University of Alberta

Development of Isotope Labeling Methods and Data Processing Program for Liquid Chromatography Mass Spectrometry-Based Metabolome Analysis

by

Ruokun Zhou

A thesis submitted to the Faculty of Graduate Studies and Research

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry

©Ruokun Zhou

Spring 2014

Edmonton, Alberta

Abstract

Metabolomics strives to gain a comprehensive picture of the metabolites in a biological system. This emerging research field gives us an insight into the discovery of disease-related biomarkers and an understanding of metabolic flux as well as mechanisms. Liquid chromatography mass spectrometry (LC-MS) is one of the leading techniques used in metabolomics owing to its high sensitivity and capability for structure elucidation.

The main objective of my research was to develop improved LC-MS based differential isotopic labeling methods to quantify and identify metabolites in biological samples with enhanced analytical performance. First of all, an improved dansylation labeling protocol was developed for profiling amine- and phenol-containing metabolites, which can be used to more efficiently label multiple samples in metabolome analysis. Secondly, a new triplex DIL reagent: 5-diethylamino-naphthalene-1-sulfonyl chloride (DensCl), was developed to increase the sample throughput by at least two-fold, compared to the duplex dansylation labeling approach. Thirdly, a new type of isotope labeling reagent was synthesized to provide a means of chemical structure analysis based on the MS/MS fragmentation patterns generated from the labeled metabolites. Finally, a data processing software, IsoMS, was developed to quickly extract quantitative information from LC-MS data generated by the isotope labeling method.

The significance of my research work is that the methods developed can be used to analyze the metabolome with high performance, thereby increasing the

application capability of metabolomics for disease biomarker discovery and biological functional studies of metabolites in systems biology.

Acknowledgements

First of all, I would like to express the deepest appreciation to my supervisor, Professor Liang Li, for his invaluable support to my Ph.D study and research, and also for his unsurpassed knowledge, patience, and motivation throughout my graduate studies. Indeed, without his guidance, I would not be able to finish the dissertation.

I also would like to thank my dissertation committee members, Professor Derrick L. J. Clive, Professor Robert E. Campbell, Professor D. Jed Harrison, Professor Sambasivarao Damaraju, and Professor Helene Perreault for insightful comments on my thesis work. My sincere gratitude also goes to Professor Todd L. Lowary, who was my supervisor in my Master of Science study, for his generous help and great patience in the past.

I am most grateful to Dr. Kevin (Kun) Guo for his training and collaborative work. I would like to thank Chiao-Li Tseng, Dr. Xianmin Hu for their extraordinary hard work and wonderful contributions to the data processing program. I would like to acknowledge Dr. Randy Whittal, Jing Zheng and Bela Reiz from the Mass Spectrometry Laboratory for their support and excellent advices. I would like to offer my gratitude to my group members for their personal help and support to my work.

Finally, I appreciate my wife Hongmei Shang and my parents very much for their support and love over the years. The especial gratitude is delivered to my son, Ethan, for everything that he brought to me.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACN	Acetonitrile
ATP	Adenosine triphosphate
BALF	Bronchoalveolar lavage fluid
BPC	Base peak chromatogram
CID	Collision-induced dissociation
CSF	Cerebrospinal fluid
DmPA	p-dimethylaminophenacyl bromide
DnCl	Dansyl chloride
EI	Electron impact ionization
ESI	Electrospray ionization
nM	Nano molarity
nm	Nano meter
FTICR	Fourier-transform ion cyclotron resonance
FWHM	Full width at half maximum
GC	Gas chromatography
GC-MS	Gas chromatography mass spectrometry
h	Hour
HFBA	heptafluorobutyric acid
HILIC	Hydrophilic interaction chromatography
HMDB	Human metabolome database
HPLC	High performance liquid chromatography
IP	Ion pairing
LC-MS	Liquid chromatography mass spectrometry
LC	Liquid chromatography
LLE	Liquid liquid extraction
m	milli- (10^{-3})
m/z	mass to charge
MeOH	Methanol
Min.	Minute

MRM	Multiple reaction monitoring
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MW	Molecular weight
PLS-DA	Partial least square-discriminant analysis
OPLS-DA	Orthogonal projections to latent structures-discriminant analysis
PCA	Principal component analysis
NMR	Nuclear magnetic resonance
PBr	Phenacyl bromide
ppm	part(s) per million
QC	Quality control
QTOF	Quadrupole time-of-flight
RF	Radio frequency
RP	Reversed-phase
RSD	Relative standard derivation
S/N	Signal to noise ratio
SPE	Solid phase extraction
SIL	stable-isotope-labeled
TEA	Triethylamine
TCA	Tricarboxylic acid
TLC	Thin layer chromatography
TOF	Time-of-flight
UV	Ultra- violet
VIP	Variable importance on the projection
μM	Micro molarity

Chapter I: Introduction

1.1   Metabolomics

Metabolomics represents a promising method to assess overall metabolites present in biological samples. Liquid chromatography mass spectrometry (LC-MS) is one of the leading techniques upon which metabolomics is built. The objective of my research is to develop novel and improved methods to study the metabolome of biological systems based on the LC-MS platform. This chapter introduces the background and instrument principles involved in my thesis research.

The word metabolome was first used by Olivier in 1998 to describe the entirety of metabolites present in cells of a particular physiological or developmental state.[1] Metabolomics was coined by Fiehn to emphasize comprehensive identification and quantification of all small molecules within cells, tissues, body fluids or organisms.[2-5] Owing to the complexities of the metabolome, it is virtually an impossible task to completely measure every metabolite in all biological samples, so most metabolomics studies have focused on qualification and quantification of metabolites belonging to a biological pathway or a certain class of compounds. This approach is known as "metabolic profiling".[6,7] In metabolic target analysis, analytical methods are established to identify and quantify selected metabolites.[8,9] Metabolic fingerprinting is an approach often used in clinical diagnosis to quickly differentiate diseased samples

from the control according to patterns of global metabolites instead of quantification of individual metabolite levels.[10-12]

1.2   Why We Need Metabolomics?

A thorough understanding of systems biology can provide valuable knowledge about human health evaluation, nutritional supplement, drug safety assessment, disease biomarker discovery, and more.[13] Several parallel functional genomics technologies, such as genomics, transcriptomics, proteomics, and metabolomics have been used to investigate the mechanistic correlation between genes and functional phenotypes of biological systems. As the latest episode of the "omics" cascade (see Figure 1-1),[6] the concentrations of metabolites reflect the ultimate response of a biological system to genetic or environmental changes, so metabolomics acts as a "glass window" that allows for observing and determining the relationship between them.[14] Furthermore, small changes occurring in the upstream of the "omics" cascade can be amplified in the metabolome which is the main driving force that people use to discover disease-related biomarkers.[15,16] As a "data-driven" strategy, metabolomics evaluates a given system as a whole and enables people to infer hypotheses based on the acquired data,[17] so the strategy is complementary to traditional metabolic studies. Whereas accuracy of a hypothesis depends on the quantity and quality of data, metabolomics needs powerful data mining capability to dig and interpret useful information.[18-20] To this end, metabolomics has displayed its utilization in the investigation of metabolic bio-functions in various metabolomes, even though the discipline is still at an early stage of development.

The "Omics" Cascade

What can happen → **GENOME**

What appears to be happening → **TRANSCRIPTOME**

What makes it happen → **PROTEOME**

What has happened and is happening → **METABOLOME**

PHENOTYPE
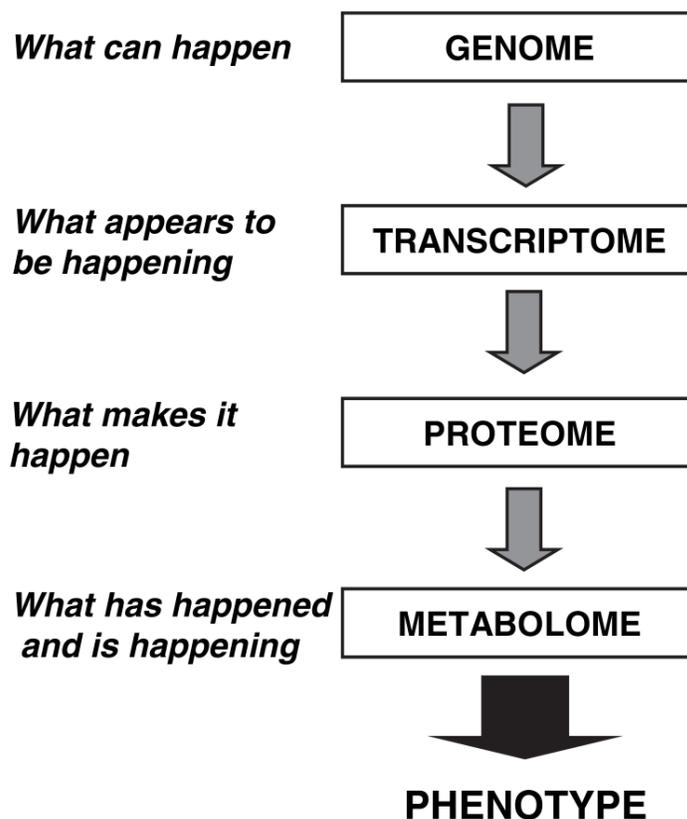
Figure 1-1.   The components of the "omics" cascade. Reproduction with permission.[6]

Metabolites have very diverse structures as well as physic-chemical properties. For instance, the human metabolome contains more than 8,000 metabolites.[21-23] In addition, the dynamic range of metabolites in a biological system is very wide (~9 orders of magnitudes in the human metabolome).[5,24] These characteristics make the global metabolomics difficult to achieve based on already known techniques. Nowadays, nuclear magnetic resonance (NMR) spectroscopy, gas chromatography mass spectrometry (GC-MS) and liquid chromatography mass spectrometry (LC-MS) are considered the most competitive

approaches. These techniques have their own strengths and shortcomings and a brief discussion is given below.

1.3   Analytical Techniques within Metabolomics

1.3.1   NMR Based Metabolomics

Nuclear magnetic resonance spectroscopy is based on the physical phenomenon that nuclei in a molecule can absorb and re-emit radio waves if an external magnetic field is applied. The frequency of absorbed radio waves indicates the chemical environment of atoms in the molecule so that the molecular structure can be deciphered. Since 1970s, NMR has been one of the most popular techniques in identification and quantification of small molecules.[25-27] NMR can simultaneously qualify and quantify medium or high level metabolites in a sample without bias,[28,29] and NMR based metabolomics often requires minimal sample preparation. It is a non-destructive technique, so valuable samples can be recycled. Since NMR spectra are in principle independent of instruments, the data are highly reproducible as long as the same sample preparation is followed. These advantages greatly facilitate the comparison and validation of NMR data inter-laboratory. However, the sensitivity of NMR is not comparable to mass spectrometry, so it is commonly used to detect metabolites in medium or high concentrations. In addition, deconvolution and interpretation of NMR spectra from a complex biological sample are still a challenging task due to a lack of separation prior to analysis.

1.3.2 GC-MS Based Metabolomics

Whereas mass spectrometry offers far better sensitivity than NMR spectroscopy, its performance is heavily restricted by ion suppression due to co-eluted analytes in complicated biological samples. A chromatographic separation prior to mass spectrometry can effectively address this issue. Nowadays, gas chromatography hyphenated to mass spectrometry is the primary technique to separate and analyze volatile compounds. In gas chromatography, the mobile phase is a carrier gas, such as He or H_2, and the stationary phase is a layer of liquid or polymer coated on a solid support (PTGC) or the wall of an open tube (WCOT). The retention of analytes in a GC column is based on the interaction between molecules and the stationary phase, and column temperature. Electron ionization (EI) and chemical ionization (CI) are two of the most popular ionization interfaces used for GC-MS. The EI-MS spectrum provides unique reliable robustness and reproducibility over other ion sources. To this end, several GC-MS spectra libraries have been established and enable users to identify unknowns by comparing their EI-MS spectra to the databases.[30,31] Contrary to EI, CI is regarded as a soft ionization technique, so the molecular ion in CI spectra is often recognized. Unfortunately, sample preparation for GC-MS is laborious and time-consuming in that chemical derivatization is commonly necessary to render volatility and thermo-stability to metabolites of interest. As biological samples have to be dried down completely before derivatization, some volatile metabolites might be lost in the process, which could result in large errors in the quantification of these species.

1.3.3 LC-MS Based Metabolomics

1.3.3.1 Principle of Reversed Phase Liquid Chromatography (RPLC)

Although liquid chromatography (LC) was introduced in 1903, it did not become a practical separation technique until the theory of modern separation science was established on gas chromatography. Since 1960s, LC has grown into a large family that consists of normal phase (NP), reverse phase (RP), ion exchange (IEC), etc. Unlike GC in which analytes only interact with the stationary phase, in LC, both the stationary phase and the mobile phase have interactions with analytes. Consequently, LC offers far greater versatility than GC due to a large range of stationary phases and mobile phases. Reverse phase liquid chromatography is the most important separation means used in LC-MS based metabolomics in that it provides premium resolution to most of medium polar and small non-polar molecules. Unless otherwise noted, the following discussion regarding columns would be limited to RPLC columns.

The stationary phase of reverse phase column is manufactured by chemical bonding non-polar C18 or shorter alkyl chains (e.g. C8, C4) on the surface of silica. The retention mechanism of molecules in RPLC is not completely understood, although it is often regarded as the result of partition and/or adsorption.[32,33] A reverse phase column is commonly packed by C-18 or C-8 narrow bore stationary phase (<30 nm) with particle size of 3-5 μm, which has been the most popular equipment for analysis of medium polar and small non-polar molecules. Recently, sub-2 μm particle columns used in ultra performance

liquid chromatography (UPLC) are rapidly expanding influences in the market. Compared to HPLC, UPLC is able to provide higher sample processing throughput and reduce solvent consumption without compromising separation efficiency. On the downside, high back-pressure (10,000 to 15,000 psi) in the UPLC column generates column heat. Thus, the maximum diameter of sub-2 μm particle columns has to be restricted to 2.1 mm in order to minimize the radial temperature gradient. Sub-2 μm particle columns not only demonstrate outstanding performance in UPLC, but also work well in the traditional HPLC (up to 6000 psi) when running at a moderate flow rate. Mobile phases of RPLC consist of water and a miscible organic solvent, typically acetonitrile or methanol. These aqueous-organic mobile phases are perfectly compatible with most of biological samples as well as hyphenated ESI-MS.

1.3.3.2   Electrospray Ionization

Development of atmospheric pressure ionization (API) interfaces is the most challenging task for successfully coupling liquid chromatography to mass spectrometry. The API interface is responsible for rapid evaporation of the mobile liquid eluted from LC and ionization of sample molecules prior to analysis by mass spectrometry. A good ionization efficiency of analytes in the interface is essential to the sensitivity of mass spectrometry. To this end, several techniques have been developed in the field, including electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), and atmospheric pressure photoionization ionization (APPI). Among these techniques, ESI is the most

widely used in metabolome analysis, especially for medium polar and small polar molecules.

Electrospray is a physical phenomenon that occurs when a liquid flows through a capillary tube under strong electric field and this potential induces liquid drops at the capillary tip to disperse into a fine aerosol. In the late 1980's, the phenomenon was first utilized by Fenn to create ionization sources for mass spectrometry.[34] To date, ESI has been extensively studied, whereas the mechanism on how analytes in droplets are converted into gas phase ions is still in discussion.[35-37] Simply speaking, the process of ESI can be divided into three major stages including: 1) generation of charged droplets at the capillary tip; 2) evolution of charged droplets as a result of solvent evaporation and repeated fusion of charged droplets until the formation of ultimate droplets with high charge density; and 3) generation of gas phase ions via ion evaporation model (IEM) or charge residue model (CRM).[38]

First of all, as shown in Figure 1-2, when a strong potential (3-6 kV) is applied on the capillary tip (+) and counter electrode (-), the electric field (E_c $\approx 10^6$ V/m) induces a charge accumulation on the droplet surface at the capillary tip.[38] The electron repulsion forces the droplet to elongate and form the so-called "Taylor cone" to stabilize the droplet. When the potential reaches an "onset voltage", the repulsion overcomes surface tension of the solvent (Rayleigh limit), and the "Taylor cone" breaks to a mist of highly charged droplets. Under the pressure of the electric filed, the released charged droplets move to the counter-electrode close to the mass spectrometer. Since ESI gives positive charges to the

leaving droplets, negative charges will accumulate in bulk liquid. To enable a steady ESI, an electrochemical reaction occurs in the region close to the capillary tip to balance the negative charges in the solvent.[38] ESI has been known as a concentration dependent source because the droplet current linearly increases with the concentration of analytes, but the current will be saturated when the total concentration of analytes is greater than 10^{-5} M.[38] At a high flow rate (>10 μL/min), a coaxially nebulizing gas is necessary to assist with spray of the charged droplets and to limit the dispersion in space as well.
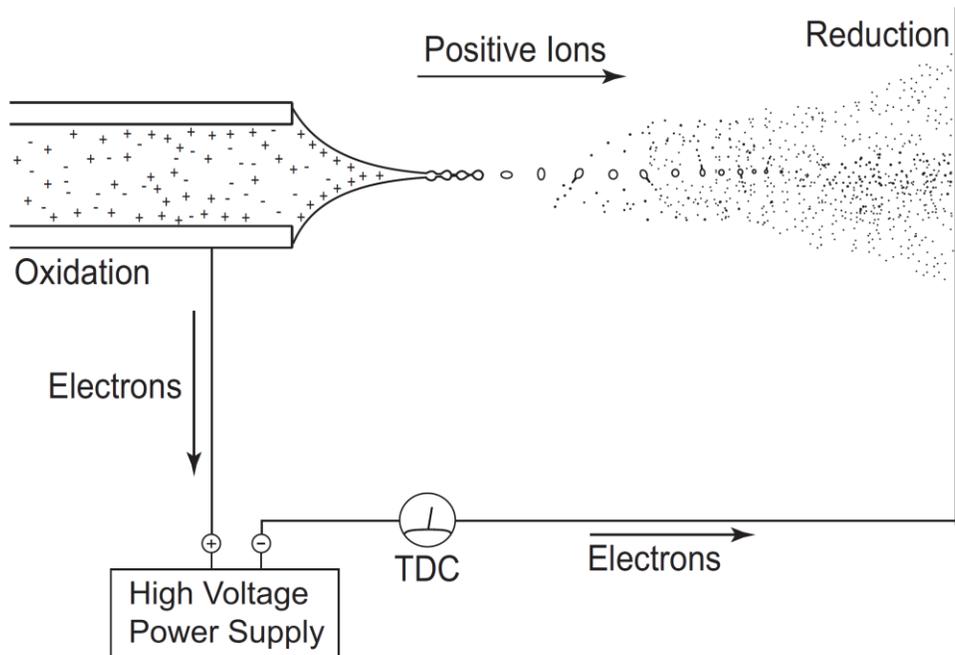
Figure 1-2.   Generation of charged droplets in the positive ESI process. TDC stands for total droplet current. An electrochemical reaction supply positive charges accumulate on leaving droplets.[38] Reproduction with permission

Secondly, when charged droplets drift to the counter-electrode, they encounter the heating gas, which promotes evaporation of the solvent while maintaining the charges constant. Under the influence of an increasing electric field, a droplet will form a new Taylor cone and then releases smaller droplets from the tip if the electric repulsion is higher than the Rayleigh limit. In the ESI processing, the first generation of droplets from the capillary typically has a diameter about 1.5 μm and carries ~50,000 charges. The volume of offspring droplets is ~2% of parent droplets, whereas they inherit ~15% charges.[38] Each parent droplet can release about 20 offspring droplets. The shrink-fusion progress is repeated until the charged droplets ultimately become adequately small to generate gas phase ions. The representative progress of droplet evolution in nano-ESI is demonstrated below (see Figure 1-3).
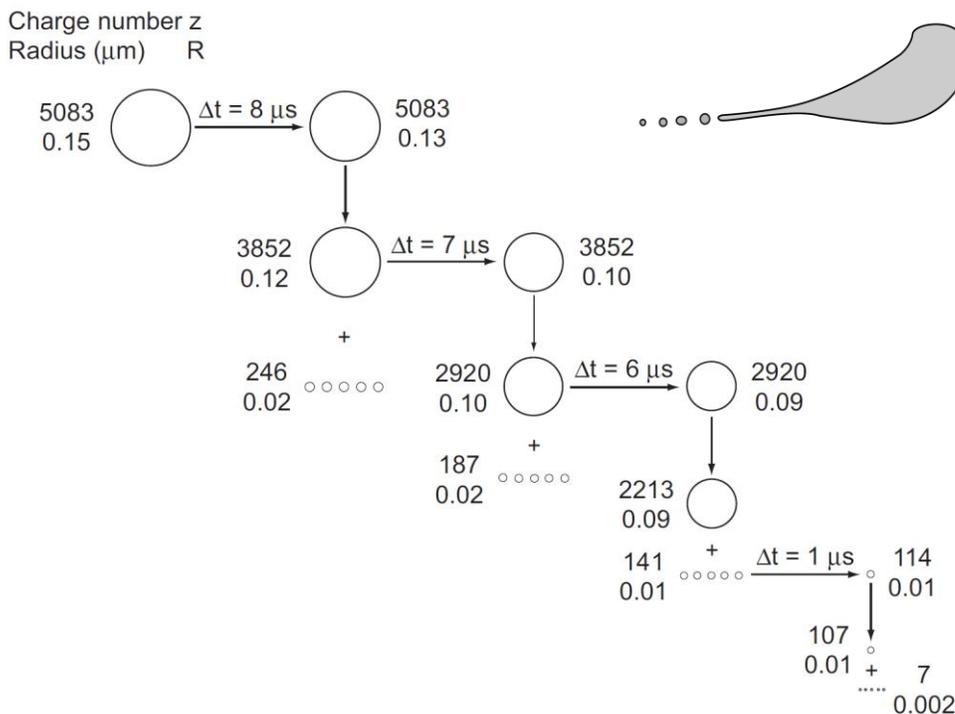
Figure 1-3.   Evolution of charged water droplets in the nano-ESI. Compare to ESI, the size of droplets generated by nano-ESI is smaller.[38] Reproduction with permission.

The last step of ESI is the formation of gas ions. To this end, two mechanisms have been proposed to account for the conversion of analytes from highly charged droplets into gas phase ions. The first mechanism was termed charged residue model (CRM) which was proposed by Dole in 1968.[39] He suggested that the consecutive shrink-evaporation process eventually produces some tiny droplets containing only one analyte molecule and multiple charges on the droplet surface. If the analyte is a large molecule, such as proteins, further evaporation will get rid of the remaining solvent molecules and lead to a gas phase ion. The gas phase ion charges are originated from the charges of the droplet. Typically, the CRM well explains the generation of multiply charged ions of macromolecules in ESI.

The other mechanism termed ion evaporation model (IEM) was proposed by Iribarne and Thomson.[40] This mechanism suggests that analyte ions can be directly desorbed from the highly charged surface of a droplet when it shrinks to a radius less than 10 nm. It also explains that Na^+ and Na^+ hydrates are main ion species produced by a droplet where other ion clusters, e.g. $[(NaCl)_n(Na)_m]^{m+}$ are dominant. In addition, they also developed a theoretical equation to predict the conditions for droplet ionization. In summary, the IEM supports particularly ion formation for small molecules.

1.3.3.3 Principle of FT-ICR-MS

FT-ICR-MS has been an important platform in metabolic analysis and is also a major tool in my research work. As it is shown in Figure 1-4, FT-ICR-MS is fundamentally based on the phenomenon known as ion cyclotron resonance (ICR). When ions perpendicularly enter into an magnetic filed B, they will experience the cyclotron motion with a frequency that is a function of their charge number q, ion mass m and strength of the magnetic field B, but independent of their initial velocity v_o. The cyclotron frequency f_c is given as Eq. 1-1.

$$f_c = \frac{qB}{2\pi m} \qquad (1\text{-}1)$$

We can analyze ions through measurement of the cyclotron frequency f_c that is inversely proportional to the mass to charge ratio. The first ICR-MS was launched in 1960s by Varian. To measure f_c of ions, the instrument applied a sweep RF electric field on trapped ions. If the frequency of the RF field was coincident with f_c of an ion, the radius of ion cyclotron motion would become larger and larger due to the excitation and finally reach the collector to give a signal.[41] The resolution and mass accuracy of the first generation of ICR-MS are dependent on the number of their cyclotron motion. MS/MS is inaccessible in the instrument because ions have been destructed during MS scan. ICR-MS has achieved glorious success since Fourier transform was introduced by Comisarow and Marshall in 1974.[42] Today, FT-ICR-MS is able to provide unsurpassed resolution (up to 1,770,000), the highest mass accuracy (~0.01mDa), excellent sensitivity (~200 ions at S/N=3) and MS^n capability.[43-45] Because the performance

of FT-ICR-MS is tightly related to the magnetic field strength, the instrument performance can be further enhanced by using more powerful superconductive magnets.
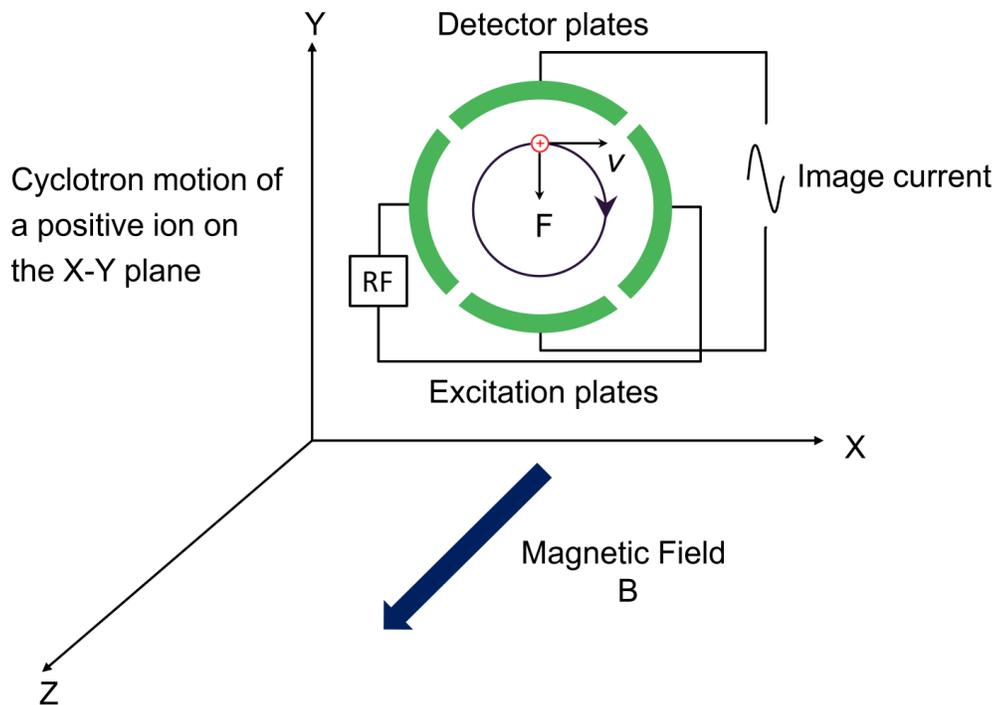
Figure 1-4.   Cyclotron resonance motion of a charged particle caused by the Lorentz force when it perpendicularly enters a magnetic field.

FT-ICR-MS employs a different concept of ICR-MS to detect ion signals. In the simplest case, when an ion rotates in an ICR cell, it induces an image current on detector plates. The image current is amplified and detected as a single RF signal. The cyclotron motion can last for a moment and finally results in a sinusoidal signal containing the cyclotron frequency and magnitude of the ion in the time domain. The sinusoidal signal is called transient or free inductive decay (FID). Fourier transform can convert the signal from the time domain into the frequency domain to give the mass to charge ratio and intensity of the ion (Figure

1-5). In real samples, the situation becomes far more complicated. FT-ICR-MS simultaneously detects all the ions present in the ICR cell (Fellgett advantage) by rapidly sweeping the whole frequency range corresponding to their m/z, so the transient is made up of all frequencies superimposed one over the other, each with its own intensity. A unique advantage of FT-ICR-MS is that researchers can acquire very high resolution data without extra costs except longer acquisition time.
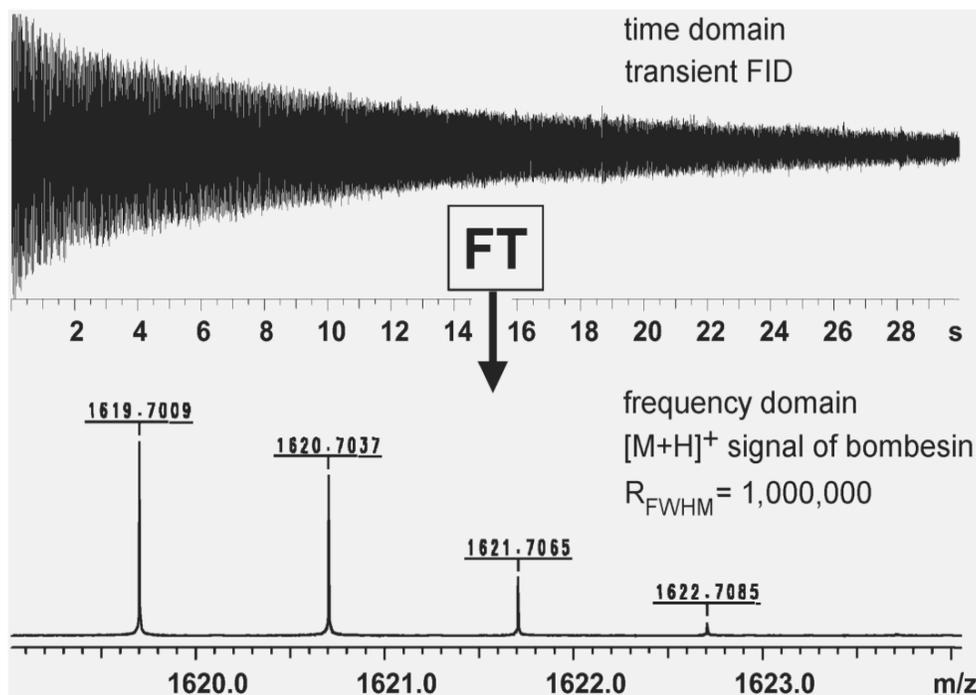
Figure 1-5 Function of Fourier transform in FT-ICR-MS. By courtesy of Bruker Daltonick, Bremen.

The mass analyzer of the FT-ICR-MS is the ICR cell, also known as Penning trap. The device is located inside a superconducting magnetic field and responsible for ion trapping, excitation and detection. The ICR cell has been designed into different geometric arrangements, but the most commonly used one

is a closed cylinder composed by three sets of electric plates (see Figure 1-6). Two trapping plates are placed at the ends of the cylinder and are perpendicularly aligned to the magnetic field. Each of them has a hole in the center allowing for ion transmission. At the beginning of analysis, external ions are transferred along the central axis into the cell, and then a small DC voltage (~1.0 V) of the same polarity as the ions is applied on the trapping plates to confine the ions in the middle of the cell. After analysis, the DC voltages of the trapping plates are changed to form a large potential difference so that the ICR cell is emptied quickly prior to the next acquisition.
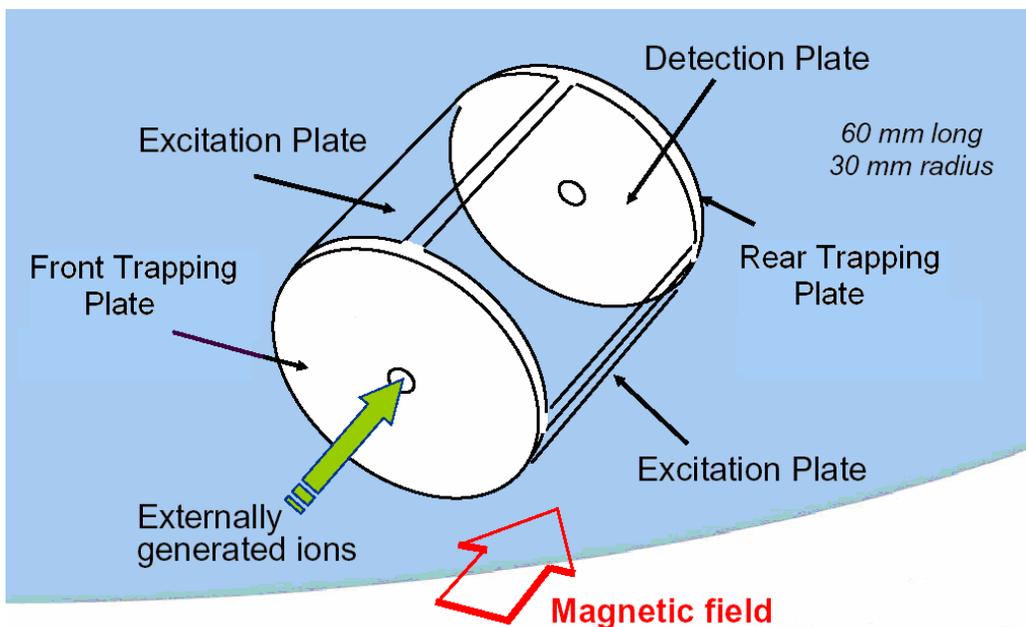
Figure 1-6.   Schematic diagram of a typical ICR cell. By courtesy of Bruker Daltonick, Bremen.

The excitation plates are oppositely situated on the ICR cell wall and connected to an RF oscillator. For most experiments, the frequency sweep (chirp) excitation waveform is used to quickly scan all frequencies corresponding to the m/z range of interest, whereas the ends of the waveform are distorted and not usable. A better option is to utilize an inverse Fourier transform to generate an optimized excitation waveform based on the mass range of ions of interest and then apply the waveform to produce RF excitation frequency. The strategy is named after stored waveform inverse Fourier transform (SWIFT). In addition, the SWIFT excitation can exclude some ions through selectively tailoring the waveform. Both of chirp and SWIFT excite all ions to the same orbit so that the ions are detected in the same environment. Notably, the excitation has to be accomplished within ~1 μs. Otherwise, the ions having the same m/z would distribute around the whole orbit and generate image currents with the same frequency but out of phase, so the total image current will be null. In an FT-ICR experiment, the large (~1 cm) cyclotron radius of active ions is essential to ensure high sensitivity. The radius is a function of the RF potential (V_0) and the time (T_{exc}) of RF executed on ions as well as magnetic field (B_0), but independent of m/z (Eq. 2).

$$r = \frac{V_0 T_{exc}}{B_0} \qquad (1\text{-}2)$$

The circuit of the detector plates is responsible for converting image currents induced by activated ions into time domain data. Compared to ion counting detectors, the image current detector has lower sensitivity. The resolving power (RP) of an ion is linearly increased with acquisition time of the transient as

shown in Eq. 3, where f is the cyclotron frequency of this ion and t is the transient time in seconds.

$$RP_{FWHM} = \frac{ft}{2} \qquad (1\text{-}3)$$

One important characteristic of the transient is that acquisition time has to be adjusted to a multiple of 2n due to the requirements imposed by the FT algorithm. Ions have a much longer lifetime in a FT-ICR cell than in a beam instrument. Compared to ~2 meters flight path in a TOF-MS, Marshall *et al.* has observed that ions of 100 Da fly approximately 30 kilometers in one second.[46] This is why FT-ICR-MS offers a higher resolution than other types of mass spectrometry. On the downside, FT-ICR-MS needs expensive equipment to maintain the ultrahigh vacuum ($< 10^{-9}$ mbar) in the ICR cell, otherwise ions would be quickly lost due to collisions between ions and residual gas molecules. Finally, the resolution and accuracy of FT-ICR-MS are restricted by the space charge effect. Because all ions are located in the same orbit for detection, they might suffer from an undesirable coulombic interaction arising from neighboring ions if too many ions are introduced into the ICR cell.

As an important instrument used in my research, the Bruker 9.4 T APEX FT-ICR-MS consists of four major sections: the ion source including ESI and MALDI, the Qh-Interface, the ion transfer optics and the detector (see Figure 1-7). In addition, six pumping stages consisting of two roughing pumps and four turbo molecular pumps enable the transport of ions from the API source to the ultrahigh vacuum ICR cell. The instrument is manipulated through Bruker apexControl and the data are processed by Bruker DataAnalysis. Bruker Hystar

acts as an interface to manage FT-ICR-MS and HPLC systems. In the LC-MS analysis of small molecules, the common mass accuracy is within 5 ppm and the resolution is around 50,000. Added to collision activated dissociation (CAD), the instrument also offers other MS/MS methods for structure identification, such as in-source collision (IS-CAD), sustained off-resonance irradiation (SORI), Infrared Multiphoton Dissociation (IRMPD), etc.
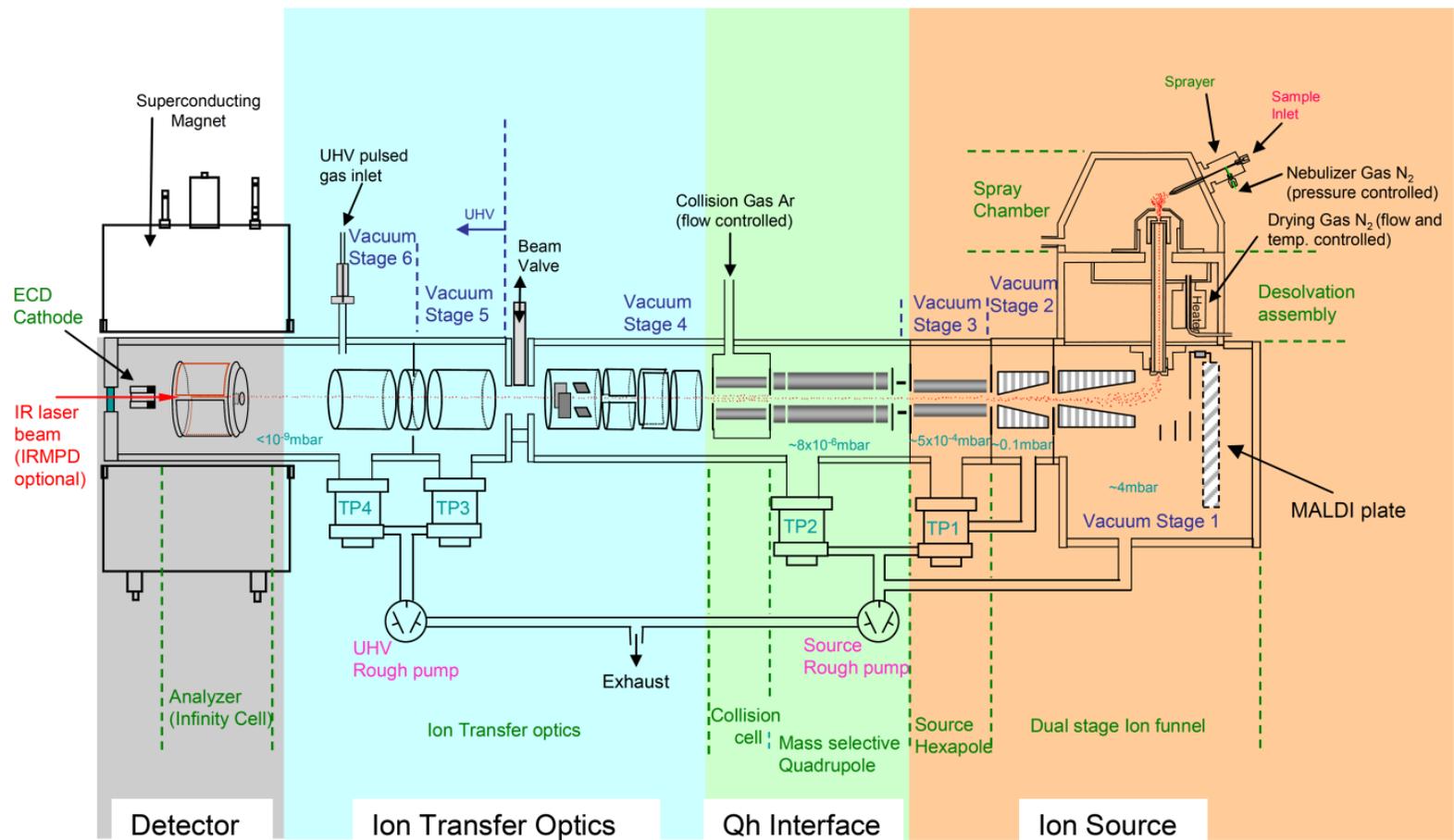
Figure 1-7.   Schematic diagram of Bruker 9.4 T Apex FT-ICR-MS. By courtesy of Bruker Daltonick, Bremen.

