)	0.60(15)	0.74(18)	(*Z*)-Resveratrol 4'-glucoside	HMDB30565	

						(*E*)-2-Glucosyl-3,4',5-trihydroxystilbene	HMDB36294	
						trans-Piceid	HMDB30564	
						cis-Piceid	HMDB31422	
26	371	365.1649	0.69(6)	0.73(13)	1.24(10)	*N*-Carbamoylputrescine	HMDB33458	
27	378	455.0949	0.52(8)	0.67(12)	1.03(12)	*S*-(3-oxo-3-carboxy-n-propyl)cysteine	HMDB02135	
28	378	602.1639	0.51(13)	0.64(11)	1.01(11)			7
29	383	446.0943	0.45(25)	0.57(17)	0.96(9)			9
30	387	558.1671	1.10(15)	0.58(11)	0.98(15)			34
31	394	364.1798	0.51(17)	0.82(7)	0.76(9)	Agmatine	HMDB01432	
						Agmatinium	HMDB60259	
32	396	396.1130	0.50(22)	0.52(24)	1.42(20)			0
33	399	432.1110	1.09(13)	0.59(13)	1.01(19)			5
34	416	365.1649	0.68(11)	0.52(18)	1.48(8)	3,4-O-Dimethylgallic acid	HMDB41662	
						3-(3,4-Dihydroxyphenyl)lactic acid	HMDB03503	
						Ethyl gallate	HMDB33836	
						Syringic acid	HMDB02085	
						Vanillylmandelic acid	HMDB00291	
						3-Hydroxy-4-methoxymandelate	HMDB29170	

35	417	594.1653	0.73(12)	0.56(14)	1.43(4)	6'-Methoxypolygoacetophenoside	HMDB35476	
36	429	367.0961	0.73(10)	0.68(13)	1.37(13)	Iminodiacetic acid L-Aspartic acid D-Aspartic acid	HMDB11753 HMDB00191 HMDB06483	
37	432	455.1677	0.59(15)	0.90(5)	0.86(17)			4
38	442	474.0694	0.31(21)	0.44(16)	0.88(15)			0
39	457	431.1402	0.62(10)	0.83(8)	0.93(9)	*N*-Acetylhistidine	HMDB32055	
40	473	337.1242	1.22(12)	0.81(12)	1.11(22)			0
41	480	540.1550	0.99(14)	0.55(8)	0.91(17)			14
42	493	423.1580	0.73(6)	0.92(5)	1.41(16)			28
43	496	364.1800	0.64(11)	1.13(13)	0.84(11)	Agmatine Agmatinium	HMDB01432 HMDB60259	
44	515	390.1593	0.56(10)	0.93(6)	0.85(13)			4
45	522	402.0865	0.41(6)	0.46(6)	0.81(16)	Uric acid	HMDB00289	
46	535	576.1896	1.54(13)	0.86(22)	0.99(16)	Myzodendrone Citrusin D	HMDB41273 HMDB39334	
47	545	378.0663	0.97(13)	0.57(21)	0.99(30)			2
48	546	522.1429	0.76(20)	0.74(19)	1.52(12)	Phlorin	HMDB35589	
49	548	574.1379	1.51(11)	1.53(12)	0.86(16)	Cichoriin Aesculin	HMDB30821 HMDB30820	

50	554	431.1159	1.16(11)	0.65(23)	1.41(19)			13
51	562	604.1651	0.66(9)	0.83(10)	1.20(10)	Kaempferol 3-(2''-rhamnosylrutinoside) Chrysophanol 1-triglucoside Kaempferol 3-(2''-rhamnosylgalactoside) 7-rhamnoside Mauritianin	HMDB37432 HMDB34571 HMDB40969 HMDB40562	
52	569	450.1149	0.76(10)	0.89(12)	1.37(11)			4
53	569	531.1411	0.68(13)	0.84(10)	1.22(9)			1
54	573	376.0840	0.74(9)	0.79(9)	1.15(11)			6
55	590	405.1166	1.06(15)	0.59(6)	1.02(20)			50
56	591	414.1219	1.07(12)	0.59(9)	0.99(17)	Paraxanthine	HMDB01860	
57	605	506.1485	0.70(11)	0.82(13)	1.26(5)	Arbutin	HMDB29943	
58	619	320.1428	0.63(16)	0.56(15)	1.56(17)	Piperazine	HMDB14730	
59	640	473.1144	1.23(14)	0.61(30)	1.37(43)	Pollenin B Alliumoside A Estragonoside Nepitrin Hesperetin 3'-O-glucuronide	HMDB39338 HMDB36120 HMDB33650 HMDB30548 HMDB41743	
60	647	369.1125	0.69(7)	0.78(16)	1.16(8)	Adenin	HMDB00034	

						(+)-threo-2-Amino-3,4-dihydroxybutanoic acid	HMDB29389	
61	654	405.1161	1.02(10)	0.54(8)	0.89(22)			48
62	655	372.1276	0.83(8)	1.13(12)	1.28(11)			19
63	656	324.0907	1.11(12)	0.56(5)	1.05(26)			87
64	658	318.1276	0.49(9)	0.43(23)	1.49(22)			2
65	659	647.1790	1.28(8)	0.66(9)	1.10(15)	(*R*)C(*R*)S-S-Propylcysteine sulfoxide	HMDB29442	
66	659	405.1166	1.04(11)	0.59(15)	0.97(17)			50
67	660	647.1826	1.24(15)	0.65(9)	1.11(13)			0
68	662	494.1520	0.57(8)	0.53(14)	1.43(13)			2
69	676	362.1531	0.66(30)	1.68(15)	0.57(19)			8
70	677	647.1765	1.17(15)	0.65(18)	1.06(17)			54
71	686	405.1180	1.13(11)	0.59(9)	1.02(21)			51
72	687	414.1222	1.13(13)	0.66(15)	1.08(16)	Paraxanthine	HMDB01860	
73	696	298.5343	0.14(29)	0.2(27)	1.03(24)			0
74	705	511.1949	0.14(19)	0.42(11)	0.74(30)			6
75	707	325.0882	1.03(14)	0.55(10)	0.90(17)	2',6'-Dihydroxy-4'-methoxyacetophenone	HMDB29646	
76	707	346.0863	0.85(8)	0.75(11)	1.52(14)	4-Carboxypyrazole Uracil	HMDB60760 HMDB00300	
77	711	315.0850	1.14(9)	0.57(15)	1.07(21)			81

78	716	324.0905	1.00(14)	0.52(10)	0.90(21)			87
79	718	546.1813	0.78(15)	1.29(14)	0.60(43)	3-Hydroxychavicol 1-glucoside	HMDB34861	
80	718	458.0911	0.47(10)	0.78(7)	0.61(23)	Dehydrochorismic acid	HMDB36314	
81	718	414.1184	1.04(12)	0.51(11)	0.89(15)			0
82	727	334.1582	0.71(14)	0.71(15)	1.33(14)			2
83	729	595.1638	0.76(10)	0.88(8)	1.18(9)			0
84	755	334.1587	0.60(14)	0.59(17)	1.42(17)			2
85	758	306.0803	1.07(15)	0.56(9)	1.03(19)			0
86	761	414.1229	1.05(14)	0.55(7)	0.98(20)	Paraxanthine	HMDB01860	
87	761	405.1179	1.11(21)	0.59(8)	1.04(21)			51
88	770	525.2090	0.16(27)	0.66(19)	0.63(28)	Desethylchloroquine	HMDB41870	
89	783	396.6131	1.20(16)	0.67(13)	1.06(17)			0
90	786	557.1596	1.14(15)	0.60(14)	1.72(43)			11
91	786	405.1171	1.19(16)	0.63(10)	1.07(21)			50
92	787	414.1216	1.10(20)	0.59(11)	1.03(20)	Paraxanthine	HMDB01860	
93	801	620.1821	0.75(8)	0.77(9)	1.27(10)			10
94	820	396.1090	1.03(15)	0.58(10)	0.99(26)	Moracin K Moracin L	HMDB41296 HMDB41297	
95	846	523.1415	0.88(10)	0.89(15)	1.54(14)			3
96	846	450.1133	0.87(5)	0.84(9)	1.55(15)	Kaempferol 3-O-alpha-L-rhamnofuranoside	HMDB40542	

						Genistein 5-glucoside	HMDB38740	
						Glucoemodin	HMDB38887	
						Apigenin 4'-O-glucoside	HMDB41591	
						Cosmosiin	HMDB37340	
						w-O-beta-D-Glucopyranosylaloeemodin	HMDB34405	
						Isogenistein 7-glucoside	HMDB33982	
						alpha-Rhamnorobin	HMDB29501	
97	858	373.0852	0.65(14)	0.62(15)	1.27(14)			9
98	867	561.1361	1.10(10)	0.65(19)	1.51(30)			8
99	867	350.0687	0.64(3)	0.67(19)	1.17(15)			17
100	883	293.1314	0.62(9)	0.66(16)	1.46(16)	1-Propylamine	HMDB34006	
101	895	404.0712	0.98(14)	1.60(10)	0.86(27)			2
102	914	396.1110	1.09(20)	0.50(8)	0.88(11)			31
103	926	364.1808	0.51(10)	0.80(9)	0.73(8)	Agmatine	HMDB01432	
						Agmatinium	HMDB60259	
104	968	514.1160	0.58(19)	0.96(10)	1.30(17)			5
105	980	396.1118	1.11(19)	0.55(11)	0.95(18)			0
106	1000	480.1589	0.71(18)	1.10(11)	0.94(19)	*N*-acetyltryptophan	HMDB13713	
						Methylphenobarbital	HMDB14987	
						cyclic 6-Hydroxymelatonin	HMDB60810	
107	1052	423.1011	0.59(20)	1.09(11)	0.88(9)	Kynurenic acid	HMDB00715	

108	1072	396.1108	1.11(16)	0.64(5)	0.97(16)			77
109	1088	441.1481	0.91(7)	1.40(9)	1.27(14)	Phenylpropionylglycine 3-Phenylpropionylglycine	HMDB00860 HMDB02042	
110	1135	285.0974	0.42(16)	0.37(23)	1.04(17)			0
111	1157	516.1317	0.58(18)	0.59(10)	1.27(18)			11
112	1169	455.1633	0.96(7)	1.65(11)	1.13(10)	Carbofuran	HMDB31770	
113	1171	378.0995	0.16(28)	0.10(20)	0.93(26)			43
114	1175	410.1531	0.57(12)	1.26(11)	0.73(2)	Serotonin	HMDB00259	
115	1208	302.0780	0.49(28)	0.45(24)	1.20(16)	Hypoxanthine Allopurinol	HMDB00157 HMDB14581	
116	1236	321.0660	0.66(17)	0.81(16)	1.27(13)			16
117	1321	335.1235	1.11(13)	1.75(15)	1.30(29)	Isoleucyl-Alanine Leucyl-Alanine Alanyl-Leucine Alanyl-Isoleucine	HMDB28900 HMDB28922 HMDB28691 HMDB28690	
118	1403	551.1773	0.59(26)	0.58(25)	1.60(17)			0
119	1404	276.0924	0.50(21)	0.55(20)	1.32(15)			2
120	1451	354.1156	0.83(14)	1.32(15)	1.00(23)			13
121	1451	530.6502	0.51(13)	1.42(25)	1.28(19)	Pelargonidin 3-sophoroside	HMDB33679	

Table D.2 Summary of putative metabolites with up-regulation or down-regulation under 24 hour MeJA treatment. The putative metabolites were identified based on accurate mass match with metabolites in HMDB or with predicted metabolic products from one metabolic reaction in EML at the website of MyCompoundID. The criteria for accurate mass match is within a mass error of less than 10 ppm. (* RT means retention time. ** The number in parenthesis refers to %RSD)

	RT* (sec)	m/z light	$^{12}C/^{13}C$ 0 HR**	$^{12}C/^{13}C$ 0.5 HR**	$^{12}C/^{13}C$ 24 HR**	Name	HMDB	# matched entries in EML
1	259	353.1174	1.16(8)	0.95(12)	0.67(12)	Threonine L-Homoserine Hydroxyethyl glycine Allothreonine	HMDB00167 HMDB00719 HMDB61148 HMDB04041	
2	270	450.1808	0.62(32)	0.89(9)	1.65(10)	*N*-a-Acetyl-L-arginine	HMDB04620	
3	327	418.1948	0.81(14)	0.93(8)	1.88(10)			0
4	338	434.1862	0.88(13)	1.11(16)	1.87(8)			3
5	366	418.1554	0.75(14)	0.89(16)	2.18(9)			2
6	*376*	*511.1758*	*1.99(28)*	*1.03(23)*	*0.30(24)*	*Queuine* *Entecavir* *S-(2-Methylpropionyl)-dihydrolipoamide-E*	*HMDB01495* *HMDB14585* *HMDB06868*	
7	388	391.1134	0.79(10)	0.96(11)	1.27(10)			0
8	*392*	*436.1550*	*3.16(32)*	*1.59(24)*	*0.23(42)*	*Serinyl-Proline* *Prolyl-Serine* *L-Coprine*	*HMDB29047* *HMDB29026* *HMDB34266*	

