Development and Applications of Chemical Isotope Labeling Methods for Metabolome Sample Normalization and Liquid Chromatography-Mass Spectrometry-Based Metabolomics

by

Yiman Wu

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

Doctor of Philosophy

Department of Chemistry
University of Alberta

Abstract

Metabolomics refers to the global study of the metabolome of a biological system where comprehensive and systematic profiling of all metabolites in a given biological sample (i.e., the metabolome) is performed. As the end product of biological processes, the status of metabolites could reflect the physiological state of an organism and provide valuable and complementary information to the genomics, transcriptomics and proteomics data. However, as a relatively new research field, there are some challenges associated with LC-MS-based metabolomics. For example, the total metabolite concentrations often vary from sample to sample, which may complicate relative quantification of the metabolome changes in comparative studies. Also, the high complexity of biological samples makes it difficult to quantitatively extract all metabolites and sensitively detect low abundance metabolites.

Towards these challenges, the objective of my research was two-fold. Firstly, a sample normalization method based on UV absorbance measurement was developed to allow quantification of the total concentration of chemically labeled metabolites (Chapters 2 and 3). This method can be readily applied to any type of biological samples for effective correction of sample concentration variations. Application of this normalization strategy was demonstrated on human urine and bacterial cell extracts with significantly reduced inter-group variations and improved statistical results. Secondly, differential isotope dansylation labeling LC-MS metabolomics workflows were developed for various biological matrices including bacterial cells, plasma and cerebrospinal fluid (Chapter 4 to 6). This labeling chemistry targets at the amine and phenol sub-metabolome with improved chromatographic separation and detection

sensitivity. The developed protocols enabled the detection of over 1000 putative metabolites in each biological sample, and have been applied to bacterial differentiation and biofluid disease biomarker discovery studies. Overall, these research activities have demonstrated enhanced analytical performance and capability of LC-MS-based metabolomics methods for real biological applications.

Preface

A version of Chapter 2 was published as: Yiman Wu and Liang Li, "Determination of Total Concentration of Chemically Labeled Metabolites as a Means of Metabolome Sample Normalization and Sample Loading Optimization in Mass Spectrometry-Based Metabolomics", *Anal. Chem.* **2012**, *84*, 10723-31. I was responsible for design of experiments, data collection and analysis, as well as manuscript preparation. L. Li supervised the project and edited the manuscript.

A version of Chapter 3 was published as: Yiman Wu and Liang Li, "Dansylation Metabolite Assay: a Simple and Rapid Method for Sample Amount Normalization in Metabolomics", *Anal. Chem.* **2014**, *86*, 9428-33. I was responsible for design of experiments, data collection and analysis, as well as manuscript preparation. L. Li supervised the project and edited the manuscript.

A version of Chapter 4 was published as: Yiman Wu and Liang Li, "Development of Isotope Labeling Liquid Chromatography-Mass Spectrometry for Metabolic Profiling of Bacterial Cells and Its Application for Bacterial Differentiation", *Anal. Chem.* **2013**, *85*, 5755-63. I was responsible for design of experiments, data collection and analysis, as well as manuscript preparation. L. Li supervised the project and edited the manuscript.

Acknowledgements

First and foremost, I would like to express my sincere gratitude and appreciation to my supervisor, Professor Liang Li, for his expert guidance, invaluable advices, understanding and encouragement, and support throughout my graduate studies. Indeed, this thesis work would not have been possible without his incredible patience and counsel.

I would also like to extend my appreciation to the members of my supervisory committee, Professor Mark T. McDermott and Professor Christopher W. Cairo, the members of the thesis examining committee, Professor Derrick L.J. Clive, and the external examiner, Professor Janusz Pawliszyn, University of Waterloo, for the time attending my oral examination and reviewing my thesis, and for their insightful comments and suggestions on my research work. I also want to thank Professor Robert E. Campbell and Professor John C. Vederas for attending my candidacy examination and providing valuable advices on my research projects.