1.3.3.4 Principle of Q-TOF-MS

Q-TOF-MS is a tandem instrument consisting of a linear quadrupole mass analyzer and an orthogonal acceleration Time-of Flight Mass spectrometry (oa TOF-MS) (see Figure 1-8). In the MS mode, all of three quadrupoles: Q0, Q1 and Q2, are operated in the RF-only mode and serve as ion guides. In the MS/MS mode, while Q0 still acts as an ion guide, Q1 is responsible for isolation and transporting ions of interest to the collision cell Q2 where the fragmentation is carried out. To transmit ions/fragments into the TOF section, ions have to be accelerated to the required energy (several tens of eV per charge) by the ion optics. In the ion modulator, a pulse voltage orthogonal to the ion trajectory pushes the ions to the accelerating zone where ions are further accelerated to several KeV prior to entering into the field-free flight tube. During drift in the flight tube, ions are resolved according to their velocities. To enhance the resolution, an ion mirror is used to minimize the distribution of initial kinetic energy and the spatial spread. Finally, ions hit at the microchannel plate (MCP) detector to give signals.
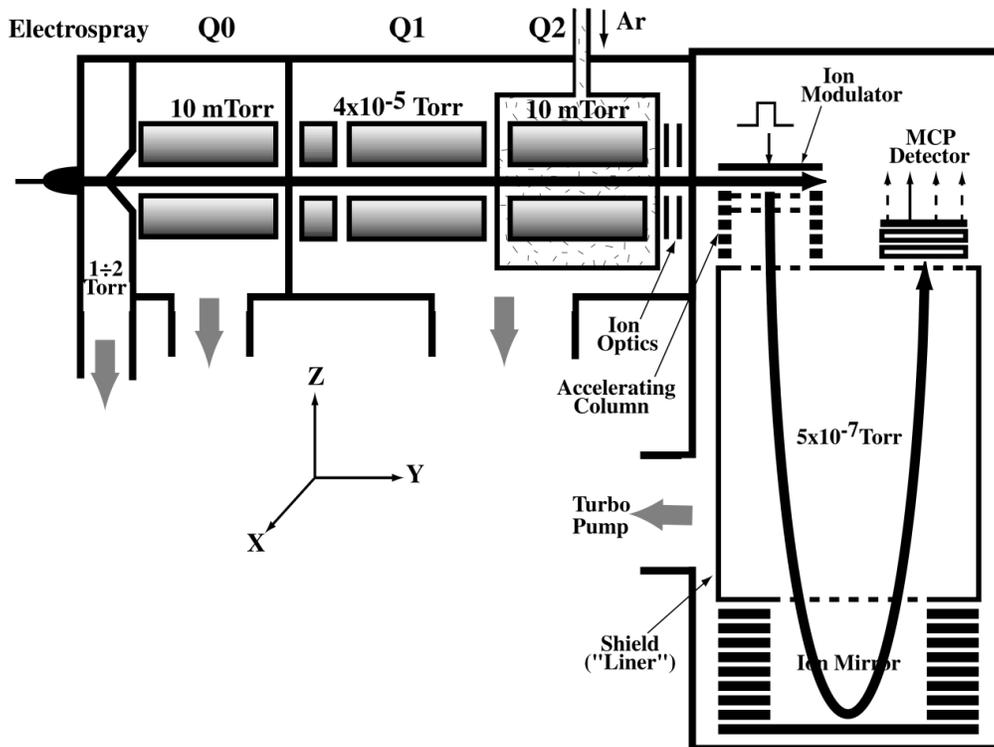
Figure 1-8.   Schematic diagram of Q-TOF-MS. Reproduction with permission.[47]

1.3.3.4.1 Principle of the Linear Quadrupole

A linear quadrupole mass analyzer is composed of four parallel mounted rod electrodes with a circular or hyperbolic cross section. To each pair of opposite electrodes is applied a potential superimposed by a RF component (V) and a DC component (U). The potentials of the two pairs have the same amplitude but opposite polarity (see Figure 1-9A). As an ion enters the quadrupole along the z-axis, it experiences an attractive force exerted by one of the rods which has the opposite charge polarity. The potential polarities on the rods switch periodically in time, the trajectory of the ion in the quadruple filed can be described by Eq. (1-4).

$$a_u = \frac{8zeU}{m\omega^2 r_0^2} \qquad\qquad q_u = \frac{4zeV}{m\omega^2 r_0^2} \qquad\qquad (1\text{-}4)$$

Where a_u and q_u are functions of the RF component (*V*) and DC component (U), respectively; z is the number of charges; *e* is the charge of an electron; ω is the angular frequency of the RF component and r_0 is the half distance because two opposite rods. When the quadrupole serves as a mass analyzer, ions have stable trajectories only within a narrow m/z window corresponding to a_u = 0.706 and q_u = 0.237.[47] In the RF-only mode, the DC component is zero and the quadrupole simultaneously transports all ions with m/z higher than a given limit. Aside from the quadrupole, hexapole and octapole are also used as ion guide and the collision cell. Compared to the quadrupole, their principles are similar, but the hexapoles and the octapoles provide more efficient ion transportation within the extended mass range at the expense of focusing power. In general, the advantages of using a quadrupole include simplicity, low costs, high transmission and great scan rates.

(A)
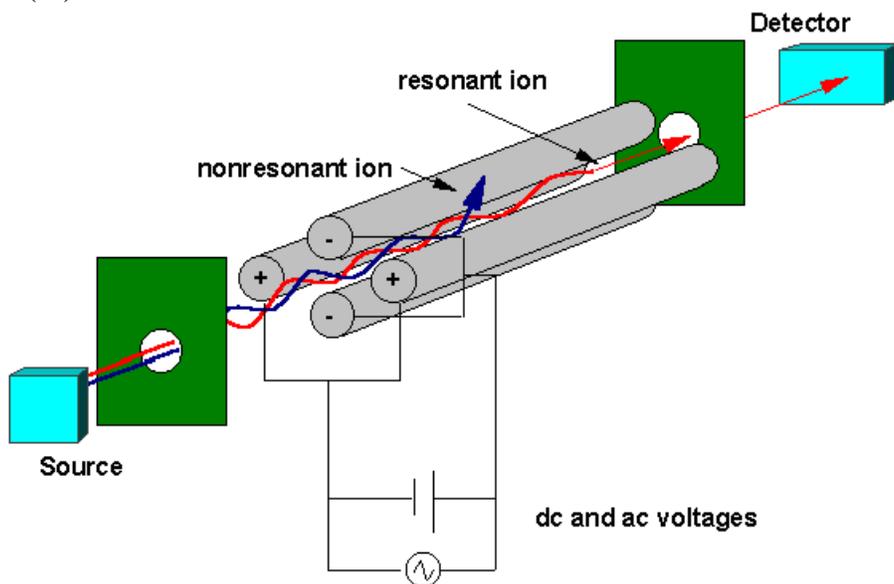
(B)
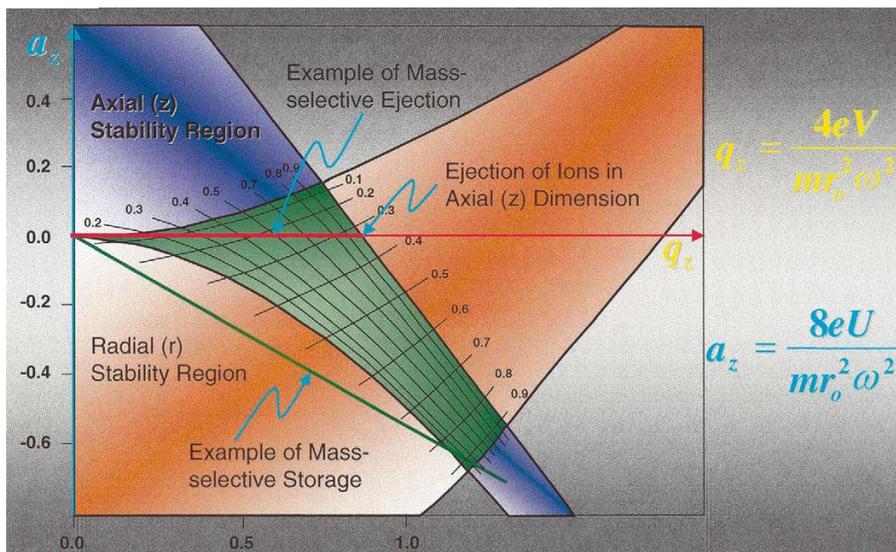
Figure 1-9.   (A) Schematic diagram of the linear quadrupole. (B) Stability diagram in the quadrupole field. Reproduction with permission48

1.3.3.4.2 Principle of Orthogonal Acceleration Time of Flight Mass spectrometry (oa-TOF-MS)

TOF-MS analyzes the mass of analytes based on the flight time that ions spend moving through the TOF tube. In principle, when a potential pulse (V_s) is exerted on an ion with charge z, the electric potential energy $E_p = zeV_s$ is converted into the kinetic energy $E_k = \frac{1}{2}mv^2$ which enables the ion to move forward to the detector. If the length of the TOF tube is L, the two equations can be arranged into Eq. (5) where m/z of the ion is proportional to the square of the fight time t.

$$(m/z)^{1/2} = \left(\frac{\sqrt{2eV_s}}{L}\right)t \qquad (1\text{-}5)$$

As a pulsed mode analyzer, TOF-MS perfectly works with MALDI and other pulsed ion sources, but it is a challenge to couple TOF-MS with a continuous ion source, e.g. ESI.[47] Especially, the initial kinetic energy and spatial distribution in the direction of flight tube seriously impact the resolution of TOF-MS. Nowadays, orthogonal acceleration is regarded as the best technique that has been widely used in ESI-TOF-MS/ESI-Q-TOF mass spectrometry. In oa-TOF-MS, the flight direction of ion beam is orthogonal to the initial direction in which ions enter the TOF tube, so the flight time is independent of initial kinetic energy and spatial distribution of ions. Furthermore, ion mirror, also known as a reflector is the most important device used in TOF-MS to minimize the unfavorable distribution present in the ions with the same m/z. A reflector consists of a series of grid electrodes and is usually situated at the end of the TOF tube opposed to that of the ion source and the detector (see Figure 1-10). In the experiment, the reflector holds a potential with the same polarity as the detected ions. When the ions approach the reflector, the more kinetic energy ions have, the deeper ions penetrate in the reflector. Consequently, the ions of high kinetic energy spend more time inside the reflector and arrive at the detector with ions of low kinetic energy simultaneously and give high spectrum resolution. Nowadays, the resolution of Q-TOF-MS can easily reach 50,000.
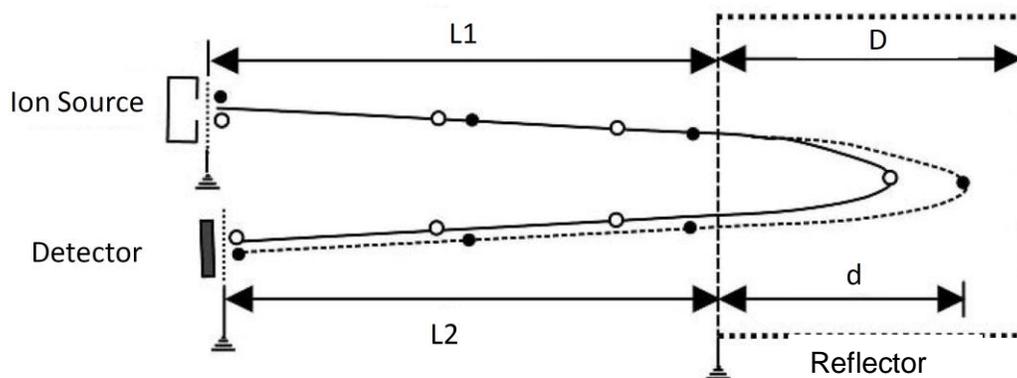
Figure 1-10.   Schematic diagram of TOF-MS equipped with a reflector. The black spots represent an ion with correct initial kinetic energy. The white spots are the ion of the same m/z but lower initial kinetic energy, so it spends shorter time in reflector than the black spot.

A mass accuracy of 5 ppm is easily achieved in the TOF-MS. The advantages of TOF-MS mainly come from the high resolution and straightforward calibration. First of all, the high resolution of TOF-MS makes it possible that the centroid of each peak can be accurately determined. Secondly, calibration of oa-TOF is fairly simple and accurate. As shown in Eq (5), the m/z is linearly increased with the square of the fight time. In theory, a simply two-point calibration of m/z vs. t can provide satisfactory results over a mass window of hundreds of m/z's.[47] Q-TOF-MS inherits the advantages of quadrupole and TOF MS, and thus it has high ion transmission, MS/MS capacity, fast scan speed, excellent mass accuracy and resolution as well.

1.3.3.5 Sample Preparation

As an important part in the LC-MS based metabolic analysis, sample preparation has a direct effect on data quality as well as data processing. Sample

preparation includes sample collection, sample storage, preliminary sample processing, sample normalization, sample extraction, and/or derivatization, etc. For the same sample, different preparation procedures can generate different, even contradictory results in metabolic analysis.[49,50] Despite its importance, sample preparation is often an ignored part of metabolomics analysis. Optimum sample preparation should achieve the following goals: 1) unbiased recovery of all metabolites of interest in high efficiency; 2) removing interferences as much as possible; 3) in a manner of high throughput, or being automated; 4) giving reproducible results; 5) compatible with LC separation. Each method has its merits and limitations, so the selection of sample preparation methods largely relies on the forms/matrices of samples and the objectives of study. Herein, the discussion focuses on typical techniques used in metabolome analysis such as "dilute and shoot", liquid-liquid extraction (LLE), and solid-phase extraction (SPE).

1.3.3.5.1 Dilute and Shoot

The strategy involves directly adding pure solvents into samples prior to LC-MS analysis without other processing. The method is often used for urine metabolome analysis.[51,52] To remove bacteria and solid particles, a short (5~10 min) centrifugation is sometimes done before the analysis is undertaken. Although some researchers advocate that a mild spin (1000-3000 g) can prevent the breakage of cells during centrifugation, the precaution has not been accepted by others.[53] A major concern to the approach is ion suppression caused by co-

eluted species in the ion source. In addition, protein precipitation sometimes occurs during sample dilution.[54]

1.3.3.5.2 Liquid-Liquid Extraction (LLE)

This method is useful method for separating hydrophobic analytes from aqueous samples based on the partition of samples between two immiscible liquids. The result of liquid-liquid extraction is dependent on the chemo-physical properties of the extraction solvents including solubility in H_2O, volatility, inertness, and especially polarity. In principle, recovery of an analyte in LLE is determined by the distribution constant K_D that is defined as the ratio of the concentration of an analyte in the organic phase over its concentration in the aqueous phase. The optimum K_D is obtained when the polarity of the extraction solvent matches that of the analytes of interest.[55] To change the K_D of ionic compounds, the pH of the sample solution needs to be adjusted so that the ionizable compounds can dissolve better in the organic phase.

In LC-MS based metabolic profiling, LLE currently plays an important role. To recover as many metabolites of interest as possible, a preliminary experiment is necessary to examine whether a solvent indeed meets the purpose of the study.[56] Furthermore, due to the concern about the possible loss of volatile metabolites during evaporation, direct injection of extracts in LC-MS analysis should be first considered if the extraction solvent was compatible with LC mobile systems. Moreover, a multiple-stage (n = 3) extraction is often performed to achieve higher recovery. Finally, to decrease the consumption of organic

solvents in the traditional LLE methods, new miniaturization methods have been developed, such as single-drop micro-extraction (SDME), hollow fiber liquid phase microextraction (HF-LPME), dispersive liquid-liquid microextraction (DLLME), etc.[57] It is worthwhile to note that all of these LLE techniques are time-consuming and require well-trained technicians.

1.3.3.5.3 Solid Phase Extraction (SPE)

SPE is the most popular method used in the pretreatment of biological samples. Contrary to the LLE method using a one-stage separation process, the mechanism of SPE is similar to that of liquid chromatography where analytes are retained based on partition or adsorption on the stationary phases (sorbents). SPE has displayed many advantages over LLE, such as: 1) higher recovery efficiency; 2) less solvent consumption; 3) cleaner sample solution; 4) better enrichment of analytes; 5) higher throughput. On the other side, SPE might cause loss of some analytes due to irreversible adsorption on the sorbents.

Several SPE configurations so far have been marketed, such as cartridges, disks, pipette tips, coated fibers, etc. The discussion herein is limited to the cartridges and disks used in our lab. The cartridge is the most popular SPE device and can be deemed to situate a sorbent bed in a syringe. The size of the sorbent varies from milligrams to grams so that it can accommodate a wide range of sample volumes and even dirty samples. Another popular configuration is the SPE disk that has a larger cross-section and thinner sorbent bed than the SPE cartridge. Their dimensions are usually 4 to 96 mm in diameter and $\leq$ 1mm thickness.[58] The

SPE disk allows for a faster flow rate and higher throughput, without the undesirable channeling effect existing in the cartridge. To date, automation SPE devices using commercial SPE devices have been made available by several manufacturers.

SPE sorbents are made of a wider range of stationary phases than that in liquid chromatography columns, including silica gels, polymers, activated charcoal, and alumina. According to the functional groups chemically bonded to the surface of silica gels or polymers, the sorbents can be classified into polar, non-polar and ion exchange.[59] To isolate particular compounds, some sorbents are prepared by coating a layer of selective covalent/non-covalent antibodies or complex-forming groups on the surface of supporting materials.[60-62] The initial SPE sorbents were made of reverse phase silica gels such as C_8 and C_{18}. Different from RPLC columns packed by type-B silica gels, inexpensive type-A silica gels are typically used for SPE sorbents.[63] The recently emerging polymeric sorbents have gained lots of attention because of their unique characteristics. The polymeric sorbent has larger surface area and causses less irreversible adsorption than silica gel sorbents. Another advantage is that the polymeric sorbents can be used over the entire pH range.

A general work flow of SPE involves four steps. First of all, a SPE device has to be conditioned prior to loading samples. Several bed-volumes of strong solvents are allowed to pass through the sorbent to wash impurities out and wet the sorbent. The device is then quickly rinsed by a weak solvent to get rid of the excess strong solvent. Secondly, the sample is dissolved in a weak solvent and

then loaded on the top of the sorbent. In the third step, a wash solvent or mixture is added to wash impurities out of the sorbent. The volume of washing solvent has to be carefully measured to minimize loss of trapped analytes. Finally, all trapped analytes are eluted by a small amount of a strong solvent ($0 < k < 1$).

1.4   LC-MS Based Quantification Analysis

In analytical chemistry, the concentration of an analyte in a sample can be determined by comparing the instrument response of the sample to that of standards. Thus, the relationship, also known as the calibration curve has to be established to indicate the instrumental response (y-axis) to the known concentrations (x-axis) of a series of standards. In the calibration curve, the slope is called sensitivity and the dynamic range is the concentration range where a measurable response is given due to changes in the analyte concentration.

Three kinds of calibration methods are often used in LC-MS based quantification. In the external calibration curve, the signal response of a sample is compared to the calibration curve developed in advance. Because there is no internal standard spiked, the matrix effect in the real sample might severely affect the accuracy and precision of the quantification.[64,65] Furthermore, the instrumental response drifts over time, so the mass spectrometric response needs to be calibrated regularly so that the error remains in an accepted range. Standard additions is another calibration method used to eliminate the matrix effect and the calibration curve is prepared by adding a series of known amounts of standard(s) into the sample solution (See Figure 1-11). The concentration of an analyte (x)

can be determined by extrapolation of the calibration curve to cross the x-axis (y = 0). Since the above two methods are time-consuming in preparing the calibration curve for each analyte, they are commonly used for metabolic target analysis rather than metabolic profiling.
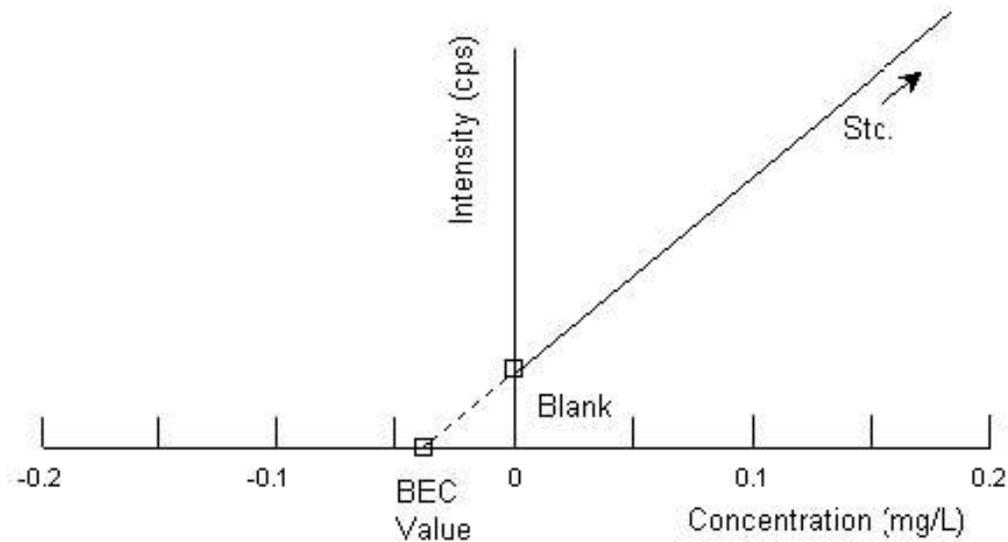
Figure 1-11.   Standard addition calibration method. The concentration of analytes can be calculated by extrapolating the curve to cross x axis (y = 0).

The third method is internal standard calibration in which the concentration of analytes of interest in a sample is measured based on the signal ratios between the sample and the internal standards added to the sample. The internal standard should behave similarly to the analyte, but provide a signal distinguishable from the analyte. Structural analogues are the most popular choices, whereas this kind of internal standards often display different chromatography performance and ionization efficiency in ESI interfaces from targeted analytes. Better internal standards would be isotope analogues which have (almost) identical behaviors as the targeted analytes in the RPLC and ESI

source but different mass peaks in the spectra.[66,67] This method is known as stable isotopic labeling (SIL) dilution quantification. In SIL, ^{1}H, ^{12}C, ^{14}N, or ^{16}O are the most common elements to be replaced by ^{2}H, ^{13}C, ^{15}N or ^{18}O, respectively. As the relative concentration, not absolute response is determined in the SIL dilution method, the SIL internal standards completely eliminate ion suppression effect from co-eluted compounds and run-to-run variations caused by sample injection and LC-MS parameters.[68]

Unfortunately, commercial sources of SIL standards are very limited and the available ones are quite expensive. As the aim of metabolomics is to measure every metabolite present in a bio-system, it is essentially impossible to use stable isotopic internal standards in analyzing all the metabolites. To this end, a differential isotopic labeling (DIL) method has been developed and it has provided a promising approach to address this issue.

1.5   LC-MS Based DIL Method

The differential isotope labeling (DIL) method uses a chemical reaction to add a labeling to the analytes in one sample and the isotopic counterpart to the same analytes in another comparative sample (or standard). The two labeled samples are then combined prior to mass spectrometric analysis. The relative quantification of the analytes in two comparative samples can be determined based on the peak abundance ratio of the labeled analyte pair.[69] An absolute quantification is also possible if the concentrations of analytes labeled by the isotopic reagent are known. Compared to *in vivo* isotopic labeling approaches in

living cells or organisms, DIL is suitable for a wide range of samples including various bio-fluid samples and extracts. Derivatization of amine-containing metabolites, organic acids and carbonyl compounds has attracted most interest since more than 80% of metabolites contain at least one of the three functional groups.[18] To date, many isotopic labeling reagents have been developed for metabolic research.[69-73] The principle of DIL is quite similar to that of SIL; while DIL uses chemical derivatization, the two methods give comparable data quality in LC-MS based quantification. However, it is be worth noting that deuterium-coded isotopologues are faster eluted than their counterparts in RPLC. The isotopic effect has been attributed to weaker dispersion interface between deuterium-coded analytes and alkyl chains in reversed-phase chromatography.[74]

In addition to giving high quality data, DIL offers additional advantages to metabolites in LC-MS analysis. First of all, chemical derivatization can convert ionic or polar compounds into more hydrophobic products so that these metabolites are better resolved in RPLC and ion suppression in the ESI source is decreased. Moreover, labeling tags shift the signal response of small molecules from the low-mass region with more noisy background to the cleaner high-mass region, as such the signal to noise ratios of labeled analytes is improved.[75,76] Finally, the hydrophobic labeling tag can greatly enhance the ionization efficiency of metabolites in the ESI source. The improvement of ESI activity can be attributed to three major factors: 1) A labeled metabolite has better droplet surface affinity owing to the hydrophobic labeling tag and prefers to stay on the surface of droplets during the electrospray process, so the labeled metabolites have more

chances to be ionized based on the ion evaporation mechanism; 2) Hydrophobic derivatization products are eluted by mobile phase with a higher proportion of the organic solvent during a RPLC gradient run, so the ionization efficiency is enhanced due to faster desolvation of the organic component; 3) Many DIL reagents contain a basic group which is easily protonated within the ESI process.[69]

In the work, twenty amino acid standards have been selected in development of DIL methods profiling amine-containing metabolites. Compared to other small molecules, amino acids have important biological functions in metabolic pathway.[77-79] Furthermore, the amino acids display diverse chemo-physical properties which allow us to examine RPLC performance and ESI activities of labeling tag. Finally, twenty amino acids are inexpensive compounds with ample commercial sources.

1.6   LC-MS Based Metabolic Qualification Analysis

LC-MS is a powerful technique for identification of metabolites due to its sensitivity, specificity, and high throughput. Nevertheless, LC-MS based metabolic qualification is not an easy task because the process needs a number of instrumental techniques and knowledge of tandem mass spectrometry based metabolite identification.[80-81] A tandem mass spectrometry consists of two or more MS analyzers mounted in tandem; each analyzer has different functions: the isolation stage, the fragmentation stage, and the scan stage. Tandem mass spectrometry can be classified into time-dependent MS and space-dependent MS.

Trapping instruments (e.g. ion trap) are typical time-dependent tandem MS while the triple quadrupole and Q-TOF-MS are mostly known as space-dependent tandem MS.

The main advantage of tandem MS is its ability to use different scan modes to acquire structure information of the targeted metabolites. The product ion scan mode utilizes the first mass analyzer to isolate ions with selected m/z. After fragmentation in the collision inducted/activated dissociation (CID/CAD) cell, all resultant product ions are determined by the second mass analyzer. The precursor ion scan is composed of a mass scan of precursor ions in the first mass analyzer and isolation of a selected product ion in the second mass analyser. In this kind of acquisition, only the precursor ions that generate the selected product ion can be acquired. A neutral loss scan includes the mass scan of precursor ions in the first mass stage and the second mass scan of product ions in the third mass stage. In the neutral loss scan, all precursor ions losing a selected neutral fragment are determined. However, time-dependent tandem mass spectrometry does not support the precursor ion scan and the neutral ion scan. Selected reaction monitoring (SRM) is a scan mode only supported by the triple quadrupole or the Qtrap. The method consists first of the isolation of selected precursor ions in Q1 and of the isolation of the selected product ions in Q3. If multiple parent ions are monitored in the experiment, the scan is called multiple reaction monitoring (MRM). It is worthwhile to note that both SRM and MRM are commonly used in LC-MS based quantification of targeted metabolites, not in the identification of unknowns.

Identification can be divided into two catalogs: putative identification and definitive identification. However, definitive identification is difficult in metabolomics because it is (almost) impossible to purchase all possible authentic standards to compare their properties with analytes of interest. To address this issue, the metabolomics community has established several databases which collect published information of known metabolites as much as possible, such as HMDB, METLIN, PubChem, KEGG, NIST, etc.[22,30,79-81] When people detect a bioactive metabolite in their samples, they can compare the accurate mass, MS/MS fragmentation patterns, retention time in GC/LC, even NMR spectra of the unknown to the data of possible candidates in the libraries. The approach is able to give a putative identification result of the unknown with high confidence. Unfortunately, fragmentation patterns of labeled metabolites are usually not the same as those of unlabeled compounds, so identification remains a big challenge for researchers who are using differential isotopic derivatization method in the LC-MS based metabolic profiling.

1.7 Overview of the Thesis

The objective of my research is to develop LC-MS based differential isotopic labeling methods and the related data processing software to quantify and qualify metabolites in biological systems. Dansyl chloride (DnsCl) is a well known chemical derivatization reagent and has been widely used to analyze amino acids and primary/secondary amines on fluorescence/ultraviolet platforms. Dr. Kevin (Kun) Guo, a previous group member, has successfully used $^{12}C_2$-/$^{13}C_2$-DnsCl as differential isotopic labeling reagents to quantify amine-containing

metabolites with LC-MS.[69] The method has displayed high sensitivity, robustness, specificity in metabolic analysis. To apply the method in a large-scale analysis, first my research was focused on the thorough optimization of the method, including the synthesis of $^{12}C_2$-/$^{13}C_2$-DnsCl, labeling reaction conditions and development of LC-MS analysis. I then developed a multiplex DnsCl-like isotopic labeling reagent, 5-Diethylamino-naphthalene-1-sulfonyl chloride (DensCl). The reagent consists of three isotope forms: except $^{12}C_4$-DensCl, $^{12}C_2{}^{13}C_2$-DensCl has two ^{13}C atoms, and $^{12}C_2$-/$^{13}C_4$-DensCl has four ^{13}C atoms, respectively. The reagent provides better sensitivity and higher sample analysis throughput than DnsCl. During the development of DensCl, we realized that data processing software is needed for the application of this method in metabolomics research. I then worked on the development of a computer tool, IsoMS that can rapidly process our DIL method datasets and generate the information for quantification as well as identification. In Chapter IV, the principle and preliminary results of the software are discussed. Using the DIL method, we can quickly quantify hundreds, even thousands of metabolites from various biological samples. It should be noted that the identification of bio-active compounds is still at an early stage of development. A major issue is that metabolites labeled by DnsCl or DensCl cannot give their own native fragment patterns, because the proton prefers to stay on the basic dimethylamino/diethylamino nitrogen of the labeling tag. To address the issue, in Chapter V, I describe a new type of DIL reagent used for the structural analysis of metabolites. My thesis and future work will be summarized in Chapter VI.

References

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

(2)   Fiehn, O. *Comp. Funct. Genomics* **2001**, *2*, 155.

(3)   German, J. B.; Watkins, S. M.; Fay, L. B. *J. Am. Diet. Assoc.* **2005**, *105*, 1425.

(4)   Brown, S. C.; Kruppa, G.; Dasseux, J. L. *Mass Spectrom. Rev.* **2005**, *24*, 223.

(5)   Becker, S.; Kortz, L.; Helmschrodt, C.; Thiery, J.; Ceglarek, U. *J. Chromatogr. B* **2012**, *883*, 68.

(6)   Dettmer, K.; Aronov, P. A.; Hammock, B. D. *Mass Spectrom. Rev.* **2007**, *26*, 51.

(7)   Schuhmacher, R.; Krska, R.; Weckwerth, W.; Goodacre, R. *Anal. Bioanal. Chem.* **2013**, *405*, 5003.

(8)   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.

(9)   Kato, H.; Izumi, Y.; Hasunuma, T.; Matsuda, F.; Kondo, A. *J. Biosci. Bioeng.* **2012**, *113*, 665.

(10)   Ellis, D. I.; Goodacre, R. *Analyst* **2006**, *131*, 875.

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

(12)   Vikram, A.; Lui, L. H.; Hossain, A.; Kushalappa, A. C. *Ann. Appl. Biol.* **2006**, *148*, 17.

(13)   Rochfort, S. *J. Nat. Prod.* **2005**, *68*, 1813.

(14)   Nicholson, J. K.; Wilson, I. D. *Nat. Rev. Drug Discovery* **2003**, 668.

(15) Soga, T.; Baran, R.; Suematsu, M.; Ueno, Y.; Ikeda, S.; Sakurakawa, T.; Kakazu, Y.; Ishikawa, T.; Robert, M.; Nishioka, T.; Tomita, M. *J. Biol. Chem.* **2006**, *281*, 16768.

(16) 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. *J. Clin. Endocrinol. Metab.* **2011**, *96*, 3775.

(17) Kell, D. B.; Oliver, S. G. *Bioessays* **2004**, *26*, 99.

(18) 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 Res.* **2007**, *35*, D521.

(19) Li, L.; Li, R.; Zhou, J.; Zuniga, A.; Stanislaus, A. E.; Wu, Y.; Huan, T.; Zheng, J.; Shi, Y.; Wishart, D. S.; Lin, G. *Anal. Chem.* **2013**, *85*, 3401.

(20) Xia, J. G.; Mandal, R.; Sinelnikov, I. V.; Broadhurst, D.; Wishart, D. S. *Nucleic Acids Res.* **2012**, *40*, W127.

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

(22) 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 Res.* **2013**, *41*, D801.

(23) 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 Res.* **2009**, *37*, D603.

(24) Ryan, D.; Robards, K. *Anal. Chem.* **2006**, *78*, 7954.

(25) Pan, J. W.; Hetherington, H. P.; Hamm, J. R.; Shulman, R. G. *Magn. Reson. Med.* **1991**, *20*, 48.

(26) Deslauriers, R.; Jarrell, H. C.; Byrd, R. A.; Smith, I. C. *FEBS Lett.* **1980**, *118*, 185.

(27) Novak, R. F.; Koop, D. R.; Hollenberg, P. F. *Mol. Pharmacol.* **1980**, *17*, 128.

(28) Lewis, I. A.; Schommer, S. C.; Hodis, B.; Robb, K. A.; Tonelli, M.; Westler, W. M.; Sussman, M. R.; Markley, J. L. *Anal. Chem.* **2007**, *79*, 9385.

(29) Slupsky, C. M. *Biomark. Med.* **2010**, *4*, 195.

(30) http://webbook.nist.gov/chemistry/.

(31) McLafferty, F. W.; Stauffer, D. B.; 9 ed. 2013; Vol. 2013.

(32) Molnar, I.; Horvath, C. *J. Chromatogr.* **1977**, *142*, 623.

(33) Canals, I.; Valko, K.; Bosch, E.; Hill, A. P.; Roses, M. *Anal. Chem.* **2001**, *73*, 4937.

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

(35) Mora, J. F.; Van Berkel, G. J.; Enke, C. G.; Cole, R. B.; Martinez-Sanchez, M.; Fenn, J. B. *J. Mass Spectrom.* **2000**, *35*, 939.

(36)   Wampler, F. M.; Blades, A. T.; Kebarle, P. *J. Am. Soc. Mass. Spectrom.* **1993**, *4*, 289.

(37)   Kebarle, P. *J. Mass Spectrom.* **2000**, *35*, 804.

(38)   Kebarle, P.; Verkerk, U. H. *Mass Spectrom. Rev.* **2009**, *28*, 898.

(39)   Dole, M.; Mack, L.; Hines, R.; Mobley, R.; Ferguson, L.; Alice, M. d. *J. Chem. Phys.* **1968**, *49*, 2240.

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

(41)   Sommer, H.; Thomas, H. A.; Hipple, J. A. *Physical Review* **1951**, *82*, 697.

(42)   Comisaro.Mb; Marshall, A. G. *Chem. Phys. Lett.* **1974**, *25*, 282.

(43)   He, F.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2001**, *73*, 647.

(44)   White, F. M.; Marto, J. A.; Marshall, A. G. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1845.

(45)   Bossio, R. E.; Marshall, A. G. *Anal. Chem.* **2002**, *74*, 1674.

(46)   Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17*, 1.

(47)   Chernushevich, I. V.; Loboda, A. V.; Thomson, B. A. *J. Mass Spectrom.* **2001**, *36*, 849.

(48)   Stafford Jr, G. *J. Am. Soc. Mass. Spectrom.* **2002**, *13*, 589.

(49)   Canelas, A. B.; ten Pierick, A.; Ras, C.; Seifar, R. M.; van Dam, J. C.; van Gulik, W. M.; Heijnen, J. J. *Anal. Chem.* **2009**, *81*, 7379.

(50)   Duportet, X.; Aggio, R. B. M.; Carneiro, S.; Villas-Boas, S. G. *Metabolomics* **2012**, *8*, 410.

(51)   Mattarucchi, E.; Baraldi, E.; Guillou, C. *Biomed. Chromatogr.* **2012**, *26*, 89.

(52)   Gika, H. G.; Macpherson, E.; Theodoridis, G. A.; Wilson, I. D. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2008**, *871*, 299.

(53) Sheikh, K. D.; Khanna, S.; Byers, S. W.; Fornace Jr, A. J.; Cheema, A. K. *Journal of biomolecular techniques: JBT* **2011**, *22*, 1.

(54) Vuckovic, D. *Anal. Bioanal. Chem.* **2012**, *403*, 1523.

(55) Snyder, L. *J. Chromatogr. A* **1974**, *92*, 223.

(56) Kim, H. K.; Verpoorte, R. *Phytochem. Anal* **2010**, *21*, 4.

(57) Pena-Pereira, F.; Lavilla, I.; Bendicho, C. *Spectrochimica Acta Part B: Atomic Spectroscopy* **2009**, *64*, 1.

(58) Hennion, M.-C. *J. Chromatogr. A* **1999**, *856*, 3.

(59) Huie, C. W. *Anal. Bioanal. Chem.* **2002**, *373*, 23.