						Hydroxyprolyl-Alanine	*HMDB28856*	
						Alanyl-Hydroxyproline	*HMDB28688*	
9	*417*	*464.1493*	*3.16(29)*	*2.06(24)*	*0.25(43)*	*(2S,3'S)-alpha-Amino-2-carboxy-5-oxo-1-pyrrolidinebutanoic acid*	*HMDB39110*	
						1-(gamma-Glutamylamino)cyclopropanecarboxylic acid	*HMDB31701*	
						Aspartyl-Proline	*HMDB28761*	
						Prolyl-Aspartate	*HMDB29013*	
						Aspartyl-L-proline	*HMDB02335*	
10	429	367.0961	1.37(13)	1.09(8)	0.75(16)	Iminodiacetic acid	HMDB11753	
						L-Aspartic acid	HMDB00191	
						D-Aspartic acid	HMDB06483	
11	*432*	*478.1652*	*2.38(25)*	*1.72(21)*	*0.26(23)*	*Carbidopa*	*HMDB14336*	
						(2S,4S)-Pinnatanine	*HMDB29439*	
						Hydroxyprolyl-Hydroxyproline	*HMDB28864*	
12	432	455.1677	0.86(17)	1.18(10)	1.73(3)			4
13	*439*	*450.1698*	*2.82(29)*	*1.68(27)*	*0.23(29)*	*Sakacin P*	*HMDB38239*	
						Prolyl-Threonine	*HMDB29027*	
						Threoninyl-Proline	*HMDB29069*	

14	447	323.1073	1.31(16)	1.04(12)	0.66(10)			0
15	457	442.1550	0.06(17)	0.1(15)	3.57(7)			26
16	*476*	*420.1584*	*3.62(29)*	*1.72(25)*	*0.20(22)*	*Prolyl-Alanine* *Alanyl-Proline*	*HMDB29010* *HMDB28695*	
17	489	469.1835	0.90(12)	1.11(10)	1.81(9)			1
18	*510*	*323.1060*	*1.26(12)*	*1.00(15)*	*0.72(9)*	*Beta-alanine* *Ethyl carbamate* *D-Alanine* *Sarcosine* *L-Alanine*	*HMDB00056* *HMDB01310* *HMDB31219* *HMDB00271* *HMDB00161*	
19	*527*	*445.1539*	*2.28(29)*	*1.20(20)*	*0.20(34)*	*Zalcitabine*	*HMDB15078*	
20	*528*	*450.1692*	*2.35(24)*	*1.27(23)*	*0.36(34)*	*Sakacin P* *Prolyl-Threonine* *Threoninyl-Proline*	*HMDB38239* *HMDB29027* *HMDB29069*	
21	563	399.0678	0.63(16)	0.79(17)	1.69(7)			7
22	580	483.1987	0.82(8)	1.02(10)	1.90(9)			1
23	595	450.1495	0.51(15)	0.89(27)	2.77(15)	L-1,2,3,4-Tetrahydro-beta-carboline-3-carboxylic acid	HMDB35665	
24	618	395.1062	0.60(32)	0.77(12)	1.74(5)	Indole-3-carboxylic acid 2-Indolecarboxylic acid 3-Formyl-6-hydroxyindole Quinoline-4,8-diol	HMDB03320 HMDB02285 HMDB31172 HMDB60289	

						4,6-Dihydroxyquinoline	HMDB04077	
25	618	557.1596	0.62(23)	0.79(14)	1.84(10)			11
26	*665*	*420.1591*	*3.03(32)*	*1.99(27)*	*0.19(30)*	*Prolyl-Alanine* *Alanyl-Proline*	*HMDB29010* *HMDB28695*	
27	691	337.1219	1.54(18)	1.14(8)	0.76(13)	*N*-Ethylglycine 3-Aminobutanoic acid *O*-Acetylethanolamine (*R*)-β-aminoisobutyric acid (*S*)-β-aminoisobutyric acid	HMDB41945 HMDB31654 HMDB38394 HMDB02299 HMDB02166	
28	*696*	*460.1551*	*2.15(31)*	*1.34(27)*	*0.10(56)*	*Porphobilinogen*	*HMDB00245*	
29	705	380.1109	0.63(22)	0.88(14)	1.83(8)			0
30	*729*	*448.1914*	*2.66(33)*	*1.26(28)*	*0.31(32)*	*Dethiobiotin* *Prolyl-Valine* *Valyl-Proline*	*HMDB03581* *HMDB29030* *HMDB29135*	
31	748	378.1000	0.63(17)	0.83(12)	1.58(15)			43
32	809	328.1218	0.83(15)	1.00(14)	1.73(12)	L-Targinine Homo-L-arginine	HMDB29416 HMDB00670	
33	811	379.1327	1.41(12)	1.01(15)	0.64(6)	L-trans-5-Hydroxy-2-piperidinecarboxylic acid L-cis-4-(Hydroxymethyl)-2-pyrrolidinecarboxylic acid Allysine	HMDB29426 HMDB29425 HMDB01263	

						(*S*)-5-Amino-3-oxohexanoate	HMDB12131	
						(*S*)-2-amino-6-oxohexanoate	HMDB59595	
34	*824*	*351.1010*	*1.68(24)*	*1.14(16)*	*0.47(10)*	*L-Aspartate-semialdehyde*	*HMDB12249*	
						L-2-Amino-3-oxobutanoic acid	*HMDB06454*	
						Acetylglycine	*HMDB00532*	
35	835	624.1902	0.62(7)	0.74(9)	1.48(10)	(*Z*)-Resveratrol 4'-glucoside	HMDB30565	
						(*E*)-2-Glucosyl-3,4',5-trihydroxystilbene	HMDB36294	
						trans-Piceid	HMDB30564	
						cis-Piceid	HMDB31422	
36	879	305.0950	1.52(13)	1.19(17)	0.59(15)	Acrylamide	HMDB04296	
37	*890*	*347.6452*	*1.65(18)*	*1.26(29)*	*0.37(29)*			*0*
38	910	624.1901	0.65(11)	0.82(14)	1.37(12)	(*Z*)-Resveratrol 4'-glucoside	HMDB30565	
						(*E*)-2-Glucosyl-3,4',5-trihydroxystilbene	HMDB36294	
						trans-Piceid	HMDB30564	
						cis-Piceid	HMDB31422	
39	926	364.1808	0.73(8)	0.83(10)	1.90(7)	Agmatine	HMDB01432	
						Agmatinium	HMDB60259	
40	974	378.1001	0.65(27)	0.84(17)	1.75(13)			43

41	*978*	*654.2014*	*0.58(11)*	*0.88(8)*	*1.49(14)*	*2',4',6'-Trihydroxydihydrochalcone 2'-glucoside* *E-3179* *Bikoeniquinone A*	*HMDB37483* *HMDB13846* *HMDB40789*	
42	1002	319.1106	1.29(8)	1.09(11)	0.69(11)			16
43	1006	371.1057	0.64(10)	0.75(9)	1.61(12)	Salicylamide *m*-Aminobenzoic acid *p*-Aminobenzoic acid 2-Aminobenzoic acid	HMDB15687 HMDB01891 HMDB01392 HMDB01123	
44	1081	341.1293	0.67(16)	0.86(19)	2.04(14)			0
45	*1147*	*510.1500*	*0.81(14)*	*1.11(12)*	*1.75(14)*			*7*
46	1175	395.1055	0.69(23)	0.93(12)	1.32(10)	Indole-3-carboxylic 2-Indolecarboxylic acid 3-Formyl-6-hydroxyindol Quinoline-4,8-diol 4,6-Dihydroxyquinoline	HMDB03320 HMDB02285 HMDB31172 HMDB60289 HMDB04077	
47	1178	336.0898	0.74(12)	1.08(4)	1.24(3)			52
48	*1202*	*335.1055*	*2.34(33)*	*1.24(23)*	*0.19(53)*	*Serinyl-Proline* *Prolyl-Serine* *L-Coprine* *Hydroxyprolyl-Alanine*	*HMDB29047* *HMDB29026* *HMDB34266* *HMDB28856*	

						Alanyl-Hydroxyproline	*HMDB28688*	
49	1276	325.1163	0.60(27)	0.98(18)	1.70(10)			5
50	*1307*	*291.0975*	*2.24(32)*	*1.58(15)*	*0.17(28)*	*3-Amino-2-piperidone*	*HMDB00323*	
51	1316	412.1208	0.71(7)	0.98(17)	1.37(8)			20
52	1335	319.1163	0.70(22)	0.90(10)	1.61(7)			1
53	1407	326.1242	0.84(11)	1.06(12)	1.77(8)			2

Appendix E

Table E.1 List of putative metabolites of Goji tea identified based on accurate mass match with the HMDB human metabolite library.

	Retention time (sec)	Intensity	Accurate mass	Name	HMDB ID
1	1284	2.31E+04	88.0997	Putrescine	HMDB01414
2	512	1.76E+04	45.0207	Formamide	HMDB01536
3	736	4.07E+04	45.0576	Dimethylamine	HMDB00087
				Ethylamine	HMDB13231
4	1356	3.38E+03	102.1160	*N*-Methylputrescine	HMDB03661
				Cadaverine	HMDB02322
5	1613	8.91E+03	110.0372	Hydroquinone	HMDB02434
				Pyrocatechol	HMDB00957
6	1660	2.48E+05	110.0368	Hydroquinone	HMDB02434
				Pyrocatechol	HMDB00957
7	1591	1.07E+05	110.0477	Imidazole-4-acetaldehyde	HMDB03905
8	1618	6.48E+04	111.0318	Pyrrole-2-carboxylic acid	HMDB04230
9	1569	7.56E+03	111.0354	Pyrrole-2-carboxylic acid	HMDB04230
10	1591	1.74E+04	112.0472	trans-1,2-Dihydrobenzene-1,2-diol	HMDB01164
11	1317	2.53E+05	114.0794	3-Amino-2-piperidone	HMDB00323
12	1304	3.17E+03	115.0641	D-Proline	HMDB03411
				L-Proline	HMDB00162
				Acetamidopropanal	HMDB12880
13	345	1.19E+04	60.0318	Urea	HMDB00294
14	368	3.23E+05	61.0522	Ethanolamine	HMDB00149
15	861	2.40E+04	132.0545	L-Asparagine	HMDB00168
				Ureidopropionic acid	HMDB00026
				Glycyl-glycine	HMDB11733
16	994	1.44E+05	132.0907	D-Ornithine	HMDB03374
				Ornithine	HMDB00214
17	1674	1.08E+05	136.0524	4-Hydroxyphenylacetaldehyde	HMDB03767

18	1559	8.79E+04	137.0841	2-Hydroxyphenethylamine	HMDB01065
				Tyramine	HMDB00306
				m-Tyramine	HMDB04989
19	1594	4.39E+04	138.0321	Gentisate aldehyde	HMDB04062
				3-Hydroxybenzoic acid	HMDB02466
				Salicylic acid	HMDB01895
				4-Hydroxybenzoic acid	HMDB00500
20	157	1.17E+04	71.0367	Acrylamide	HMDB04296
21	1370	5.07E+03	143.0590	Vinylacetylglycine	HMDB00894
22	1546	3.78E+04	146.0427	Coumarin	HMDB01218
23	561	1.66E+04	73.0523	3-Aminopropionaldehyde	HMDB01106
				Aminoacetone	HMDB02134
24	1063	1.58E+05	146.1063	(3*S*,5*S*)-3,5-Diaminohexanoate	HMDB12115
				(3*S*)-3,6-Diaminohexanoate	HMDB12114
				D-Lysine	HMDB03405
				L-Lysine	HMDB00182
25	402	6.78E+05	75.0315	Glycine	HMDB00123
26	445	1.05E+04	75.0677	1-Amino-propan-2-ol	HMDB12136
27	508	2.58E+04	75.0678	1-Amino-propan-2-ol	HMDB12136
28	1482	6.77E+03	151.0503	2-Hydroxyadenine	HMDB00403
				Guanine	HMDB00132
				8-Hydroxyadenine	HMDB00542
29	1485	3.00E+04	154.0271	3,5-Dihydroxyphenylbenzoic acid	HMDB13173
				Protocatechuic acid	HMDB01856
				2-Pyrocatechuic acid	HMDB00397
				Gentisic acid	HMDB00152
				2,6-dihydroxybenzoic acid	HMDB13676
				3,5-dihydroxybenzoic acid	HMDB13677
30	1099	1.26E+04	155.0709	L-Histidine	HMDB00177
31	1548	1.19E+04	159.0679	Indoleacetaldehyde	HMDB01190
32	1253	5.37E+03	160.1223	*N*6-Methyllysine	HMDB02038
				Isoputreanine	HMDB06009