My deep gratefulness also goes to all the people with whom I have worked with during my graduate study. In particular, I want to thank Dr. Avalyn E. Stanislaus and Dr. Azeret Zuniga, for their valuable training at the beginning of my study. I would also like to express my gratitude to Professor Kelvin S. Y. Leung and his team, for providing me plasma samples and much advice, and to Dr. Chiao-Li Tseng for her extraordinary hard work and great contributions to the autism project (Chapter 5). Heartfelt gratitude is also given to Professor Brian Kwon and Dr. Femke Streijger, for providing me serum and cerebrospinal fluid samples, as well as much advice on the spinal cord injury project (Chapter 6). Moreover, I want to thank all group members in Professor Li's research

group for their helpful discussions and assistance throughout these years. Working with them made my graduate experience enjoyable.

In addition, I would like to thank all the staffs in the Department of Chemistry, University of Alberta. I'm especially grateful to Gareth Lambkin from the Biological Services Facility for his kind assistance in cell culture and use of equipment. I also want to acknowledge Dr. Randy Whittal, Jing Zheng and Bela Reiz from the Mass Spectrometry Facility for their helpful suggestions and professional support. Thanks are also given to all other departmental staffs for their kind assistance and to all those who helped me with my work.

Finally and most importantly, I wish to express my special thanks to my parents, Mr. Leping Wu, Mrs. Qin He, and my husband Wei Sun for their unconditional love and support over the years, and to my son Andrew, for the endless happiness he has brought to me.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AC	Alternating Current
ACN	Acetonitrile
AIS	American Spinal Injury Association Impairment Scale
ANOVA	Analysis of Variance
APCI	Atmospheric Pressure Chemical Ionization
API	Atmospheric Pressure Ionization
APPI	Atmospheric Pressure Photoionization Ionization
ASD	Autism Spectrum Disorder
AUC	Area Under the Curve
BCA	Bicinchoninic Acid
BPC	Base Peak Chromatogram
BSA	Bovine Serum Albumin
CE	Capillary Electrophoresis
CID	Collision-Induced Dissociation
CRM	Charge Residue Model
CSF	Cerebrospinal Fluid
CV	Coefficient of Variation
Da	Dalton
DC	Direct Current
DIL	Differential Isotopic Labeling
DMA	Dansyl Metabolite Assay
DNA	Deoxyribonucleic acid
Dsn	Dansyl
EI	Electron Impact Ionization
ESI	Electrospray Ionization
FC	Fold Change
FT-ICR-MS	Fourier Transform Ion Cyclotron Resonance Mass Spectrometry
GC	Gas Chromatography
GC-MS	Gas Chromatography Mass Spectrometry
GGT	Gamma-glutamyl Transferase
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HMDB	Human Metabolome Database
HILIC	Hydrophilic Interaction Liquid Chromatography
HPLC	High Performance Liquid Chromatography
ICR	Ion Cyclotron Resonance
IEM	Ion Evaporation Model
LC	Liquid Chromatography
LC-MS	Liquid Chromatography Mass Spectrometry
LC-UV	Liquid Chromatography Ultraviolet
LLE	Liquid -liquid Extraction
LOESS	Locally weighted scatter plot smoothing
MALDI	Matrix-assisted Laser Desorption Ionization
MeOH	Methanol
MRM	Multiple Reaction Monitoring

MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MSTUS	MS Total Useful Signal
m/z	Mass to Charge
nm	Nano meter
NMR	Nuclear Magnetic Resonance
OD	Optical Density
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis
PDA	Photo Diode Array
PLS-DA	Partial Least Square Discriminant Analysis
ppm	part(s) per million
QC	Quality Control
Q-TOF-MS	Quadrupole Time-Of-Flight Mass Spectrometry
RF	Radio Frequency
ROC	Receiver Operating Characteristic
RT	Retention Time
RPLC	Reversed Phase Liquid Chromatography
RSD	Relative standard derivation
SCI	Spinal Cord Injury
SDS	Sodium Dodecyl Sulfate
S/N	Signal to noise ratio
SPE	Solid Phase Extraction
SVM	Support Vector Machine
TOF	Time-Of-Flight
UPLC	Ultra Performance Liquid Chromatography
UV	Ultra-violet
VIP	Variable Importance on the Projection
W	Watts
μM	Micro molarity