(60) Muldoon, M. T.; Stanker, L. H. *Anal. Chem.* **1997**, *69*, 803.

(61) Wolfe, K. A.; Breadmore, M. C.; Ferrance, J. P.; Power, M. E.; Conroy, J. F.; Norris, P. M.; Landers, J. P. *Electrophoresis* **2002**, *23*, 727.

(62) Farjam, A.; Brugman, A. E.; Lingeman, H.; Brinkman, U. A. *Analyst* **1991**, *116*, 891.

(63) Liška, I. *J. Chromatogr. A* **2000**, *885*, 3.

(64) Matuszewski, B.; Constanzer, M.; Chavez-Eng, C. *Anal. Chem.* **1998**, *70*, 882.

(65) Van Eeckhaut, A.; Lanckmans, K.; Sarre, S.; Smolders, I.; Michotte, Y. *J. Chromatogr. B* **2009**, *877*, 2198.

(66) 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. *Anal. Biochem.* **2005**, *336*, 164.

(67) Goodlett, D. R.; Keller, A.; Watts, J. D.; Newitt, R.; Yi, E. C.; Purvine, S.; Eng, J. K.; Haller, P. v.; Aebersold, R.; Kolker, E. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1214.

(68) Julka, S.; Regnier, F. *J. Proteome Res.* **2004**, *3*, 350.

(69) Guo, K.; Li, L. *Anal. Chem.* **2009**, *81*, 3919.

(70)   Guo, K.; Li, L. *Anal. Chem.* **2010**, *82*, 8789.

(71)   Shortreed, M. R.; Lamos, S. M.; Frey, B. L.; Phillips, M. F.; Patel, M.; Belshaw, P. J.; Smith, L. M. *Anal. Chem.* **2006**, *78*, 6398.

(72)   Yang, W. C.; Regnier, F. E.; Jiang, Q.; Adamec, J. *J. Chromatogr. A* **2010**, *1217*, 667.

(73)   Ji, C.; Li, L. *J. Proteome Res.* **2005**, *4*, 734.

(74)   Turowski, M.; Yamakawa, N.; Meller, J.; Kimata, K.; Ikegami, T.; Hosoya, K.; Tanaka, N.; Thornton, E. R. J. Am. Chem. Soc. 2003, 125, 13836.

(75)   Ikonomou, M. G.; Blades, A. T.; Kebarle, P. *J. Am. Soc. Mass. Spectrom.* **1991**, *2*, 497.

(76)   Hewavitharana, A. K.; Herath, H. M.; Shaw, P. N.; Cabot, P. J.; Kebarle, P. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 3510.

(77)   Yost, R. A.; Enke, C. G. *Anal. Chem.* **1979**, *51*, 1251.

(78)   Yost, R. A.; Boyd, R. K. *Methods Enzymol.* **1990**, *193*, 154.

(79)   McClellan, J. E.; Quarmby, S. T.; Yost, R. A. *Anal. Chem.* **2002**, *74*, 5799.

(80)   Stokes, H. L.; Easton, A. J.; Marriott, A. C. *J. Gen. Virol.* **2003**, *84*, 2679.

(81)   Wang, W. W.; Qiao, S. Y.; Li, D. F. *Amino Acids* **2009**, *37*, 105.

(82)   Li, P.; Yin, Y. L.; Li, D.; Kim, S. W.; Wu, G. Br. *J. Nutr.* **2007**, *98*, 237.

(83)   Zhu, Z. J.; Schultz, A. W.; Wang, J.; Johnson, C. H.; Yannone, S. M.; Patti, G. J.; Siuzdak, G. *Nat. Protoc.* **2013**, *8*, 451.

(84)   Wang, Y.; Xiao, J.; Suzek, T. O.; Zhang, J.; Wang, J.; Zhou, Z.; Han, L.; Karapetyan, K.; Dracheva, S.; Shoemaker, B. A.; Bolton, E.; Gindulyte, A.; Bryant, S. H. *Nucleic Acids Res.* **2012**, *40*, D400.

(85)   Kanehisa, M.; Goto, S.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. *Nucleic Acids Res.* **2013**.

Chapter II: Quantitative Metabolic profiling Using Dansylation

Isotope Labeling and LC-MS*

2.1   Introduction

Differential isotopic labeling (DIL) LC-MS uses a chemical reaction to introduce an isotopic tag to an analyte in one sample and a mass-difference isotopic tag to the same analyte in another comparative sample (or an authentic standard with known concentration), followed by mixing the two labeled samples for mass spectrometric analysis. The peak intensity ratio of the isotope labeled analyte pairs provides the relative or absolute quantitative information on the analyte in the two comparative samples (*1-3*). DIL LC-MS is commonly used for targeted metabolite quantification, such as quantifying a drug in a bio-fluid, because of its high sensitivity of detection and high precision and accuracy of quantification. This concept of DIL LC-MS can be expended into metabolic profiling where a sub-metabolome, i.e., all the metabolites sharing the same functional group, instead of one targeted metabolite, can be simultaneously labeled with a proper isotope reagent to introduce a mass tag for quantification. High performance DIL LC-MS methods can be developed based on careful selection of the labeling reagents. They can improve separation, detection, quantification and identification of the labeled metabolites, compared to analyzing their unlabeled counterparts.

Due to the diverse chemo-physical properties and broadly distributed concentrations of metabolites in biological samples, many unlabeled metabolites cannot be readily separated by reversed phase liquid chromatography (RPLC), which is commonly interfaced to a mass spectrometer. They also suffer from serious ion suppression in electrospray ionization mass spectrometry (ESI-MS). After applying chemical derivatization, the labeled metabolite scan has enhanced chromatographic properties in RPLC for efficient separation, as well as higher ionization efficiency in ESI-MS for improved detection sensitivity (*4*).

One of the high performance labeling reagents for DIL LC-MS is $^{12}C_2$- and $^{13}C_2$-dansyl chloride (DnsCl). DnsCl has been extensively used to derivatize primary, secondary amines and phenolic compounds for liquid chromatography with UV and fluorescence detection (*9-12*). DnsCl is well qualified to be a high performance DIL reagent for LC-MS-based metabolic profiling. It is easy to introduce the isotope tags on the amine group of DnsCl by simple chemical reactions in high labeling efficiency (*4*). To avoid the isotope effect on RPLC, ^{13}C-methyl iodine has been chosen to introduce the isotope tags in two steps (**Figure 2-1**). DnsCl prefers to react with primary/secondary amines and phenolic compounds even in solvent with high water content (**Figure 2-2**). The aromatic structure of the dansyl group allows an increase in hydrophobicity of the labeled metabolite, thereby rendering it easier to retain on an RPLC column. The dimethylamino moiety can be readily protonated in ESI. The detection sensitivity of the dansyl labeled metabolites is usually enhanced by 10 to 1000-fold, compared to the unlabeled metabolites.
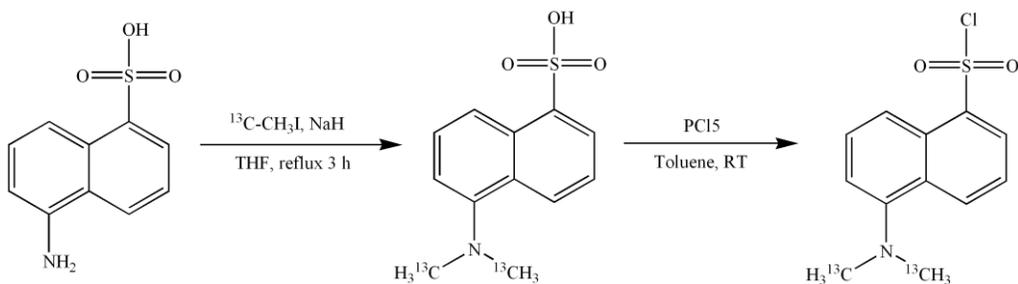
Figure 2-1.   Reaction scheme for the synthesis of $^{13}C_2$-DnsCl.
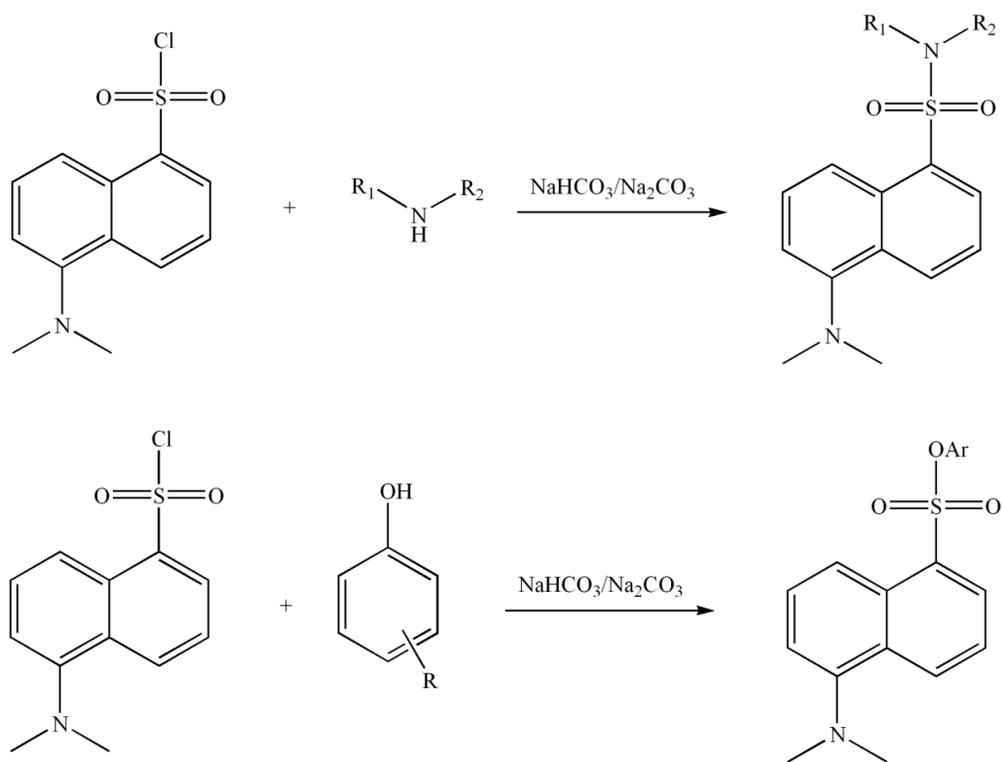
Figure 2-2.   Reaction scheme for dansylation of primary/secondary amines and phenolic compounds.

High resolution MS, such as Fourier transform ion cyclotron resonance (FT-ICR) MS, can be used to detect the peak pairs of differentially labeled metabolites and determine their intensity ratios accurately. However, it is time consuming to manually pick the peak pairs from a large number of LC-MS

datasets. To address this critical issue, we have developed a computer program called IsoMS to automatically pick the peak pairs, filter out the false or redundant pairs, calculate the peak ratios, and then save the data into an Excel file with a data format compatible with statistical software. Using this high performance dansylation isotope labeling LC-MS method, we can profile hundreds to thousands of metabolites in various biological samples (*4, 13-15*). The quantitative information can be further processed by a bio-statistical tool. The overall workflow can be used for comparative metabolomics in various research areas such as discovery of disease biomarkers (**Figure 2-3**).
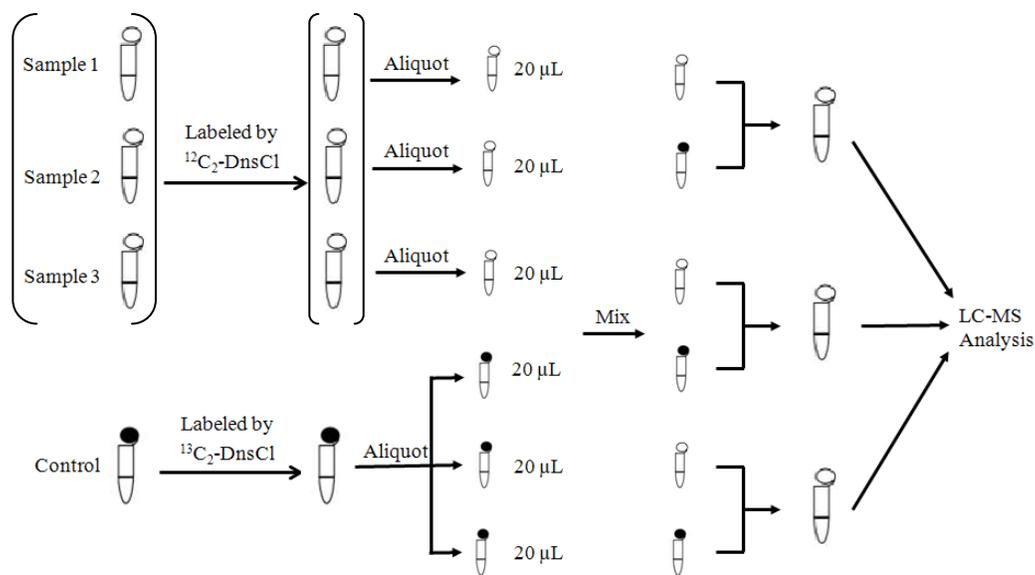
Figure 2-4.   Workflow for relative and absolute quantification of metabolites using dansylation isotope labeling LC-MS.

In this chapter, an optimized experimental protocol is described for dansylation isotope labeling LC-MS.

2.2 Materials

All chemicals and reagents were purchased from Sigma-Aldrich Canada (Markham, ON, Canada) unless otherwise noted. ^{13}C-methyl Iodide was obtained from Cambridge Isotope Laboratories Inc. (MA, USA).LC-MS grade water, methanol, and acetonitrile (ACN) were purchased from Thermo Fisher Scientific (Edmonton, AB, Canada). Buffer salts including sodium carbonate and sodium bicarbonate were stored at room temperature.^{12}C$_2$-/^{13}C$_2$-DnsCl and amine standards were stored in -20 ℃ freezer. All solutions and HPLC mobile phases were prepared using LC-MS grade solvents (*see* **Note 1**), and the solutions were stored at room temperature. Urine and blood samples were collected from healthy individuals with standard operating procedures (SOPs, see below) after they reviewed and signed informed consents. All human samples were processed in a bio-safety Level II laboratory and stored in a Level I or regular chemistry laboratory after being processed. The waste was disposed following the waste disposal regulations. All the studies involving human subject have been reviewed and approved by the University of Alberta Health Research Ethics Board.

2.3 Methods

2.3.1 Preparation of $^{13}C_2$-DnsCl

To a solution of 1.73 g 5-amino-1-naphthalenesulfonic acid in 100 mL anhydrous THF, 1.60 g60% NaH and 4.26 g ^{13}C-MeI was added to the mixture at 0 °C (*see* **Note 2**). After refluxing for 3h, 10 mLH$_2$O was slowly added to quench the excess NaH. The reaction solution was acidified by 36% HCl to give a yellow precipitate, 5-dimethylamino-naphthalene-1-sulfonicacid (*See* **Note 3**). The intermediate was mixed with 6.8 g PCl$_5$ in 10 mL toluene and stirred at room temperature until the solid disappears (*See* **Note 4**).The solution was poured into 50 mL ice water and extracted twice with 50mL EtOAc. The organic layer was dried with anhydrous Na$_2$SO$_4$. After evaporation of the solvent, the residue was purified by flash chromatography (95/5 hexane–EtOAc) to give 2.19 g $^{13}C_2$-DnsCl in a 76% yield over the two steps. The structure of $^{13}C_2$-DnsCl was verified by NMR and the purity was validated by LC-MS.

2.3.2 0.5 M NaHCO$_3$/Na$_2$CO$_3$ Buffer Solution

0.5 M Na$_2$CO$_3$ solution: Weigh 5.3 g anhydrous Na$_2$CO$_3$ in a clean beaker. Measure 100 mL water in 100 mL volumetric flask and transfer to the beaker to dissolve the solid.

0.5 M NaHCO$_3$ solution: Weigh 4.2 g anhydrous NaHCO$_3$ and prepare 0.5M solution as in previous step.

Mix the two solutions with a ratio of 1:1 (*v:v*) in a Pyrex storage bottle.

2.3.3 250 mM NaOH Solution

Weigh 1.0 g NaOH and transfer to a Nalgene lab quality bottle. Measure 100 mL water in 100 mL volumetric flask and transfer to the plastic bottle.

2.3.4 425 mM Formic Acid Solution

Dilute 1.60 mL formic acid to 100 mL with ACN/H_2O (50/50,*v/v*).

2.3.5 50 mM DnsCl Solution

Dissolve 13.5 mg DnsCl in 1 mL LC-MS grade ACN (*see* **Note 5**).

2.3.6 Human Urine Samples (*see* note 6)

The second morning urinations were collected from each of healthy volunteers who have fasted for 8 hours. Within one hour after the sample collection, the urine was centrifuged at 4,000 rpm for 10 minutes. The supernatant was filtered twice through 0.22 μM-pore-size millipore filters (Millipore Corp., MA) and aliquoted into 0.5 mL vials. A pooled urine sample was prepared by mixing the aliquots of each urine sample. All equipment was cleaned by 10% bleach solution (*v/v*) and 70% alcohol (*v/v*) before and after use. The urine samples were stored at -80 ℃.

A 10 mL blood sample was taken from each of the healthy individuals who have fasted for 8 hours. The samples were allowed to clot spontaneously at room temperature in 1 hour, and then centrifuged at 3,500 rpm for 15 minutes to separate serum and cells. After the supernatant was decanted into a new clean

tube, methanol was added into the tube to precipitate proteins from the serum in a ratio of 3:1 (*v/v*). The mixture was centrifuged at 14,000 g for 15 minutes and the supernatant was divided into 250 μL aliquots. The serum samples were dried by speed vacuum prior to being stored in-80 ℃ freezer. A pooled serum sample was prepared by mixing the aliquots of each serum sample. All equipment was cleaned by 10% bleach solution (*v/v*) and 70% alcohol (*v/v*) before and after use.

2.3.7   Dansylation

An aliquot of 50 μL of sample solution (*see* **Note 7**)was mixed with 25 μL of 500 mM $NaHCO_3$/Na_2CO_3 buffer, and then added 75 μL of 50 mM$^{12}C_2$-DnsCl for light labeling, or50mM$^{13}C_2$-DnsClfor heavy labeling to reaction vials (*see* **Note 8**). The derivatization reactions were incubated at 60 ℃ for 60 minutes. After cooling down, the vials were spun down and 10 μL of 250 mM NaOH solution were added to quench the excess labeling reagents (*see* **Note 9**). The reaction solutions were acidified by adding 50 μL of 425 mM formic acid solution, and then equal volumes of $^{12}C_2$-/$^{13}C_2$-DnsCl labeled samples were mixed to give the final samples for analysis (*see* **Notes 10, 11**) (**Figure 2-4**).
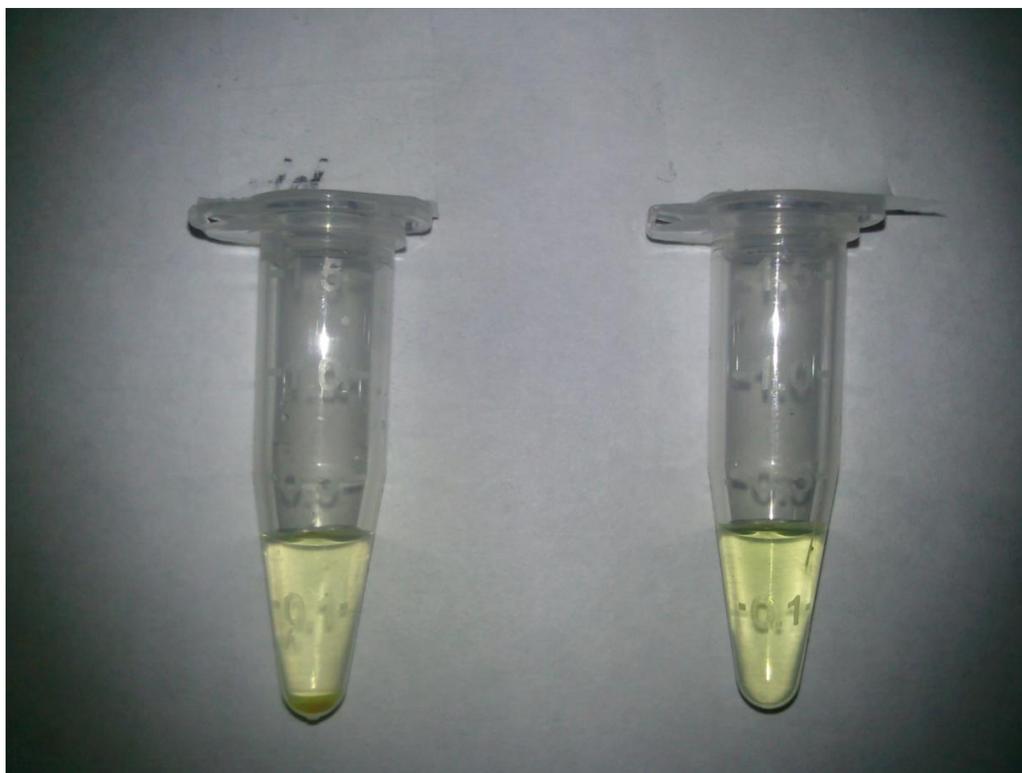
Figure 2-5.   Comparison of the dansylated human urine samples prepared with the original condition (left vial) and the optimized condition (right vial). The original condition gave a two-layer solution and the unknown in the left vial was difficult

2.3.8   LC-FT-ICR-MS

The LC-MS analysis was performed using an Agilent 1100 series binary system (Agilent Palo Alto, CA) connected to a 9.4 T Apex-Qe FT-ICR-MS (Bruker, Billerica, MA). The MS data were acquired in the positive ion mode. An Agilent Zorbax Eclipse Plus C18 column (2.1 mm × 100 mm, 1.8 μm particle size, New Castle, DE) was used in the chromatographic separations. The mobile phase A was 0.1% formic acid in ACN/H$_2$O (5/95, v/v), and the mobile phase B was 0.1% formic acid in ACN. All samples were analyzed with a 28-min

gradient: 0 min (20% B), 0-3.5 min (20-35% B), 3.5-18 min (45-65% B), 18-21 min (65-95% B), 21-24 min (95-99% B), and 24-28 min (99% B). The column was re-equilibrated with the initial mobile phase conditions for 15 min before injecting the next sample. The flow rate was 180 μL/min and the injection volume was 2.0 μL (**Figure 2-5**).
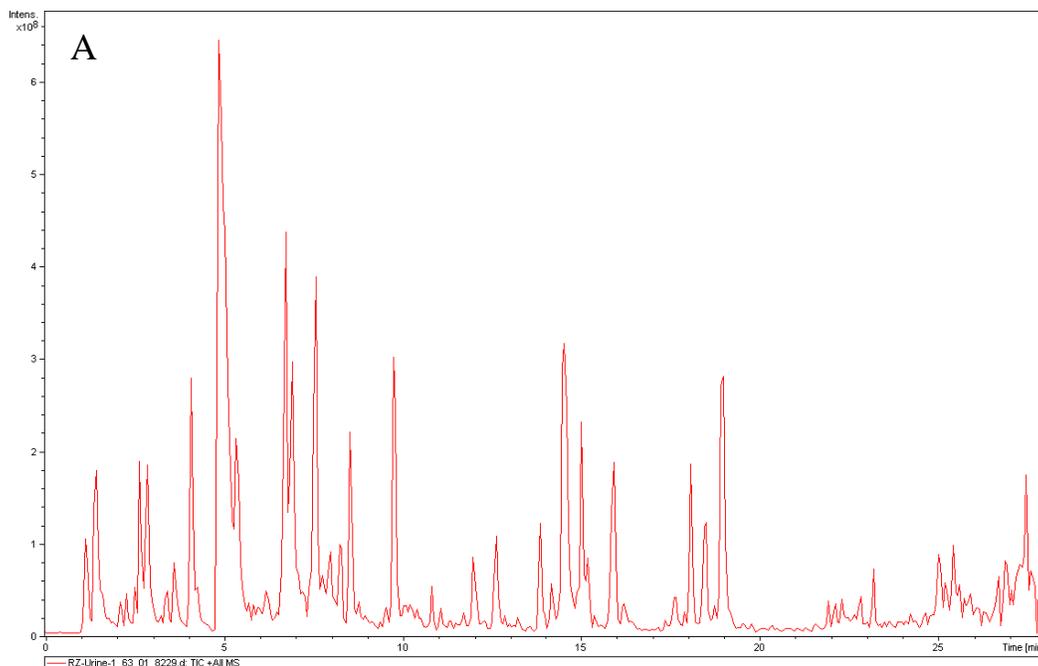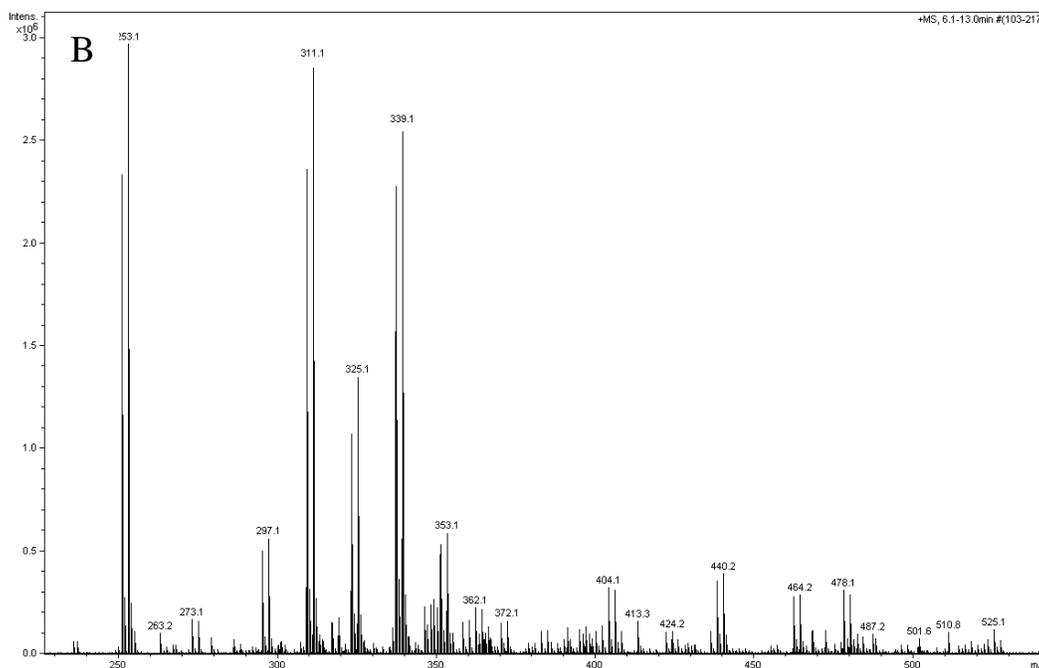
Figure 2-6.   Example of dansylated human urine sample: (A) Base-peak
ion chromatogram of the labeled urine and (B) Expanded mass window of
an average mass spectrum from 6.1to 13.0 min in chromatogram A.

2.3.9   Data Processing

The LC-FT-ICR-MS datasets were processed by an in-house developed software, IsoMS, which will be available publicly upon the publication of the paper describing the IsoMS program (manuscript in preparation). The peak pairs list and the corresponding quantitative information were analyzed by SIMCA P+ 12.

2.4   Notes

1.   HPLC grade organic solvents or purified deionized water are found to be adequate for preparing solutions for dansylation.

2.   It is recommended to use an inert gas protection for the reaction.

54

3. The solid was washed sequentially with water (2 x 10 mL) and ethanol (2 x 10 mL) without further purification. Before the next step, the intermediate should be completely dried.

4. The solid usually dissolves in 20 min and thus the reaction mixture becomes clear. If PCl_5 has partially decomposed during storage, some residues may be observed.

5. DnsCl solution in acetonitrile is stable at least one week at -20 ℃.

6. The SOP protocols developed by Dr. A.D. Souza and Dr. L. Li are available upon request from Dr. L. Li.

7. Thawed urine sample was centrifuged at 14,000 g for 10 minutes and the supernatant was used for dansylation. The dried serum sample was dissolved in 250 μL water.

8. The percentage of ACN of the reaction solution and the concentration of the buffer solution are very important for dansylation. At the optimized condition, the solution should be homogeneous which gives higher labeling efficiency.

9. After add NaOH solution, the reaction vial was incubated at 60 ℃ for another 10 minutes.

10. To reduce the impact of real samples to a C18 column, all samples should be centrifuged at 14,000 g for 10 minutes before analysis. A pre-column filter might be helpful to blood samples.

11. In the original protocol reported (**4**), a vial of the labeled sample solution and a vial of the labeled control solution were mixed to generate the

sample mixture for LC-MS analysis. In the current protocol, the product solution is homogenous and thus can be aliquoted for mixing. There is no need to prepare multiple $^{13}C_2$-labeled control solutions for mixing with individual $^{12}C_2$-labeled samples. Only one $^{13}C_2$-control solution needs to be prepared from which an aliquot can be taken for mixing with an individual $^{12}C_2$-labeled sample. This method can reduce the consumption of the valuable $^{13}C_2$-DnsCl reagent.

2.5   Acknowledgement

References

1.   Goodlett, D. R., Keller, A., Watts, J. D. et al. (2001) *Rapid Commun. Mass Spectrom.* 15, 1214–1221.

2.   Lemmel, C.,Weik, S.,Eberle, U., et al. (2004) *Nat. Biotechnol.* **22**, 450–454.

3.   Fricker, L. D.,Che, F-Y (2002) *Anal. Chem.* **74**, 3190–3198.

4.   Guo, K., Li, L. (2009) *Anal. Chem.* **81**, 3919–3932.

5.   Fukusaki, E., Harada, K.,Bamba, T.; et al. (2005) *J. Biosci. Bioeng.* **99**, 75–77.

6.   Shortreed, M.R., Lamos, S.M., Frey, B.L.,et al. (2006) *Anal. Chem.* **78**, 6398–6403.

7.   Yang, W.-C., Sedlak, M.,Regnier, et al. (2008) *Anal. Chem.* **80**, 9508–9516.

8.   Guo, K., Li, L. (2010) *Anal. Chem.* **82**, 8789–8793.

9. Chen, R. F. (1968) *Anal. Biochem,* **25***,* 412–416.

10. Venn, R. F., Basford, J. M., Curtis, C. C. (1978) *Anal. Biochem.* **87***,* 278–282.

11. Tapuhi, Y., Schmidt, D.E., Linder, W., et al. (1981) *Anal. Biochem,* **115***,*123–129.

12. Flengsrud, R. (1976) *Anal. Biochem,* **76***,* 547–550.

13. Guo, K., Peng, J., Zhou, R. et al (2011) *J. Chromatogr. A* **1218**, 3689-3694.

14. Zheng, J., Dixon, R.A, Li, L. (2012) *Anal. Chem.* **84**, 10802–10811.

15. Wu, Y., Li, L. (2013) *Anal. Chem.* **85**, 5755–5763.

Chapter III: 5-Diethylamino-naphthalene-1-sulfonyl chloride

(DensCl): a Novel Triplex Isotope Labeling Reagent for Quantitative

Metabolome Analysis by Liquid Chromatography Mass Spectrometry*

3.1   Introduction

Metabolomics involves the global identification and quantification of small molecules in a given biological system. Despite significant efforts in developing new and improved methods for metabolome analysis, high throughput quantitative metabolic profiling with comprehensive metabolome coverage (i.e., thousands of metabolites) is still a technical challenge.[1-3] Among different reported methods, high performance differential isotope labeling combined with liquid chromatography mass spectrometry (LC-MS) is a promising method for quantitative analysis of a large number of metabolites.[4,5] This method is based on the use of rationally designed chemical labeling reactions to introduce an isotope tag to the common functional group of metabolites with concurrent enhancement in metabolite separation, MS detection and data processing. While a number of isotope labeling reagents have been reported for metabolite or metabolome analysis,[6-8] including multiplex isotope reagents[9], only a few reagents can provide high performance separation and detection for comprehensive metabolome analysis.[4,5,10-14].

In a previous study, we reported a LC-MS strategy that utilized $^{12}C_2$-/$^{13}C_2$-isotope dansyl chloride (DnsCl) to selectively introduce a differential isotope tag to amines and phenolic compounds for metabolic profiling.[4] After dansylation, the

hydrophobic aromatic structure of the dansyl group improves the separation of labeled metabolites on a reversed phase liquid chromatography (RPLC) column. This allows the analysis of polar and even ionic metabolites using RPLC, rendering the possibility of analyzing the amine and phenol sub-metabolome with diverse physiochemical properties using only one LC-MS condition. Additionally, the dimethylamino moiety can be readily protonated in electrospray ionization (ESI). The detection sensitivity of the dansyl labeled metabolites is usually enhanced by 10 to 1000-fold, compared to the unlabeled counterparts. This dansyl-labeling LC-MS method has been used in metabolic profiling of a variety of biological samples in discovering potential metabolite biomarkers of diseases and cellular metabolomics for biological studies.[15-18]

In applying the isotope labeling LC-MS method for quantitative metabolome profiling, the bottleneck of sample throughput lies in the LC-MS analysis step, as the sample preparation process including isotope labeling can be multiplexed. For dansylation labeling, there are only two possible isotope forms of DnsCl (i.e., ^{12}C-methyl and ^{13}C-methyl group) that can be readily synthesized; deuterium-code isotopologues are not ideal as they would display isotopic effect that they elute faster than their counterparts in RPLC separation.[19] In this work, we report a new dansyl-like reagent, 5-diethylamino-naphthalene-1-sulfonyl chloride (DensCl), for profiling the amine and phenol sub-metabolome. DensCl is structurally similar to DnsCl, in which the two methyl groups attached to the naphthyl amine are replaced by two ethyl groups (see Figure 3-1). Compared to DnsCl, three isotope-encoded forms of DensCl with ^{12}C$_2$-ethyl, ^{12}C^{13}C-ethyl and

$^{13}C_2$-ethyl groups can be synthesized, producing three differential isotope-encoded reagents, i.e., $^{12}C_4$-, $^{12}C_2{}^{13}C_2$-, and $^{13}C_4$-DensCl, respectively. Like the duplex isotope reagents of DnsCl having a nominal mass difference of 2-Da between the two reagents, the triplex isotope reagents of DensCl also have a 2-Da mass difference from one reagent to another. The use of the DensCl reagents for labeling metabolites increases the sample throughput by allowing the analysis of a mixture prepared from mixing the $^{13}C_4$-Dens-labeled control with $^{12}C_4$-Dens-labeled sample #n and $^{12}C_2{}^{13}C_2$-Dens-labeled sample #(n+1) in one LC-MS run. Essentially, the triplex reagents double the analysis speed. Additionally, Dens-labeling increases the detection sensitivity or the number of detectable metabolites in a biological sample, compared to Dns-labeling LC-MS. The application and analytical performance of this Dens-labeling LC-MS method for metabolome profiling are demonstrated in the analysis of urinary metabolite concentration changes in urine samples collected from a healthy individual over a period of 14 days. To our knowledge, this is the first report of multiplex reagents that offer simultaneous enhancement in sample throughput, chromatographic separation and MS detection with no isotopic effect.

3.2   Experimental Section

3.2.1   Chemicals and reagents

All chemicals and standards were purchased from Sigma-Aldrich (Oakville, ON) unless otherwise noted. LC-MS grade water and acetonitrile (ACN) were purchased from Thermo Fisher Scientific (Edmonton, AB). $^{12}C_2{}^{13}C_2$-

ethyl iodide and [13]C_2-ethyl iodide were from Cambridge Isotopes Laboratories (Andover, MA). Twenty primary amino acid stock solutions (20 mM each) were prepared in water. A pooled amino acid solution (1 mM each) was obtained through mixing aliquots of the stock solutions. Stock solutions of five dansylated amino acid standards (1 mM each) were prepared in ACN/H2O (50:50, *v/v*). A pooled dansylated amino acid standard solution (200 μM each) was prepared by mixing aliquots of the stock solutions. A calibration solution for FTICR-MS was prepared by 400-fold dilution of the pooled dansylated amino acid standard solution to 0.5 μM of each standard. All standard solutions were stored at -20 ℃.

3.2.2   Synthesis of diethylamino-1-naphthalenesulfonyl chloride (DensCl)

Figure 3-1A shows the two-step reactions for the synthesis of DensCl according to the literature with some modifications.[20] To a solution of 1.73 g 5-amino-1-naphthalenesulfonic acid in 100 mL anhydrous THF, 1.60 g 60% NaH and 4.68 g [12]C_2-ethyl iodide or its isotopic counterparts were added to the mixture at 0 ℃. After refluxing for 3 h, 10 mL H_2O was slowly added to quench the excess NaH. The reaction solution was acidified by 36% HCl to give a yellow precipitate, 5-diethylamino-naphthalene-1-sulfonic acid. This intermediate was mixed with 6.8 g PCl_5 in 10 mL toluene and stirred at room temperature until the solid disappeared. The solution was poured into 50 mL ice water and extracted twice with 50 mL EtOAc. The organic layer was dried with anhydrous Na_2SO_4. After evaporation of the solvent, the residue was purified by flash chromatography (95/5 hexane–EtOAc) to give 2.19 g final product with an overall yield of 76% over the two steps. The structures of [13]C_4-DensCl and the isotopic

analogs were verified by NMR and the purity was confirmed by LC-MS. All products were stored at -20 ℃.