33	1186	1.32E+04	161.0697	Aminoadipic acid	HMDB00510
34	1258	3.29E+03	161.0712	Aminoadipic acid	HMDB00510
35	1540	1.91E+04	167.0951	4-Methoxytyramine	HMDB12162
				p-Synephrine	HMDB04826
				Phenylephrine	HMDB02182
				3-Methoxytyramine	HMDB00022
36	1630	2.04E+04	170.0604	3,4-Dihydroxyphenylglycol	HMDB00318
37	809	5.69E+05	85.0531	2-Pyrrolidinone	HMDB02039
38	1379	1.85E+04	172.0863	Glycylproline	HMDB00721
				L-prolyl-L-glycine	HMDB11178
39	565	3.08E+04	87.0317	2-Aminoacrylic acid	HMDB03609
40	685	1.70E+04	87.0323	2-Aminoacrylic acid	HMDB03609
41	676	7.09E+04	174.1123	*N*-Methyltryptamine	HMDB04370
				D-Arginine	HMDB03416
				L-Arginine	HMDB00517
42	574	1.93E+04	87.0680	4-Aminobutyraldehyde	HMDB01080
43	684	1.68E+04	87.0684	4-Aminobutyraldehyde	HMDB01080
44	597	3.63E+05	87.0683	4-Aminobutyraldehyde	HMDB01080
45	854	9.68E+04	87.0686	4-Aminobutyraldehyde	HMDB01080
46	1499	1.57E+04	176.0952	Serotonin	HMDB00259
				Canavanine	HMDB02706
47	1638	9.28E+03	88.1000	Putrescine	HMDB01414
48	570	5.84E+03	89.0459	Sarcosine	HMDB00271
				L-Alanine	HMDB00161
				Beta-Alanine	HMDB00056
				D-Alanine	HMDB01310
49	455	3.42E+06	89.0486	Sarcosine	HMDB00271
				L-Alanine	HMDB00161
				Beta-Alanine	HMDB00056
				D-Alanine	HMDB01310
50	461	5.07E+06	89.0493	Sarcosine	HMDB00271
				L-Alanine	HMDB00161

				Beta-Alanine	HMDB00056
				D-Alanine	HMDB01310
51	933	1.25E+05	179.0804	Fructosamine	HMDB02030
				Glucosamine	HMDB01514
				7-Aminomethyl-7-carbaguanine	HMDB11690
52	873	1.49E+04	179.0809	Fructosamine	HMDB02030
				Glucosamine	HMDB01514
				7-Aminomethyl-7-carbaguanine	HMDB11690
53	750	5.92E+03	179.0807	Fructosamine	HMDB02030
				Glucosamine	HMDB01514
				7-Aminomethyl-7-carbaguanine	HMDB11690
54	961	9.02E+03	179.0818	Fructosamine	HMDB02030
				Glucosamine	HMDB01514
				7-Aminomethyl-7-carbaguanine	HMDB11690
55	1289	7.15E+03	180.0428	3,4-Dihydroxy-trans-cinnamate	HMDB03501
				Caffeic acid	HMDB01964
56	1492	5.68E+05	180.0431	3,4-Dihydroxy-trans-cinnamate	HMDB03501
				Caffeic acid	HMDB01964
57	1373	3.84E+05	181.0752	Beta-Tyrosine	HMDB03831
				L-Threo-3-Phenylserine	HMDB02184
58	1470	2.46E+04	182.0588	3,4-Dihydroxyhydrocinnamic acid	HMDB00423
59	1507	1.52E+04	184.0339	3,4-Dihydroxymandelic acid	HMDB01866
				4-O-Methylgallic acid	HMDB13198
60	1406	1.00E+04	94.0420	Phenol	HMDB00228
61	1039	1.09E+04	95.0374	2-hydroxypyridine	HMDB13751
62	820	1.96E+05	101.0841	5-Aminopentanal	HMDB12815
63	554	1.03E+05	103.0624	L-Alpha-aminobutyric acid	HMDB00452
				Gamma-Aminobutyric acid	HMDB00112
				Dimethylglycine	HMDB00092
				D-Alpha-aminobutyric acid	HMDB00650
				2-Aminoisobutyric acid	HMDB01906
				(*S*)-β-aminoisobutyric acid	HMDB02166

				(*R*)-β-aminoisobutyric acid	HMDB02299
				3-Aminoisobutanoic acid	HMDB03911
64	246	2.53E+04	103.0624	L-Alpha-aminobutyric acid	HMDB00452
				Gamma-Aminobutyric acid	HMDB00112
				Dimethylglycine	HMDB00092
				D-Alpha-aminobutyric acid	HMDB00650
				2-Aminoisobutyric acid	HMDB01906
				(*S*)-β-aminoisobutyric acid	HMDB02166
				(*R*)-β-aminoisobutyric acid	HMDB02299
				3-Aminoisobutanoic acid	HMDB03911
65	276	5.07E+06	105.0438	L-Serine	HMDB00187
				D-Serine	HMDB03406
66	851	3.62E+03	217.1447	Gamma-glutamyl-L-putrescine	HMDB12230
67	1080	4.81E+04	109.0527	4-Aminophenol	HMDB01169
68	1108	5.99E+03	109.0542	4-Aminophenol	HMDB01169
69	1576	1.22E+04	220.0853	5-Hydroxy-L-tryptophan	HMDB00472
70	455	1.17E+04	110.0469	Imidazole-4-acetaldehyde	HMDB03905
71	423	3.01E+04	110.0473	Imidazole-4-acetaldehyde	HMDB03905
72	1576	4.79E+04	110.0474	Imidazole-4-acetaldehyde	HMDB03905
73	966	4.90E+03	221.0902	2'-Deoxysepiapterin	HMDB00389
74	1202	7.06E+03	222.0744	Salsolinol 1-carboxylate	HMDB13068
75	958	2.97E+04	222.0740	Salsolinol 1-carboxylate	HMDB13068
				Allocystathionine	HMDB00455
				L-Cystathionine	HMDB00099
76	1184	3.37E+03	222.0744	Salsolinol 1-carboxylate	HMDB13068
77	1010	7.18E+03	222.0753	Salsolinol 1-carboxylate	HMDB13068
78	167	7.08E+03	111.0776	Histamine	HMDB00870
79	1380	7.72E+03	111.0801	Histamine	HMDB00870
80	1651	1.28E+04	223.0809	*N*-Acetyl-L-tyrosine	HMDB00866
81	1484	3.76E+03	226.1002	Porphobilinogen	HMDB00245
				Carnosine	HMDB00033

				(6*R*)-6-(L-Erythro-1,2-Dihydroxypropyl)-5,6,7,8-tetrahydro-4a-hydroxypterin	HMDB12482
82	427	8.94E+03	114.0776	3-Amino-2-piperidone	HMDB00323
83	475	5.07E+06	114.0803	3-Amino-2-piperidone	HMDB00323
84	736	2.19E+04	115.0622	D-Proline	HMDB03411
				L-Proline	HMDB00162
				Acetamidopropanal	HMDB12880
85	1659	1.39E+04	115.0622	D-Proline	HMDB03411
				L-Proline	HMDB00162
				Acetamidopropanal	HMDB12880
86	794	1.12E+04	115.0627	D-Proline	HMDB03411
				L-Proline	HMDB00162
				Acetamidopropanal	HMDB12880
87	375	1.81E+04	115.0625	D-Proline	HMDB03411
				L-Proline	HMDB00162
				Acetamidopropanal	HMDB12880
88	274	7.60E+03	115.0626	D-Proline	HMDB03411
				L-Proline	HMDB00162
				Acetamidopropanal	HMDB12880
89	806	1.07E+04	115.0628	D-Proline	HMDB03411
				L-Proline	HMDB00162
				Acetamidopropanal	HMDB12880
90	835	9.98E+03	115.0628	D-Proline	HMDB03411
				L-Proline	HMDB00162
				Acetamidopropanal	HMDB12880
91	1295	1.34E+04	115.0629	D-Proline	HMDB03411
				L-Proline	HMDB00162
				Acetamidopropanal	HMDB12880
92	671	9.82E+04	115.0629	D-Proline	HMDB03411
				L-Proline	HMDB00162
				Acetamidopropanal	HMDB12880
93	767	1.41E+04	115.0630	D-Proline	HMDB03411

				L-Proline	HMDB00162
				Acetamidopropanal	HMDB12880
94	1132	6.77E+03	115.0632	D-Proline	HMDB03411
				L-Proline	HMDB00162
				Acetamidopropanal	HMDB12880
95	883	7.75E+03	115.0638	D-Proline	HMDB03411
				L-Proline	HMDB00162
				Acetamidopropanal	HMDB12880
96	920	6.38E+03	115.0639	D-Proline	HMDB03411
				L-Proline	HMDB00162
				Acetamidopropanal	HMDB12880
97	628	5.07E+06	115.0650	D-Proline	HMDB03411
				L-Proline	HMDB00162
				Acetamidopropanal	HMDB12880
98	813	1.53E+05	117.0424	L-2-Amino-3-oxobutanoic acid	HMDB06454
				Acetylglycine	HMDB00532
				L-Aspartate-semialdehyde	HMDB12249
99	254	3.48E+04	119.0571	L-Homoserine	HMDB00719
				L-Threonine	HMDB00167
				L-Allothreonine	HMDB04041
100	1307	3.25E+04	122.0364	4-Hydroxybenzaldehyde	HMDB11718
				4-Hydroxybenzaldehyde	HMDB11718
101	955	7.44E+03	249.0835	*S*-Acetyldihydrolipoamide	HMDB01526
				S-Acetyldihydrolipoamide-E	HMDB06878
102	883	4.16E+03	249.0864	*S*-Acetyldihydrolipoamide	HMDB01526
				S-Acetyldihydrolipoamide-E	HMDB06878
103	1075	1.50E+04	249.0853	*S*-Acetyldihydrolipoamide	HMDB01526
				S-Acetyldihydrolipoamide-E	HMDB06878
104	1052	9.41E+03	249.0858	*S*-Acetyldihydrolipoamide	HMDB01526
				S-Acetyldihydrolipoamide-E	HMDB06878
105	955	7.07E+03	249.0857	*S*-Acetyldihydrolipoamide	HMDB01526
				S-Acetyldihydrolipoamide-E	HMDB06878

106	1288	3.85E+03	249.0868	*S*-Acetyldihydrolipoamide	HMDB01526
				S-Acetyldihydrolipoamide-E	HMDB06878
107	1073	3.49E+03	251.0971	Deoxyadenosine	HMDB00101
				Muramic acid	HMDB03254
				5'-Deoxyadenosine	HMDB01983
108	994	1.11E+04	251.1004	Deoxyadenosine	HMDB00101
				Muramic acid	HMDB03254
				5'-Deoxyadenosine	HMDB01983
109	1131	8.25E+03	251.1018	Deoxyadenosine	HMDB00101
				Muramic acid	HMDB03254
				5'-Deoxyadenosine	HMDB01983
110	1186	4.15E+03	251.1028	Deoxyadenosine	HMDB00101
				Muramic acid	HMDB03254
				5'-Deoxyadenosine	HMDB01983
111	856	1.21E+04	126.0317	1,3,5-trihydroxybenzene	HMDB13675
				1,2,3-trihydroxybenzene	HMDB13674
112	1651	5.22E+03	126.0315	1,3,5-trihydroxybenzene	HMDB13675
				1,2,3-trihydroxybenzene	HMDB13674
113	1453	4.40E+03	126.0319	1,3,5-trihydroxybenzene	HMDB13675
				1,2,3-trihydroxybenzene	HMDB13674
114	1217	5.85E+03	126.0329	1,3,5-trihydroxybenzene	HMDB13675
				1,2,3-trihydroxybenzene	HMDB13674
115	310	1.16E+04	128.0569	Dihydrothymine	HMDB00079
116	443	6.64E+05	128.0579	Dihydrothymine	HMDB00079
117	579	2.44E+05	129.0419	*N*-Acryloylglycine	HMDB01843
				Pyrrolidonecarboxylic acid	HMDB00805
118	770	4.84E+03	129.0777	*N*4-Acetylaminobutanal	HMDB04226
				L-Pipecolic acid	HMDB00716
				Pipecolic acid	HMDB00070
				D-Pipecolic acid	HMDB05960
119	710	1.67E+04	129.0786	*N*4-Acetylaminobutanal	HMDB04226
				L-Pipecolic acid	HMDB00716

				Pipecolic acid	HMDB00070
				D-Pipecolic acid	HMDB05960
120	809	1.36E+06	129.0789	*N*4-Acetylaminobutanal	HMDB04226
				L-Pipecolic acid	HMDB00716
				Pipecolic acid	HMDB00070
				D-Pipecolic acid	HMDB05960
121	405	1.11E+04	131.0569	5-Amino-2-oxopentanoic acid	HMDB06272
				4-Hydroxy-L-proline	HMDB06055
				3-Hydroxy-L-proline	HMDB02113
				L-Glutamic-gamma-semialdehyde	HMDB02104
				5-Aminolevulinic acid	HMDB01149
				Propionylglycine	HMDB00783
				N-Acetyl-L-alanine	HMDB00766
				Hydroxyproline	HMDB00725
122	705	5.95E+04	131.0578	5-Amino-2-oxopentanoic acid	HMDB06272
				4-Hydroxy-L-proline	HMDB06055
				3-Hydroxy-L-proline	HMDB02113
				L-Glutamic-gamma-semialdehyde	HMDB02104
				5-Aminolevulinic acid	HMDB01149
				Propionylglycine	HMDB00783
				N-Acetyl-L-alanine	HMDB00766
				Hydroxyproline	HMDB00725
123	797	5.85E+05	131.0946	Aminocaproic acid	HMDB01901
				L-Norleucine	HMDB01645
				L-Leucine	HMDB00687
				L-Alloisoleucine	HMDB00557
				L-Isoleucine	HMDB00172
				Beta-Leucine	HMDB03640
				D-Leucine	HMDB13773
124	188	5.07E+06	132.0542	L-Asparagine	HMDB00168
				Ureidopropionic acid	HMDB00026
				Glycyl-glycine	HMDB11733

				N-Carbamoylsarcosine	HMDB12265
				Cinnamaldehyde	HMDB03441
125	1171	1.05E+04	265.1068	Thiamine	HMDB00235
126	881	1.45E+04	265.1143	Thiamine	HMDB00235
127	945	5.44E+03	265.1151	Thiamine	HMDB00235
128	426	1.76E+04	133.0366	L-Aspartic acid	HMDB00191
				D-Aspartic acid	HMDB06483
				Iminodiacetate	HMDB11753
129	850	2.52E+04	267.0950	Adenosine	HMDB00050
				Neuraminic acid	HMDB00830
				Deoxyguanosine	HMDB00085
130	1048	5.04E+03	267.1004	Adenosine	HMDB00050
				Neuraminic acid	HMDB00830
				Deoxyguanosine	HMDB00085
131	1580	8.98E+03	268.0589	L-Homocystine	HMDB00676
				DL-Homocystine	HMDB00575
132	584	5.28E+03	135.0349	Methylcysteine	HMDB02108
				Homocysteine	HMDB00742
133	590	4.88E+04	135.0538	Adenine	HMDB00034
134	642	2.03E+05	135.0542	Adenine	HMDB00034
135	1317	9.82E+03	136.0518	4-Hydroxyphenylacetaldehyde	HMDB03767
136	1077	1.89E+05	138.0316	Gentisate aldehyde	HMDB04062
				3-Hydroxybenzoic acid	HMDB02466
				Salicylic acid	HMDB01895
				4-Hydroxybenzoic acid	HMDB00500
137	1007	1.07E+04	276.1312	Saccharopine	HMDB00279
138	605	1.44E+04	276.1317	Saccharopine	HMDB00279
139	1491	8.95E+03	138.0686	Tyrosol	HMDB04284
140	1065	1.05E+04	277.1154	*S*-(2-Methylpropionyl)-dihydrolipoamide-E	HMDB06868
				Queuine	HMDB01495
141	1222	1.68E+04	277.1157	*S*-(2-Methylpropionyl)-dihydrolipoamide-E	HMDB06868
				Queuine	HMDB01495

142	1123	3.69E+04	277.1157	*S*-(2-Methylpropionyl)-dihydrolipoamide-E	HMDB06868
				Queuine	HMDB01495
143	1182	2.62E+04	277.1159	*S*-(2-Methylpropionyl)-dihydrolipoamide-E	HMDB06868
				Queuine	HMDB01495
144	1281	1.55E+05	277.1162	*S*-(2-Methylpropionyl)-dihydrolipoamide-E	HMDB06868
				Queuine	HMDB01495
145	1295	3.86E+05	277.1163	*S*-(2-Methylpropionyl)-dihydrolipoamide-E	HMDB06868
				Queuine	HMDB01495
146	1085	2.63E+04	277.1174	*S*-(2-Methylpropionyl)-dihydrolipoamide-E	HMDB06868
				Queuine	HMDB01495
147	1154	6.89E+04	277.1175	*S*-(2-Methylpropionyl)-dihydrolipoamide-E	HMDB06868
				Queuine	HMDB01495
148	848	2.13E+04	139.0263	6-Hydroxynicotinic acid	HMDB02658
				4-Nitrophenol	HMDB01232
				3-Hydroxypicolinic acid	HMDB13188
149	1421	1.23E+04	278.1267	L-phenylalanyl-L-hydroxyproline	HMDB11176
150	902	2.03E+04	281.1118	2'-*O*-Methyladenosine	HMDB04326
				3'-*O*-Methyladenosine	HMDB06023
151	1554	3.68E+05	141.0185	*O*-Phosphoethanolamine	HMDB00224
152	609	8.96E+03	143.0577	Vinylacetylglycine	HMDB00894
153	1487	1.59E+04	290.0791	Catechin	HMDB02780
				Epicatechin	HMDB01871
154	450	9.28E+03	145.0718	2-Keto-6-aminocaproate	HMDB12151
				(*S*)-5-Amino-3-oxohexanoate	HMDB12131
				4-Acetamidobutanoic acid	HMDB03681
				Allysine	HMDB01263
				N-Butyrylglycine	HMDB00808
				Isobutyrylglycine	HMDB00730
155	407	1.49E+04	145.0727	2-Keto-6-aminocaproate	HMDB12151
				(*S*)-5-Amino-3-oxohexanoate	HMDB12131
				4-Acetamidobutanoic acid	HMDB03681
				Allysine	HMDB01263

				N-Butyrylglycine	HMDB00808
				Isobutyrylglycine	HMDB00730
156	808	4.64E+04	145.0731	2-Keto-6-aminocaproate	HMDB12151
				(*S*)-5-Amino-3-oxohexanoate	HMDB12131
				4-Acetamidobutanoic acid	HMDB03681
				Allysine	HMDB01263
				N-Butyrylglycine	HMDB00808
				Isobutyrylglycine	HMDB00730
157	251	1.29E+04	146.0674	L-Glutamine	HMDB00641
				Ureidoisobutyric acid	HMDB02031
				D-Glutamine	HMDB03423
				Alanylglycine	HMDB06899
158	287	4.55E+03	146.0692	L-Glutamine	HMDB00641
				Ureidoisobutyric acid	HMDB02031
				D-Glutamine	HMDB03423
				Alanylglycine	HMDB06899
159	447	5.59E+04	147.0518	D-Glutamic acid	HMDB03339
				O-Acetylserine	HMDB03011
				N-Acetylserine	HMDB02931
				N-Methyl-D-aspartic acid	HMDB02393
				L-Glutamic acid	HMDB00148
				L-4-Hydroxyglutamate semialdehyde	HMDB06556
160	507	7.46E+03	147.0520	D-Glutamic acid	HMDB03339
				O-Acetylserine	HMDB03011
				N-Acetylserine	HMDB02931
				N-Methyl-D-aspartic acid	HMDB02393
				L-Glutamic acid	HMDB00148
				L-4-Hydroxyglutamate semialdehyde	HMDB06556
161	313	9.40E+05	147.0522	D-Glutamic acid	HMDB03339
				O-Acetylserine	HMDB03011
				N-Acetylserine	HMDB02931
				N-Methyl-D-aspartic acid	HMDB02393

				L-Glutamic acid	HMDB00148
				L-4-Hydroxyglutamate semialdehyde	HMDB06556
162	375	1.74E+04	147.0522	D-Glutamic acid	HMDB03339
				O-Acetylserine	HMDB03011
				N-Acetylserine	HMDB02931
				N-Methyl-D-aspartic acid	HMDB02393
				L-Glutamic acid	HMDB00148
				L-4-Hydroxyglutamate semialdehyde	HMDB06556
163	667	7.99E+04	149.0507	5,6-Dihydroxyindole	HMDB04058
				L-Methionine	HMDB00696
164	375	5.97E+03	150.0547	1-Methylhypoxanthine	HMDB13141
165	1480	1.72E+04	150.0673	4-Coumaryl alcohol	HMDB03654
166	1458	1.71E+04	151.0634	Acetaminophen	HMDB01859
				2-Phenylglycine	HMDB02210
				Dopamine quinone	HMDB12219
				Leukoaminochrome	HMDB12992
167	1297	1.76E+04	152.0470	3-Hydroxyphenylacetic acid	HMDB00440
				p-Hydroxyphenylacetic acid	HMDB00020
				Ortho-Hydroxyphenylacetic acid	HMDB00669
				3-Cresotinic acid	HMDB02390
				3,4-Dihydroxyphenylacetaldehyde	HMDB03791
				4-Hydroxy-3-methylbenzoic acid	HMDB04815
168	1034	9.68E+03	152.0475	3-Hydroxyphenylacetic acid	HMDB00440
				p-Hydroxyphenylacetic acid	HMDB00020
				Ortho-Hydroxyphenylacetic acid	HMDB00669
				3-Cresotinic acid	HMDB02390
				3,4-Dihydroxyphenylacetaldehyde	HMDB03791
				4-Hydroxy-3-methylbenzoic acid	HMDB04815
169	1420	8.81E+03	152.0488	3-Hydroxyphenylacetic acid	HMDB00440
				p-Hydroxyphenylacetic acid	HMDB00020
				Ortho-Hydroxyphenylacetic acid	HMDB00669
				3-Cresotinic acid	HMDB02390

				3,4-Dihydroxyphenylacetaldehyde	HMDB03791
				4-Hydroxy-3-methylbenzoic acid	HMDB04815
170	1113	9.02E+03	153.0452	3-Aminosalicylic acid	HMDB01972
				3-Hydroxyanthranilic acid	HMDB01476
171	368	1.68E+04	155.0679	L-Histidine	HMDB00177
172	341	1.18E+05	155.0682	L-Histidine	HMDB00177
173	1100	5.58E+05	155.0698	L-Histidine	HMDB00177
174	310	6.57E+03	157.0724	3-Methylcrotonylglycine	HMDB00459
				Tiglylglycine	HMDB00959
175	252	6.64E+04	157.0726	3-Methylcrotonylglycine	HMDB00459
				Tiglylglycine	HMDB00959
176	537	6.90E+03	157.0764	3-Methylcrotonylglycine	HMDB00459
				Tiglylglycine	HMDB00959
177	275	4.96E+04	158.0429	Allantoin	HMDB00462
178	1643	6.34E+03	159.0722	Indoleacetaldehyde	HMDB01190
179	969	1.60E+04	159.0892	2-Methylbutyrylglycine	HMDB00339
				Isovalerylglycine	HMDB00678
				Valerylglycine	HMDB00927
				N-Acetylvaline	HMDB11757
				5-Acetamidovalerate	HMDB12175
180	1637	5.53E+03	162.0326	3 Hydroxycoumarin	HMDB02149
181	675	1.47E+04	164.0474	2-Hydroxycinnamic acid	HMDB02641
182	642	1.37E+04	164.0482	*m*-Coumaric acid	HMDB01713
				2-Hydroxycinnamic acid	HMDB02641
183	1122	1.12E+06	164.0476	*m*-Coumaric acid	HMDB01713
				2-Hydroxycinnamic acid	HMDB02641
184	644	2.28E+04	164.0486	*m*-Coumaric acid	HMDB01713
				2-Hydroxycinnamic acid	HMDB02641
185	590	1.45E+04	164.0484	*m*-Coumaric acid	HMDB01713
				2-Hydroxycinnamic acid	HMDB02641
186	681	9.85E+03	164.0486	*m*-Coumaric acid	HMDB01713
				2-Hydroxycinnamic acid	HMDB02641