3.2.3 DensCl labeling reaction

Figure 3-1B shows the reaction schemes for derivatization of amines and phenols using DensCl. For the labeling reaction, a 50 µL aliquot of a sample solution was mixed with 25 µL of 500 mM $NaHCO_3/Na_2CO_3$ buffer, and then added 75 µL of 50 mM $^{12}C_4$-DensCl or other isotope reagents in ACN. The derivatization reaction solution was incubated at 60 ℃ for 60 min. After cooling down, the solution in the vial was spun down and 10 µL of 250 mM NaOH were added and incubated for another 10 min to quench the excess labeling reagent. The reaction solution was acidified by adding 50 µL of 425 mM formic acid. Finally, equal volumes of $^{12}C_4$-, $^{12}C_2{}^{13}C_2$-, and $^{13}C_4$-Dens-labeled samples were mixed to give the final solution for LC-MS analysis.
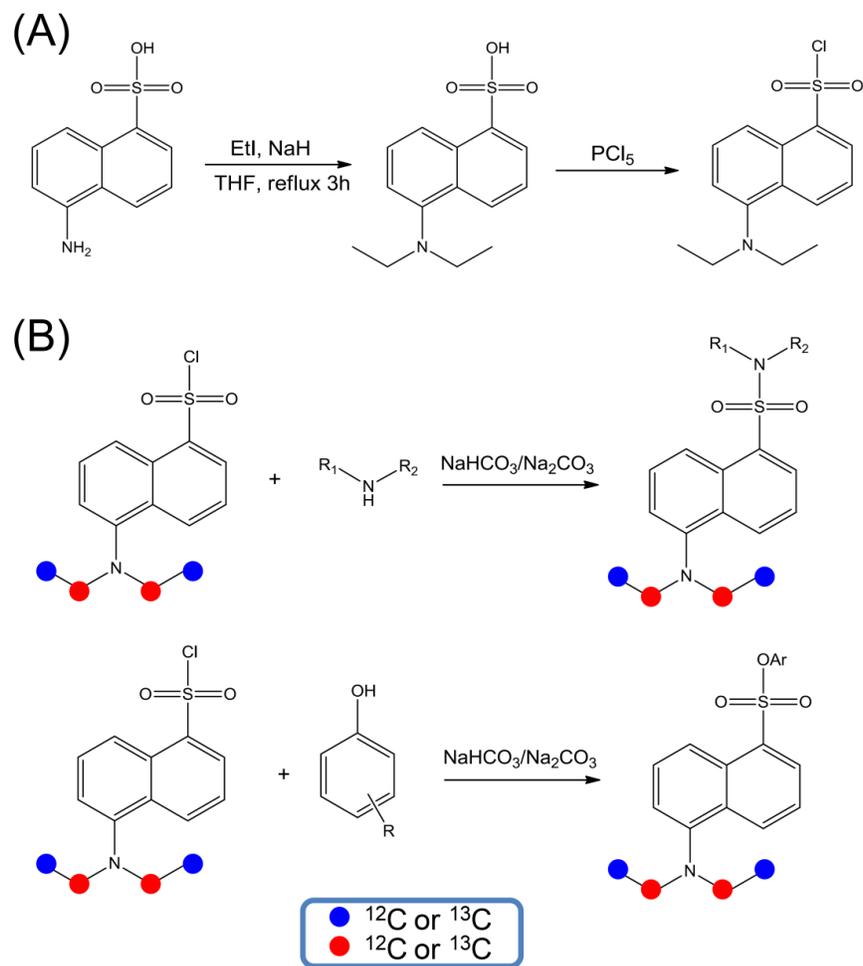
Figure 3-1.   Reaction schemes for (A) synthesis of the isotope labeling reagent, 5-(diethylamino)-naphthalene-1-sulfonyl chloride (DensCl), and (B) Dens-derivatization with amines and phenols. The triplex reagents consist of three possible isotopic forms: $^{12}C_4$-DensCl with all "red" and "blue" carbons in the ^{12}C form, $^{13}C_2{}^{12}C_2$-DensCl with two "red" carbons as ^{12}C and two "blue" carbons as ^{13}C, and $^{13}C_4$-DensCl with all the carbons in the ^{13}C form.

3.2.4 Linear dynamic range study

To study the linear dynamic range for relative quantification of metabolites using Dens-labeling LC-MS, three amino acid standard solutions were prepared by diluting the pooled stock solution 1-, 5-, or 10-fold. Each solution was then divided into three aliquots. The amino acid solutions were derivatized by the triplex DensCl reagents. The labeled solutions were combined in the concentration ratios of 1:10:10, 1:5:5, 1:1:1, 5:1:1, and 10:1:1 to form a series of mixtures for LC-MS analysis.

3.2.5 Urine sample analysis

A second morning urine sample was collected daily from a healthy volunteer in 14 consecutive days. An informed consent was obtained from the individual volunteer and ethics approval was obtained from the University of Alberta. Within one hour of urine collection, 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. A pooled urine sample was prepared by combining the aliquots of daily urine samples. The urine samples were stored at -80 ℃.

For relative quantification of metabolites in daily urine samples, the individual samples were separately derivatized by using $^{12}C_4$-DensCl or $^{12}C_2{}^{13}C_2$-DensCl in alternation (i.e., day #n urine was labeled with $^{12}C_4$-DensCl where n is an odd number and day #(n+1) urine was labeled with $^{12}C_2{}^{13}C_2$-DensCl). The pooled urine sample was labeled with $^{13}C_4$-DensCl. After the labeling reactions, a

mixture of $^{13}C_4$-Dens-labeled pooled sample, $^{12}C_4$-Dens-labeled day #n sample and $^{12}C_2^{13}C_2$-Dens-labeled day #(n+1) sample was prepared for LC-MS analysis.

3.2.6 LC-MS

LC-MS analysis was performed using an Agilent 1100 series binary system (Palo Alto, CA) connected to a Bruker Apex-Qe 9.4-T Fourier transform ion cyclotron resonance (FTICR) MS (Bruker, Billerica, MA) equipped with electrospray ionization (ESI). Chromatographic separation was performed on an Agilent Zorbax Eclipse Plus C18 column (2.1 mm × 100 mm, 1.8 μm particle size). The mobile phase A was 0.1% formic acid in acetonitrile/water (5/95, *v/v*) and the mobile phase B was 0.1% formic acid in acetonitrile. All the samples were analyzed using a 25-min gradient: 0 min (20% B), 0-3.5 min (20-45% B), 3.5-18 min (45-65% B), 18-21 min (65-95% B), 21-24 min (95-99% B), 24-25 min (99% B). The column was re-equilibrated with the initial mobile phase conditions for 15 min before injecting the next sample. The flow rate was 180 μL/min and the injection volume was 2.0 μL. The eluted solution from the LC column was split at a ratio of 3:1 and was introduced into MS at 60 μL/min. All MS spectra were obtained in the positive ion mode. The MS conditions used for FTICR-MS were as follows: nitrogen nebulizer gas: 2.3 L/min, dry gas flow: 7.0 L/min, dry temperature: 190 °C, capillary voltage: 4200 V, end plate: 3700 V, acquisition size: 256 k, scan range: 200-1000, and ion accumulation time: 1 sec.

3.2.7 Data Analysis

The LC-MS data were processed using an in-house developed software called IsoMS. First, the mass spectral peaks were picked using the Bruker DataAnalysis software 4.0. The $^{12}C/^{13}C$-peak pairs were found by their accurate mass difference of 2.0067 Da within the 10 ppm mass tolerance window. Redundant peaks of the same metabolite including natural isotope peaks and adducts peaks, such as sodium adducts, dimer peaks, and multiply charged peaks, were removed by the IsoMS software. The retention time and accurate mass were employed to align the extracted peak-pair data from LC-MS. Heatmap comparison and multivariate analysis were performed by Metaboanalyst (www.metaboanalyst.ca).[21]

For putative metabolite identification based on accurate mass matches with the metabolites in a database, the MyCompoundID program (www.mycompoundid.org)[22] was used to search against the Human Metabolome Database (www.hmdb.ca) consisting of 8,021 known human endogenous metabolites and the Evidence-based Metabolome Library (EML) consisting of about 375,809 predicted human metabolites. The mass accuracy window used for database search was set at 5 ppm.

3.3 Results and Discussion

3.3.1 Derivatization Reaction of DensCl

The chemistry of dansyl chloride (DnsCl) for labeling amino acids has been well known for several decades.[23, 24] DensCl has a similar structure as DnsCl

except that the naphthyl amine of DensCl is attached to two ethyl groups, instead of two methyl groups (see Figure 3-1). Despite the structural similarity, the reactivity of DensCl was found to be different from DnsCl. At the beginning of this work, the original dansylation labeling protocol was followed in DensCl derivatization. However, the MS signal intensities of Dens-labeled standard metabolites (e.g., amino acids) were much lower than those from dansylation labeling.

To improve the DensCl labeling, different experimental conditions were carefully examined. First of all, while 1 M $NaHCO_3/Na_2CO_3$ (v/v 1:1) buffer was used in the original dansylation reaction, we found that the use of 250 mM buffer was sufficient for both Dns- and Dens-labeling reactions. Lower buffer concentration decreases the contamination of salts in the ESI interface. Secondly, we found that the water content of the reaction solution plays an important role in the labeling reaction. In both Dns- and Dens-labeling reactions, an unknown precipitate was generated using the original dansylation labeling conditions and the organic phase partially separated from the aqueous layer. After adjusting the water content from 67% to 50%, the resulting solution became clear and homogenous. Because of the formation of precipitates during the labeling, in the original protocol, the same volume of the pooled sample as that of the individual sample was taken for $^{12}C_2$-DnsCl labeling to reduce the volume effect on labeling. With the new protocol, homogenous solution is formed and thus we can label the pooled sample in a larger volume and the aliquot of the labeled pooled sample can be taken to mix with the differentially labeled individual sample for LC-MS

analysis. This reduces the isotope reagent consumption and the overall time for sample labeling. Finally, for DensCl labeling, we used aqueous NaOH to quench the excess DensCl, instead of methyl amine used in the original dansylation protocol. In this way, the quenched product, 5-(diethylamino)-naphthalene-1-sulfonic acid, is polar and eluted at the beginning of the chromatogram, which avoids the suppression of the analyte in the separation.

Panels A and B of Figure 3-2 show the base-peak ion chromatograms of Dens-labeled amino acid standards obtained by following the original dansylation conditions and the new protocol, respectively. It is clear that signal intensities for most amino acids are significantly increased using the new protocol. It should be noted that the new protocol is also applicable for dansylation labeling and a comparison of two base-peak ion chromatograms of Dns-labeled amino acid standards derived from the different protocols is shown in panels C and D of Figure 3-2. As shown by panels B and D, the Dens-labeled products generally give higher intensities than those of the corresponding Dns-labeled amino acids. More discussion on sensitivity comparison of the two labeling methods will be described in the last section.
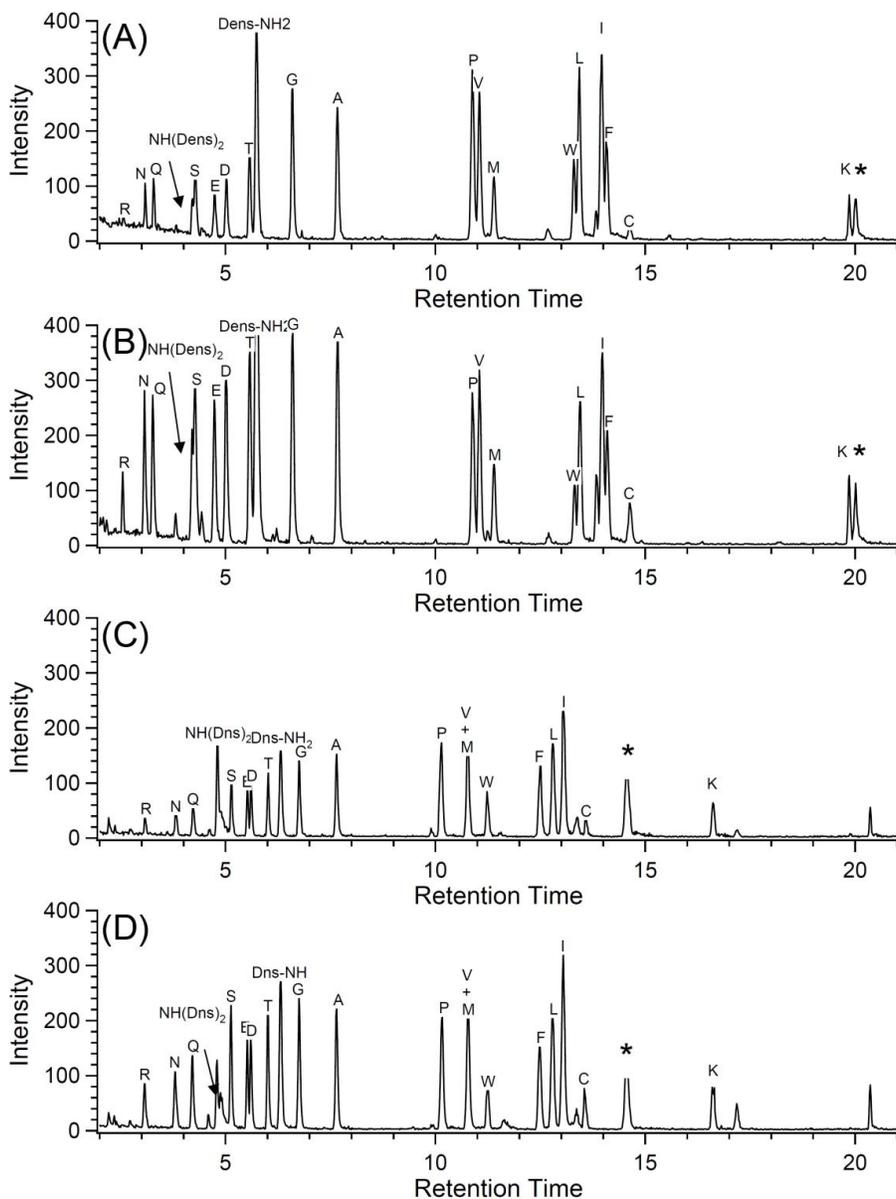
Figure 3-2. Base-peak ion chromatograms of (A) Dens-labeled amino acids prepared by using the original dansylation labeling condition, (B) Dens-labeled amino acids prepared by using the new protocol, (C) Dns-labeled amino acids prepared by using the original dansylation condition, and (D) Dns-labeled amino acids prepared by using the new protocol. Peaks labeled with "*" are the by-products of the labeling reactions.

3.3.2 Relative Quantification

Relative quantification of an individual metabolite in two comparative samples is based on the intensity ratio of the peak pair of the same metabolite differentially labeled in the corresponding samples. The overall dynamic range of the peak pair ratio is dependent on the detection dynamic range of the mass spectrometer. This range is typically about 100 in FTICR MS. We have examined the dynamic range of linear response for the Dens-labeling LC-MS method using amino acid (AA) standards as the model compounds. In our experiment, $^{12}C_4$-DensCl derivatized amino acid standards (i.e., $^{12}C_4$-DensAAs) were added into $^{12}C_2{}^{13}C_2$-DensCl and $^{13}C_4$-DensCl derivatized amino acid standards to make the mixture solutions at the ratios of 1:10:10, 1:5:5, 1:1:1, 5:1:1, and 10:1:1 ($^{12}C_4$-DensAA:$^{12}C_2{}^{13}C_2$-DensAA:$^{13}C_4$-DensAA). Each solution was injected into LC-MS for analysis.

A labeled amino acid displays characteristic triplet peaks in the mass spectrum; one example (glutamine or Gln) is shown in Figure 3-3. The ratios of two peak pairs, $^{12}C_4$-DensGln:$^{12}C_2{}^{13}C_2$-DensGln referred as the 1st pair and $^{12}C_4$-DensGln:$^{13}C_4$-DensGln as the 2nd pair, can be determined by the intensities of the mass spectral peaks from the differentially labeled amino acid. The average ratios of the 1st and 2nd pairs from experimental triplicate runs were determined to be 0.010 (1st pair), 0.0099 (2nd pair); 0.194, 0.191; 0.947, 0.935; 5.11, 4.94; and 10.1, 9.78, with relative standard derivations of 5.2%, 5.3%; 7.9%, 9.5%; 8.0%, 8.8%; 10.4%, 9.8%; and 8.9%, 8.3%, for the 1:10:10, 1:5:5, 1:1:1, 5:5:1, and 10:1:1 mixtures, respectively. Figure 3-4 shows the linear regression plots of the

measured values *vs.* the expected values. The average *R*-squared values obtained for 20 amino acids were 0.999, indicating a good correlation of the experimental data with the theoretical or expected ratios.
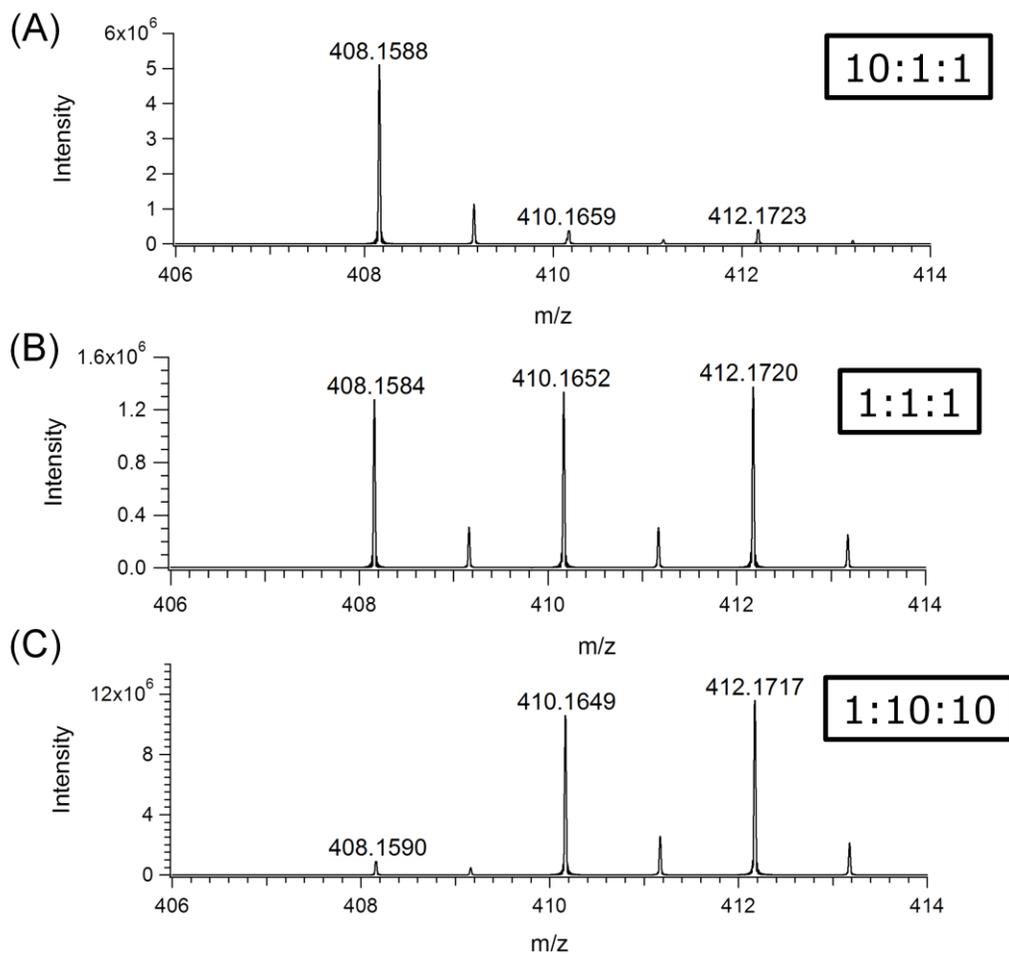
Figure 3-3. Expanded mass spectra of isotope-differentially Dens-labeled glutamine from samples with concentration ratios of 10:1:1, 1:1:1 and 1:10:10.
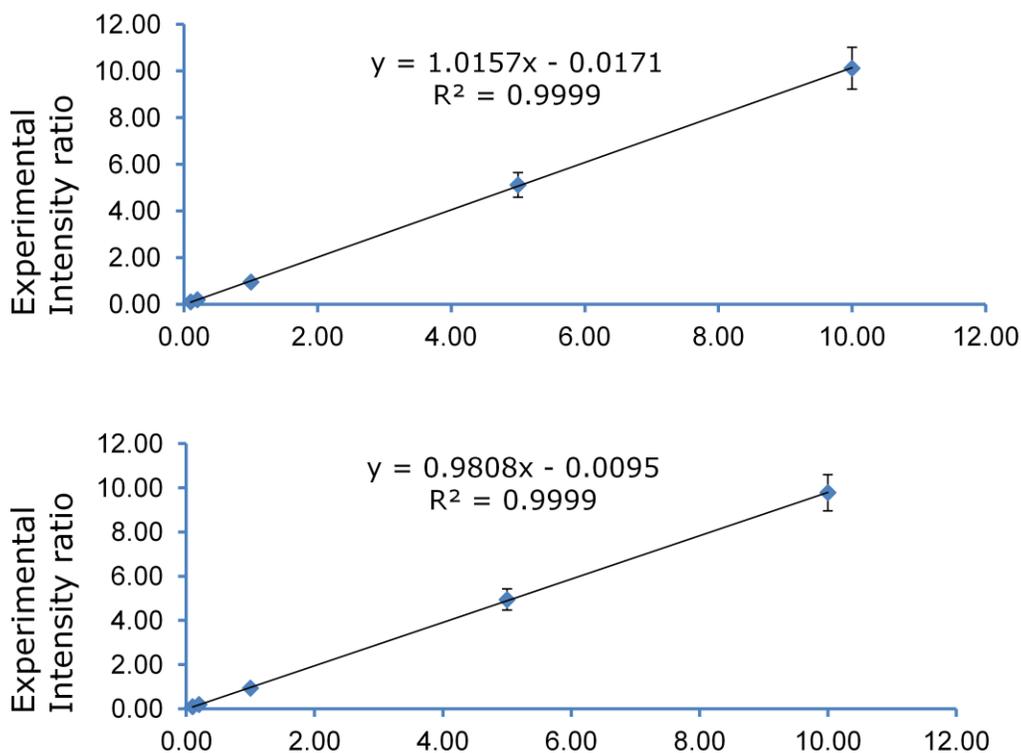
Figure 3-4. Linear regression plots of (A) the ratio of $^{12}C_2{}^{12}C_2$- and $^{12}C_4$-peak intensities and (B) the ratio of $^{13}C_4$- and $^{12}C_4$- peak intensities.

The above results indicate that the Dens-labeling LC-FTICR MS method can be used to provide relative quantification of metabolites in comparative samples with relative concentration changes of 10:1 to 1:10. This is adequate for most comparative metabolomics applications where the metabolite concentration changes are usually below 10-fold. It should be noted that the absolute concentrations of different between metabolites can be much larger than 10-fold. Because of this wide range of absolute concentration difference, in our metabolome profiling workflow, a pooled sample prepared by combining a small aliquot of individual samples is used as the reference sample serving as a global internal standard. This pooled sample and the individual samples are subjected to

differential isotope labeling, followed by mixing and LC-MS analysis. In this way, the concentration of a given metabolite in an individual sample is not too different from that in the pooled sample.

3.3.3 Urine Metabolome Profiling: Sample Normalization

Quantitative metabolome profiling involves the determination of metabolite changes among different comparative samples. To illustrate the applicability of the Dens-labeling LC-MS method in this important area of research, we have examined the urine metabolite changes of a healthy volunteer over a period of two weeks. Urine samples were collected on a daily basis for 14 days without control of diet. Because the concentrations of endogenous metabolites in urine samples are highly dependent on water consumption, normalization of the total urine concentration is needed to more accurately measure the concentration changes of individual metabolites. Creatinine levels in urine have been commonly used to normalize the urine samples. Recently our group has developed a rapid LC-UV method that provides a more accurate and convenient means of normalizing urine concentrations for MS-based metabolomics work.[25] In this work, we applied this method to normalize all the urine samples prior to sample mixing and LC-MS analysis.

3.3.4 Urine Metabolome Profiling: Relative Concentration Changes

To determine the relative metabolite concentration changes among the daily urine samples, a pooled urine sample was prepared by mixing equal volumes of all normalized daily urine samples. This pooled sample was labeled

with $^{13}C_4$-DensCl. Each daily urine sample was separately labeled with $^{12}C_4$-DensCl or $^{12}C_2{}^{13}C_2$-DensCl. We then mixed an aliquot of $^{13}C_4$-Dens-labeled pooled sample with $^{12}C_4$-Dens-labeled day #n urine and $^{12}C_2{}^{13}C_2$-Dens-labeled day #(n+1) urine, followed by injecting this mixture into LC-MS for analysis. A representative base-peak ion chromatogram generated from the triple-labeled mixture is shown in Figure 3-5A. A number of chromatographic peaks are obtained across the entire elution time window, indicating that a wide separation time in RPLC can be used for separating and analyzing the complex sample.

Panels B and C in Figure 3-5 show two representative mass spectra displaying the molecular ion regions of two triple-labeled metabolites (Panel B from a metabolite, threonine, with one tag attached and Panel C from an unknown metabolite labeled with two tags). The ratios of the 1st pair (day #3 *vs.* pool) and 2nd pair (day #4 *vs.* pool) can be determined from the mass spectra, while only one ratio value would be determined if the duplex reagents were used. The triplex reagents can at least halve the LC-MS run time required for analyzing all the samples in a metabolomics study. If the column equilibrium time between two runs is accounted for, the time saving is even more significant. As the LC-MS analysis is the bottleneck in the metabolome profiling workflow, the overall sample throughput can be more than doubled.
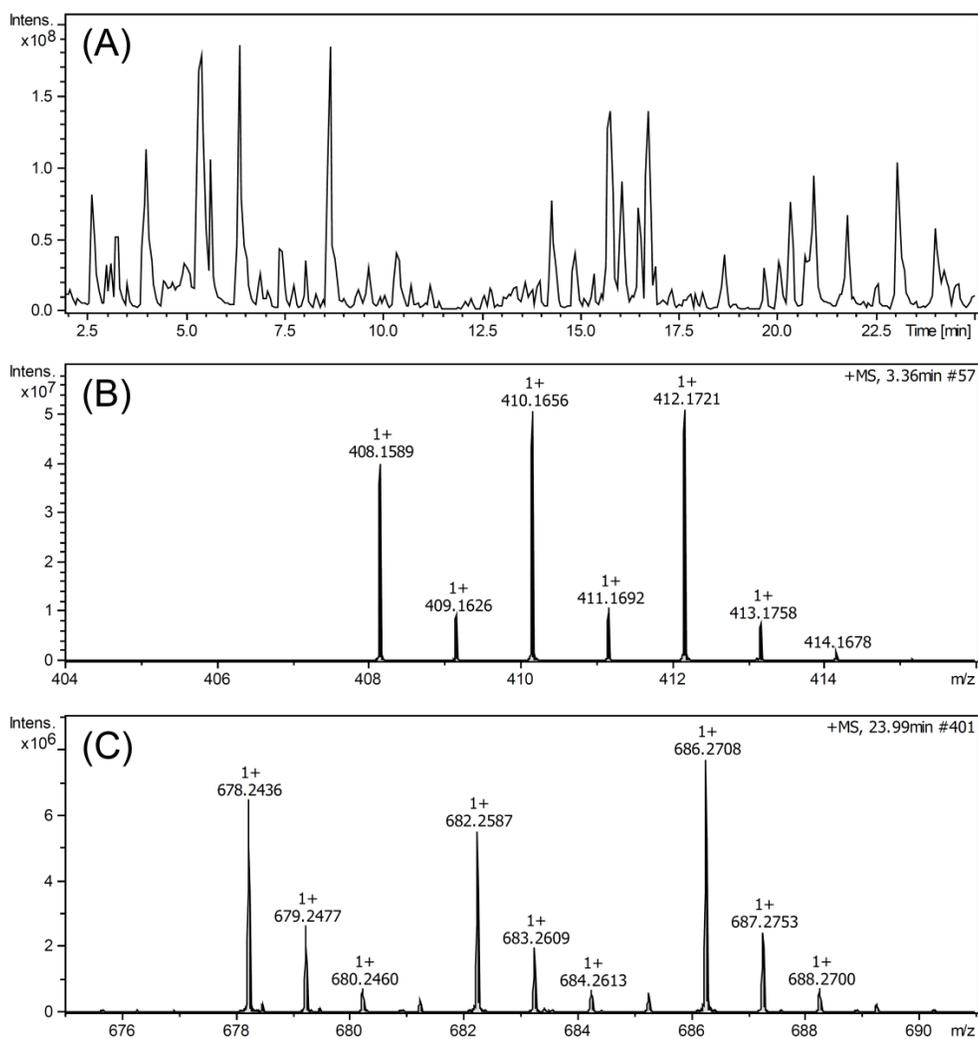
Figure 3-5.   (A) Base-peak ion chromatogram of a mixture consisting of $^{13}C_4$-labeled pooled-urine, $^{12}C_4$-labeled "Day-3" urine and $^{12}C_2{}^{13}C_4$-labeled "Day-4" urine of an healthy individual. Expanded mass spectra of (B) Dens-labeled threonine that has one-tag and one-charge and (C) (Dens)$_2$-labeled unknown metabolite that has two-charge and one-tag from the chromatogram (A).

For the 14-day urine samples, we did an experimental triplicate for each sample in order to gauge the reproducibility of the method. In this regard, principal components analysis (PCA) is a useful tool to identify the distribution patterns in the datasets and determine the extent of clustering from similar observations (e.g., replicate runs of the same sample). In the peak list table generated from all the LC-MS runs, 303 peak pairs or putative metabolites were consistently observed with ratios detectable in at least 50% of the samples. These peak pairs were analyzed by PCA. Figure 3-6A shows the score plot displaying the distribution of the data. As Figure 3-6A shows, the samples collected from different days are randomly scattered, indicating day-to-day variations of the metabolite profile of an individual. More importantly, the experimental triplicate data of the same daily sample are clustered together, indicating that the triplex reagents can be used to generate reproducible results. A heatmap of the individual metabolites is shown in Figure 3-6B. Since a heatmap is able to display concentration changes of detectable metabolites in all observations, it allows us to rapidly inspect variations of the quantification in experimental triplicate data. Although the relative concentrations of individual metabolites have some variations from different days, most of them are reproducible within the replicate dataset, which is consistent with the PCA score plot results.

(A) Score Plot

(B)
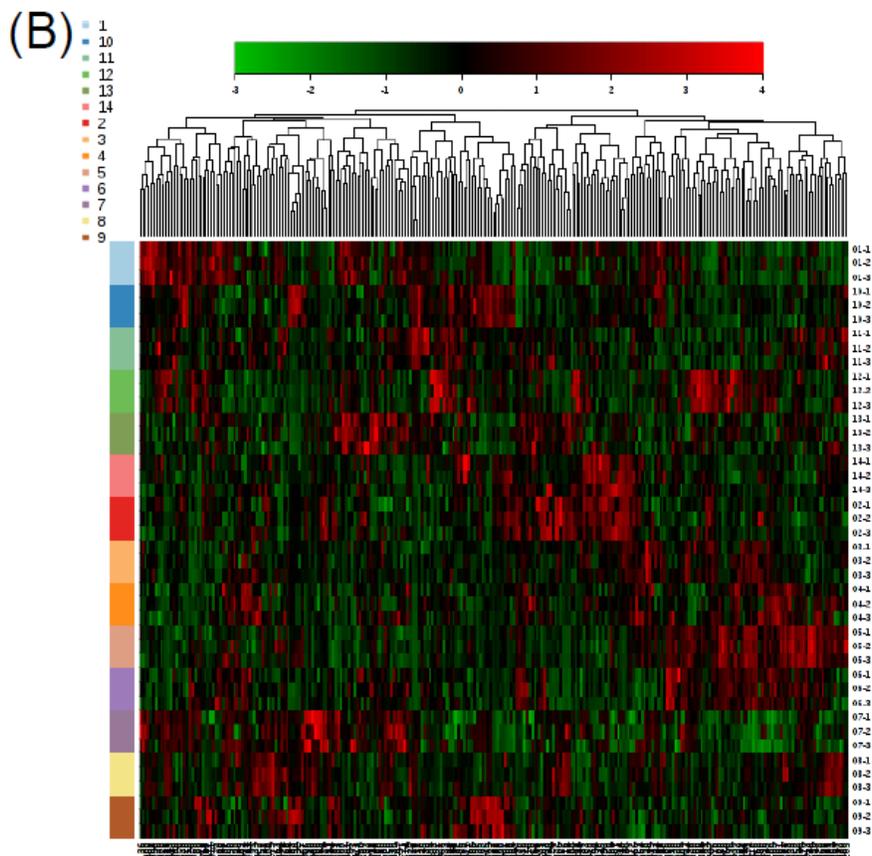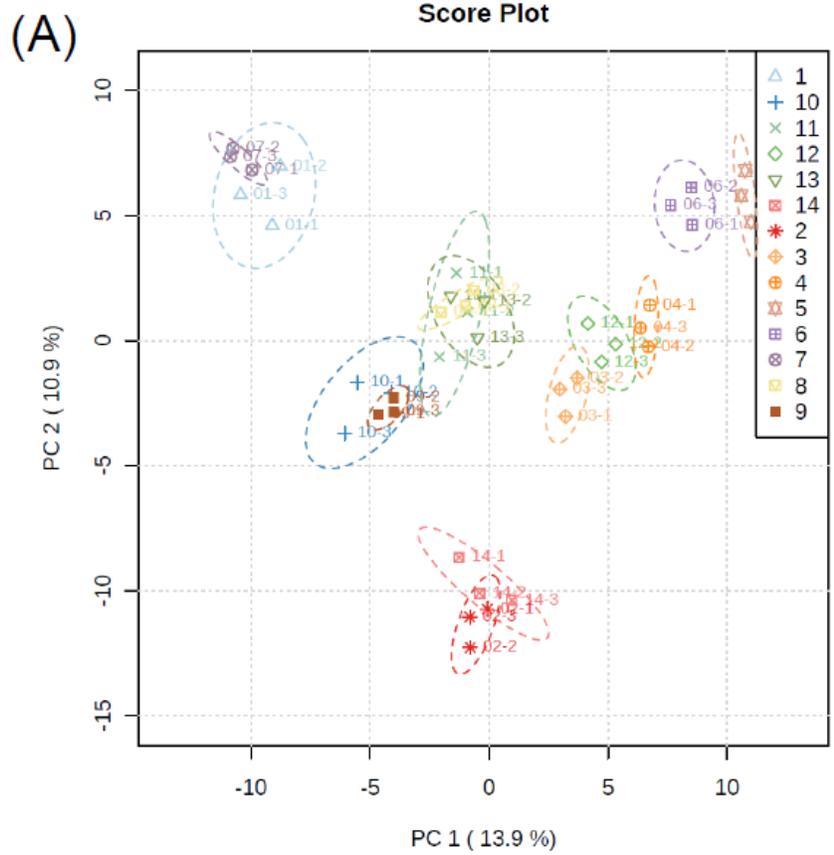
Figure 3-6.   (A) PCA score plot and (B) heatmap generated by the peak

ratios of peak pairs detected from human urine samples collected from a

healthy individual in 14 consequent days.

3.3.5   Urine Metabolome Profiling: Absolute Concentration Changes

If one is interested in determining the absolute concentration of an

individual metabolite in a sample, a labeled standard with known concentration

can be spiked into the labeled pooled sample.[4] The peak ratio of the analyte in the

pooled sample and the spiked standard can be used for determining the

concentration of the analyte. Because of a limited dynamic range for relative

quantification (i.e., 10:1 to 1:10), we found that it was necessary to adjust the

absolute concentration of the spiked standard to ensure that the peak ratio of the

analyte and the standard fell into the linear range for accurate absolute

quantification. In this work, we measured the absolute concentrations of 20 amino

acids in the pooled sample in order to examine how the absolute concentration

changes over the 14-day period for a healthy individual. In principle, all the

known metabolites, as long as pure standards are available, can be analyzed in the

same manner as in the analysis of the amino acids. Table 3-1 shows the results

obtained from the measurement of the absolute concentrations of 20 amino acids

in the pooled urine. The concentrations of the amino acids range from 2.68 μM to

1.70 μM. The run-to-run RSD ranges from 1.5% to 22.9% with an average of

6.3%.

Table 3-1. Absolute concentration of 20 amino acid labeled by DensCl.

Amino acids	Spiked AA (μM)	Retention Time	AA conc. In pooled urine	Average of AA in urine	RSD (triplicate)
Dns-His	284.37	19.49	93.26	87.43	6.72%
			81.59		
			88.63		
Dns-Asn	49.86	3.66	48.02	48.33	3.95%
			48.65		
			45.07		
Dns-Gln	178.56	4.02	163.48	160.58	4.61%
			157.68		
			148.80		
Dns-Ser	194.02	5.17	118.97	118.27	5.87%
			117.57		
			106.29		
Dns-Glu	9.42	5.83	7.29	7.26	1.53%
			7.23		
			7.07		
Dns-Asp	73.64	6.01	2.74	2.68	4.74%
			2.61		
			2.49		
Dns-Thr	55.88	6.61	57.45	57.75	3.10%
			58.06		
			54.70		
Dns-Gly	1048.59	7.58	1820.80	1703.66	10.41%
			1586.51		
			1473.04		
Dns-Ala	76.38	8.66	55.31	59.15	8.49%
			62.99		
			53.54		
Dns-Pro	3.60	11.61	3.40	3.38	1.49%
			3.35		
			3.30		
Dns-Val	14.52	12.33	12.61	12.56	5.15%
			12.51		
			11.45		
Dns-Met	6.95	12.45	4.13	4.11	4.23%
			4.08		

			3.81		
Dns-Ile	3.12	14.62	3.31	3.26	4.73%
			3.21		
			3.01		
Dns-Leu	9.35	14.92	9.40	9.50	5.78%
			9.60		
			8.56		
Dns-Phe	17.74	14.26	17.38	17.26	2.21%
			17.13		
			16.63		
Dns-Trp	22.21	13.0	26.21	25.18	7.52%
			24.15		
			22.42		
Dns-Cys	67.60	15.7	145.16	121.78	22.93%
			98.40		
			95.36		
Dns2-Lys	57.01	18.95	48.08	46.01	7.94%
			43.94		
			40.80		

After the measurement of the absolute concentrations of 20 amino acids in the pooled urine sample, their concentrations in the individual urine samples can be readily determined based on the peak ratios of the triple-labeled peak pairs and the amino acid concentrations in the pooled sample. Figure 3-7 shows a plot of day-to-day concentration changes of the amino acids in urine samples. The error bar shows the standard derivation from the experimental triplicates. The concentration of 20 amino acids ranges from a few μM to a few mM, representing over 1000-fold difference in absolute concentration. On the other hand, as shown in Figure 3-7, the concentration variations of individual amino acids over the 14-day period were usually less than 2-fold with the exception of the Day 7 sample (as large as ~6-fold change), which is still covered by the linear dynamic range of

the Dens-labeling LC-MS method (i.e., 1:10 to 10:1). The run-to-run RSD of all the measurements ranges from 3.9% to 10.1% with an average of 6.2%, demonstrating the good reproducibility of the method.
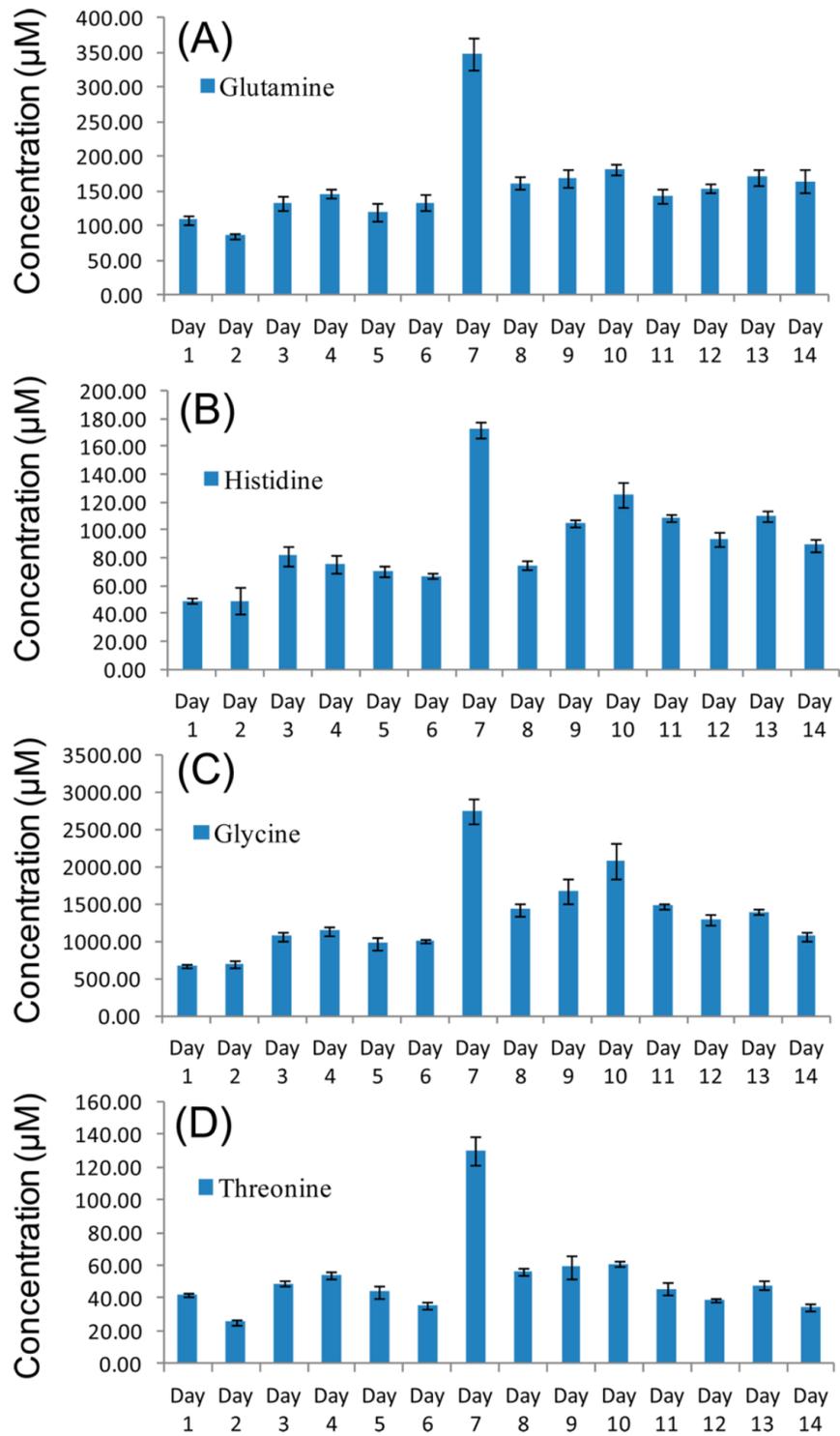
Figure 3-7. Day-to-day absolute concentration variations of four amino acids in urine samples collected from a healthy individual in 14 consequent days.