187	981	8.49E+03	165.0437	Formylanthranilic acid	HMDB04089
				Methionine sulfoxide	HMDB02005
188	264	7.21E+03	165.0465	Formylanthranilic acid	HMDB04089
				Methionine sulfoxide	HMDB02005
189	778	9.95E+04	165.0785	L-Phenylalanine	HMDB00159
				Benzocaine	HMDB04992
				Norsalsolinol	HMDB06044
190	1103	3.31E+04	166.0629	Desaminotyrosine	HMDB02199
				3-(3-Hydroxyphenyl)propanoic acid	HMDB00375
191	297	4.44E+03	167.0922	4-Methoxytyramine	HMDB12162
				p-Synephrine	HMDB04826
				Phenylephrine	HMDB02182
				3-Methoxytyramine	HMDB00022
192	409	1.81E+04	167.0938	4-Methoxytyramine	HMDB12162
				p-Synephrine	HMDB04826
				Phenylephrine	HMDB02182
				3-Methoxytyramine	HMDB00022
193	1062	3.85E+04	168.0422	Homogentisic acid	HMDB00130
				Vanillic acid	HMDB00484
				3-Hydroxymandelic acid	HMDB00750
				p-Hydroxymandelic acid	HMDB00822
				3,4-Dihydroxybenzeneacetic acid	HMDB01336
				5-Methoxysalicylic acid	HMDB01868
				3,4-Dihydroxymandelaldehyde	HMDB06242
194	1285	6.61E+03	168.0420	Homogentisic acid	HMDB00130
				Vanillic acid	HMDB00484
				3-Hydroxymandelic acid	HMDB00750
				p-Hydroxymandelic acid	HMDB00822
				3,4-Dihydroxybenzeneacetic acid	HMDB01336
				5-Methoxysalicylic acid	HMDB01868
				3,4-Dihydroxymandelaldehyde	HMDB06242
195	899	5.28E+03	168.0424	Homogentisic acid	HMDB00130

				Vanillic acid	HMDB00484
				3-Hydroxymandelic acid	HMDB00750
				p-Hydroxymandelic acid	HMDB00822
				3,4-Dihydroxybenzeneacetic acid	HMDB01336
				5-Methoxysalicylic acid	HMDB01868
				3,4-Dihydroxymandelaldehyde	HMDB06242
196	1451	1.37E+04	168.0431	Homogentisic acid	HMDB00130
				Vanillic acid	HMDB00484
				3-Hydroxymandelic acid	HMDB00750
				p-Hydroxymandelic acid	HMDB00822
				3,4-Dihydroxybenzeneacetic acid	HMDB01336
				5-Methoxysalicylic acid	HMDB01868
				3,4-Dihydroxymandelaldehyde	HMDB06242
197	1524	4.74E+03	169.0340	2-Furoylglycine	HMDB00439
198	733	4.37E+04	341.1318	Lactosamine	HMDB06591
				6-(alpha-D-Glucosaminyl)-1D-myo-inositol	HMDB11668
199	474	5.72E+04	172.0838	Glycylproline	HMDB00721
				L-prolyl-L-glycine	HMDB11178
200	1608	1.84E+04	172.0835	Glycylproline	HMDB00721
				L-prolyl-L-glycine	HMDB11178
201	701	1.31E+04	173.0652	N-Acetyl-L-glutamate 5-semialdehyde	HMDB06488
202	451	5.68E+03	173.1045	Isovalerylalanine	HMDB00747
				Hexanoylglycine	HMDB00701
				N-Acetylleucine	HMDB11756
203	1113	1.49E+04	173.1050	Isovalerylalanine	HMDB00747
				Hexanoylglycine	HMDB00701
				N-Acetylleucine	HMDB11756
204	1070	7.20E+03	173.1052	Isovalerylalanine	HMDB00747
				Hexanoylglycine	HMDB00701
				N-Acetylleucine	HMDB11756
205	364	9.70E+03	174.0995	*N*-Acetylornithine	HMDB03357
206	300	1.29E+04	174.0996	*N*-Acetylornithine	HMDB03357

207	184	1.41E+04	174.1099	L-Arginine	HMDB00517
				D-Arginine	HMDB03416
208	153	1.08E+06	174.1106	L-Arginine	HMDB00517
				D-Arginine	HMDB03416
209	1531	7.80E+03	175.0630	Indoleacetic acid	HMDB00197
				Guanidinosuccinic acid	HMDB03157
				5-Hydroxyindoleacetaldehyde	HMDB04073
210	502	1.05E+04	175.0831	*N*-Carboxyethyl-g-aminobutyric acid	HMDB02201
211	143	2.61E+04	175.0943	Argininic acid	HMDB03148
				Citrulline	HMDB00904
212	222	8.22E+05	175.0948	Argininic acid	HMDB03148
				Citrulline	HMDB00904
213	1248	6.07E+03	177.0480	*N*-Formyl-L-methionine	HMDB01015
214	1297	6.43E+03	177.0486	*N*-Formyl-L-methionine	HMDB01015
215	1194	8.82E+03	177.0489	*N*-Formyl-L-methionine	HMDB01015
216	1681	5.10E+03	358.1424	2-Phenylaminoadenosine	HMDB01069
217	125	4.51E+05	179.0782	Fructosamine	HMDB02030
				Glucosamine	HMDB01514
				7-Aminomethyl-7-carbaguanine	HMDB11690
218	180	2.05E+05	179.0783	Fructosamine	HMDB02030
				Glucosamine	HMDB01514
				7-Aminomethyl-7-carbaguanine	HMDB11690
219	1495	6.71E+03	358.1722	Malabaricone C	HMDB05798
220	1371	5.59E+03	358.1725	Malabaricone C	HMDB05798
221	1417	7.75E+03	358.1739	Malabaricone C	HMDB05798
222	532	1.14E+04	180.0415	3,4-Dihydroxy-trans-cinnamate	HMDB03501
				Caffeic acid	HMDB01964
				Aspirin	HMDB01879
				4-Hydroxyphenylpyruvic acid	HMDB00707
				2-Hydroxy-3-(4-hydroxyphenyl)propenoic acid	HMDB06915
				3-Hydroxyphenylpyruvic acid	HMDB11663

223	448	6.23E+03	180.0566	Nicotinuric acid	HMDB03269
224	668	7.59E+03	181.0737	Beta-Tyrosine	HMDB03831
				L-Threo-3-Phenylserine	HMDB02184
				L-Tyrosine	HMDB00158
				o-Tyrosine	HMDB06050
225	1248	5.19E+03	182.0582	3,4-Dihydroxyhydrocinnamic acid	HMDB00423
				Isohomovanillic acid	HMDB00333
				Homovanillic acid	HMDB00118
				Hydroxyphenyllactic acid	HMDB00755
				3-(3-hydroxyphenyl)-3-hydroxypropanoic acid	HMDB02643
				3-Methoxy-4-hydroxyphenylglycolaldehyde	HMDB04061
226	858	1.46E+04	182.0578	3,4-Dihydroxyhydrocinnamic acid	HMDB00423
				Isohomovanillic acid	HMDB00333
				Homovanillic acid	HMDB00118
				Hydroxyphenyllactic acid	HMDB00755
				3-(3-hydroxyphenyl)-3-hydroxypropanoic acid	HMDB02643
				3-Methoxy-4-hydroxyphenylglycolaldehyde	HMDB04061
227	883	6.29E+03	182.0583	3,4-Dihydroxyhydrocinnamic acid	HMDB00423
				Isohomovanillic acid	HMDB00333
				Homovanillic acid	HMDB00118
				Hydroxyphenyllactic acid	HMDB00755
				3-(3-hydroxyphenyl)-3-hydroxypropanoic acid	HMDB02643
				3-Methoxy-4-hydroxyphenylglycolaldehyde	HMDB04061
228	509	5.72E+03	182.0602	3,4-Dihydroxyhydrocinnamic acid	HMDB00423
				Isohomovanillic acid	HMDB00333
				Homovanillic acid	HMDB00118
				Hydroxyphenyllactic acid	HMDB00755
				3-(3-hydroxyphenyl)-3-hydroxypropanoic acid	HMDB02643

				3-Methoxy-4-hydroxyphenylglycolaldehyde	HMDB04061
229	1580	9.95E+03	183.0873	Epinephrine	HMDB00068
				Normetanephrine	HMDB00819
				Methylnoradrenaline	HMDB02832
230	1544	1.35E+04	183.0896	Epinephrine	HMDB00068
				Normetanephrine	HMDB00819
				Methylnoradrenaline	HMDB02832
231	1341	7.25E+03	184.0735	Vanylglycol	HMDB01490
232	1479	5.00E+03	372.1884	Biocytin	HMDB03134
233	1441	3.83E+03	372.1888	Biocytin	HMDB03134
234	796	6.15E+03	187.0836	2-Keto-6-acetamidocaproate	HMDB12150
235	1216	9.12E+03	187.1198	*N*-Heptanoylglycine	HMDB13010
236	499	6.34E+03	188.1134	*N*6-Acetyl-L-lysine	HMDB00206
				N-Alpha-acetyllysine	HMDB00446
				Glycyl-L-leucine	HMDB00759
237	173	1.11E+04	188.1263	Homo-L-arginine	HMDB00670
238	170	4.86E+03	188.1494	*N*6,*N*6,*N*6-Trimethyl-L-lysine	HMDB01325
239	193	1.81E+04	190.0941	Diaminopimelic acid	HMDB01370
240	1129	5.99E+05	194.0579	Isoferulic acid	HMDB00955
				trans-Ferulic acid	HMDB00954
241	1105	5.99E+04	194.0581	Isoferulic acid	HMDB00955
				trans-Ferulic acid	HMDB00954
242	431	5.26E+03	197.1024	Metanephrine	HMDB04063
243	1023	4.37E+03	198.0520	Vanillylmandelic acid	HMDB00291
244	700	2.84E+05	204.0896	L-Tryptophan	HMDB00929
245	382	5.77E+03	205.0573	Lipoamide	HMDB00962
246	454	6.88E+03	205.0586	Lipoamide	HMDB00962
247	831	6.96E+03	207.0872	*N*-Acetyl-L-phenylalanine	HMDB00512
				Phenylpropionylglycine	HMDB00860
				3-Phenylpropionylglycine	HMDB02042
248	551	2.56E+04	212.1148	L-prolyl-L-proline	HMDB11180
249	565	1.23E+05	212.1152	L-prolyl-L-proline	HMDB11180

250	733	8.30E+04	214.1310	Dethiobiotin	HMDB03581
251	1579	1.43E+04	220.1062	*N*-Acetyl-b-glucosaminylamine	HMDB01104
252	217	3.78E+04	221.0861	*N*-Acetylmannosamine	HMDB01129
				N-Acetyl-b-D-galactosamine	HMDB00853
				Beta-N-Acetylglucosamine	HMDB00803
				N-Acetyl-D-glucosamine	HMDB00215
				N-Acetylgalactosamine	HMDB00212
253	252	9.79E+03	221.0878	*N*-Acetylmannosamine	HMDB01129
				N-Acetyl-b-D-galactosamine	HMDB00853
				Beta-N-Acetylglucosamine	HMDB00803
				N-Acetyl-D-glucosamine	HMDB00215
				N-Acetylgalactosamine	HMDB00212
				2'-Deoxysepiapterin	HMDB00389
254	310	1.63E+04	221.0879	*N*-Acetylmannosamine	HMDB01129
				N-Acetyl-b-D-galactosamine	HMDB00853
				Beta-N-Acetylglucosamine	HMDB00803
				N-Acetyl-D-glucosamine	HMDB00215
				N-Acetylgalactosamine	HMDB00212
				2'-Deoxysepiapterin	HMDB00389
255	329	4.71E+04	221.0886	*N*-Acetylmannosamine	HMDB01129
				N-Acetyl-b-D-galactosamine	HMDB00853
				Beta-N-Acetylglucosamine	HMDB00803
				N-Acetyl-D-glucosamine	HMDB00215
				N-Acetylgalactosamine	HMDB00212
				2'-Deoxysepiapterin	HMDB00389
256	296	1.02E+04	221.0890	*N*-Acetylmannosamine	HMDB01129
				N-Acetyl-b-D-galactosamine	HMDB00853
				Beta-N-Acetylglucosamine	HMDB00803
				N-Acetyl-D-glucosamine	HMDB00215
				N-Acetylgalactosamine	HMDB00212
				2'-Deoxysepiapterin	HMDB00389
257	434	5.80E+04	222.0722	Salsolinol 1-carboxylate	HMDB13068

258	428	6.56E+04	222.0726	Salsolinol 1-carboxylate	HMDB13068
259	406	5.54E+03	223.0832	*N*-Acetyl-L-tyrosine	HMDB00866
260	1092	2.71E+04	224.0684	sinapic acid	HMDB13728
261	531	1.43E+04	226.0584	3-Nitrotyrosine	HMDB01904
262	691	3.59E+04	226.0947	Porphobilinogen	HMDB00245
263	858	3.32E+04	228.1466	L-isoleucyl-L-proline	HMDB11174
				L-leucyl-L-proline	HMDB11175
264	855	2.54E+04	228.1474	L-isoleucyl-L-proline	HMDB11174
	855	2.54E+04	228.1474	L-leucyl-L-proline	HMDB11175
265	421	9.86E+04	230.0892	Aspartyl-L-proline	HMDB02335
266	506	1.03E+04	231.0732	*N*2-Succinyl-L-glutamic acid 5-semialdehyde	HMDB01180
267	441	1.31E+04	231.0730	*N*2-Succinyl-L-glutamic acid 5-semialdehyde	HMDB01180
268	437	1.42E+04	231.0737	*N*2-Succinyl-L-glutamic acid 5-semialdehyde	HMDB01180
269	266	6.63E+03	231.1063	Isovalerylglutamic acid	HMDB00726
				Suberylglycine	HMDB00953
270	506	6.98E+03	231.1089	Isovalerylglutamic acid	HMDB00726
				Suberylglycine	HMDB00953
271	346	6.70E+03	231.1097	Isovalerylglutamic acid	HMDB00726
				Suberylglycine	HMDB00953
272	1667	6.66E+03	236.1020	*S*-aminomethyldihydrolipoamide	HMDB06239
273	1506	1.28E+04	236.1019	*S*-aminomethyldihydrolipoamide	HMDB06239
274	203	1.69E+05	237.0837	8-[(aminomethyl)sulfanyl]-6-sulfanyloctanoic acid	HMDB13639
				Sepiapterin	HMDB00238
				Biopterin	HMDB00468
				D-Biopterin	HMDB00633
				Orinapterin	HMDB00817
				Dyspropterin	HMDB01195
				Primapterin	HMDB02263