While we cannot draw much biological information from the measurement of amino acid concentrations of only one individual, the results of the 14-day data, such as those shown in Figure 3-5, suggest that the Dens-labeling method has precision required to reveal subtle changes in metabolite concentration. In this case, we do not know the cause of a great variation of amino acid concentrations from the average values in the Day 7 sample. In future work, it would be useful to investigate and discover the key factors (e.g., type of food, drinks, nutritional supplements, etc.) that would cause significant concentration perturbations to their "normal" or "average" values of an individual.[26-30] Considering that only one daily urine sample is normally collected for most metabolomics studies, controlling these factors during urine sample collection would greatly reduce the sampling effect on metabolome profile, thereby increasing the probability of determining the metabolite concentration changes due to biological or other important factors.

3.3.6 Comparison of Metabolite Detectability of Dns- and Dens-Labeling

It has been shown that dansylation labeling can enhance the MS response of polar amines and phenolic compounds by 1-3 orders of magnitude. We have compared the overall metabolite detectability of the dansylation labeling and the new Dens-labeling. For direct comparison, a portion of a urine sample was divided into two aliquots with each labeled with either $^{12}C_4$-DensCl or $^{12}C_2{}^{13}C_2$-DensCl. The labeled samples were mixed for LC-MS analysis. Another portion of the same urine sample was divided into two aliquots with each labeled with $^{12}C_2$-DnsCl or $^{13}C_2$-DnsCl. The resultant labeled samples were mixed for analysis. In

total, an average of 1083±18 (n=3) peak pairs were found in the mixture labeled by DensCl, while 601±9 (n=3) peak pairs were found in the sample labeled by DnsCl. There were 789 common peak pairs found from the triplicate runs with DensCl, compared to 446 common pairs found from the triplicate data of the DnsCl labeled samples. These results are consistent with our observation that the signal intensities of Dens-labeled amino acid standards were generally higher than those of the Dns-labeled counterparts (see, for example, comparison of Panels B and D in Figure 3-2). The gain in detection sensitivity may be attributed to the presence of ethyl chains of Dens, instead of the methyl groups of Dns, which may enhance the ionization in the ESI process. Compared to the methyl groups, the ethyl chains offer higher hydrophobicity and stronger electron donating propensity, hence, higher basicity of the amine group, resulting in higher ionization efficiency.

Positive identification of the peak pairs detected is beyond the scope of this work. However, based on the accurate mass information, we can search the Human Metabolome Database (HMDB) and the Evidence-based Metabolome Library (EML) using MyCompoundID to generate a list of putative metabolites from the peak pairs detected. The putative identification results are shown in Table T2. Among the 789 commonly found peak pairs, 506 matched with the metabolites in HMDB and 240 matched with the predicted metabolites in EML with one metabolic reaction. Thus, 94% of the common peak pairs can be matched.

Table 3-2.   Putatively identified metabolites in human urine.

Molecular Weight	Name	HMDB ID
45.05742	Dimethylamine	HMDB00087
46.04125	Ethanol	HMDB00108
59.07272	Trimethylamine	HMDB00906
62.01837	Dimethylsulfide	HMDB02303
130.1094	N-Acetylputrescine	HMDB02064
70.0523	Beta-Aminopropionitrile	HMDB04101
70.05268	Beta-Aminopropionitrile	HMDB04101
70.05284	Beta-Aminopropionitrile	HMDB04101
70.05284	Beta-Aminopropionitrile	HMDB04101

Finally, we have tried MS/MS analysis of the Dens-labeled compounds, such as the labeled amino acids, and found that their fragmentation behaviors were similar to those of the Dns labeled counterparts. The major fragment ions generated were from the Dens moiety and only a few neutral loss peaks ($-H_2O$ and $-CO_2$) were observed from the metabolite itself. For generating diagnostic peaks for metabolite structure analysis, an improved method, such as the pseudo-MS3 strategy as described earlier for structural analysis of Dns-labeled compounds, will need to be developed. [31]

3.4   Conclusions

We have developed a novel triplex isotope labeling reagent, DensCl, for quantitative profiling of amines and phenolic metabolites in biological samples. In contrast to the well established dansylation reagents where two isotope forms can be prepared, the new labeling reagents can provide three isotope forms so that two samples relative to a control can be simultaneously analyzed in one LC-MS run.

We have shown that relative quantification of metabolites can be performed by analyzing a mixture consisting of the $^{13}C_4$-Dens-labeled pooled sample which serves as a reference or control, the $^{12}C_4$-Dens-labeled sample #n and the $^{12}C_2{}^{13}C_2$-Dens-labeled sample #(n+1). The utility of this method for metabolome profiling has been demonstrated in the analysis of human urine samples collected daily for a period of 14 days. Besides an obvious increase in sample throughput, the Dens-labeling LC-MS method provides high detection sensitivity over the Dns-labeling LC-MS method. Over 1000 peak pairs could be detected using Dens-labeling in a human urine sample, compared to about 600 peak pairs detected using Dns-labeling. We believe that the triplex isotope DensCl reagents should be particularly useful for metabolome profiling of a large number of biological samples, such as in discovery of disease biomarkers where a great number of bio-fluids are needed to be analyzed to increase the statistical power.

3.5   Acknowledgement

References

(1)   Beckonert, O.; Keun, H. C.; Ebbels, T. M. D.; Bundy, J. G.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Nat. Protoc.* **2007**, *2*, 2692-2703.

(2)   Da Silva, L.; Godejohann, M.; Martin, F. P. J.; Collino, S.; Burkle, A.; Moreno-Villanueva, M.; Bernhardt, J.; Toussaint, O.; Grubeck-Loebenstein, B.; Gonos, E. S.; Sikora, E.; Grune, T.; Breusing, N.;

Franceschi, C.; Hervonen, A.; Spraul, M.; Moco, S. *Anal. Chem.* **2013**, *85*, 5801-5809.

(3)   Patti, G. J.; Yanes, O.; Siuzdak, G. *Nat Rev Mol Cell Biol* **2012**, *13*, 263-269. (4) Guo, K.; Li, L. *Anal. Chem.* **2009**, *81*, 3919-3932.

(5)   Guo, K.; Li, L. *Anal. Chem.* **2010**, *82*, 8789-8793.

(6)   Bueschl, C.; Krska, R.; Kluger, B.; Schuhmacher, R. *Anal. and Bioanal. Chem.* **2013**, *405*, 27-33.

(7)   Bruheim, P.; Kvitvang, H. F. N.; Villas-Boas, S. G. *J. of chromatogra.. A* **2013**, *1296*, 196-203.

(8)   Mueller, D.; Heinz, E. *Current Opinion in Biotechnology* **2013**, *24*, 54-59.

(9)   Abello, N.; Geurink, P. P.; Toorn, M. v. d.; Oosterhout, A. J. M. v.; Lugtenburg, J.; Marel, G. A. v. d.; Kerstjens, H. A. M.; Postma, D. S.; Overkleeft, H. S.; Bischoff, R. *Anal. Chem.* **2008**, *80*, 9171-9180.

(10)   Dai, W. D.; Huang, Q.; Yin, P. Y.; Li, J.; Zhou, J.; Kong, H. W.; Zhao, C. X.; Lu, X.; Xu, G. W. *Anal. Chem.* **2012**, *84*, 10245-10251.

(11)   Zhang, S. J.; You, J. M.; Ning, S. J.; Song, C. H.; Suo, Y. R. *J. of chromatogra. A* **2013**, *1280*, 84-91.

(12)   Mazzotti, F.; Benabdelkamel, H.; Di Donna, L.; Athanassopoulos, C. M.; Napoli, A.; Sindona, G. *J. Mass Spectrom.* **2012**, *47*, 932-939.

(13)   Wang, H.; Wang, H. Y.; Zhang, L.; Zhang, J.; Guo, Y. L. *Anal. Chim. Acta* **2011**, *690*, 1-9. (14) Huang, Y. Q.; Liu, J. Q.; Gong, H. Y.; Yang, J.; Li, Y. S.; Feng, Y. Q. *Analyst* **2011**, *136*, 1515-1522.

(15)   Guo, K.; Peng, J.; Zhou, R.; Li, L. *J. Chromatogr. A* **2011**, *1218*, 3689-3694.

(16)   Zheng, J.; Dixon, R. A.; Li, L. *Anal. Chem.* **2012**, *84*, 10802-10811.

(17)   Wu, Y.; Li, L. *Anal. Chem.* **2013**, *85*, 5755-5763.

(18)   Wu, M.; Yates, S.; Zurek, G.; Barsch, A.; Bruker, 2013.

(19)   Guo, K.; Ji, C.; Li, L. *Anal. Chem.* **2007**, *79*, 8631-8638.

(20)   Mendel, A. *J. of Chem. Eng. Data* **1970**, *15*, 340-341.

(21)   Xia, J.; Mandal, R.; Sinelnikov, I. V.; Broadhurst, D.; Wishart, D. S. *Nucleic Acids Res.* **2012**, *40*, W127-133.

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

(23)   Gros, c.; Labouesse, B. *Euro J of Biochem.* **1969**, *7*, 462-470.

(24)   Parris, N.; Gallelli, D. *J. of Liq. Chromatogra.* **1984**, *7*, 917-924.

(25)   Wu, Y.; Li, L. *Anal. Chem.* **2012**, *84*, 10723-10731.

(26)   Lampe, J. W.; Navarro, S. L.; Hullar, M. A. J.; Shojaie, A. *Proc. of Nutri. Soc.* **2013**, *72*, 207-218.

(27)   Llorach, R.; Urpi-Sarda, M.; Jauregui, O.; Monagas, M.; Andres-Lacueva, C. *J. of Proteome. Res.* **2009**, *8*, 5060-5068.

(28)   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. *J. of Proteome Res.* **2013**, *12*, 1645-1659.

(29)   Tulipani, S.; Llorach, R.; Jauregui, O.; Lopez-Uriarte, P.; Garcia-Aloy, M.; Bullo, M.; Salas-Salvado, J.; Andres-Lacueva, C. *J. of Proteome Res.* **2012**, *10*, 5047- 5058.

(30)   Jacobs, D. M.; Fuhrmann, J. C.; van Dorsten, F. A.; Rein, D.; Peters, S.; van Velzen, E. J. J.; Hollebrands, B.; Draijer, R.; van Duynhoven, J.; Garczarek, U. *J. of Agric. and Food Chem.* **2012**, *60*, 3078-3085.

(31)   Zheng, J.; Li, L. *Int. J. of Mass Spectrom.* **2012**, *316*, 292-299.

Using IsoMS to Process Liquid Chromatography Mass Spectrometry

Based Differential Isotope Labeling Data*

4.1   Introduction

LC-MS based metabolomics has been considered to be one of the most important platforms for discovering bioactive metabolites in complex biological samples.[1-4] In a typical workflow, samples are collected and prepared in a manner appropriate to study purpose. Their LC-MS data are processed to generate comprehensive quantification information that is then analyzed by bio-statistical tools. If a bioactive metabolite was detected, multiple analytical techniques could be utilized to identify it as well as determine its roles in the biological system. Metabolome quantification serves as a key step in the overall workflow, but high quality quantification is difficult to be achieved on LC-MS platform due to several factors, such as retention time shift in chromatography, instrument performance drift, matrix effect in the ESI source, etc. LC-MS based differential isotope labeling (DIL) method is a promising method to quantify metabolites belonging to the same class or metabolic pathway.[5-7] The method uses chemical derivatization to introduce a labeling tag to all metabolites with the same functional group in one sample, and an isotopic counterpart to the same compounds in another comparative sample acting as an internal standard. After combination of the two samples, relative quantitation can be determined by the peak intensity ratios of isotopic peak pairs.

DIL is the most reliable and accurate method to profile metabolites in real samples because each metabolite is quantified by its isotopic internal standard. To date, several high performance DIL reagents have been reported and applied to metabolic studies,[8-10] but the data processing lags far behind the application. In LC-MS based metabolomics, we would like to detect as many peaks as possible, so the traditional visual inspection cannot meet the emerging requirements of high throughput analysis in many metabolomics research areas. However, some issues need to be addressed in order to develop such a computer program. For instance, how do we convert the experimental data collected using proprietary software from the instrument manufacturer into the required format for further analysis? How do we differentiate noise and signal? How do we pick the peak pairs reliably?

The metabolomics community has made large efforts to develop a network so that researchers can use similar tools to process and report their data in a comparable format. Consequently, it is very important to convert raw data into the forms accepted by these tools. To this end, mzData, mzXML and mzML have been proposed by different organizations [11].[12,13] In addition, the csv format is also popular for various tools. Knowing the importance of data sharing, almost all the instrument manufacturers now support these formats in their proprietary software, at least partially right now.

Without proper peak detection and de-noising (i.e., removing noise peaks), it is impossible to mine useful quantification information from LC-MS data. An LC-ESI-MS experiment comes with considerable noise that can mask

and distort real signals and lead to incorrect quantification. Several algorithms have been reported to address the issues.[14-17] Windig *et al.* developed component detection algorithm (CODA) to enhance signal to noise ratio (SNR) in chromatograms.[18,19] This method uses the mass chromatographic quality index to evaluate extracted ion chromatograms (EICs). Only the high score (low noise) EICs are used for peak picking and subsequent processing. Matched filtration is the other popular de-noising algorithm which applies a pre-defined model to differentiate chromatographic peaks from background depending on the similarity of peak shapes between the model and chromatographic peak.[20,21] Gaussian distribution is the most common shape in the algorithm. Driven by the previous success, Andreev *et al.* also described a modified algorithm known as matched filtration with experimental noise determination (MEND) which employs an experimentally determined model instead of theoretical models in the peak picking.[22] MEND has been adopted by XCMS to process LC-MS data.[23-26] In contrast to LC-MS data of unlabeled metabolites, data from a DIL method have a unique feature in that all targeted analytes must exist in pairs, so any peak without an isotopologue could be removed in the data processing. However, even with chemical derivatization, undesirable chemical noise and background peaks can still be present, which require specific filters to remove them.

Although the data processing has been a major bottleneck in the DIL metabolic profiling workflow, to the best of our knowledge, no research effort has been devoted to developing this kind of computer tool. Herein, we report a

91

new computer program, IsoMS, tailored to process data generated by Isotope Labeling Mass Spectrometry. IsoMS can efficiently process LC-MS data acquired by a DIL method to mine the quantification information in the form of peak pairs. This program has been examined on many LC-MS datasets ranging from $^{12}C_2$-/$^{13}C_2$-DnsCl labeled standards to real world samples. The results obtained indicate that IsoMS is able to properly quantify all labeled metabolites present in complex samples. In addition, the results also contain accurate mass, charge status and tag numbers that are very useful for the follow-up work aimed at identification of unknowns (see Table. 4-1).

Table 4-1.　　Format of IsoMS results.

sample	scanNum	rt	mz_light	mz_heavy	distance	int_light	int_heavy	ratio_light _to_heavy	sn_light	sn_heavy	nCharge	nTag	pairLev	isSaturated
1	194	696.3	424.0517	426.0583	2.0066	1.48E+07	1.46E+07	1.01	5601	5521	1	1	1	0
1	194	696.3	481.1460	483.1528	2.0068	4.65E+05	4.39E+05	1.06	192	182	1	1	1	0
1	195	699.9	399.1234	401.1302	2.0068	4.25E+05	4.09E+05	1.04	170	163	1	1	1	0
1	195	699.9	496.0729	498.0799	2.0070	1.73E+05	1.54E+05	1.12	69	61	1	1	1	0
1	196	703.5	415.0961	417.1023	2.0062	7.79E+04	8.85E+04	0.88	30	34	1	1	2	0
1	197	707.1	381.5998	383.6068	2.0069	1.12E+05	8.49E+04	1.32	46	34	2	2	1	0
1	198	710.6	442.1320	444.1384	2.0064	1.76E+05	1.67E+05	1.05	73	69	1	1	1	0
1	198	710.6	522.0884	524.0953	2.0069	1.29E+06	1.21E+06	1.07	515	483	1	1	1	0
1	200	717.8	369.1127	371.1194	2.0067	1.33E+06	1.28E+06	1.04	573	552	1	1	1	0
1	200	717.8	411.5839	413.5906	2.0067	3.14E+05	2.93E+05	1.07	127	119	2	2	1	0
1	201	721.5	379.1324	381.1390	2.0066	1.15E+05	1.14E+05	1.01	47	47	2	2	2	0
1	201	721.5	429.1226	431.1291	2.0064	3.31E+05	3.32E+05	1.00	138	139	1	1	1	0
1	201	721.5	454.1432	456.1501	2.0069	1.60E+05	1.37E+05	1.16	67	57	1	1	1	0
1	202	725.1	355.1111	357.1175	2.0064	1.29E+05	9.73E+04	1.32	54	40	1	1	2	0
1	202	725.1	396.0780	398.0848	2.0068	2.68E+05	2.39E+05	1.12	110	98	1	1	1	0
1	203	728.6	361.1327	363.1396	2.0068	2.23E+05	2.35E+05	0.95	95	99	1	1	1	0
1	203	728.6	393.1353	395.1418	2.0065	2.90E+05	2.71E+05	1.07	120	112	2	2	1	0
1	204	732.2	370.096	372.103	2.00735	6.94E+05	7.23E+05	0.96	286	295	1	1	1	0
1	204	732.2	388.137	390.144	2.00691	7.65E+05	7.64E+05	1.00	316	315	2	2	1	0

4.2 Experimental Section

4.2.1 Reagents and Chemicals

All chemicals and reagents were purchased from Sigma-Aldrich (Oakville, ON) unless otherwise noted. 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 solvents were purchased from Fisher Scientific Canada (Edmonton, AB, Canada). $^{13}C_2$-Dansyl chloride was prepared in-house according to the method described by Bergmann under modified conditions.[27]

4.2.2 Sampling of Human Urine

Urine was collected three times per day from one healthy volunteer not under dietary control for 5 days. The first sample was collected from the second morning urine before drinking coffee, and the second and third urine samples were obtained after drinking coffee. Within an hour of the collection, samples were 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 1.5 ml vials. A pooled urine sample was prepared by mixing the aliquots of the fifteen urine samples. All processed urine samples were stored at -80 °C.

4.2.3 Derivatization

Amine standards and urine samples were derivatized by $^{12}C_2$- and $/^{13}C_2$-DnsCl. There were two groups of labeled urine samples prepared in this work. The first group of samples (Group I) was prepared by separately labeling two

identical urine samples with $^{12}C_2$- and $^{13}C_2$-DnsCl, and then mixing the two differentially labeled solutions. The group I urine samples were prepared for exploring the maximum injection volume of urine in our system in order to optimize the detection of peak pairs found in human urine. The second sample group (Group II) was made by incorporation of the $^{13}C_2$-DnsCl labeled pooled urine and the $^{12}C_2$-DnsCl labeled individual urine samples. Group II was used to investigate the impact of drinking coffee on the human urine metabolome.

4.2.4   LC-MS

LC-MS analysis was performed using an Agilent 1100 series binary LC system (Agilent Palo Alto, CA) connected to a 9.4 T Apex-QE FT-ICR-MS (Bruker, Billerica, MA) or Maxis Q-TOF MS (Bruker, Billerica, MA). All data were acquired in the positive ion mode. All chromatographic separations were performed on an Agilent Zorbax Eclipse Plus C18 (2.1 mm × 100 mm, 1.8 μm, New Castle, DE) column. The mobile phase A was 0.1% formic acid in ACN/H_2O (5/95, v/v) and the mobile phase B was 0.1% formic acid in ACN. Two LC-MS methods were employed in the research. Method I was a 25-min gradient: 0 min (20% B), 0-3.5 min (20-35% B), 3.5-18 min (45-65% B), 18-21 min (65-95% B), 21-24 min (95-99% B), 24-28 min (99% B). The column was re-equilibrated with the initial mobile phase conditions for 15 min before injecting the next sample. The flow rate was 180 μL/min and the injection volume is 2.0 μL. This gradient was used for all the samples except that in the dynamic range experiment. Method II was a 7-min isocratic elution in 50%

95

mobile phase B which was only carried out in the extended dynamic range experiment.

4.2.5 LC−UV

ACQUITY UPLC system (Waters Corporation, Milford, MA) and Waters ACQUITY UPLC BEH C18 column (2.1 mm ×50 mm, 1.7 μm) were used to perform normalization of total metabolite concentrations in urine samples as described by Wu and Li.[28]

4.2.6 Extended Dynamic Range

In the study of the dynamic range of detection, 10 mM dansylated methionine (DnsMet) and valine (DnsVal) stock solutions were prepared by dansylation of methionine and valine stock solutions with $^{12}C_2$-/$^{13}C_2$-dansyl chloride. Their dilution series (25, 2.5, 0.25, 0.125, 0.063, 0.042, 0.031, 0.025, 0.021, 0.018, 0.016 μM) were prepared by diluting an appropriate volume of DnsMet and DnsVal in ACN/H_2O (50/50, *v/v*). The evaluation of in-scan dynamic range was accomplished by mixing two series of dilutions in concentration ratios ranging from 1600:1 to 1:1600 (Met/Val). For each sample, two volumes (2, 12 μL) were injected into Bruker Maxis Q-TOF MS. The acquired data were processed by IsoMS.

4.2.7 Multivariate Analysis

PCA and heatmap were obtained by processing aligned peak lists of IsoMS using MetaboAnalysis (www.metaboanalysis.ca). During the analysis,

variables were removed if they were missing in more than 50% of observations. The remaining variables were filled by the weighted K-Nearest Neighbors (KNN) option which first determines similarity between observations and uses a variable of the closest (neighbor) one to fill the missing variable in another observation.[29] The datasets were normalized via median and then autoscaled.

4.3   Results and Discussion

4.3.1   Workflow of IsoMS

The workflow of IsoMS for data processing is shown in Figure 4-1, which is composed of peak list input, peak pairing, de-noising and pair grouping. Some of these components are discussed below to illustrate the underlying principles of the program. IsoMS cannot directly process the raw data obtained from an LC-MS instrument and thus the raw datasets need to be converted into csv format that contains centroid peak list including retention time, m/z and peak intensity. In the case of processing the FT-ICR-MS data, spectral resolution is also required as discussed below. Data conversion can be done somewhat differently depending on the manufacturer. Unlike other data processing programs, IsoMS does not apply a de-noising or deconvolution algorithm to the spectral peak list based on two major considerations. Firstly, the resolution of mass spectrometry used in the metabolic profiling is quite high (typically >20,000), the influence of noise on the detection of centroid peaks has been substantially minimized. Secondly, when noise are filtered, many low abundant peaks may also be falsely filtered. In our approach, only an intensity

cut-off is used to remove the background peaks after the data conversion and most features are still preserved in the peak list. In subsequent analysis, the noise in these features can be filtered by the peak pairing algorithms, i.e., only the peak pairs are retained while the singlet noise peaks are removed. Thus, IsoMS displays very good sensitivity during the subsequent data processing.
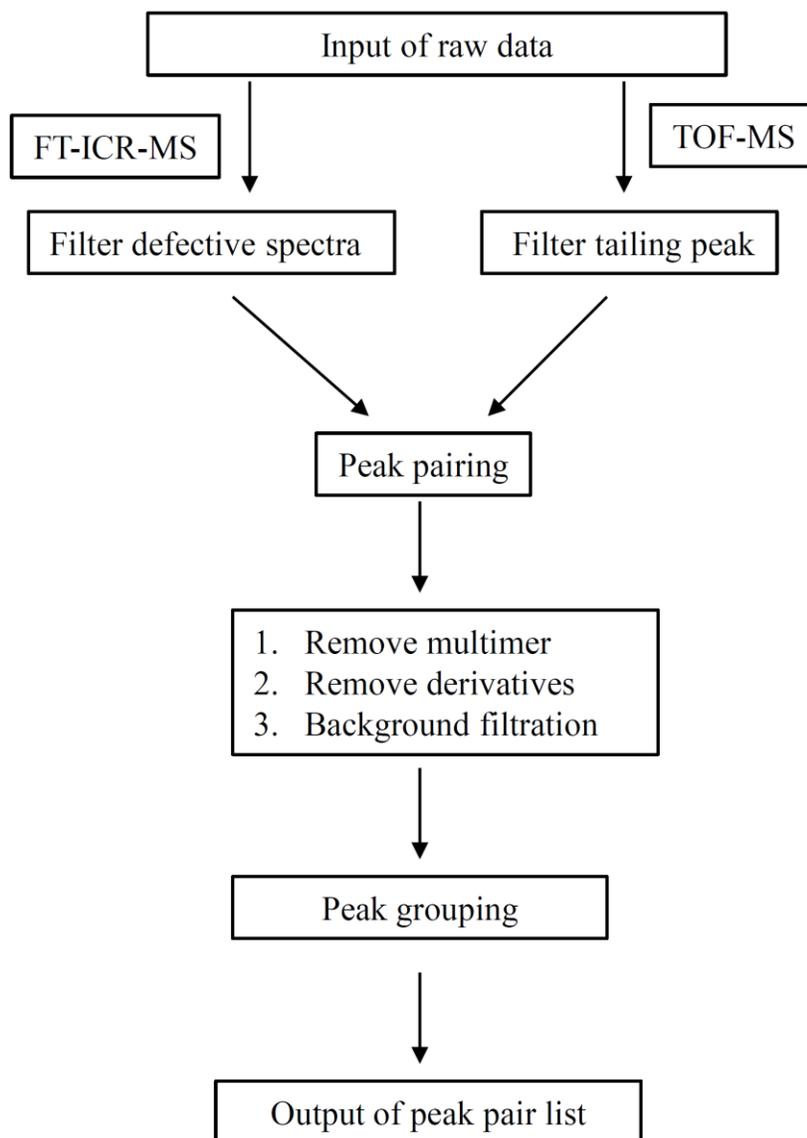
Figure 4-1.   Workflow of IsoMS. A user can choose between TOF mode and FT mode in the program, depending on the type of instrument used.

4.3.2 Peak Pairing

Because ^{12}C-labeled compounds and their ^{13}C-isotopic counterparts are perfectly co-eluted in RPLC, all the targeted isotopologues should produce their MS responses in pairs. Consequently, the isotopic peak pairing not only generates the relative quantification information, but also serves as an efficient de-noising tool for removing most of noise. The mass difference of the isotopologue is the typical criterion used in the visual inspection of DIL data to determine whether two peaks belong to a peak pair or not. In the initial development of IsoMS, we found that using the mass difference could not produce acceptable outcomes. Many unrelated peaks are mismatched by the coincidental mass difference. However, the isotopic distributions of the labeled peak pairs can provide very useful information. In principle, a pair of isotopologues should display similar isotopic patterns, which enable us to use the isotopic distributions for peak picking in IsoMS, as explained below.

When reading a peak from the peak list, IsoMS first examines its isotopic pattern and calculates the charge state. Since all the theoretical mass differences of the isotopic pairs according to the charge status and tag number have been stored in the program, IsoMS can check whether a peak and its corresponding counter-peak exist or not in the peak list. For example, in Figure 4-2A, as a new peak at m/z 455.13 is read, the program scans all the predicted regions where the mass is from 455.38 Da to 456.89 Da in 0.25 Da increment within 10 ppm of mass tolerance. Once the peak at m/z 456.1 is detected, IsoMS further determines whether it is a real isotopic peak of ion at m/z 455.13 by

examining the intensity ratio of m/z 456.1 to m/z 455.1. If so, IsoMS can infer charge status of the ion and continue to search potential isotopic analogues based on the acquired information. In this case, the peak at m/z 457.14 is detected because it matches the condition of one charge and one tag. Then the program examines the isotopic distribution of m/z 457.1 including peaks at m/z 458.1 and 459.1 in the same manner. Because two peaks at m/z 455.1 and m/z 457.1 have comparable patterns, they are paired with very high confidence. The pairs detected under those conditions are assigned "level 1" by the program. Sometimes, a peak cannot display its isotopic pattern due to low abundance like the peak at m/z 457.1 shown in Figure 4-2B. This peak does not give the similar isotopic pattern as its isotopologue, m/z 455.1. However, the isotopic distribution of the ion at m/z 455.1 has already indicated that it is a real peak, rather than noise. Thus, IsoMS still pairs them based on the proper mass difference. In principle, this type of pairs has higher probability of being mismatched, and thus they are labeled as "level two" in the peak pair result. In our experience (see below), the number of false positive peak pairs found in levels 1 and 2 pairs is usually less than 5%, which should be adequate for statistic analysis. In the last case (see Figure 4-2C), even though both peaks at m/z 436.1 and m/z 438.1 have no detectable isotopic patterns, IsoMS still pairs them based on their mass difference. These pairs are assigned as "level 3", which are not as reliable as the other two types, but they can be an important resource to fill missing values in the alignment result prior to statistical analysis.
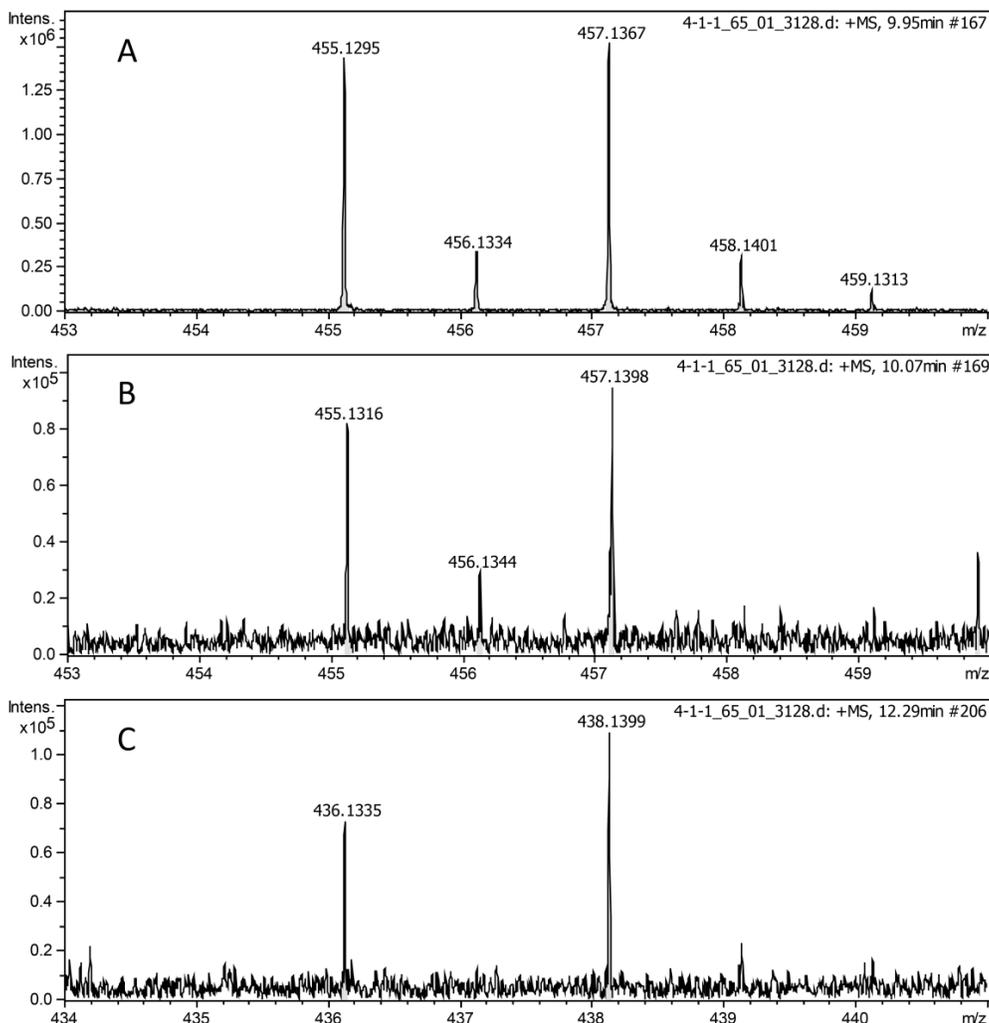
Figure 4-2. Examples of peak pairs of (A) Level 1, (B) Level 2 and (C) Level 3. At Level 1,each peak in the pair displays its natural isotopic peak. At Level 2, the natural isotopic peak of m/z 457 is too low to be detected. At Level 3, neither peak of the pair has the natural isotopic peak.

IsoMS has several filters built-in to remove different noise. These filters can be divided into two groups. The first group is responsible for filtering noise before peak pairing to decrease mismatched pairs, while the filters in the second group are designed to remove specific chemical noise. As an example, some FT-ICR-MS spectra have extraordinary backgrounds. In addition, their mass

accuracy and resolution are much lower than other spectra. Figure 4-3A shows a normal spectrum and Figure 4-3B shows the expanded spectrum of a defective scan. In contrast to the spectrum shown in Figure 4-3A, the defective scan has strong noise peaks. These defective spectra are not suitable for peak pairing as they can cause many false peak pairs. Thus, it is necessary to remove them before peak pairing. The built-in filtering algorithm works in the same manner as "Boxcar averaging" in that the resolution of each spectrum is evaluated by the average resolution of the neighboring spectra. If the relative resolution of the spectrum is too low, IsoMS would delete it from the peak list as a defective spectrum. The cause of these defective spectra in LC-MS may be related to the ion signal saturation and they occur at a rate of x-y defective spectra per LC-MS run.
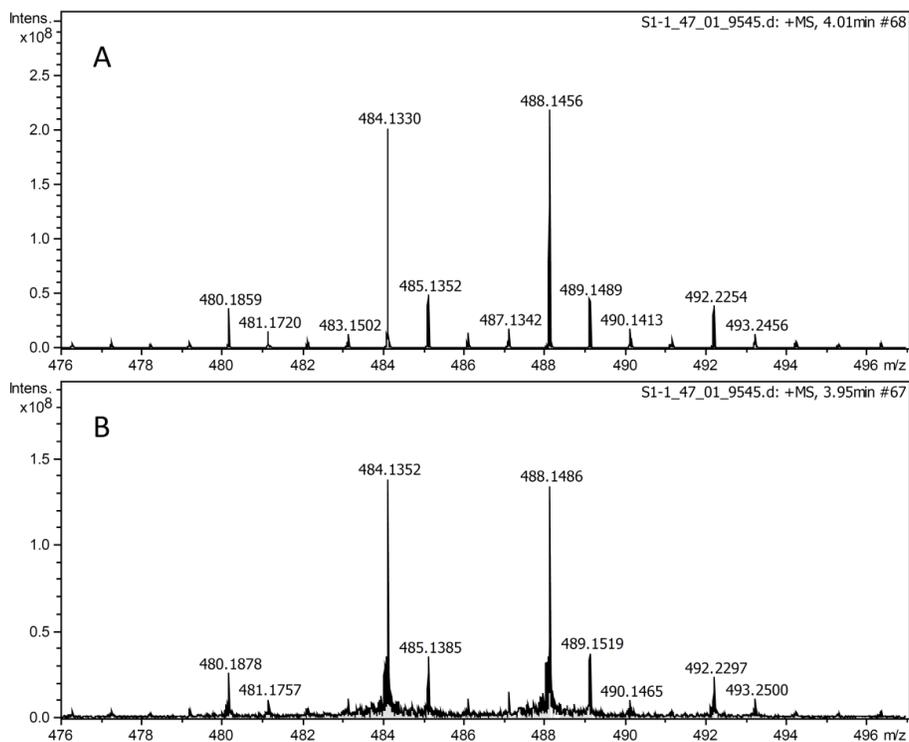
Figure 4-3.   Examples of a normal spectrum (A) and a defective spectrum (B).

In LC-MS analysis of complicated metabolome sample, noise peaks are always present in the LC-MS data. They can be in the form of adduct ions of the analytes, in-source fragments, dimers, trimers, and other multimers, etc. Since these ions are derived from the labeled metabolites, the m/z differences between a metabolite signal and its related noise peaks are predictable. For instance, a sodium adduct peak is 22 Da more than a protonated metabolite. And a protonated ion at m/z 350 can form a dimer at m/z 699 and a dimer of sodium adduct at m/z 721. In IsoMS, these specific values have been entered and thus the program can check all the paired peaks one by one to remove the noise peaks. Moreover, we have found some unexpected dimers in the experiments, in which they were formed by two different species of ions. After determining the

noise source, IsoMS adopted a new filter to calculate all the possible combinations of ions in each spectrum to remove these inter-species dimers. Furthermore, chemical derivatization also produces some chemical noise at fixed m/z's due to side reactions. IsoMS allows a user to edit a pair list containing this kind of noise, which would then be deleted to produce the final results.