275	478	3.31E+04	244.0683	Uridine	HMDB00296
				Pseudouridine	HMDB00767
276	936	2.86E+04	244.0761	3,3',4'5-Tetrahydroxystilbene	HMDB04215
277	1075	8.74E+03	244.0767	3,3',4'5-Tetrahydroxystilbene	HMDB04215
278	1117	6.57E+03	244.0770	3,3',4'5-Tetrahydroxystilbene	HMDB04215
279	1091	8.14E+03	244.0770	3,3',4'5-Tetrahydroxystilbene	HMDB04215
280	1143	5.45E+03	244.0776	3,3',4'5-Tetrahydroxystilbene	HMDB04215
281	1061	4.37E+03	244.0779	3,3',4'5-Tetrahydroxystilbene	HMDB04215
282	1043	1.54E+04	246.0878	L-alpha-Aspartyl-L-hydroxyproline	HMDB11160
283	1638	7.82E+03	248.1148	6-Hydroxymelatonin	HMDB04081
284	222	1.20E+04	249.0834	*S*-Acetyldihydrolipoamide	HMDB01526
				S-Acetyldihydrolipoamide-E	HMDB06878
285	290	4.34E+04	249.0838	*S*-Acetyldihydrolipoamide	HMDB01526
				S-Acetyldihydrolipoamide-E	HMDB06878
286	133	1.20E+05	249.0836	*S*-Acetyldihydrolipoamide	HMDB01526
				S-Acetyldihydrolipoamide-E	HMDB06878
287	186	8.29E+04	249.0839	*S*-Acetyldihydrolipoamide	HMDB01526
				S-Acetyldihydrolipoamide-E	HMDB06878
288	207	5.73E+04	249.0840	*S*-Acetyldihydrolipoamide	HMDB01526
				S-Acetyldihydrolipoamide-E	HMDB06878
289	356	1.34E+04	249.0840	*S*-Acetyldihydrolipoamide	HMDB01526
				S-Acetyldihydrolipoamide-E	HMDB06878
290	373	1.32E+04	251.0994	Deoxyadenosine	HMDB00101
				Muramic acid	HMDB03254
				5'-Deoxyadenosine	HMDB01983
291	318	8.75E+05	251.0994	Deoxyadenosine	HMDB00101
				Muramic acid	HMDB03254
				5'-Deoxyadenosine	HMDB01983
292	1029	5.95E+04	265.1135	Thiamine	HMDB00235
293	899	4.82E+04	265.1136	Thiamine	HMDB00235
294	842	6.28E+04	265.1145	Thiamine	HMDB00235
295	771	2.37E+04	265.1143	Thiamine	HMDB00235

296	939	3.08E+04	265.1142	Thiamine	HMDB00235
297	792	2.35E+04	265.1143	Thiamine	HMDB00235
298	887	4.54E+05	265.1147	Thiamine	HMDB00235
299	1022	3.48E+04	265.1144	Thiamine	HMDB00235
300	947	4.35E+05	265.1148	Thiamine	HMDB00235
301	1009	1.63E+05	265.1146	Thiamine	HMDB00235
302	884	6.37E+05	265.1150	Thiamine	HMDB00235
303	760	4.21E+04	267.0566	2-Amino-4-oxo-6-(1',2'-dioxoprolyl)-7,8-dihydroxypteridine	HMDB01410
304	219	4.78E+03	267.0928	Adenosine	HMDB00050
				Neuraminic acid	HMDB00830
				Deoxyguanosine	HMDB00085
305	163	4.39E+03	275.1113	Norophthalmic acid	HMDB05766
				Gamma-Glutamylglutamine	HMDB11738
306	297	1.54E+04	276.0935	Gamma-Glutamylglutamic acid	HMDB11737
307	315	2.59E+04	276.0941	Gamma-Glutamylglutamic acid	HMDB11737
308	124	4.68E+03	276.0950	Gamma-Glutamylglutamic acid	HMDB11737
309	1020	3.56E+03	276.0993	Gamma-Glutamylglutamic acid	HMDB11737
310	280	2.76E+04	276.1308	Saccharopine	HMDB00279
311	260	5.46E+04	276.1313	Saccharopine	HMDB00279
312	447	4.27E+04	277.1151	*S*-(2-Methylpropionyl)-dihydrolipoamide-E	HMDB06868
				Queuine	HMDB01495
313	350	6.39E+04	277.1146	*S*-(2-Methylpropionyl)-dihydrolipoamide-E	HMDB06868
				Queuine	HMDB01495
314	193	1.55E+04	277.1149	*S*-(2-Methylpropionyl)-dihydrolipoamide-E	HMDB06868
				Queuine	HMDB01495
315	371	7.20E+05	277.1153	*S*-(2-Methylpropionyl)-dihydrolipoamide-E	HMDB06868
				Queuine	HMDB01495
316	401	2.87E+05	277.1153	*S*-(2-Methylpropionyl)-dihydrolipoamide-E	HMDB06868
				Queuine	HMDB01495
317	334	1.87E+04	277.1154	*S*-(2-Methylpropionyl)-dihydrolipoamide-E	HMDB06868
				Queuine	HMDB01495

318	1101	1.25E+04	278.1306	L-phenylalanyl-L-hydroxyproline	HMDB11176
				Pantetheine	HMDB03426
319	1073	9.39E+03	278.1306	L-phenylalanyl-L-hydroxyproline	HMDB11176
				Pantetheine	HMDB03426
320	158	2.20E+05	281.1102	*N*6-Methyladenosine	HMDB04044
				2'-*O*-Methyladenosine	HMDB04326
				3'-*O*-Methyladenosine	HMDB06023
321	140	2.00E+04	283.0908	8-Hydroxy-deoxyguanosine	HMDB03333
				Guanosine	HMDB00133
322	454	4.59E+03	283.0912	8-Hydroxy-deoxyguanosine	HMDB03333
				Guanosine	HMDB00133
323	691	2.20E+04	287.1183	(*R,S*)-Norlaudanosoline	HMDB12486
324	829	1.03E+04	290.0789	Catechin	HMDB02780
				Epicatechin	HMDB01871
325	867	9.77E+04	290.0793	Catechin	HMDB02780
				Epicatechin	HMDB01871
326	143	2.32E+04	290.1098	*N*-Succinyl-L,L-2,6-diaminopimelate	HMDB12267
327	316	6.38E+03	290.1096	*N*-Succinyl-L,L-2,6-diaminopimelate	HMDB12267
328	1464	1.30E+04	290.1195	Argininosuccinic acid	HMDB00052
329	1505	3.15E+04	290.1191	Argininosuccinic acid	HMDB00052
330	1601	5.68E+04	290.1196	Argininosuccinic acid	HMDB00052
331	425	9.28E+03	291.1294	*S*-(3-Methylbutanoyl)-dihydrolipoamide-E	HMDB06867
				S-(2-Methylbutanoyl)-dihydrolipoamide	HMDB06869
332	430	5.40E+03	297.0890	5'-Methylthioadenosine	HMDB01173
333	438	6.14E+03	297.1072	2-Methylguanosine	HMDB05862
	438	6.14E+03	297.1072	3'-O-Methylguanosine	HMDB06038
334	305	1.94E+04	298.1126	7-Methylguanosine	HMDB01107
335	324	2.31E+04	298.1126	7-Methylguanosine	HMDB01107
336	222	4.06E+03	314.1105	7,8-Dihydropteroic acid	HMDB01412
337	324	4.72E+03	354.0908	Biflorin	HMDB02336
				Chlorogenic acid	HMDB03164
338	824	6.22E+03	358.1374	2-Phenylaminoadenosine	HMDB01069

339	737	1.94E+04	358.1730	Malabaricone C	HMDB05798
340	585	6.67E+03	358.1738	Malabaricone C	HMDB05798
341	1636	2.13E+04	126.0318	1,3,5-trihydroxybenzene	HMDB13675
				1,2,3-trihydroxybenzene	HMDB13674
342	1446	1.14E+04	131.0582	5-Amino-2-oxopentanoic acid	HMDB06272
				L-Glutamic-gamma-semialdehyde	HMDB02104
				5-Aminolevulinic acid	HMDB01149
343	503	4.12E+03	372.1860	Biocytin	HMDB03134
344	682	3.83E+03	372.1884	Biocytin	HMDB03134
345	851	8.89E+03	372.1890	Biocytin	HMDB03134
346	125	4.42E+03	381.1228	*N*-Acetyl-9-O-lactoylneuraminic acid	HMDB00768
347	1637	3.22E+04	383.0710	Pantoprazole	HMDB05017
348	1098	1.11E+05	155.0699	L-Histidine	HMDB00177
349	1531	5.48E+03	390.1067	Dopaxanthin	HMDB12221
350	1584	2.40E+04	168.0894	Pyridoxamine	HMDB01431
351	149	8.52E+03	409.1566	Queuosine	HMDB11596
352	1287	8.20E+03	180.0430	3,4-Dihydroxy-trans-cinnamate	HMDB03501
				Caffeic acid	HMDB01964
353	942	4.22E+03	181.0685	Beta-Tyrosine	HMDB03831
				L-Threo-3-Phenylserine	HMDB02184
				L-Tyrosine	HMDB00158
				o-Tyrosine	HMDB06050
354	349	3.34E+03	439.1660	18-Carboxy-dinor-LTE4	HMDB12607
355	238	3.87E+03	439.1672	18-Carboxy-dinor-LTE4	HMDB12607
356	1295	4.10E+04	278.1195	L-phenylalanyl-L-hydroxyproline	HMDB11176
357	1097	3.87E+03	278.1207	L-phenylalanyl-L-hydroxyproline	HMDB11176
358	360	4.69E+04	516.1461	Dermatan	HMDB00632
359	441	9.69E+04	516.1460	Dermatan	HMDB00632
360	1551	6.29E+03	286.0473	Kaempferol	HMDB05801
				Luteolin	HMDB05800
361	1344	4.93E+03	314.0771	Valdecoxib	HMDB05033

Table E.2 List of putative metabolites of Goji tea identified based on accurate mass match with the EML predicted human metabolite library. *Click on the link of the matches to see the matched metabolites in EML at the website www.mycompoundid.org.

	Retention time (sec)	Intensity	Accurate mass	Number of matched entries
1	983	3.82E+04	32.0261	37*
2	1462	7.12E+03	82.0524	4
3	1600	1.07E+04	100.0167	16
4	1414	7.20E+03	104.0115	16
5	1549	6.17E+03	106.0429	12
6	1635	2.85E+04	114.0317	28
7	1016	4.59E+04	57.0214	5
8	1671	6.23E+04	125.0476	15
9	1585	4.87E+04	125.0478	15
10	1516	2.26E+04	128.0109	8
11	1520	8.60E+03	128.0474	32
12	1470	1.50E+04	130.0270	40
13	475	2.06E+05	69.0572	2
14	346	9.18E+03	69.0573	2
15	1596	2.31E+04	142.0274	6
16	1460	3.33E+04	142.0276	6
17	1400	5.68E+03	142.0279	6
18	157	1.17E+04	71.0367	9
19	1047	1.02E+05	71.0737	6
20	1534	8.09E+03	143.0949	19
21	730	5.18E+03	72.0572	22
22	561	1.66E+04	73.0523	18
23	906	2.92E+05	74.0371	40
24	600	8.53E+04	76.0157	15
25	1597	1.28E+04	156.0433	17
26	481	9.93E+03	78.0310	8
27	1414	3.77E+04	158.0223	17

28	1512	1.89E+04	160.0372	39
29	1407	3.54E+04	160.0374	39
30	1325	1.19E+05	160.0377	39
31	1308	1.62E+04	162.0527	81
32	711	1.02E+04	162.0533	81
33	1611	5.33E+04	162.0532	81
34	1272	1.28E+04	162.0537	81
35	1417	5.35E+05	162.0537	81
36	1323	4.59E+05	162.0539	81
37	1423	1.43E+04	162.0546	5
38	1316	2.99E+05	162.0545	5
39	1256	7.42E+03	87.1049	10
40	775	1.82E+04	88.0161	30
41	897	3.17E+04	88.0164	30
42	1499	1.57E+04	176.0952	20
43	1638	9.28E+03	88.1000	5
44	1581	4.36E+04	178.0271	11
45	873	1.49E+04	179.0809	42
46	1579	2.35E+04	180.0206	1
47	1289	7.15E+03	180.0428	39
48	1492	5.68E+05	180.0431	39
49	1188	5.02E+03	180.0518	17
50	719	2.26E+04	180.0637	87
51	972	4.59E+04	180.0638	87
52	1079	4.19E+04	180.0645	87
53	1208	6.17E+03	180.0655	10
54	1008	9.10E+04	180.0650	87
55	1470	2.46E+04	182.0588	45
56	1237	5.28E+03	185.1537	2
57	1529	3.03E+04	188.0329	15
58	397	1.44E+04	96.0203	4
59	450	3.46E+05	96.0207	4

60	1355	1.28E+04	192.0636	38
61	1120	1.07E+04	192.0643	38
62	1578	1.75E+04	194.1050	26
63	1402	2.99E+04	198.0308	23
64	1242	1.27E+04	99.1049	1
65	960	7.06E+04	100.0161	16
66	389	1.87E+05	100.0632	3
67	539	3.61E+04	101.0474	29
68	1202	1.15E+04	101.0481	29
69	746	6.27E+03	102.0319	52
70	1271	2.00E+04	205.0960	32
71	1388	5.67E+03	205.0972	10
72	554	1.03E+05	103.0624	45
73	246	2.53E+04	103.0624	45
74	357	4.55E+05	103.0991	5
75	1173	7.78E+03	208.0592	49
76	941	2.32E+04	210.0742	24
77	1311	3.04E+04	216.1112	15
78	1584	7.96E+04	222.0526	12
79	783	1.60E+04	112.0158	8
80	697	3.00E+04	112.0269	9
81	1377	8.13E+03	226.0476	13
82	1076	3.01E+04	114.0315	28
83	937	1.23E+04	114.0322	28
84	441	3.97E+05	114.0423	12
85	1222	5.47E+03	228.1237	2
86	1041	1.17E+04	229.1434	3
87	1042	5.09E+03	230.1404	5
88	1619	1.21E+05	236.1196	3
89	1673	7.56E+03	237.1272	2
90	1560	2.07E+04	239.0794	17
91	1570	5.80E+03	239.1169	5

92	898	2.68E+04	240.0853	29
93	1517	1.86E+04	120.0572	13
94	1568	9.07E+03	241.0955	16
95	1666	6.36E+03	241.0951	16
96	1659	9.71E+03	241.0957	16
97	1087	6.45E+03	243.1588	2
98	1220	9.10E+03	123.0314	10
99	1416	9.17E+03	247.1055	24
100	1538	5.91E+03	247.1428	20
101	899	6.28E+03	248.1375	16
102	687	9.69E+03	250.1312	10
103	867	1.32E+04	250.1315	10
104	137	1.25E+04	125.0700	1
105	802	1.39E+04	250.1515	2
106	810	1.79E+04	250.1528	2
107	994	1.11E+04	251.1004	39
108	856	1.21E+04	126.0317	15
109	1651	5.22E+03	126.0315	15
110	1217	5.85E+03	126.0329	15
111	1213	1.27E+04	252.1217	9
112	757	3.20E+04	252.1468	6
113	1212	5.17E+03	253.1221	3
114	1587	3.59E+04	253.1317	9
115	755	1.33E+04	253.1482	1
116	1669	5.60E+03	127.0388	6
117	468	7.53E+03	128.0106	8
118	761	1.82E+04	128.0469	32
119	674	2.86E+04	128.0469	32
120	672	7.35E+04	128.0944	8
121	1237	1.00E+04	258.1217	8
122	1345	9.45E+04	259.1054	14
123	1282	3.66E+04	259.1054	14

124	1424	6.24E+04	259.1059	14
125	902	1.01E+04	130.0269	40
126	1474	3.12E+04	261.1204	20
127	1043	3.21E+04	263.1009	44
128	1091	2.19E+04	263.1008	44
129	512	2.68E+04	132.0411	65
130	455	9.19E+03	132.0414	65
131	658	1.62E+04	132.0412	65
132	188	5.07E+06	132.0542	12
133	1171	1.05E+04	265.1068	7
134	1599	6.00E+03	266.0425	1
135	1522	5.63E+03	269.1251	10
136	736	1.44E+04	271.1634	1
137	678	9.79E+03	271.1638	1
138	931	6.71E+03	273.1333	6
139	1489	6.32E+03	275.1355	23
140	1489	6.98E+03	275.1374	23
141	1537	9.01E+03	275.1369	23
142	999	1.85E+04	276.1312	17
143	1181	2.45E+04	277.1164	26
144	1185	4.66E+04	277.1168	26
145	1072	7.12E+03	278.1172	3
146	1612	8.02E+04	283.1202	1
147	568	5.76E+04	142.0259	6
148	819	8.34E+03	142.0261	6
149	1609	2.57E+04	142.0260	6
150	1058	6.92E+04	142.0264	6
151	1604	6.25E+03	284.1456	1
152	1285	5.53E+03	285.0673	17
153	1511	1.38E+05	143.0343	5
154	609	8.96E+03	143.0577	25
155	764	6.12E+03	143.0942	19

156	589	1.14E+04	144.0413	43
157	758	2.84E+04	144.0417	43
158	1398	1.95E+04	144.0421	43
159	1242	5.45E+03	288.0852	9
160	651	1.53E+04	144.0424	43
161	1487	1.59E+04	290.0791	18
162	407	1.49E+04	145.0727	54
163	808	4.64E+04	145.0731	54
164	650	1.30E+04	290.1466	14
165	694	5.14E+03	294.1074	19
166	674	7.84E+03	294.1416	11
167	608	3.00E+04	294.1413	11
168	1283	2.07E+04	297.1209	21
169	1266	1.24E+04	299.1371	12
170	588	6.84E+03	301.1278	14
171	1321	1.74E+04	306.0743	6
172	1111	5.16E+03	306.0934	14
173	1567	2.65E+04	154.0185	1
174	1678	7.35E+03	308.1419	4
175	668	6.55E+04	308.1568	12
176	648	2.62E+04	308.1566	7
177	649	2.62E+04	308.1582	12
178	1438	1.97E+04	311.1365	14
179	1608	7.34E+04	313.1311	8
180	1238	1.21E+04	157.0858	12
181	688	9.44E+03	158.0208	17
182	419	4.78E+04	158.1043	6
183	739	9.54E+03	317.1215	13
184	790	1.04E+04	317.1224	13
185	1621	1.57E+05	317.1461	8
186	1571	9.12E+04	317.1465	8
187	1324	1.15E+04	159.0291	3

188	580	1.58E+04	318.1532	5
189	1603	9.04E+04	318.1533	5
190	1507	1.48E+04	318.1537	5
191	1466	5.40E+03	318.1539	5
192	541	1.67E+04	318.1541	5
193	1570	1.69E+05	318.1546	5
194	773	1.30E+04	319.1370	6
195	138	3.82E+04	160.0360	39
196	753	2.94E+04	160.0365	39
197	827	1.41E+04	320.0977	5
198	1500	3.26E+04	320.1258	5
199	1498	6.29E+03	160.0636	2
200	1637	5.53E+03	162.0326	14
201	931	1.35E+04	324.0891	4
202	933	1.41E+04	324.0908	4
203	1017	8.71E+03	324.0933	3
204	896	6.66E+03	324.0964	13
205	548	4.08E+04	162.0513	76
206	477	3.68E+04	162.0516	76
207	583	8.34E+04	162.0517	76
208	1573	6.42E+03	162.0521	77
209	888	1.01E+04	162.0517	76
210	1027	5.71E+03	162.0519	76
211	1097	1.14E+04	162.0523	77
212	925	9.66E+03	162.0524	77
213	939	1.23E+04	162.0534	81
214	975	1.21E+04	324.1073	33
215	847	6.68E+03	162.0547	5
216	845	5.43E+03	324.1111	3
217	1116	8.17E+03	326.1014	19
218	1428	1.89E+04	327.1317	12
219	1429	9.60E+03	328.1311	9

220	1663	7.10E+03	329.1460	6
221	510	1.32E+04	166.1095	5
222	1288	1.27E+04	336.0838	6
223	1452	2.07E+04	336.0838	6
224	1444	1.75E+04	336.0855	6
225	1288	1.18E+04	336.0863	5
226	1671	5.26E+03	336.1684	5
227	1549	1.08E+04	337.0624	3
228	1452	6.94E+03	338.0875	20
229	590	5.13E+03	170.1043	7
230	957	9.10E+03	342.0954	25
231	956	7.42E+03	171.0477	3
232	1018	9.23E+03	342.0955	25
233	952	1.20E+04	342.0978	27
234	877	7.73E+03	342.1007	9
235	795	1.66E+04	342.1119	7
236	443	8.92E+03	342.1147	46
237	545	2.95E+04	342.1148	46
238	599	1.01E+04	342.1147	46
239	775	4.43E+04	342.1159	50
240	673	9.93E+04	342.1158	50
241	711	9.67E+04	342.1159	50
242	735	1.68E+04	342.1278	2
243	838	7.58E+03	171.0892	30
244	733	9.45E+03	343.1306	8
245	544	5.92E+03	344.1192	21
246	1608	1.84E+04	172.0835	9
247	1391	8.95E+03	173.0906	1
248	1269	2.38E+04	173.0906	1
249	1201	1.25E+04	173.0908	1
250	1391	6.11E+03	347.1841	2
251	1531	7.80E+03	175.0630	13

252	1476	5.22E+03	350.1368	11
253	1247	7.55E+03	354.0938	2
254	1287	3.36E+05	354.0952	1
255	1188	1.94E+05	354.0954	1
256	550	9.04E+03	178.0468	56
257	595	3.51E+04	178.0468	56
258	696	2.87E+04	178.0471	56
259	879	6.83E+03	180.0193	5
260	532	1.14E+04	180.0415	39
261	537	8.28E+03	180.0428	39
262	202	3.86E+05	180.0623	77
263	768	5.65E+03	180.0627	77
264	198	6.90E+05	180.0624	77
265	144	1.45E+06	180.0625	77
266	356	5.03E+06	180.0631	87
267	1099	1.21E+04	180.0786	26
268	1545	1.08E+04	367.1813	1
269	158	1.64E+04	185.0678	21
270	558	1.22E+04	186.0996	15
271	589	6.52E+03	373.1794	1
272	1564	1.03E+04	375.1532	1
273	1041	4.46E+04	188.0318	15
274	1005	3.75E+04	188.0322	15
275	1539	7.28E+03	376.1604	5
276	1514	5.45E+03	380.1657	7
277	444	5.70E+03	192.0082	3
278	1225	9.85E+04	192.0421	14
279	407	6.28E+04	192.0620	33
280	325	5.53E+03	192.0626	36
281	371	1.14E+04	192.0628	36
282	1520	5.56E+03	192.0799	19
283	283	1.27E+04	193.0938	15

284	1129	5.99E+05	194.0579	39
285	302	1.49E+04	196.0577	47
286	1075	1.37E+05	196.0735	25
287	516	7.26E+03	199.0837	21
288	1320	2.47E+04	398.1728	17
289	1368	2.42E+04	199.0865	1
290	1410	6.41E+03	398.1733	15
291	1505	1.35E+04	199.0867	1
292	1493	1.21E+04	398.1734	15
293	1440	1.16E+04	398.1744	16
294	1319	2.91E+04	398.1750	19
295	1368	1.89E+04	399.1793	10
296	1366	1.40E+04	399.1807	1
297	1310	4.85E+04	399.1806	1
298	502	1.13E+05	200.1151	11
299	718	1.20E+04	200.1161	11
300	395	1.11E+06	202.0944	19
301	849	5.81E+03	203.0793	36
302	464	3.73E+04	205.0931	22
303	266	1.22E+04	205.0930	22
304	400	2.63E+04	205.0936	22
305	266	1.29E+04	205.0946	31
306	567	5.87E+03	206.0426	21
307	645	1.12E+04	206.0563	8
308	586	5.13E+03	206.0590	8
309	681	6.17E+03	412.1840	3
310	796	1.84E+05	412.1844	3
311	795	8.12E+04	206.0921	22
312	867	6.24E+04	412.1851	3
313	866	8.65E+04	206.0926	20
314	509	8.12E+03	206.1013	2
315	505	8.16E+03	206.1003	1