4.3.3   Chromatographic Peak Detection

Thanks to perfect co-elution of ^{12}C-labeled light peak and ^{13}C-labeled heavy peak in RPLC, it is straightforward to use the light ion of a pair only to do chromatographic peak detection in the time domain. After peak pairing and de-noising, IsoMS sorts the pair list in m/z, and then divides it into slices with one m/z unit apart. Within each slice of an m/z mass window, the most intensive peak is regarded as the metabolite signal and IsoMS extracts its ion chromatogram in a certain m/z width (e.g., 10 ppm) which is termed as the extracted ion-pair chromatogram (EIPC). If there are more than one EIPC in a slice, IsoMS would repeat the operation for the remaining pairs. In the metabolic profiling of real samples, an EIPC might contain several isomers. To detect the isomers, IsoMS uses different methods depending on the spectrum number in an EIPC. As an example, if an EIPC contains less than five FT-ICR spectra within about 15 seconds of retention time window, IsoMS uses the most abundant signal to annotate the metabolite, because RPLC cannot resolve two isomers in such a short time using our LC-MS conditions. On the other hand, if an EIPC lasts over dozens of spectra, the trends of abundance in the EIPC are calculated.

As the results, the slope of $(n+1)^{th}$ spectrum is positive if its intensity is larger than that in the nth spectrum, and vice versa. In IsoMS, a real peak must have at least three consecutive spectra having positive slopes on one side as well as the same number of spectra with negative slopes on another side.

A potential concern in the chromatographic peak detection method discussed above is whether an EIPC can enclose all ion-pairs of a compound and its isomers. If not, the ion-pairs might be improperly divided into two EIPCs which would result in picking additional peaks. However, we have used the IsoMS program to analyze many datasets and have not found this coincident division of EIPCs is a problem; peak pair division rarely occurs which should be attributed to very high resolutions of instruments.

4.3.4 Validation

Visual inspection of the collected mass spectra for peak pair picking can provide detail information regarding the false positive pairs. For the development of IsoMS, all the data from testing runs of standards and biofluid samples were manually checked so that the peak pairs picked manually could be compared to those from IsoMS. However, the manual inspection method is very time consuming and laborious. To improve the efficiency of validation, we developed a method to use the intensity ratio to quickly differentiate false pairs from correctly picked pairs. In a sample prepared by combining two identical aliquots but labeled by ^{12}C and ^{13}C reagents, respectively, the theoretical intensity ratio for each peak pair is 1.0. Considering the possible experimental

errors and technical variations, the range of the ratio deemed to be acceptable should be 0.67-1.5, i.e., within the variation of ±50% of the mean (1.0). This range is also used as the threshold to analyze the biological variations in real world metabolomics applications (i.e., peak pairs with ratios beyond this range are considered to be caused by biological variations). Since a 95% confidence level is often used to judge the quality of the data, we proposed that at least 95% of the peak pairs picked by IsoMS should have their intensity ratio values within the range of 0.67-1.5. It means any pair with an intensity ratio outside the range is regarded as mismatched or false pair and the number of false pairs can be used to calculate false positive ratios (FPR). We note that some pairs outside the range were actually correct based on manual check of the dataset. However, almost all the peak pairs found within the range of 0.67-1.5 were the correct pairs that could be validated by manual inspection.

4.3.5 Optimization of Sample Injection Amount

False positive ratio (FPR) increases with the increase of sample injection volume or injection amount. To optimize and standardize a workflow for metabolome analysis (e.g., analyzing human urine metabolomes), it is useful to determine the maximum injection volume or amount of a labeled sample in order to achieve the highest metabolite detectability while maintaining an acceptable FPR, i.e., 5%, for peak pair picking. In this work, we analyzed the labeled urine samples in the LC-QTOF mass spectrometer. As shown in Figure 4-4, when the injection volume of dansylated human urine sample was increased from 2 μL to 12 μL, the detected level 1 peak pairs increase from 960 to 1422,

106

while the combined number of levels 1 and 2 pairs went from 1495 to 2345. In Table 4-2, the FPRs of level 1 pairs only increase by 1% even though the injection volume was changed by six-fold. However, if level 2 pairs are also considered, the FPR increases from 3.48% to 6.67%. Considering the FPR and the number of peak pairs found, 5 μL injection appears to be optimal for human urine metabolome analysis in order to strike a good balance between pair number and false pair number. As Figure 4-4 and Table 4-1 show, with 5 μL injection, 1301 level 1 peak pairs can be detected with FPR of 1.97% and 2064 levels 1 and 2 peak pairs can be detected with FPR of 4.86%.
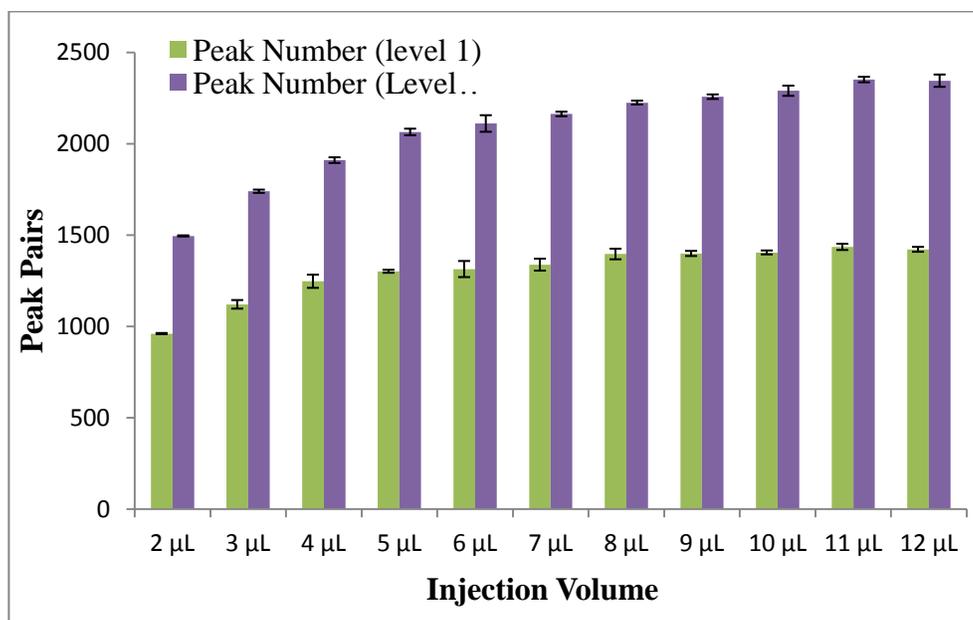

Figure 4-4.   The relationship between the injection volume or amount and the detected pair number. The green bar is for the level 1 pairs and the purple bar is for the levels 1 and 2 pairs.

Table 4-2.　The relationship between the injection volume and the false positive ratio (FPR).

| Injection Vol.	FPR Lev 1	SD 1	FPR Lev. 1& 2	SD 2
2 μL	1.77%	0.37%	3.48%	0.31%
3 μL	1.76%	0.31%	4.29%	0.30%
4 μL	2.16%	0.33%	4.69%	0.25%
5 μL	1.97%	0.16%	4.86%	0.17%
6 μL	2.20%	0.16%	5.45%	0.15%
7 μL	2.26%	0.42%	5.69%	0.03%
8 μL	2.53%	0.32%	5.86%	0.31%
9 μL	2.24%	0.61%	6.01%	0.46%
10 μL	2.33%	0.16%	6.20%	0.47%
11 μL	2.46%	0.32%	6.16%	0.64%
12 μL	2.72%	0.02%	6.67%	0.13%

4.3.6　Extended Dynamic Range of Detection

Metabolites in biological systems cover a tremendous concentration dynamic range.[30] Global metabolic quantification of biological samples is a major challenge for LC-MS because the dynamic ranges of LC-MS methods are often inadequate to simultaneously detect and quantify all the metabolites, especially very low abundance ones. Increasing sample injection amount can effectively enhance signals of low abundance metabolites, but also saturate signals of high abundance metabolites. IsoMS has a function to extend the dynamic range of detection. For saturated signals, if intensities of their natural isotopic peaks are below the upper dynamic range limit of the instrument, IsoMS would use the natural isotopic peaks to generate peak ratio information. Therefore, the program

allows a user to increase the sample injection amount to quantify the low level metabolites without compromising the quantification of high level metabolites. This extended dynamic range detection is best performed with no-trap-based mass spectrometers, such as QTOF. Ion trap MS including FT-MS has an upper limit of ion storage and thus injecting a larger amount may not benefit the detection. In fact, space charge caused by large numbers of ions in a trap can degrade the instrument performance such as lowering the resolution and mass measurement accuracy.

The function and performance of IsoMS for extending the dynamic range of QTOF-MS are demonstrated in the assay of standard solutions prepared by mixing DnsMet and DnsVal. The two compounds have similar retention and ESI response, but different m/z values (see Figure 4-6). In this work, we employed a series of dilutions to explore the in-scan dynamic range. In the first half of the standard solutions, DnsVal was fixed at 25 μM and the concentration ratios of DnsMet/DnsVal ranged from 1/1 to 1/1600. The other half of the samples were prepared in the same manner, but the two standards exchanged their concentrations in the samples. As shown in Figure 4-5, the normal in-scan dynamic ranges of the two standards plotted in blue are up to 10^3 for DnsVal and 8×10^2 for DnsMet. Within the dynamic range, IsoMS was able to generate quantitative information for the two standards, based on their LC-MS data. However, outside the range, injecting a larger sample volume was necessary. When 12 μL samples were injected, the dynamic range was extended to 1.6×10^3 for both of standards (red points in Figure 4-5A). Peak ratios of the samples in the

whole dynamic range are shown in Figure 4-7. The average ratio for DnsMet and DnsVal is 0.96 with standard deviations of 0.07. These data illustrate that IsoMS can generate consistent quantitation results in the whole extended range. We envisage that the combination of IsoMS with LC-QTOF-MS will be a very useful tool to analyze the complex metabolome samples with extended dynamic range of detection and quantification.
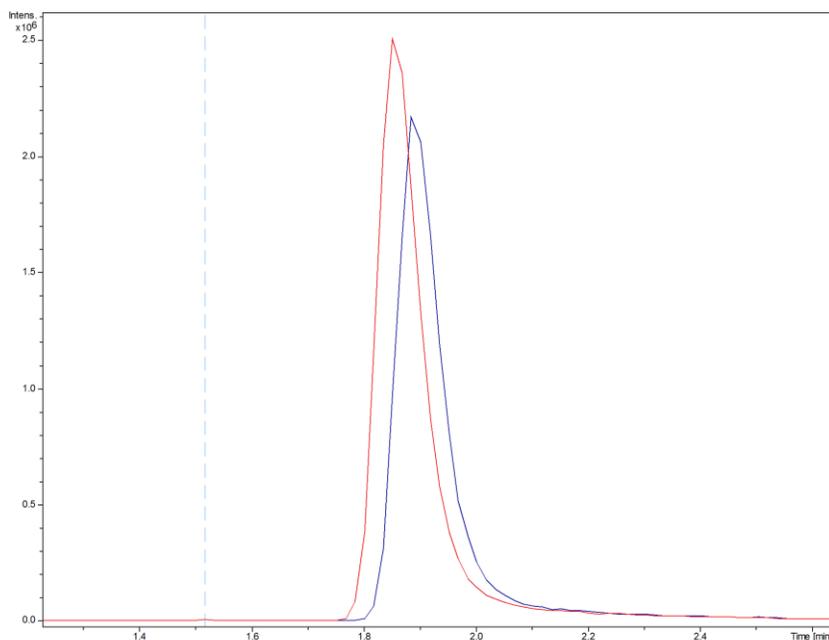
EIC of two standards: Blue trace is DnsVal, and red trace is DnsMet. The concentration ratio is 1:1 (25 μM each).

Figure 4-5.   EIC of DnsVal and DnsMet with a concentration ratio of 1:1 (25 μM each).

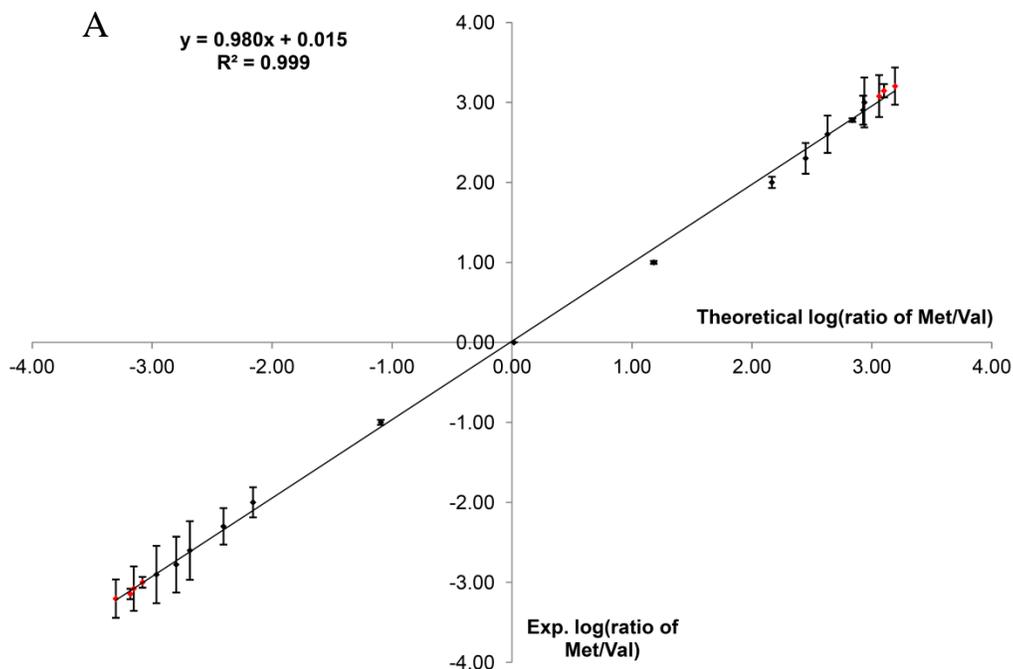Figure 4-6.   Extended dynamic range demonstrated with the calibration curves prepared by analyzing a series of dilutions of mixtures of DnsMet and DnsVal; up to a linear range of 1.6×10^3 for the two standards can be obtained. The dark spots represent a normal dynamic range for DnsMet (8×10^2) and DnsVal (10^3) with 2 μL injection. The red spots were obtained from12 μL of sample injection.
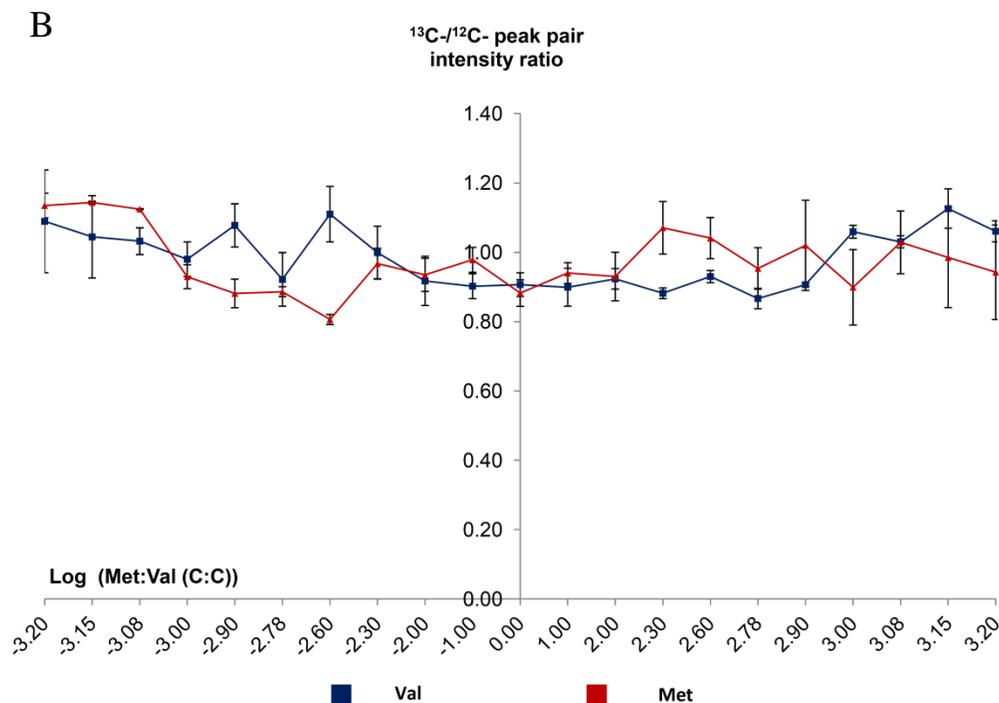
B

Figure 4-7. ^{13}C-/^{12}C- peak ratios of DnsMet and DnsVal in the extended dynamic range. Red line is for DnsMet which has an average peak ratio of 0.97 and a standard deviation of 0.07. The blue line is for DnsVal which has the same average peak ratio and standard deviation as DnsMet.

4.3.7   Influence of Drinking Coffee on Human Urine Metabolome

To illustrate an application of IsoMS LC-MS for quantitative metabolomics, we have analyzed urine samples collected before and after drinking coffee. Coffee beans contain more than 800 detectable metabolites, and the consumption of coffee is known to change the metabolome in the human body.[31,32] In recent years, the relationship between coffee and human health has been extensively studied.[33-35] As the end product of the human metabolism, urine samples can provide important insight to the influence of drinking coffee. In this work, the urine samples were collected from a healthy volunteer over five

consecutive days. Each day, one urine sample before drinking coffee and two urine samples after drinking coffee were sampled and processed by using a standard operation procedure (SOP).[10] Individual urine sample was derivatized by $^{12}C_2$-DnsCl, and the pooled urine sample was derivatized by $^{13}C_2$-DnsCl. Their FT-ICR-MS data were processed by IsoMS to generate peak pair lists and the relative quantitation of metabolites. The peak lists were aligned and then uploaded to MetaboAnalyst where multivariate analysis is performed.

As shown in Figure 4-8, these samples have been classified into two groups in PCA: drinking coffee (Green) and without drinking coffee (Red). Furthermore, two urine samples collected after drinking coffee in the same day were differentiated to some extent. This result indicates that the consumption of coffee indeed impacts the human urine metabolome, and the effect also changes with time. In Figure 4-9, the top 50 metabolites differentiating the two groups are shown in the heatmap. It is clear that a number of individual metabolites have different concentrations in the two groups. Identification of these compounds and their functions in the coffee-induced metabolism will need to be investigated in the future.
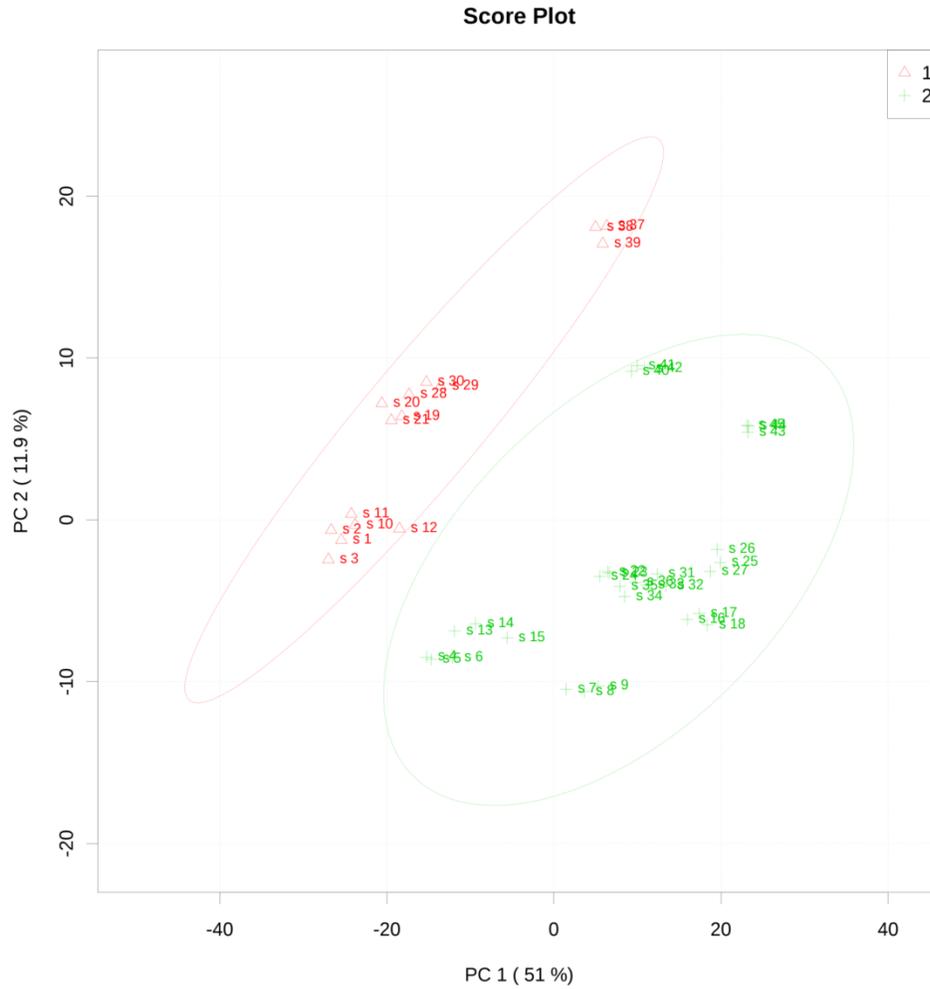
Figure 4-8.   PCA of the human urine samples. Samples labeled in red are urine samples collected before consumption of coffee. Green points represent the urine samples collected after consumption of coffee.
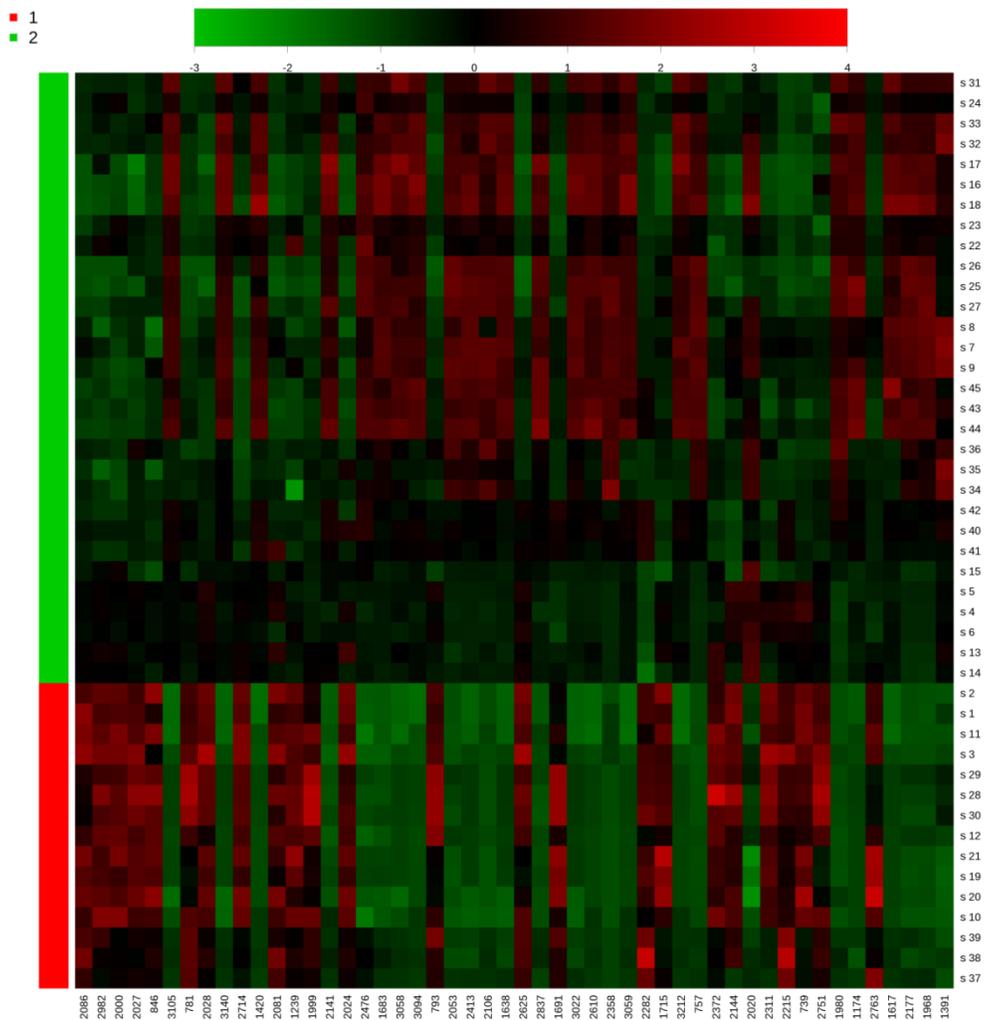
Figure 4-9.   Heatmap of the human urine samples. The top 50 compounds causing the classification are listed.

4.4   Conclusions

We have developed an automated data processing program, IsoMS, to handle the LC-MS data produced by an isotope labeling metabolomic profiling technique. IsoMS can process DIL LC-MS data to generate quantitative information of metabolites in comparative samples. The program includes de-noising filters, peak pairing and chromatographic peak detection. To date, the

performance of IsoMS has been tested on dansylated standards and biological samples. Pair lists generated by IsoMS can be further processed by multivariate analysis programs, such as MetaboAnalyst and SIMCA P+. IsoMS also has a function to extend the dynamic range of mass spectrometry. Finally, the influence of coffee consumption on the human urine metabolome has been preliminarily examined that illustrates IsoMS as a useful tool for quantitative metabolome profiling.

4.5 Acknowledgement

<div align="center">

References

</div>

(1)   Weckwerth, W. *Annu. Rev. Plant Biol.* **2003**, *54*, 669.

(2)   Oldiges, M.; Lutz, S.; Pflug, S.; Schroer, K.; Stein, N.; Wiendahl, C. *Appl. Microbiol. Biotechnol.* **2007**, *76*, 495.

(3)   Verpoorte, R.; Choi, Y. H.; Mustafa, N. R.; Kim, H. K. *Phytochem. Rev.* **2008**, *7*, 525.

4)   Clarke, C. J.; Haselden, J. N. *Toxicol. Pathol.* **2008**, *36*, 140.

(5)   Guo, K.; Liang, L. *Anal. Chem.* **2010**, *82*, 8789.

(6)   Yang, W. C.; Regnier, F. E.; Jiang, Q.; Adamec, J. *J. Chromatogr. A* **2010**, *1217*, 667.

(7)   Yang, W. C.; HMirzaei, H.; Liu, X. P.; FRegnier, F. E. *Anal. Chem. (Washington, DC, U. S.)* **2006**, *78*, 4702.

(8)   Guo, K.; Liang, L. *Anal. Chem.* **2009**, *81*, 3919.

(9)   Guo, K.; Peng, J.; Zhou, R.; Li, L. *J. Chromatogr. A* **2011**, *1218*, 3689.

(10)   Zhou, R.; Li, L. *Mass Spectrometry Methods in Metabolomics* **2013**.

(11)   Orchard, S.; Montechi-Palazzi, L.; Deutsch, E. W.; Binz, P. A.; Jones, A. R.; Paton, N.; Pizarro, A.; Creasy, D. M.; Wojcik, J.; Hermjakob, H. *Proteomics* **2007**, *7*, 3436.

(12)   Pedrioli, P. G.; Eng, J. K.; Hubley, R.; Vogelzang, M.; Deutsch, E. W.; Raught, B.; Pratt, B.; Nilsson, E.; Angeletti, R. H.; Apweiler, R.; Cheung, K.; Costello, C. E.; Hermjakob, H.; Huang, S.; Julian, R. K.; Kapp, E.; McComb, M. E.; Oliver, S. G.; Omenn, G.; Paton, N. W.; Simpson, R.; Smith, R.; Taylor, C. F.; Zhu, W.; Aebersold, R. *Nat. Biotechnol.* **2004**, *22*, 1459.

(13)   Deutsch, E. *Proteomics* **2008**, *8*, 2776.

(14)   Lee, T. A.; Headley, L. M.; Hardy, J. K. *Anal. Chem.* **1991**, *63*, 357.

(15)   Visentini, J.; Kwong, E. C.; Czrrier, Z.; Zidarov, D.; Bertrand, M. J. *J. Chromatogr. A* **1995**, *712*, 31.

(16)   Hindmarch, P.; Demir, C.; Brereton, R. G. *Analyst* **1996**, *121*, 993.

(17)   Barclay, V. J.; Bonner, R. F.; I.P. Hamilton *Anal. Chem.* **1997**, *69*, 78.

(18)   Windig, W.; Phalp, J. M.; Payne, A. W. *Anal. Chem.* **1996**, *68*, 3602.

(19)   Rech, D. H.; Linskens, S.; Paquola, A.; Frohilich, A. A.

(20)   Heuvel, E. J. V. D.; Malssen, K. F. V.; Smit, H. C. *Anal. Chim. Acta* **1990**, *235*, 1990.

(21)   Bogaert, B. V. D.; Boelens, H. F. M.; Smit, H. C. *Anal. Chim. Acta* **1993**, *274*, 71.

(22)   Andreev, V. P.; Rejtar, T.; Chen, H.-S.; Moskovets, E. V.; Ivanov, A. R.; Karger, B. L. *Anal. Chem.* **2003**, *75*, 6314.

(23)   Smith, C. A.; Want, E. J.; O'Maille, G. A., R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779.

(24)   Fraga, C. G.; Clowers, B. H.; Moore, R. J.; Zink, E. M. *Anal. Chem.* **2010**, *82*, 4165.

(25)   Liu, Q. F.; Shi, Y.; Guo, T.; Wang, Y.; Cong, W. J.; Zhu, J. J. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2012**, *907*, 146.

(26)   Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G. *Anal. Chem.* **2012**, *84*, 5035.

(27)   Bergmann, F.; Pfleiderer, W. *Helv. Chim. Acta* **1994**, *77*, 203.

(28)   Wu, Y.; Li, L. *Anal. Chem.* **2013**, *85*, 5755.

(29)   Kim, K. Y.; Kim, B. J.; Yi, G. S. *BMC Bioinformatics* **2004**, *5*, 160.

(30)   Becker, S.; Kortz, L.; Helmschrodt, C.; Thiery, J.; Ceglarek, U. *J. Chromatogr. B* **2012**, *883*, 68.

(31)   Altmaier, E.; Kastenmuller, G.; Romisch-Margl, W.; Thorand, B.; Weinberger, K. M.; Adamski, J.; Illig, T.; Doring, A.; Suhre, K. *Mol. Nutr. Food Res.* **2009**, *53*, 1357.

(32)   Gomez-Ruiz, J. A.; Leake, D. S.; Ames, J. M. *J. Agric. Food. Chem.* **2007**, *55*, 6962.

(33)   Redeuil, K.; Smarrito-Menozzi, C.; Guy, P.; Rezzi, S.; Dionisi, F.; Williamson, G.; Nagy, K.; Renouf, M. *J. Chromatogr. A* **2011**, *1218*, 4678.

(34)   Lardeau, A.; Poquet, L. *J. Pharm. Biomed. Anal.* **2013**, *76*, 134.

(35)   Renouf, M.; Guy, P.; Marmet, C.; Longet, K.; Fraering, A. L.; Moulin, J.; Barron, D.; Dionisi, F.; Cavin, C.; Steiling, H.; Williamson, G. *Br. J. Nutr.* **2010**, *104*, 1635.

Chapter V:

Development of Novel Isotopic Labeling Reagents for LC-MS Based Metabolic Quantification and Identification*

5.1 Introduction

Differential isotopic labeling (DIL) method introduces a chemical tag and its isotopic analogue into targeted analytes present in a sample and a comparative control (or standard) through a chemical reaction, respectively. After mixing the two labeled aliquots that are analyzed by LC-MS, relative quantification can be easily performed based on the abundance ratio of the peak pair of the differentially labeled individual metabolite found in a mass spectrum. Absolute quantification is also possible if the concentration of the analytes in the control (or standard) is known. An LC-MS based DIL approach has shown high accuracy and precision in the metabolomics study.[1] The optimal data quality obtained is attributed to simultaneous measurements of the analytes and their isotopic internal standards, so ion suppression and instrument drift can be neglected during the analysis. Furthermore, chemical derivatization has been accepted as a useful means for improving the resolution of polar metabolites in the reversed phase liquid chromatography (RPLC) and also enhancing the ionization efficiency in electrospray ionization (ESI) by introducing hydrophobic ESI-active tags.[2-5] In general, hydrophobic molecules are preferably retained by the reversed phase stationary phase and enriched on the surface of droplets within the ESI process. To date, a number of DIL reagents have been developed to quantify organic amines, carboxylic acids and carbonyl compounds owing to their important biological functions in metabolic pathways.[6-8] An interesting

fact is that the number of DIL reagents developed for LC-MS is more than that for GC-MS, even though derivatization is a necessary step in the latter.[9] This reflects the concerns of matrix effects present in the LC-MS based metabolic analysis.

Development of chemical labeling reagents is a tricky task because multiple factors have to be comprehensively considered, including synthesis and cos. A synthesis of isotopic labeling shouldn't to longer (<5 steps). Otherwise, yield is greatly decreased. Moreover, most of ^{13}C coded compounds are quite expensive so that it is not practical to be used in the preparation. Contrary to the emerging LC-MS based DIL methods, ultraviolet (UV) and fluorescence (FL) chemical derivatization have been well developed in the past several decades. A possible route in the development of DIL reagents for MS is to introduce isotopic atoms into an ultraviolet or fluorescence (UV/FL) labeling reagent so that the derivatization chemistry and LC methods can be transferred into the LC-MS method. Guo *et al.* developed a DIL method based on $^{12}C_2$- and $^{13}C_2$-dansyl chloride (DnsCl). Although the method allows for rapid and accurate quantification of amine-containing metabolites in various biological samples,[2,10] identification of bioactive compounds presents a special challenge due to the lack of structural information in MS/MS spectra of labeled metabolites. In the same manner, Tsukamoto *et al.* have developed H_6-/D_6-7-(N,N-dimethylaminosulfonyl) -4-(aminoethyl)-piperazino-2,1,3-benzoxadiazole (H_6-/D_6-DBD-PZ-NH$_2$) to profile fatty acids in rat plasma samples.[11,12] As the derivatization reaction used the deuterium-labeled reagents, the isotope effect on retention time was observed. Abello *et al.* have described new multiplex reagents for analyzing amine-

containing metabolites in human cells based on pentafluorophenyl-activated ester of ^{13}C-containing poly(ethylene glycol) chains (PEG-OPFP). The family of these DIL reagents can be used to quantify three samples in parallel, but synthesis of PEG-OPFP is fair challenge to many analytical laboratories.[13] In addition, the reagents do not provide much signal enhancement in LC-MS. Yang *et al*. developed H$_3$-/D$_3$-N-hydroxysuccinimide ester of N-alkylnicotinic acid (H$_3$/D$_3$-C$_n$-NA-NHS) to measure concentrations of amino acids in rat urine by which the sensitivity of labeled amino acids was enhanced by up to 1,000-fold.[14] Besides the deuterium isotope effect, the ionizable tag also limits the efficient separation of the labeled amino acids on RPLC.

Results of these studies have indicated that using DIL reagents for LC-MS can be a promising approach for metabolic quantification. Unfortunately, these reagents are not commercial available. In addition, many reagents require time-consuming and costly synthesis pathway. Although deuterium coded reagents are relatively inexpensive, the deuterium isotope effect in RPLC makes them undesirable for the metabolic profiling of complicated biological samples. In contrast, ^{13}C-coded reagents have undistinguishable chromatographic performance from that of the natural counterpart in RPLC.

In this work, we report a strategy which uses ^{12}C$_2$-glycine and ^{13}C$_2$-glycine to construct a family of DIL reagents which can derivatize primary and secondary amine-containing metabolites. Two DIL reagents, ^{12}C$_2$-/^{13}C$_2$-4-dimethylamino-benzoylamido acetic acid *N*-hydroxysuccinimide ester (^{12}C$_2$-/^{13}C$_2$-DBAA-NHS) and ^{12}C$_2$-/^{13}C$_2$-4-methoxybenzoylamido acetic acid *N*-hydroxysuccinimide ester

($^{12}C_2$-/$^{13}C_2$-MBAA-NHS) have been synthesized (see Figure 5-1). DBAA-NHS is designed as a sensitive reagent because of a basic amino group on the aromatic ring.[3] Since MBAA-NHS has no an basic functional group, the labeling is expected to give structural information about the labeled moiety, i.e., the metabolite. Beside low cost and simple preparation, the method developed is also designed to provide derivatives additional benefits as the followings: 1) improve separation of labeled polar compounds on the reversed phase column as a result of increasing hydrophobicity, 2) enhance ESI response of derivatives compared to underivatized compounds, and 3) move the signals of small molecules out of the noisy, low m/z region. All of these properties greatly improve the accuracy and precision in LC-MS metabolic analysis. The labeling protocol has been optimized based on LC-UV analysis of amino acids standards. In addition, LC-MS methods using MBAA-NHS and DBAA-NHS have been investigated. Finally, the MS/MS experiments on MBAA-NHS labeled amino acids have been performed which indicate that the MS/MS fragment ions of MBAA-NHS derivatives provide useful structural information for metabolite identification.
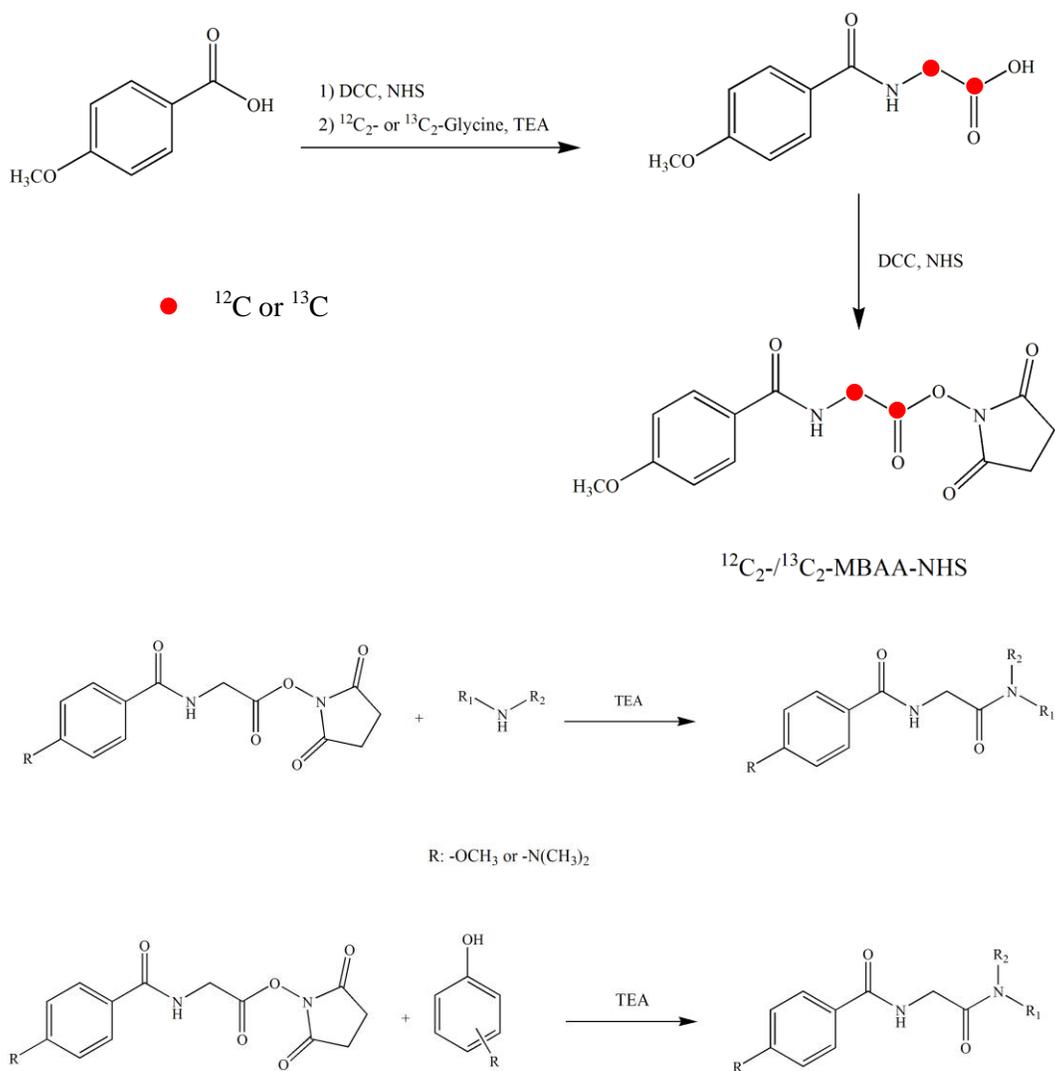
Figure 5-1. Reaction schemes for (A) synthesis of the isotope labeling reagents MBAA and DBAA. Red spots represent ^{12}C or ^{13}C and (B) derivatization of amine-containing metabolites by MBAA or DBAA.

5.2 Experimental Section

5.2.1 Reagents and Chemicals

All chemicals and reagents were purchased from Sigma-Aldrich (Oakville, ON) unless otherwise noted. ^{13}C$_2$-glycine was purchased from Cambridge Isotopes Laboratories (Andover, MA). Stock solutions of twenty amino acid

standards (20 mM each) were prepared in H_2O:ACN (50:50 (*v/v*)) and stored at 4 °C. A pooled amino acid (1 mM each) was prepared by mixing aliquots of twenty stock solutions. LC-MS grade solvents were purchased from Fisher Scientific Canada (Edmonton, AB, Canada). 300 mM triethylamine (TEA) buffer solution was prepared by dilution of 500 μL LC grade TEA in 28.5 mL acetonitrile. 300 mM formic acid (FA) solution was prepared by adding 1.13 mL formic acid into 100 mL H_2O. Both solutions were stored at 4 °C.

5.2.2 Synthesis of 4-methoxybenzoylamido acetic acid *N*-hydroxy-succinimide Ester (MBAA-NHS) and 4-dimethylamino-benzoylamido acetic acid *N*-hydroxy-succinimide ester (DBAA-NHS)

Figure5-1A shows the three-step synthesis pathway to prepare MBAA-NHS and its analogs. *N*, *N*-dicyclohexylcarbodiimide (DCC) (1.25 g) and *N*-hydroxyl succinimide (HOSu) (0.67 g) were added into a solution of 4-methoxybenzoic acid 0.92 g (or 4-dimethylaminobenzoic acid (0.99 g)) in 40 mL DMF at 25 °C. The mixture was stirred for 24 hr, and then the mixture was filtered to give a clear solution without further purification. glycine (0.75 g) or its isotopic counterpart was dissolved in a mixture of 20 mL H_2O and 2.7 mL triethylamine. Then the solution was quickly added into the filtrate from the first reaction. After 30 min, the mixture was neutralized by formic acid to give a white solid, 4-methoxybenzoylamido acetic acid (or 4-dimethylbenzoylamido acetic acid), which was filtered out and washed by cold acetone (3×10 mL). The product could be directly used in the next step without purification. The third step was exactly the same as the first reaction. 4-methoxybenzoylamido acetic acid (0.90 g) (or 4-

dimethylbenzoylamido acetic acid (0.98 g)) was dissolved in 40 mL DMF. HOSu (0.50 g) and DCC (0.89 g) were added into the solution which was stirred for 24 hr at 25 ℃. The final product was purified by recrystallization with Hexane:*i*-PrOH (3:1 *v/v*). The total yield in the three steps was 68% and the product was stored at -20 ℃.

5.2.3   Derivatization and Mixing

As shown in Figure 5-2, 50 μL of the pooled amino acid standards or the urine sample was mixed with 25 μL of 0.3 M TEA buffer, and then added to freshly prepared 100 μL of 20 mM $^{12}C_2$-MBAA solution in ACN:DMSO (1:1 v/v) for light labeling, or an equal volume of $^{13}C_2$-MBAA for heavy labeling in reaction vials. The solutions were incubated at 25 ℃ for 10 min prior to being acidified by adding 50 μL of 0.3 M FA solution. After centrifugation at 14,000 g for 5 min, the aliquots of $^{12}C_2$-/$^{13}C_2$-MBAA-NHS labeled samples were mixed to give the final sample for LC-MS analysis.
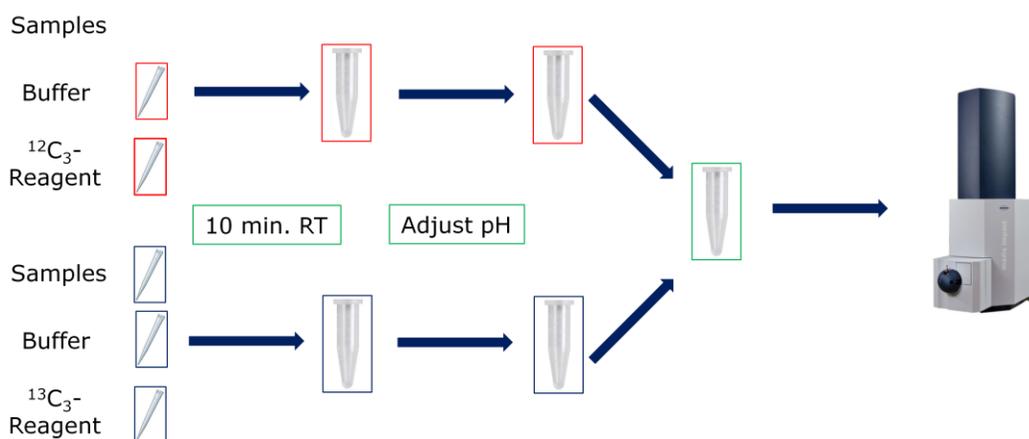

Figure 5-2. Workflow of derivatization of amines in standard or biological samples.

5.2.4　Absolute Quantification of Amino Acids in Human Urine

Twenty aliquots of amino acid standard stock solutions were derivatized by $^{13}C_2$-MBAA-NHS. Based on previous results[15], the appropriate volume of labeled amino acid standard solution was spiked into a pooled human urine labeled by $^{12}C_2$-MBAA-NHS prior to LC-MS analysis.

5.2.5　Direct Flow Injection MS/MS and Pseudo MS^3 Analysis

A Bruker maXis Impact Q-TOF-MS (Bruker, Bremen, GE) was used to perform MS/MS and pseudo MS^3 analysis. The sample solutions were infused directly by a syringe pump at a flow rate of 3 μL/min. The MS instrument was operated under the following conditions: Nebulizer gas 1.8 bar, Dry gas 8.0 L/min, dry gas heater 220 ℃, capillary voltage 4500 V, end plate offset -500 V, the mass range was set at m/z 50-500. A mass scan was first performed to find the protonated molecular ion. In the MS/MS scan, the precursor ion of a labeled amino acid was isolated within ± 1 Da mass window. After collision induced dissociation (CID) in the CID cell, the product ions were detected by the TOF analyzer. In a pseudo MS^3 scan, the first generation fragments were generated in the skimmer region by raising the in-source CID voltage. Then the protonated unlabeled amino acid ion was isolated by the first quadrupole (Q0) within ± 1 Da of mass window and then subject to collision induced dissociation in the CID cell. Finally, the second generation of product ions was analyzed in the TOF. To efficiently transfer the low mass fragments to the TOF, the RF voltage of the CID cell, accumulation time and pulse time were tuned for each scan. All of the MS

scan, MS/MS scan and pseudo MS^3 experiments were performed in the positive mode. The MS/MS spectra and pseudo MS/MS spectra were compared to the MS^2 spectra of amino acid standards in the Human Metabolome Database (HMDB) that were acquired in a triple quadrupole mass spectrometer.

5.2.6 LC-MS

LC-MS analysis was performed using an Agilent 1100 series binary LC system (Agilent Palo Alto, CA) hyphenated to an Impact Maxis Q-TOF MS (Bruker, Billerica, MA). All data were acquired in the positive ion mode. Chromatographic separations were carried out on an Agilent Zorbax Eclipse Plus C18 (2.1 mm $\times$ 100 mm, 1.8 µm, New Castle, DE) column. The mobile phase A was 0.1% formic acid in ACN/H$_2$O (5/95, v/v) and the mobile phase B was 0.1% formic acid in ACN. A 20-min gradient: 0 min (10% B), 0-0.5 min (10% B), 0.5-7 min (10-25% B), 7-18 min (25-99% B), and 18-20 min (99% B). The column was re-equilibrated with the initial mobile phase conditions for 15 min prior to the next sample run. The flow rate was 180 µL/min and the injection volume was 2.0 µL.

5.3   Results and Discussion

5.3.1   Synthesis of MBAA-NHS and DBAA-NHS

A stable isotope labeling reagent can be divided into three parts according to their functions: 1) reaction group, 2) isotope group, and 3) tuning group. The reaction group is responsible for the derivatization reaction between the labeling

reagent and the targeted analytes. The isotope group containing isotope atoms provides the mass difference in a peak pair found in a mass spectrum. The tuning group improves the separation of the labeled polar metabolites in the LC and enhances the ionization in ESI due to its hydrophobicity feature. Selection of an isotope group is particularly important for the development of a DIL method, because it determines the manner of coding isotope atoms to the molecule and the cost of the method. In this work, glycine was chosen as an isotope group in the proposed reagents. Beside the reasonable price, the structure of glycine makes it very promising for doing the following tasks. First of all, glycine has an amino group which can react with various compounds including carbonyl chlorides or alkyl bromide, so that the isotope group can be easily combined into numerous tuning compounds in the preparation. Secondly, the amide bond is sufficiently strong to resist other reaction conditions. Finally, the carboxylic acid of glycine is easily converted into the active *N*-hydroxysuccinimide (NHS) ester which is well known to label amines under mild conditions. Since glycine is a polar compound, aromatic compounds are good candidates of the tuning group to increase the hydrophobicity of the labeling reagents. Contrary to other DIL reagents, glycine serves as both the isotope group and the coupling group, while the hydrophobic group is only used to change the performance of the final products. This feature allows us to optimize the performance of these DIL reagents by simply changing the structure of the aromatic compound.

The synthesis of the reagents is quite straightforward. The two-step reactions are performed under very mild conditions and the final compounds are

purified by a simple recrystallization. Thus, the procedure should be accessible to many analytical laboratories. As shown in Figure 5-1B, the amino group of glycine (or $^{13}C_2$-glycine) reacts with a carboxylic acid NHS ester to form an amide intermediate. Then the carboxylic acid of the glycine moiety in the intermediate is converted into an active NHS ester which can selectively derivatize primary and secondary amines in aqueous solutions. Two kinds of DIL reagents have been synthesized based on the common pathway, 4-methoxybenzoylamido acetic acid *N*-hydroxysuccinimide ester (MBAA-NHS) and 4-dimethylaminobenzoylamido acetic acid *N*-hydroxysuccinimide ester (DBAA-NHS). MBAA-NHS is a neutral molecule which was expected to provide structural information in the MS/MS mode for metabolite identification. DBAA-NHS has a basic amino group in the tuning group and thus it should provide better ionization efficiency in the ESI positive ion mode, which is useful for quantification. Although we have not tried other types of tuning groups, we expect that they would be workable under the same pathway.

5.3.2 Optimization of Labeling Efficiency

A reproducible and efficient chemistry is essential for any stable isotope labeling method. In addition, the reaction should be carried out under mild conditions to avoid any undesirable side-reactions. Considering the similarity between MBAA-NHS and DBAA-NHS, the optimization of labeling efficiency was examined by MBAA-NHS only, but the conditions could be transferrable to DBAA-NHS. A mixture of 20 amino acids served as the standards to react with MBAA-NHS solution. A UPLC-UV technique was used to determine the labeling

129

efficiency. The total UV peak area of 20 amino acids under different conditions was compared (see Figure 5-3A).

To optimize the derivatization conditions, the effects of MABB-NHS concentration, reaction time and solvent composition were investigated at room temperature. First of all, the labeling efficiency was compared at five different concentrations of labeling reagents (20 mM to 60 mM) with a fixed concentration (20 mM) of the pooled amino acid standards. As shown in Figure 5-3B, the UV peak area of derivatives increases with the concentration of MBAA-NHS and reaches a plateau when MBAA-NHS is 2.5-fold more than that of the standards. Secondly, the reaction time was examined by comparing five reactions which were carried out under the same conditions but finished at different times from 10 min to 50 min. As shown in Figure 5-3C, the UV peak area is independent of the reaction time, which means the labeling reaction would be completed within 10 min. This observation is consistent with the previous report of similar reactions.[16] Contrary to the reaction time, the solvent composition has a large influence on the labeling efficiency. As shown in Figure 5-3D, the UV absorption was inversely proportional to water content in the solvent. It is understood that water is not a favorable solvent for the nucleophilic substitution reaction. Since the formation of H-bonds, the molecular interaction around the amino group is reduced. Furthermore, water also causes hydrolysis of the labeling reagents. Finally, 10 min reaction at room temperature in 25% aqueous solution was selected for the derivatization of 20 mM amino acid standard or undiluted urine samples in the presence of 50 mM MABB-NHS.
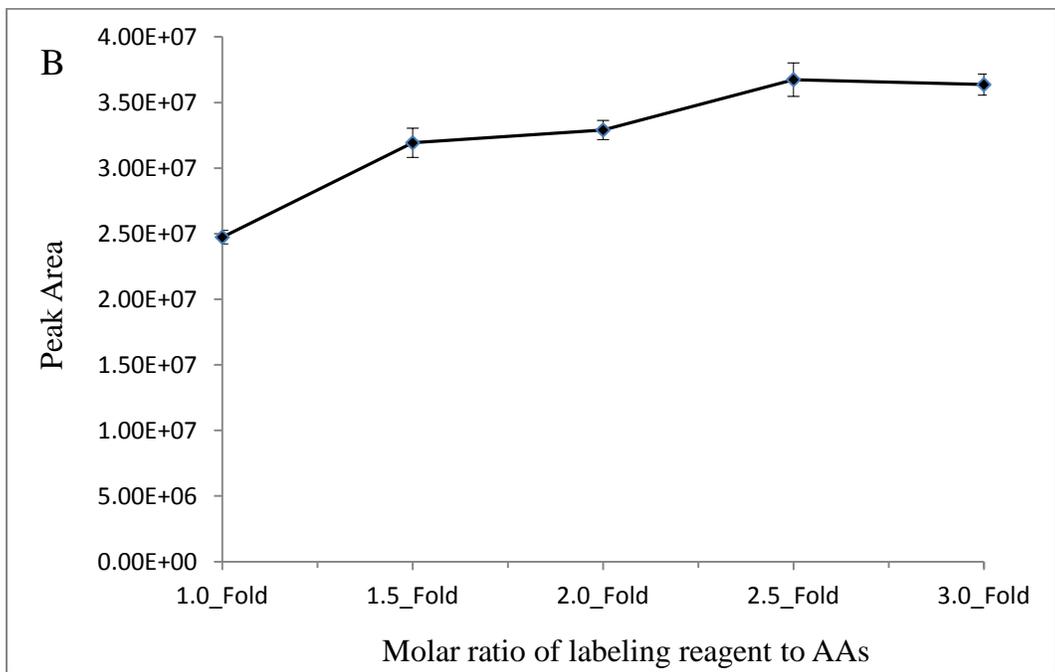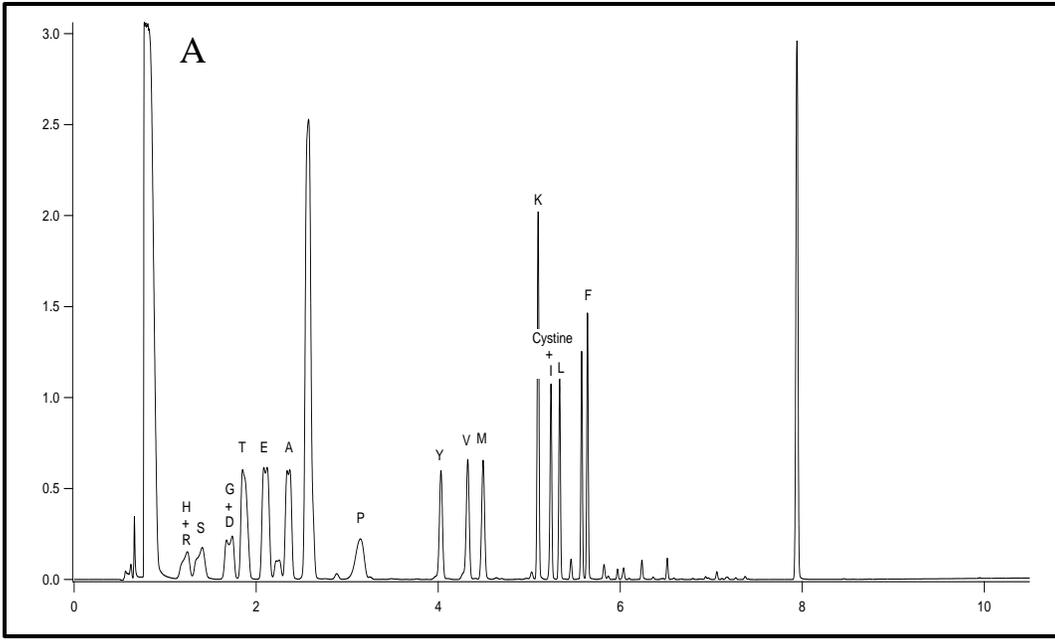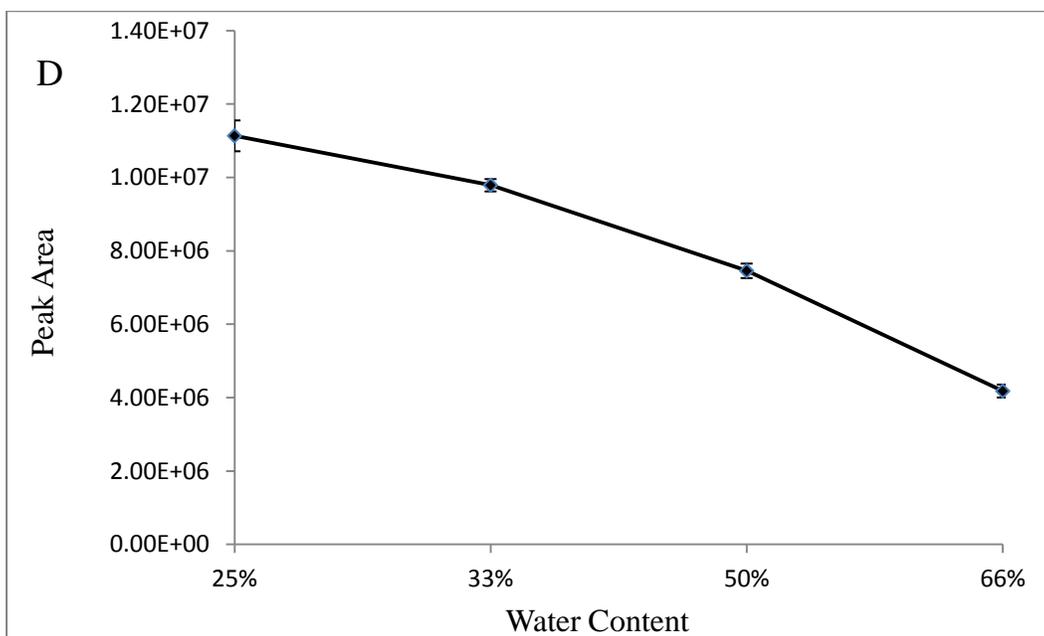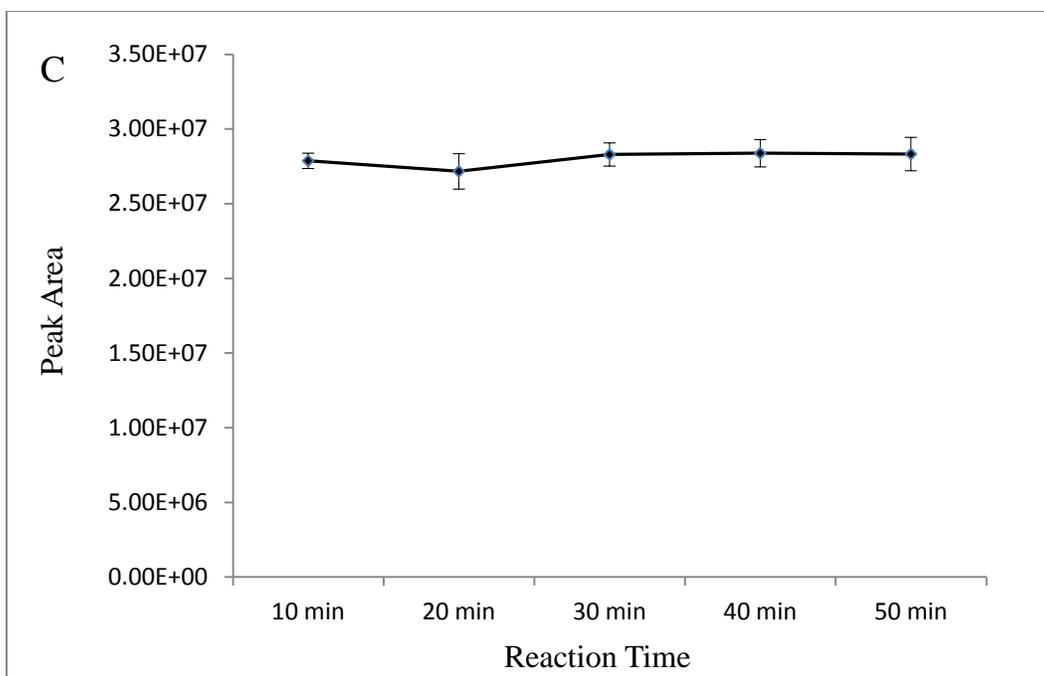
Figure 5-3. Influence of reaction conditions on MBAA derivatization. (A) LC-UV chromatogram of 20 amino acids (absorbance vs. retention time). (B) Effect of MBAA concentration. (C) Effect of reaction time. (D) Effect of water content.

5.3.3 Analytical Performance of DBAA-NHS and MBAA-NHS Labeling

We have carried out a set of experiments to gauge the analytical performance of DBAA-NHS and MBAA-NHS labeling for metabolite analysis. Figure 5-4A shows the base peak ion chromatogram of amino acid standards labeled by $^{13}C_2$-/$^{12}C_2$-DBAA. Most of the amino acids are well separated. Some background peaks from the labeled sample solution (unlabeled peaks in Figure 5-4A) are detected and they are likely from impurities in the sample or reaction by-products. Figure 5-4B shows the expanded mass spectrum of DBAA labeled Val, Met and Tyr. Beside the protonated molecular ion peak pair found for each labeled amino acid, some low intensity peaks from sodium adduct ions are also observed.
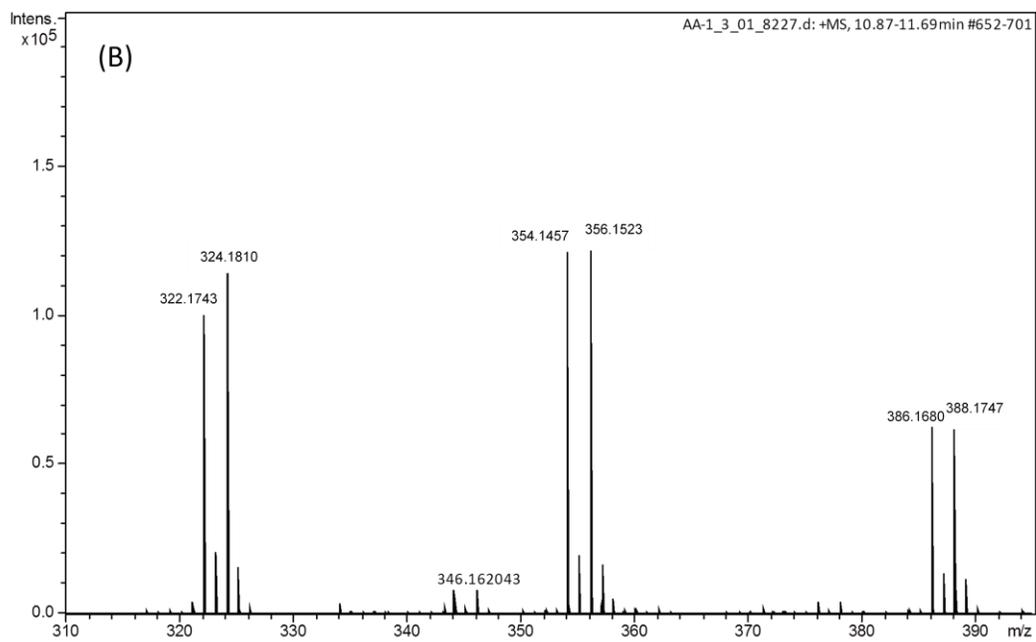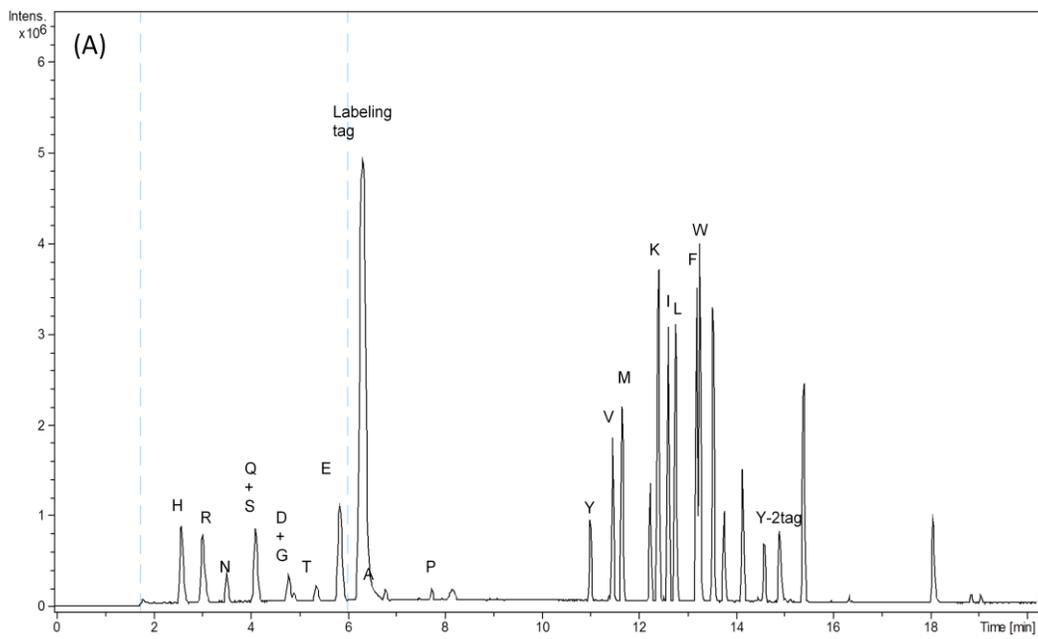
Figure 5-4. (A) LC-MS chromatogram of amino acid standards labeled by DBAA-NHS. (B) Expanded spectrum of Val, Met and Tyr labeled by DBAA-NHS.

Figure 5-5A shows the base peak ion chromatogram of a labeled human urine sample prepared by mixing an aliquot of the $^{13}C_2$-DBAA pooled sample and aliquot of a $^{12}C_2$-DBAA individual sample. The labeled sample was not diluted and 2 μL of the labeled sample, equivalent to 10 μL of the original urine sample, was injected for generating this chromatogram. There are many peaks detected in this sample, compared to almost no peaks detectable using the same equivalent volume of the original urine sample. This indicates that DBAA labeling can improve the detectability of urine metabolites significantly. Figure 5-5B shows the expanded mass spectrum of several unknown metabolites detected in the labeled urine sample within the retention time of 12 to 12.3 min.
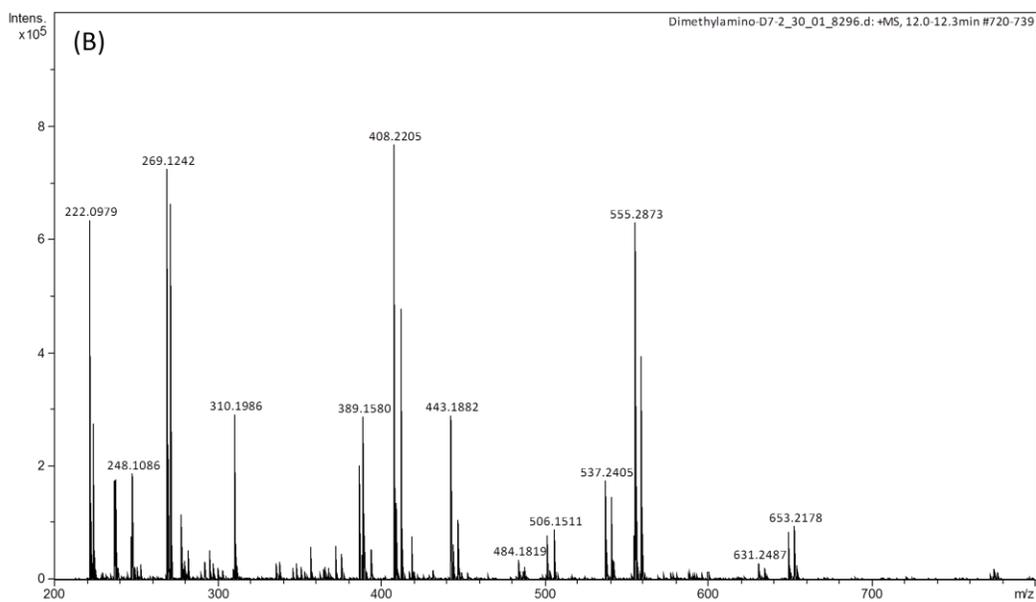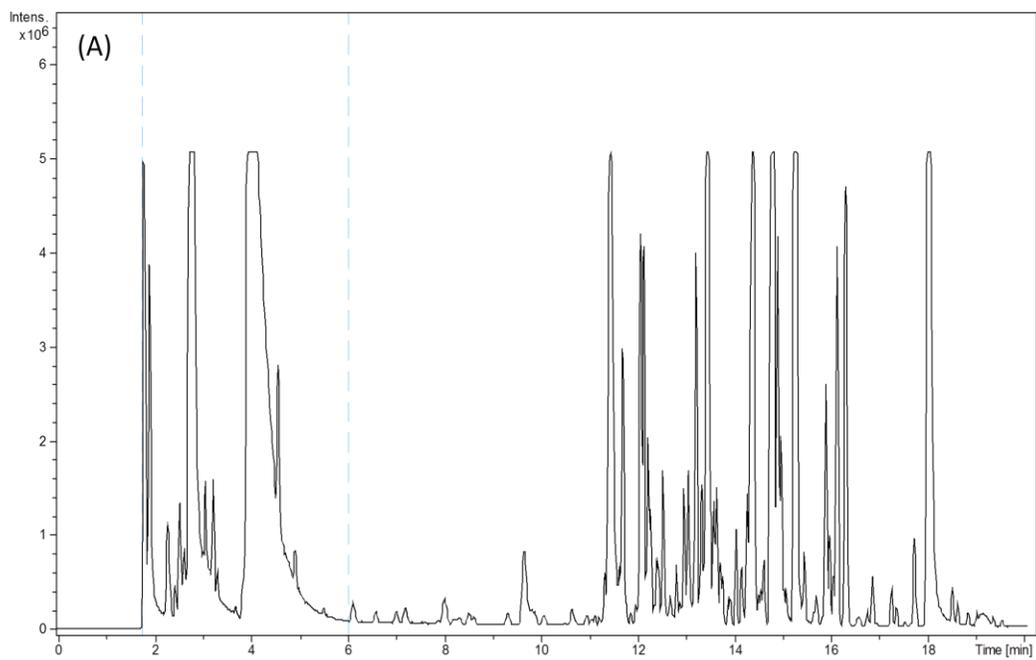
Figure 5-5. (A) LC-MS chromatogram of a human urine sample labeled by
DBAA-NHS. (B) Expanded spectrum of the urine sample.

Figure 5-6A shows the base peak ion chromatogram of amino acid standards labeled by $^{13}C_2$-/$^{12}C_2$-MBAA. In this case, the amount of sample injected was 4 times more than that used for analyzing the DBAA labeled sample. Comparing the intensities of chromatographic peaks obtained in Figures 5-4A and 5.6A, it is clear that the detection sensitivity of MBAA labeled amino acids is about 4 times lower than that of DBAA labeled amino acids except histidine and arginine, where the amino acid itself is more readily charged and thus has higher ESI ionization efficiency. These two amino acids have a lower retention time in RPLC. Figure 5-6B shows the chromatogram obtained from the MBAA labeled human urine sample. The same volume or amount of the labeled urine sample was injected as that used for the DBAA labeled sample. The intensities of the chromatographic peaks are lower in the MBAA labeled sample, compared to those from the DBAA labeled sample, which also indicates that MBAA labeling is less sensitive than DBAA labeling. Nevertheless, using IsoMS to process the LC-MS data, over 1000 peak pairs could be detected from the MBAA labeled sample. Thus, the MBAA labeling can still be very useful for metabolome quantification, while, as it will be demonstrated below, MBAA labeling offers the opportunity of generating useful fragment ion spectra for characterization or identification of unknown metabolites.
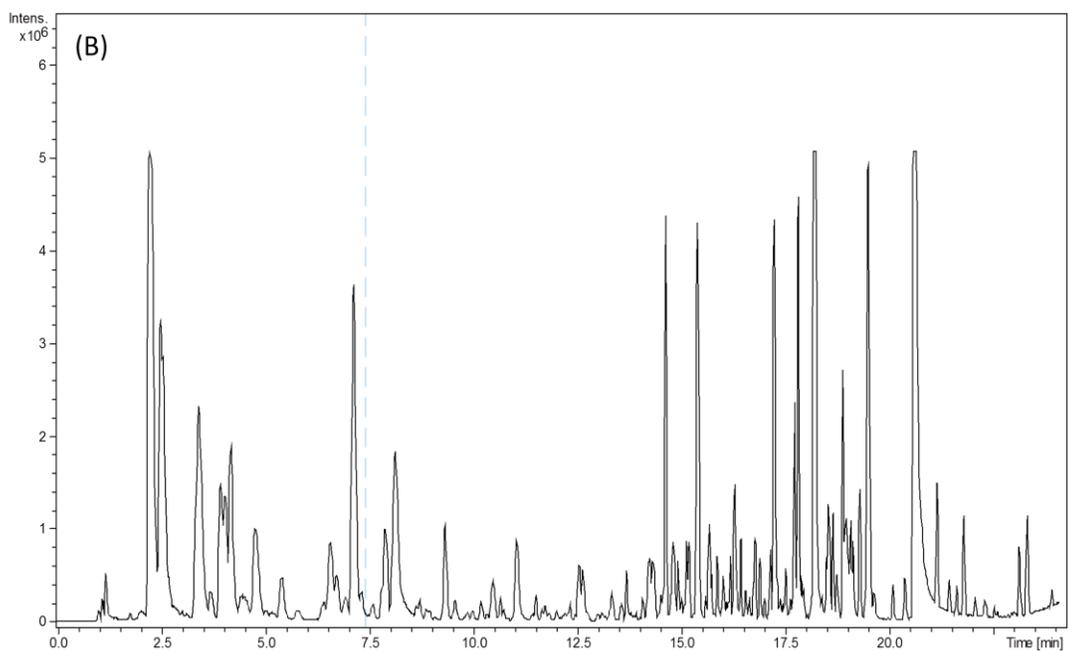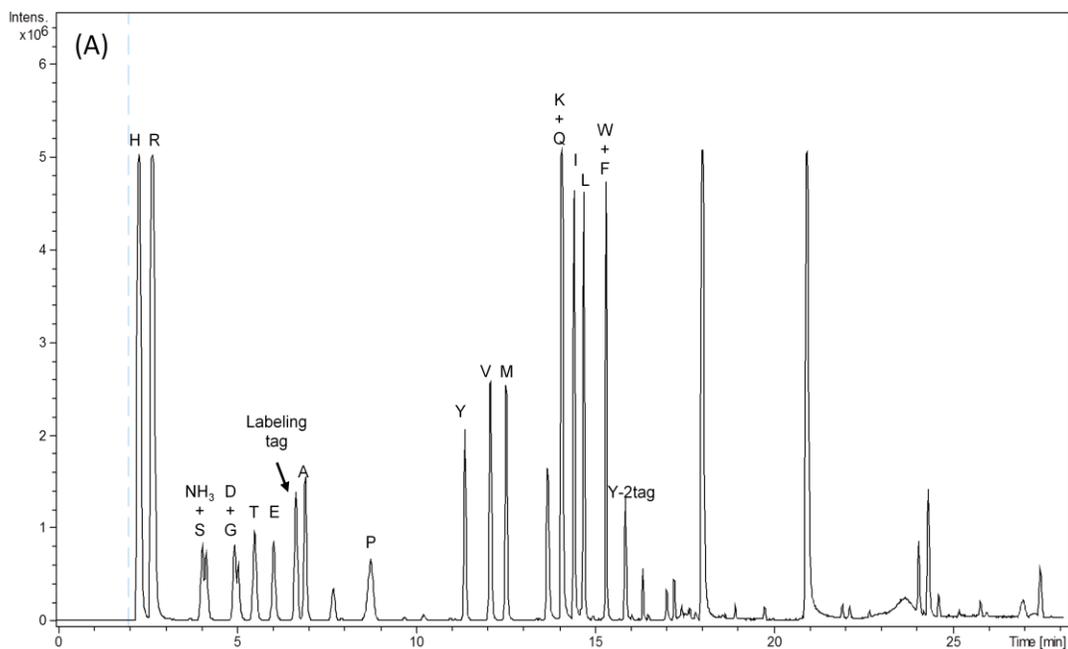
Figure 5-6. (A) LC-MS chromatogram of amino acid standards labeled by MBAA-NHS. (B) LC-MS chromatogram of a urine sample labeled by MBAA-NHS.