316	834	1.70E+04	206.1417	4
317	271	1.33E+04	207.0729	25
318	831	6.96E+03	207.0872	1
319	758	1.61E+04	414.1997	5
320	758	1.29E+04	207.1003	10
321	1269	6.01E+03	208.0733	24
322	391	1.15E+04	208.0955	26
323	509	3.00E+04	208.0959	26
324	344	5.64E+03	208.0962	26
325	192	3.02E+04	210.0725	9
326	211	3.08E+04	210.0729	10
327	493	7.54E+04	211.0943	29
328	526	8.37E+04	211.0946	29
329	1326	7.62E+03	211.0971	30
330	1008	5.50E+04	211.0968	29
331	883	1.48E+05	211.0967	29
332	851	1.48E+04	211.0970	30
333	798	8.08E+03	211.0973	30
334	948	1.09E+05	211.0970	30
335	904	1.73E+04	211.0979	1
336	1288	3.66E+04	213.1005	11
337	1220	4.03E+04	213.1005	11
338	1349	6.85E+03	213.1008	11
339	1286	4.08E+04	427.2052	6
340	1053	6.21E+03	216.0906	12
341	527	5.47E+03	216.1099	15
342	441	2.04E+05	216.1101	15
343	1401	8.45E+03	434.1830	1
344	1590	9.65E+03	220.1826	3
345	296	1.02E+04	221.0890	31
346	178	1.69E+05	223.1043	11
347	766	7.54E+03	224.0136	3

348	514	2.51E+04	225.1097	28
349	557	6.15E+03	225.1097	28
350	701	8.25E+03	226.1316	3
351	465	6.55E+03	456.1944	6
352	735	6.85E+03	229.0941	16
353	356	5.50E+05	229.1050	5
354	224	2.16E+04	233.0888	10
355	366	7.50E+03	233.0889	10
356	1628	9.99E+03	234.1984	2
357	390	4.25E+04	235.1012	4
358	963	1.22E+04	236.0673	10
359	203	1.69E+05	237.0837	27
360	1156	2.19E+04	476.2368	14
361	1672	5.30E+03	239.0453	34
362	1339	8.21E+03	479.2004	5
363	1275	7.48E+03	479.2005	5
364	1205	1.04E+04	479.2006	5
365	1218	5.20E+03	480.2036	7
366	649	1.40E+04	241.0944	16
367	561	6.78E+03	241.1301	2
368	359	4.06E+05	243.1208	8
369	1030	7.24E+03	488.1517	6
370	939	3.06E+04	488.1536	10
371	1032	6.30E+03	488.1550	10
372	1143	5.45E+03	244.0776	1
373	432	3.25E+04	244.1048	12
374	938	8.52E+03	490.1566	12
375	168	1.44E+04	245.0887	7
376	557	5.97E+03	245.0897	10
377	506	1.44E+04	246.0187	12
378	390	1.71E+04	247.1034	13
379	309	4.39E+04	247.1039	13

380	263	1.20E+05	247.1044	12
381	343	1.15E+04	247.1050	24
382	244	5.82E+05	247.1046	24
383	381	2.15E+04	249.1192	20
384	941	9.08E+03	250.1161	11
385	1213	9.06E+03	252.1215	9
386	1144	5.86E+03	508.2287	11
387	948	1.12E+04	254.1175	6
388	1080	6.18E+03	254.1178	6
389	757	8.96E+03	255.1195	5
390	962	1.68E+04	258.0557	5
391	966	1.99E+04	516.1120	4
392	974	2.07E+04	258.0562	1
393	683	5.42E+03	516.1462	5
394	827	6.15E+03	258.0731	3
395	752	5.21E+03	516.1466	5
396	784	1.28E+04	516.1468	5
397	786	1.08E+04	258.0741	3
398	945	6.65E+03	258.0737	3
399	826	5.79E+03	516.1472	5
400	685	6.36E+03	258.0740	3
401	639	5.06E+03	258.0743	3
402	751	5.54E+03	258.0746	6
403	874	5.34E+03	516.1499	7
404	304	1.53E+05	258.1203	8
405	938	9.30E+03	259.0814	24
406	687	5.36E+03	259.1037	11
407	434	1.17E+04	259.1037	11
408	500	7.18E+04	259.1046	14
409	585	1.10E+04	259.1049	14
410	550	1.98E+04	259.1059	14
411	1048	1.37E+04	260.1173	10

412	608	1.58E+04	261.1169	2
413	671	1.56E+04	261.1200	20
414	805	2.37E+04	261.1208	20
415	674	1.50E+04	263.0896	6
416	251	1.14E+04	263.0988	23
417	203	3.11E+04	263.0989	23
418	321	1.04E+04	263.0989	23
419	297	1.03E+05	263.0992	23
420	365	6.87E+04	263.0999	44
421	408	2.65E+04	263.1003	44
422	993	4.00E+04	266.1203	2
423	761	3.91E+04	267.0559	8
424	954	2.08E+04	34.0052	8
425	411	1.38E+04	267.0731	5
426	318	1.64E+04	271.1041	4
427	806	6.45E+03	271.1180	12
428	288	1.32E+05	271.1631	1
429	601	1.85E+04	272.0886	7
430	398	6.00E+03	272.1361	8
431	359	3.30E+04	272.1473	1
432	506	1.08E+04	273.0838	7
433	389	4.58E+04	273.0836	7
434	930	1.63E+04	548.2597	2
435	637	8.16E+03	275.1356	23
436	942	1.46E+04	275.1369	23
437	495	5.59E+03	276.0299	2
438	1172	1.24E+04	277.0945	7
439	561	2.33E+04	278.1127	23
440	882	6.70E+03	283.0856	8
441	436	1.64E+04	284.0516	6
442	966	5.08E+03	289.0969	4
443	517	3.47E+04	293.1462	26

444	225	5.69E+03	294.1047	8
445	126	1.45E+04	294.1405	4
446	171	3.69E+04	295.0891	35
447	132	1.04E+04	295.0896	35
448	905	1.81E+04	297.8006	1
449	885	1.10E+04	298.1368	8
450	382	1.73E+04	299.0973	1
451	501	8.30E+03	300.0832	20
452	654	1.06E+04	300.1204	14
453	163	2.83E+04	301.1254	13
454	596	8.76E+03	302.1002	7
455	926	9.25E+03	304.1197	4
456	866	1.58E+04	304.1198	4
457	732	5.05E+03	305.1102	14
458	446	4.58E+04	306.0942	14
459	197	1.20E+04	307.1246	20
460	647	5.36E+03	308.0894	9
461	222	5.34E+03	308.1186	1
462	171	1.06E+04	308.1200	8
463	141	2.31E+04	308.1570	12
464	451	9.68E+03	311.1348	12
465	349	3.54E+05	317.1213	13
466	126	3.88E+04	318.1517	5
467	771	5.61E+04	319.1317	1
468	848	1.28E+05	319.1318	1
469	908	1.25E+04	319.1321	1
470	185	1.06E+04	319.1371	6
471	370	1.29E+04	320.0730	6
472	582	6.56E+03	320.0820	7
473	858	1.59E+04	320.0888	10
474	690	1.70E+04	320.1362	6
475	853	1.66E+04	323.1197	10

476	910	7.52E+03	323.1211	23
477	365	1.20E+04	324.1039	31
478	223	2.91E+04	324.1041	31
479	324	1.51E+04	324.1041	31
480	354	1.42E+04	324.1045	31
481	409	5.89E+03	324.1041	31
482	839	2.19E+04	324.1266	3
483	441	5.68E+03	325.1155	10
484	506	2.24E+04	326.0993	19
485	575	1.66E+04	326.0989	19
486	773	6.84E+04	326.1002	19
487	732	1.84E+04	326.1010	19
488	228	5.83E+03	328.1238	14
489	548	6.74E+03	330.0940	17
490	490	1.07E+04	330.0948	17
491	132	6.06E+03	330.1025	7
492	1601	7.97E+03	100.0157	16
493	1523	5.66E+03	337.0621	3
494	798	6.96E+03	337.1432	5
495	147	1.04E+05	337.1472	7
496	164	1.34E+04	337.1478	7
497	202	9.38E+03	338.0837	16
498	282	2.78E+04	338.0841	16
499	615	2.30E+04	339.1571	15
500	441	1.20E+05	342.0943	21
501	532	9.74E+04	342.0944	21
502	186	2.61E+04	342.1141	40
503	317	5.00E+04	342.1148	46
504	283	2.63E+04	342.1146	46
505	126	8.96E+04	342.1148	46
506	296	4.36E+04	342.1151	48
507	163	2.47E+04	342.1150	48

508	212	8.69E+04	342.1154	48
509	1569	1.00E+05	110.0483	8
510	426	4.07E+04	343.1255	19
511	732	9.31E+03	343.1450	6
512	1619	2.76E+04	111.0316	8
513	509	1.56E+04	344.1091	10
514	822	5.07E+03	345.0825	16
515	858	9.24E+03	345.0837	16
516	239	7.95E+03	346.1314	3
517	168	2.01E+04	346.1356	4
518	1637	2.33E+04	114.0314	28
519	273	1.92E+04	348.1521	11
520	715	1.16E+04	348.1519	11
521	403	1.11E+04	354.0795	12
522	381	5.10E+03	354.0804	12
523	740	1.20E+04	356.1100	9
524	789	6.50E+03	356.1113	10
525	1586	3.02E+04	125.0479	15
526	708	7.77E+03	358.1261	9
527	296	8.84E+03	362.1666	3
528	1541	7.34E+03	362.2093	43
529	132	2.64E+04	364.0969	3
530	235	2.12E+04	364.0969	3
531	199	6.43E+03	364.0968	3
532	163	1.42E+04	364.1470	14
533	406	2.34E+04	366.1415	11
534	816	2.36E+04	373.1489	14
535	768	4.99E+04	373.1495	14
536	790	2.42E+04	374.1559	6
537	468	7.31E+03	376.1817	2
538	578	7.02E+03	378.1048	13
539	829	8.11E+03	388.1578	8

540	804	6.08E+03	391.1228	22
541	1410	7.89E+03	160.0371	39
542	581	1.53E+04	393.1106	2
543	513	1.06E+04	404.1315	14
544	656	1.02E+04	406.1299	7
545	726	1.82E+04	180.0630	87
546	518	1.12E+04	434.1415	5
547	285	6.14E+03	454.2058	8
548	664	1.12E+04	458.1414	3
549	505	6.76E+03	462.1374	8
550	593	1.42E+04	462.1727	1
551	441	8.98E+03	467.1286	4
552	451	5.85E+03	468.1260	4
553	1526	1.35E+04	236.1204	1
554	733	5.84E+03	470.1415	5
555	1467	9.02E+03	237.1289	1
556	1577	7.56E+03	239.0784	17
557	524	2.27E+04	474.1370	4
558	145	9.44E+03	481.1776	5
559	594	6.60E+04	488.1521	6
560	581	1.58E+05	488.1523	6
561	647	1.82E+05	488.1528	6
562	1568	5.39E+03	263.0760	2
563	472	8.87E+03	504.1451	7
564	433	6.80E+03	504.1459	9
565	449	1.74E+04	506.1618	10
566	404	7.24E+03	506.1619	10
567	422	6.34E+03	515.2174	3
568	544	2.45E+04	518.1615	9
569	638	1.77E+04	518.1624	9
570	580	5.58E+04	518.1616	9
571	662	5.31E+04	518.1622	9

572	1582	8.76E+03	306.0788	2
573	1361	5.50E+03	314.0797	11
574	652	1.66E+04	548.1760	8
575	1616	4.40E+04	317.1467	8

Table E.3 List of the results from the volcano plot showing the putative metabolites identified based on accurate mass match with the HMDB and METLIN metabolites

	Retention time (sec)	Fold change	p value	Intensity	Accurate mass	Name	HMDB ID (METLIN ID)
1	140	2.66	0.039	1.05E+06	169.0837	3-Methylhistidine	HMDB00479
						1-Methylhistidine	HMDB00001
						Nalpha-Methylhistidine	(65951)
2	159	4.48	0.015	3.07E+04	240.1203	Homocarnosine	HMDB00745
						Anserine	HMDB00194
						Balenine	HMDB05769
						N(alpha)-gamma-L-Glutamylhistamine	(66121)
3	440	2.54	0.029	6.80E+05	124.0624		
4	749	2.14	0.049	1.33E+04	133.0186		
5	1178	2.46	0.015	1.16E+04	102.0777		
6	1456	2.98	0.045	3.44E+04	151.0621	Acetaminophen	HMDB01859
						2-Phenylglycine	HMDB02210
						Dopamine quinone	HMDB12219
						Leukoaminochrome	HMDB12992
						Methyl 2-aminobenzoate	(86434)
						2-Amino-3-methylbenzoate	(71241)
						N(alpha)-gamma-L-Glutamylhistamine	(66121)