5.3.4 Absolute Quantification of Amino Acids in Human Urine

Guo *et al.* have developed a strategy to quantify the absolute concentration of metabolites in a biological sample by using a pooled sample as an internal standard. As long as the authentic standard is available, the absolute concentration of the analyte in the pooled sample can be determined by spiking the $^{13}C_2$-dansylated standard into the $^{12}C_2$-dansylated sample. Consequently, the absolute concentration in the individual samples is determined by comparison of the $^{12}C_2$-labeled individual sample with the $^{13}C_2$-labeled internal standard.[2] In this work, the absolute concentration of 20 amino acids in a pooled urine sample was also examined by MBAA in the same manner. A pooled urine sample was labeled by $^{12}C_2$-MMBA-NHS and the amino acid standard solutions were labeled with $^{13}C_2$-MMBA-NHS. Based on our previous results,[15] appropriate volumes of labeled amino acids were diluted and spiked into the labeled pooled urine prior to LC-MS analysis. As shown in Figure 5-7., except aspartic acid, concentrations of other amino acids in the urine sample have been determined. In future work, the pooled urine sample can act as the internal standard to examine the absolute quantities of amino acids in other individual urine samples.
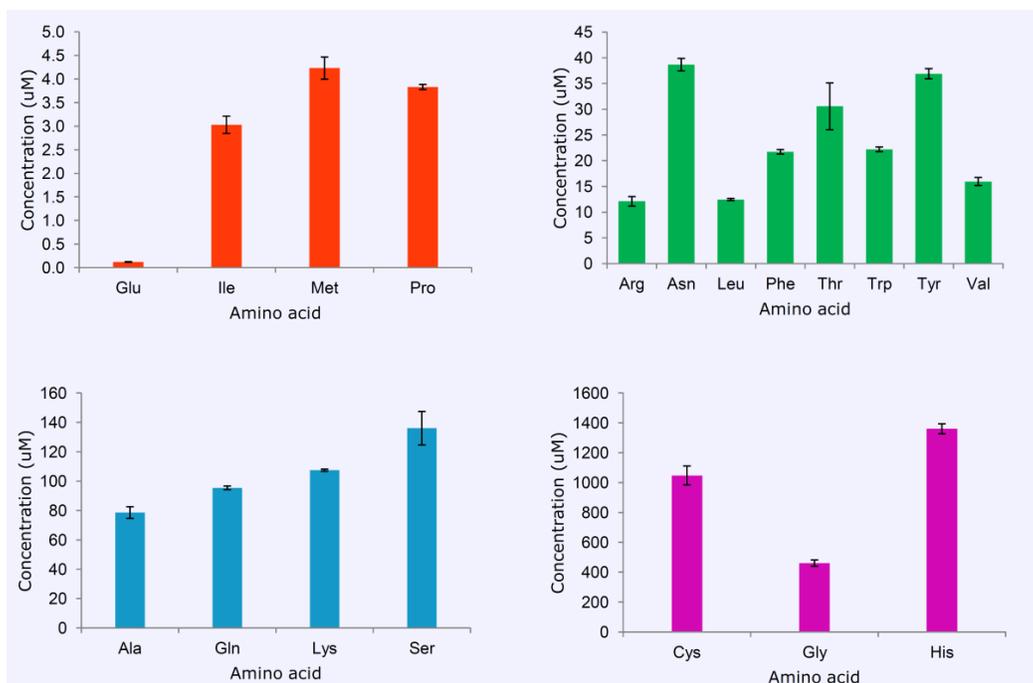
Figure 5-7. Absolute concentrations of 19 amino acids (except aspartic acid) in a pooled human urine sample.

5.3.5 Fragmentation Analysis of Amino Acids Labeled by MBAA-NHS

Although MS-based metabolic identification has employed various techniques to elucidate the structure of an unknown, such as accurate mass, MS/MS, MS^n, retention time in chromatography, metabolic identification lags far behind the progress for the metabolic quantitative analysis.[17] The challenge is mainly attributed to the fact that a range of elemental compositions can be derived from one monoisotopic mass, even within 1 ppm mass tolerance, which results in many ambiguous entries in database research.[18] Fragmentation techniques including MS/MS and MS^n can narrow down possible candidates from the long list, if the spectra of standards or *in-silico* simulated spectra are available in databases.[19,20] Compared to non-derivatization metabolic profiling, the DIL

method is confronted with a special challenge in metabolite identification. To date, few labeled compounds, if any, can give comparable fragmentation patterns to the native standard because of the influence of the labeling tag. To enhance the ionization efficiency, a basic amine or a permanently charged quaternary ammonium is often added to many labeling tags on purpose. These labeling tags and their secondary generation of fragments would hold the positive charge in the fragmentation experiment. As a result, there are very few fragment ions produced for structure diagnosis of the targeted moiety, i.e., the metabolite.[21,22]

As the major objective in the development of the MBAA-NHS reagent, the fragmentation analysis of MBAA labeled amino acids has been examined. In the MS/MS experiment, we have observed that the fragmentation behavior of MBAA labeled amino acids largely derive from the properties of the original amino acids. In contrast to small and hydrophilic amino acids, large and hydrophobic amino acids prefer to generate more fragments related to the unlabeled moiety. As an example, Figure 5-8 A shows the MS/MS spectrum of MBAA labeled tryptophan generated by Q-TOF-MS. At least three fragment ions detected in Figure 5-8 A can be assigned to the unlabeled tryptophan. The fragment with m/z 205 is the protonated ion of tryptophan. The ion of m/z 188 is from tryptophan losing a NH_3 group. The peak at m/z 159 is generated after decarboxylation of tryptophan. The fragmentation pattern can give us fair confidence in the compound identification. On the other end, Figure 5-9 A shows the MS/MS spectrum of labeled serine. The peak at m/z 297 is assigned to the molecular ion. The fragment with m/z 279 is from the molecular ion after loss of

H_2O from the labeled serine and the most abundant peak at m/z 192 is from the MBAA tag. However, the spectrum cannot tell us anything about the structure of serine. This result indicates that chargeability of the side chain of amino acids strongly influences the formation of protonated amino acid moiety and its fragment ions in the CID cell. The side chain of serine is neutral and polar, which makes it difficult to compete with the hydrophobic aromatic ring of the labeling tag for the positive charge. Furthermore, in the MS/MS mode, the instrument parameters have been optimized to detect the high mass region where the labeled molecular ion is located and thus the detectability of the low mass ions is reduced (e.g., for the protonated ion of serine at m/z 106).

For fragmentation analysis, it is important to generate as many fragment ions as possible, in order to provide adequate information for structural diagnosis of unknowns or comparison with the fragment ion spectral library. MS/MS analysis of MBAA labeled amino acids cannot give us definite identification, even for tryptophan. Fortunately, most of the MBAA labeled amino acids can produce the fragment ion of the unlabeled amino acids in the MS/MS experiment, which enables us to resort to the MS^3 technique to produce fragment ions directly from the unlabeled amino acids. Although Q-TOF-MS does not possess the capability of real MS^3, it offers a function known as in-source CID for generating the first generation fragments in the skimmer region. Thus, the unlabeled amino acid fragment ion can be isolated in the first quadrupole, then subjected to the second CID in the second quadrupole to release the second generation of fragment ions. Contrary to MS/MS, this pseudo MS^3 method offers an additional

advantage, as a user can optimize the instrument parameters in regard to ion transmission between the CID cell and the TOF analyzer for low mass ions.

The first example of the pseudo MS^3 spectrum is shown in Figure 5-8 B, acquired from the precursor ion with m/z 205. Compared to Figure 5-8 A, more fragment ions are produced by pseudo MS^3 in the low mass region. In addition to the above annotated fragmentation ions, the peak of m/z 146 is formed by decarboxylation and loss of NH_2. The ion peak at m/z 132 is generated by the ion of m/z 146 losing CH_2. The fragment ion with m/z 118 is the result from the ion at m/z 132 losing another CH_2 after the breakage of aromatic ring. The smallest fragment ion with m/z 74 might be formed by rearrangement of the aromatic ring. Figure 5-8 B is then compared to the MS/MS spectrum of unlabeled tryptophan (see Figure 5-8 C) in the human metabolome database (HMDB),[23] and the two patterns display the same types of fragment ions. Thus, in this case, the pseudo-MS^3 spectrum would provide the fragment ion information required to identify this compound, if this was an unknown metabolite.
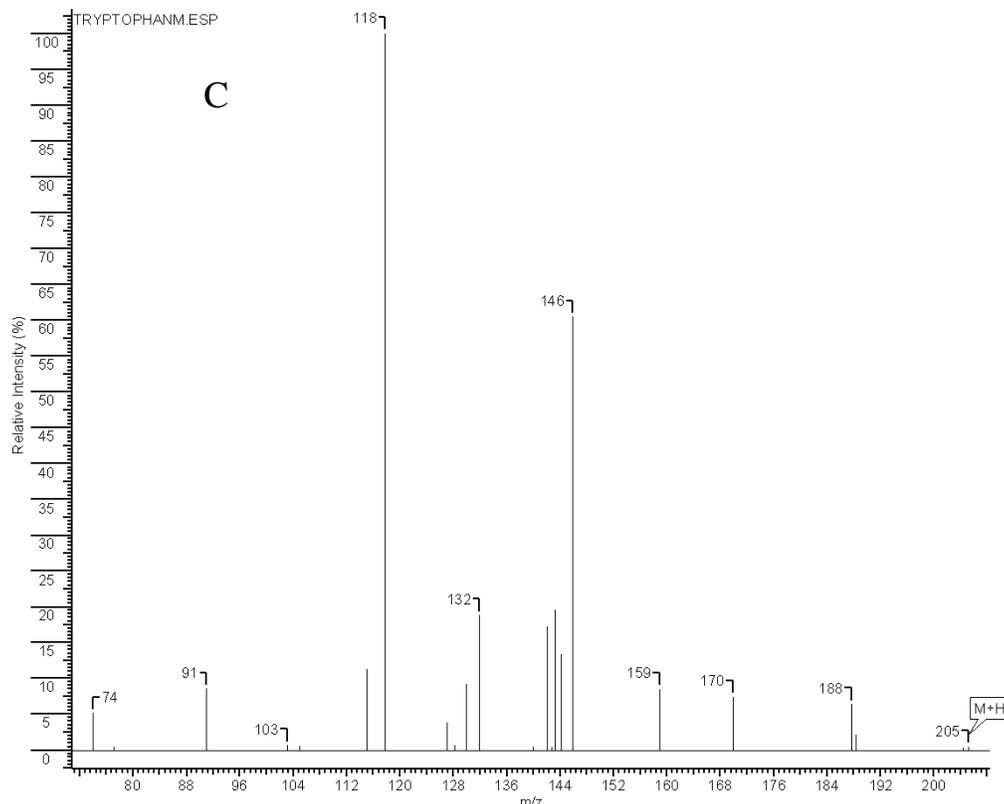
144
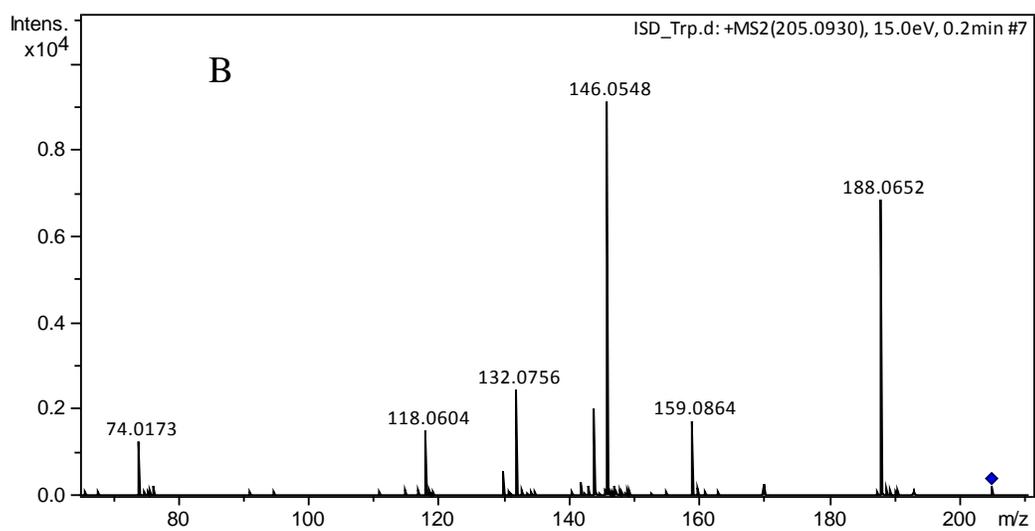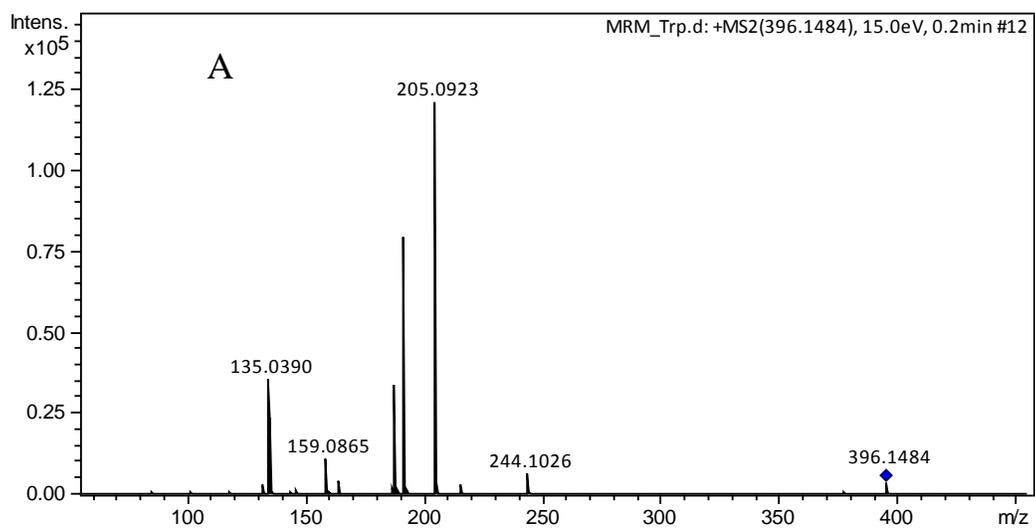
Figure 5-8. (A) MS/MS spectrum of MBAA labeled tryptophan. (B) Pseudo-MS^3 spectrum of MBAA labeled tryptophan. (C) MS/MS spectrum of tryptophan from HMDB.

As discussed earlier, in contrast to tryptophan, MBAA labeled serine provides little useful structural information in the MS/MS analysis. However, the spectrum of pseudo MS^3 produces the two major fragment ions (see Figure 5-9 B): the peak with m/z 88 is formed after serine losing H_2O, and the peak with m/z 60 is generated by decarboxylation. The two fragment ions are also detected in the MS/MS spectrum of unlabeled serine in HMDB as shown in Figure 5-9 C. Thus, the pseudo MS^3 fragmentation can be used for identification of this compound. Even though the pseudo MS^3 spectral quality of serine is not as good as that of tryptophan, the limited number of fragment ions detected can be compensated by

the fact that database search always gives fewer entries for the smaller molecules. We have searched both molecular weights in METLIN within 5 ppm mass tolerance.[24] The database gives two chiral isomers to serine only, but two chiral isomers and five structure isomers to tryptophan. Furthermore, the inferior spectrum quality only occurs for very small molecules, such as glycine and serine. Thus, for most metabolites with higher molecular weights, it would not be a major barrier for metabolic identification.

Sensitivity of the pseudo-MS^3 strategy is mainly dependent of the abundance of protonated unlabeled analyte ions in the skimmer region. As pointed out above, the production of protonated unlabeled analyte ions relies on the chargeability of the moiety itself. In addition to tryptophan, another example is shown in Figure 5-10 A. The pseudo MS^3 spectrum of glutamic acid gives definite information for comparison of fragmentation pattern with the authentic standard (see Figure 5-10 B), even though it has two carboxylic acids and the molecular weight is only at m/z 147. We speculate that the size of targeted compound played a significant role in the dissociation. The concentration of analytes in samples is also important to the sensitivity. In the direct infusion experiment, the concentration of amino acid solutions was diluted to 5 μM, which is a typical abundance of many metabolites found in the human metabolome.[25] Among these amino acid standards, except glycine, all of them give similar fragmentation patterns in their pseudo-MS^3.
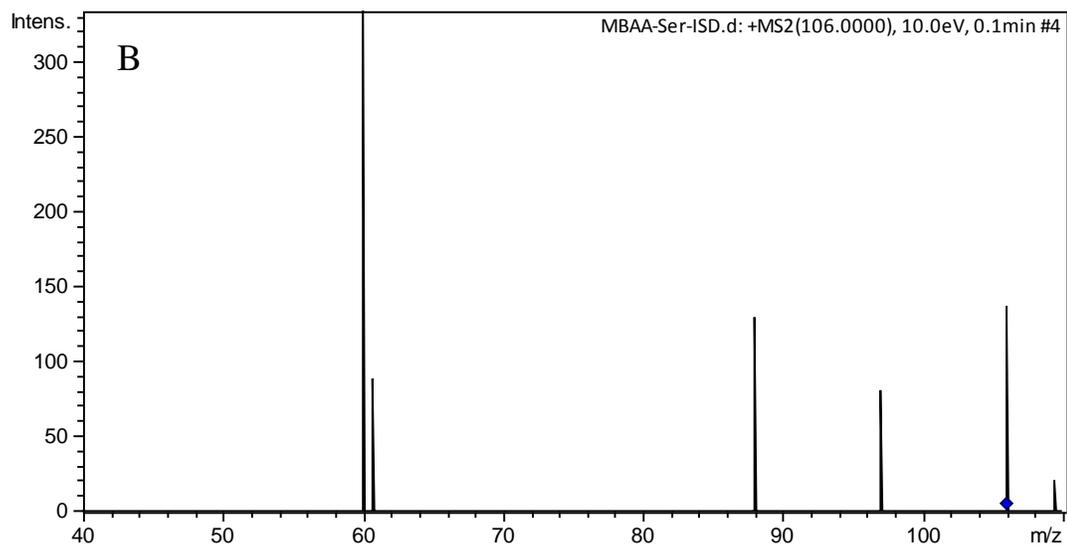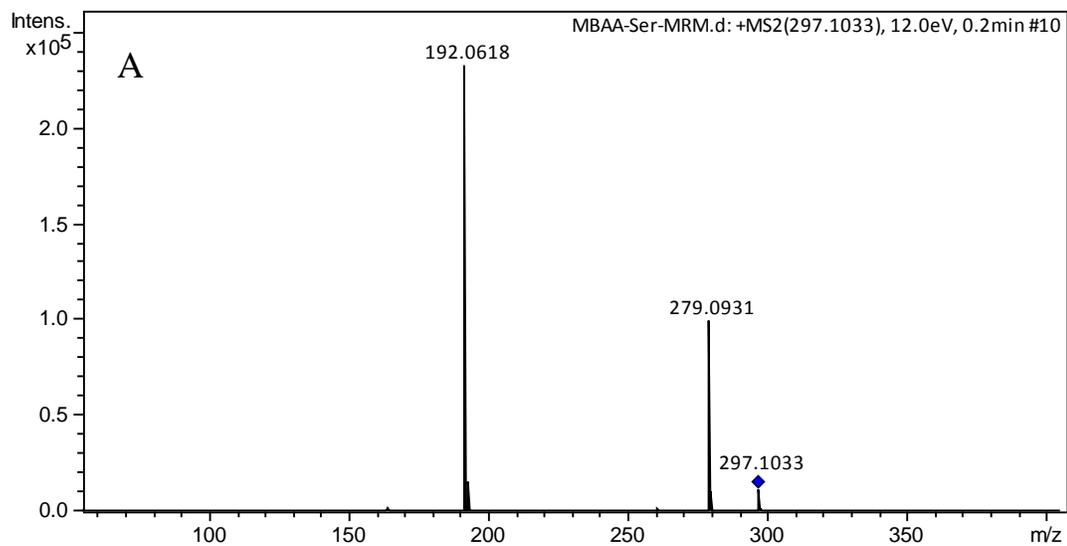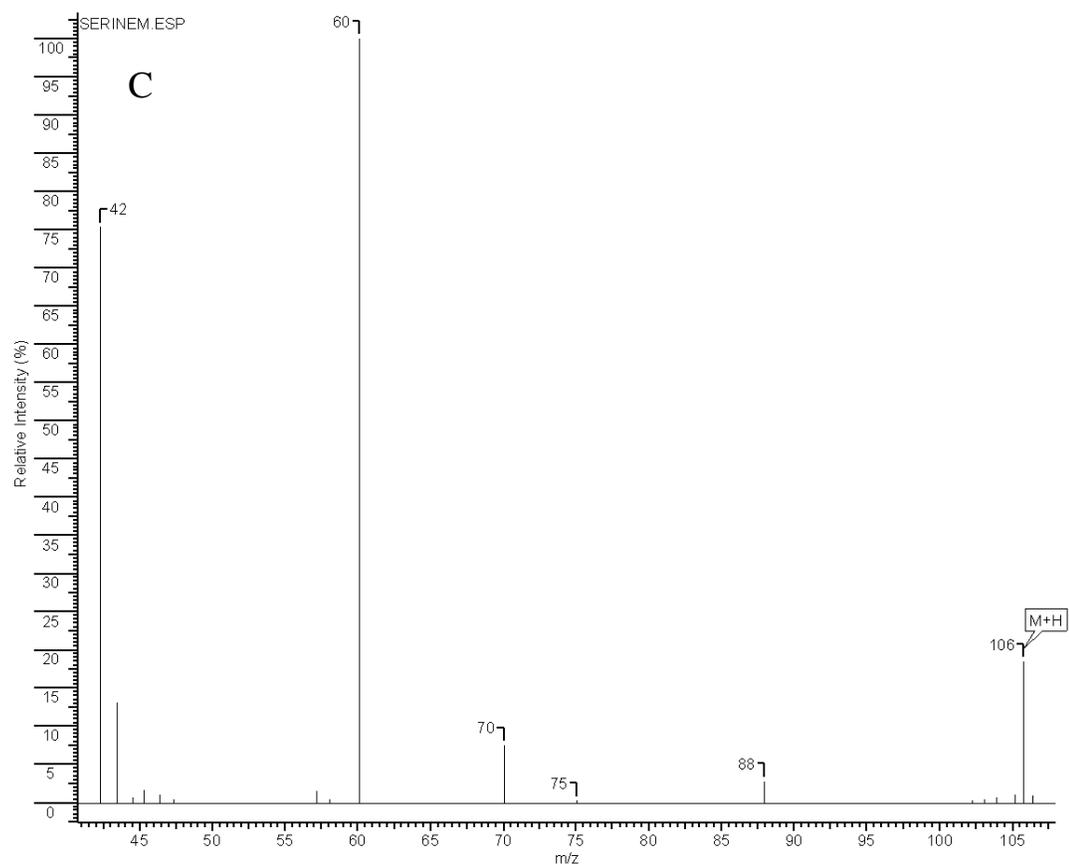
Figure 5-9. (A) MS/MS spectrum of MBAA labeled serine. (B) Pseudo-MS^3 spectrum of MBAA labeled serine. (C) MS/MS spectrum of serine from HMDB.
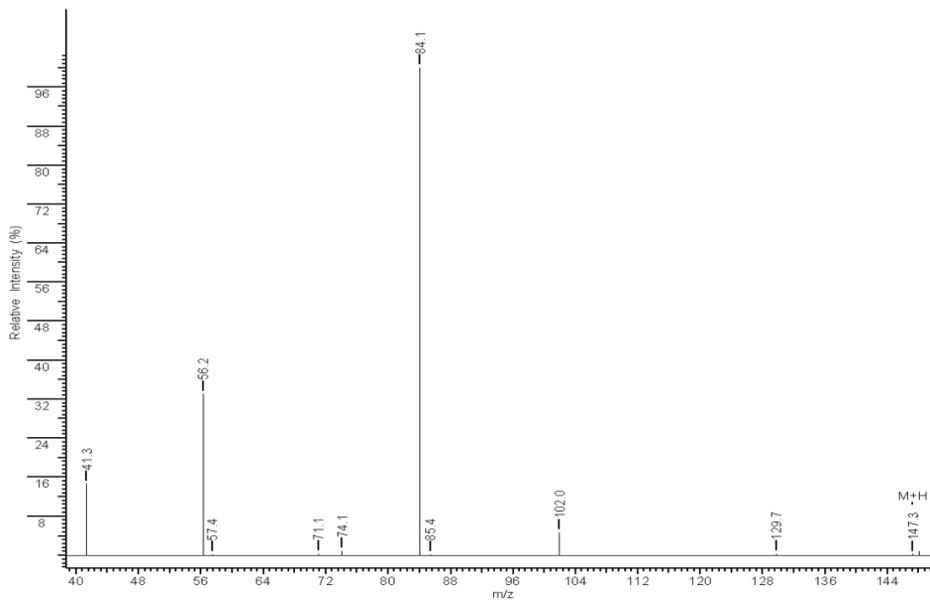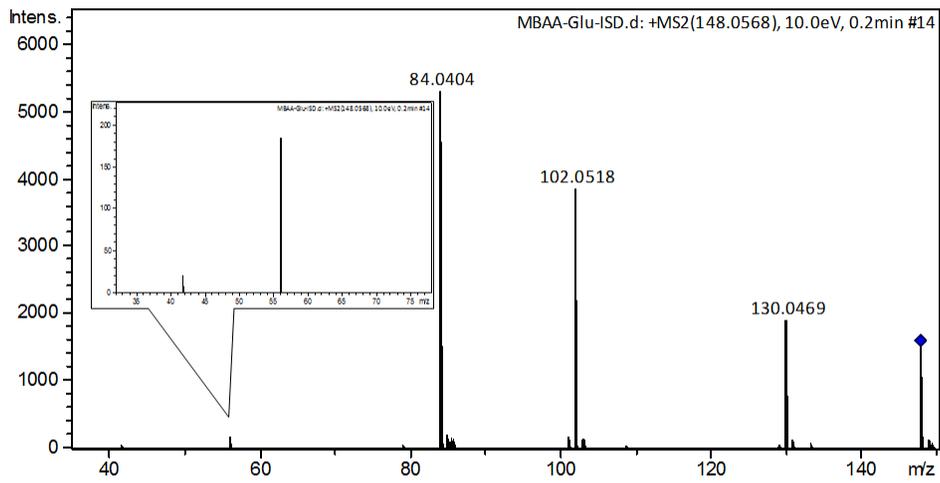
Figure 5-10. (A) Pseudo MS^3 of glutamic acid labeled by MBAA. (B) MS/MS analysis of glutamic acid from HMDB.

5.4   Conclusions

We have developed a strategy to prepare differential isotopic labeling reagents for profiling amine-containing metabolites by which a DIL reagent is divided into three components: isotope group, coupling group and tuning group. We used isotopologues of glycine to serve as an isotope core and a coupling group in DIL reagents so that we can easily modify the properties of DIL reagents by simply changing the tuning group. In addition, the cost of the reagents is quite low due to inexpensive $^{12}C_2$-/$^{13}C_2$-glycine. Thus, these reagents can be used for large-scale applications. We have prepared two DIL reagents using the strategy, $^{12}C_2$-/$^{13}C_2$-MBAA-NHS and $^{12}C_2$-/$^{13}C_2$-DBAA-NHS. MBAA-NHS is specially designed for metabolic identification. We have investigated the fragmentation behaviors of 20 MBAA labeled amino acids. Their pseudo-MS^3 spectra were then compared to the MS/MS library spectra of unlabeled standards. The results indicate that the pseudo-MS^3 spectrum of most MBAA labeled molecules can provide almost identical fragmentation patterns as unlabeled amino acids. We believe that the pseudo-MS^3 strategy based on MBAA NHS method can be used for metabolic identification. The application of this approach for metabolome analysis of real world sample will be reported in the future.

References

(1)   Toyo'oka, T. *J. Pharm. Biomed. Anal.* **2012**, *69*, 174.

(2)   Guo, K.; Li, L. *Anal. Chem.* **2009**, *81*, 3919.

(3)   Guo, K.; Li, L. *Anal. Chem*. **2010**, *82*, 8789.

(4)   Shortreed, M. R.; Lamos, S. M.; Frey, B. L.; Phillips, M. F.; Patel, M.; Belshaw, P. J.; Smith, L. M. *Anal. Chem.* **2006**, *78*, 6398.

(5)   Yang, W. C.; Adamec, J.; Regnier, F. E. *Anal. Chem.* **2007**, *79*, 5150.

(6)   Wu, Y.; Li, L. *Anal. Chem.* **2013**, *85*, 5755.

(7)   Yang, W. C.; Regnier, F. E.; Jiang, Q.; Adamec, J. *J. Chromatogr. A* **2010**, *1217*, 667.

(8)   Yuan, W.; Zhang, J. X.; Li, S. W.; Edwards, J. L. *J. Proteome Res.* **2011**, *10*, 5242.

(9)   Bruheim, P.; Kvitvang, H. F. N.; Villas-Boas, S. G. *J. Chromatogr. A* **2013**, *1296*, 196.

(10)   Guo, K.; Peng, J.; Zhou, R.; Li, L. *J. Chromatogr. A* **2011**, *1218*, 3689.

(11)   Tsukamoto, Y.; Santa, T.; Yoshida, H.; Miyano, H.; Fukushima, T.; Hirayama, K.; Imai, K.; Funatsu, T. *Biomed. Chromatogr.* **2006**, *20*, 358.

(12)   Toyo'oka, T.; Ishibashi, M.; Terao, T. *Biomed. Chromatogr.* **1992**, *6*, 143.

(13)   Abello, N.; Geurink, P. P.; Toorn, M. V. D.; Oosterhout, A. J. M. V.; Lugtenburg, J.; Marel, G. A. V. D.; Kerstjens, H. A. M.; Postma, D. S.; Overkeeft, H. S.; Bischoff, R. *Anal. Chem.* **2008**, 9171.

(14)   Yang, W. C.; Regnier, F. E.; Sliva, D.; Adamec, J. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2008**, *870*, 233.

(15)   Zhou, R.; Guo, K.; Li, L. *Anal. Chem.* **2013**.

(16)   Yang, W.; Mirzaei, H.; Liu, X.; Regnier, F. E. *Anal. Chem.* **2006**, 4702.

(17)   Giavalisco, P.; Hummel, J.; Lisec, J.; Inostroza, A. C.; Catchpole, G.; Willmitzer, L. *Anal. Chem.* **2008**, *80*, 9417.

(18)   Kind, T.; Fiehn, O. BMC *Bioinformatics* **2006**, *7*, 234.

(19)   Xiao, J. F.; Zhou, B.; Ressom, H. W. TrAC, Trends *Anal. Chem.* **2012**, *32*, 1.

(20)   Li, L.; Li, R.; Zhou, J.; Zuniga, A.; Stanislaus, A. E.; Wu, Y.; Huan, T.; Zheng, J.; Shi, Y.; Wishart, D. S.; Lin, G. *Anal. Chem.* **2013**, *85*, 3401.

(21)   Yang, W. C.; Regnier, F. E.; Adamec, *J. Electrophoresis* **2008**, *29*, 4549.

(22)   Guo. K; Ji, C. J.; Li, L. *Anal. Chem.* **2007**, *79*, 8631.

(23)   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 Res*. **2013**, *41*, D801.

(24)   Zhu, Z. J.; Schultz, A. W.; Wang, J.; Johnson, C. H.; Yannone, S. M.; Patti, G. J.; Siuzdak, G. *Nat. Protoc*. **2013**, *8*, 451.

(25)   Becker, S.; Kortz, L.; Helmschrodt, C.; Thiery, J.; Ceglarek, U. *J Chromatogra. B* 2012, 883, 68.

Chapter VI:

Conclusions and Future Work

The aim of metabolomic profiling is to achieve a comprehensive identification and quantification of metabolites in a cell or organism. However, great diversities of physiochemical properties and huge concentration dynamic range of metabolites in a sample make comprehensive metabolomic profiling a very challenging task. Over the past several years, our research group has been involved in the development of a divide-and-conquer strategy to increase the metabolome coverage for comprehensive metabolomics. This strategy divides the entire metabolome into several sub-metabolomes with each containing a common functional group such as an amine or acid group. We then develop analytical methods focusing on sensitive detection and quantification of metabolites belonging to a certain class of compounds. An ideal metabolome analysis can be ultimately accomplished by the combination of all metabolic profiling results.

The analytical methods developed in my thesis work are based on liquid chromatography mass spectrometry (LC-MS), one of the leading platforms used in metabolic analysis because of the unrivaled sensitivity and powerful structural elucidation capacity. However, reversed phase liquid chromatography (RPLC), while offering the highest efficiency for metabolite separation among different LC forms, is not suitable for separating ionizable and very polar metabolites due to little or no retention of these compounds in an RP column. Furthermore, analytes easily encounter ion suppression and matrix effect which results in poor

sensitivity due to co-eluted analytes or background in electrospray ionization (ESI) source. Finally, instrumental drift also impacts the accuracy and precision of quantitative analysis.

To overcome the limitations of LC-MS for metabolome analysis, my research uses a differential isotopic labeling (DIL) approach, which is dfferent from the conventional stable isotopic labeling (SIL) internal standard method that determines the concentration of a targeted analyte based on the signal intensity ratio between the analyte and spiked isotope internal standard. The SIL method has been widely used for targeted quantification analysis in that the internal standard effectively compensates the matrix effect and instrument drift. However, it is not applicable in the metabolomic profiling due to lack of commercial isotopologue source for all the metabolites present in a biological sample. An emerging method known as the differential isotopic labeling (DIL) method has shown the great potential in metabolomic profiling. This method uses chemical derivatization to introduce isotopic internal standards for all targeted metabolites in comparative samples. In addition to high accuracy and precision, the DIL method also improves the separation of ionizable and very polar metabolites in RPLC as well as enhances the ionization efficiency in ESI of many metabolites.

The main objective of my thesis work is to improve the DIL LC-MS method by developing novel DIL approaches and related data processing software. In Chapter I, methods and instruments related to MS-basedmetabolomics studies are briefly introduced.

In Chapter II, an optimized approach based on the $^{12}C_2$-/$^{13}C_2$-DnsCl labeling method developed by Guo and Li has been presented. The original method was used to quantify amine-containing metabolites in biological samples with high sensitivity. The new protocol provides an overall improvement in the synthesis of the DIL reagents, derivatization conditions and sample preparation processes. Compared to the original method, the improved method can at least halve the sample preparation time and save up to 90% of the isotope reagents in the analysis.

In Chapter III, a new set of DIL reagents, $^{12}C_4$-, $^{12}C_2{}^{13}C_2$-, and $^{13}C_4$-5-diethylamino-naphthalene-1-sulfonyl chloride (DensCl), in combination with liquid chromatography Fourier-transform ion cyclotron resonance mass spectrometry (LC-FTICR-MS), are presented for improved analysis of the amine- and phenol-containing sub-metabolome. Synthesis of the reagents is reported and an optimized derivatization protocol for labeling amines and phenols is described. The overall workflow is straightforward, including differential isotope labeling of individual samples and a pooled sample that serves a global internal standard, mixing of the isotope-differentially labeled samples and LC-MS analysis for relative metabolome quantification. The new triplex DensCl reagents offer the advantages of improved metabolite detectability due to enhanced sensitivity (i.e., about 1000 peak pairs detected by DensCl labeling vs. about 600 peak pairs detected by DnsCl labeling) and analysis speed (i.e., simultaneous analysis of two comparative samples by DensCl vs. only one comparative sample analyzed by DnsCl). To demonstrate the utility of the triplex reagents for metabolome

profiling of biological samples, urine samples collected daily from a healthy volunteer over a period of 14 days were analyzed.

Lack of computer tools for processing the DIL method data can be a bottleneck in the application of the method in metabolic profiling. In Chapter IV, a computer program, IsoMS, is reported. The software can automatically detect peak pairs and generate their quantitative information as well as filter various noise. The pair list exported by IsoMS is compatible with different statistical tools, such as MetaboAnalysis. The performance of IsoMS has been tested by various samples labeled with $^{12}C_2$-/$^{13}C_2$-dansyl chloride (DnsCl). Overall, the false positive ratio (FDR) of peak pairs detected by IsoMS is less than 5%. In a pilot study, IsoMS was used to process the LC-MS data generated from the dansylated human urine samples collected before and after the intake of coffee. Based on the relative quantitative information, multivariate analysis (PCA & heatmap) has been used to classify these samples and determine the significant metabolites. This work shows that IsoMS can be an important tool for discovering disease biomarkers and/or classify samples based on metabolome profiling results.

Identification of metabolites still remains a major challenge in metabolomics study. The DIL method encounters an additional challenge in metabolite identification because metabolites labeled with most of the DIL reagents cannot produce informative MS/MS spectra for structural analysis. In Chapter V, a new strategy to develop isotopic labeling reagents is described in which a family of DIL reagents can be prepared by a glycine (or its isotopic

counterpart) and aromatic acids through a simple two-step synthesis. In addition to the low costs, the strategy allows us to adjust the performance of the reagents in accordance with the research objective. To date, two DIL reagents, 4-dimethylaminoBenzoylamido acetic acid *N*-hydroxylsuccinimide ester (DBAA-NHS) and 4-methoxybenzoylamido acetic acid *N*-hydroxylsuccinimide ester (MBAA-NHS), have been synthesized. Their labeling conditions and the related LC-MS method have been optimized. DBAA-NHS labeling increases the metabolite detectability because of the presence of an ESI-active dimethylaminobenzoyl group. Another reagent, MBAA-NHS, is shown to provide excellent fragmentation behavior in MS/MS and pseudo MS^3 experiments which make MBAA-NHS as a useful reagent for structure analysis of metabolites of interest in a metabolome sample. We envisage the use of MBAA-NHS for both quantitative metabolome analysis and metabolite identification.

Future research will involve further development of MBAA-NHS for metabolomics. The real world applications of MBAA-labeling LC-MS for a variety of metabolomics projects includes serum and urine metabolomics of disease states and controls. Cellular metabolomics will also be explored. In particular, we will need to compare the analytical performance of Dns-labeling LC-MS and MBAA-labeling LC-MS in terms of metabolome coverage. While Dns-labeling provides better detection sensitivity than MBAA-labeling, it may still be possible that a similar detectability can be achieved if a larger amount of MBAA-labeled sample is injected for analysis. A variety of experimental conditions including the optimal sample injection will need to be studied. The

applications of MBAA labeling LC-MS for structural analysis of unknown metabolites will need to be demonstrated. In addition, while my thesis research has used one-dimensional LC for separation of the labeled metabolites, two-dimensional LC methods should be explored in the future to increase the metabolome coverage.