Chapter 1

Introduction

1.1 Overview of Metabolomics

The term "metabolome" was first coined by Olivier in 1998 to refer to the complete set of metabolites synthesized by an organism.[1] Consequently, the word "Metabolomics" describes the comprehensive and systematic analysis, including both identification and quantification, of the collection of all metabolites in a given biological system.[2,3] A related term, "metabonomics", is defined as the quantitative measurement of metabolic responses to pathophysiological stimuli or genetic modification,[4] with emphasis on differentiation between population groups. While metabonomics is a subset of metabolomics, the two terms are often used as synonyms.[5] Metabolomics is now a rapidly growing field that allows high-throughput profiling of metabolites at the global level. As the end product of all biological processes, the status of metabolites could be used to reflect the physiological state of an organism as a result of interactions between gene/protein expression, and the environment changes (Figure 1.1).[6,7] Therefore metabolomics often provides valuable and complementary information to the genomics, transcriptomics and proteomics data, and it is the "omics" approach closest to the phenotype.[8] For this reason, metabolomics has gained a growing number of research interests in biological applications,[9] such as discovery of disease-related biomarkers,[10-15] plant biotechnology [16,17] and toxicology.[5,18,19]

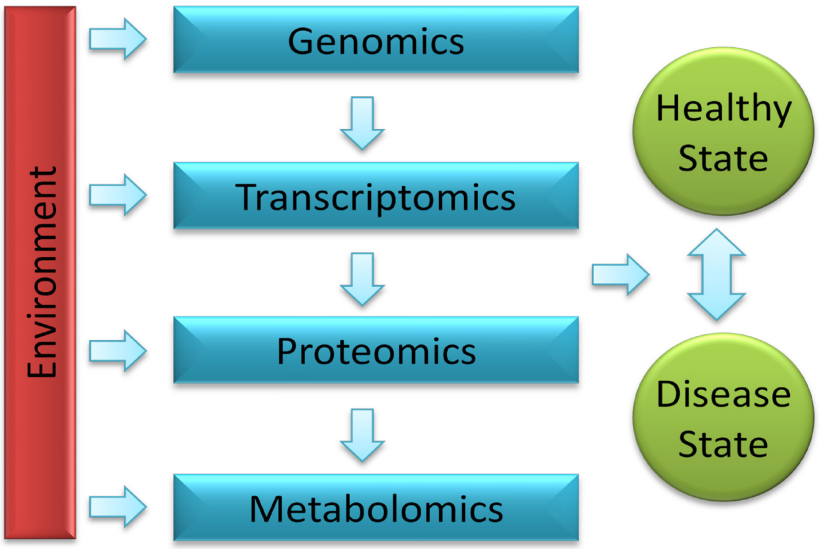


Figure 1.1 Relationship between metabolomics, other "omics" approaches, environment and the physiological state. Adapted from Reference 7.

There are two types of metabolomics studies: targeted and untargeted. In targeted analysis, selected metabolites or metabolite classes are examined, which are usually associated with one or more particular pathways.[20] This approach is usually hypothesis-driven and is often used to study particular biological questions. Since the properties of the investigated metabolites are known, sample preparation protocols and analysis methods can be optimized to minimize interferences from the matrix.[21] Selected reaction monitoring is also often performed to improve sensitivity for compounds of interest.[22] In contrast, the purpose of untargeted metabolomics is to simultaneously measure as many metabolites as possible in a biological system. This approach is hypothesis-generating which aims to detect metabolic perturbations in different biological groups and to identify pathways responsible for the phenotype variations. Untargeted metabolic profiling is more challenging compared to the targeted approach in terms of experimental designs, instrument capabilities and data complexity. A typical metabolomics study workflow is

shown in Figure 1.2. It starts with an untargeted metabolic profiling to identify metabolites that are significantly altered, followed by targeted analysis of selected metabolites for verification, quantification and pathway analysis. Since this thesis focuses on untargeted metabolomics, all the following discussions are made with an emphasis on the untargeted approach unless otherwise specified.

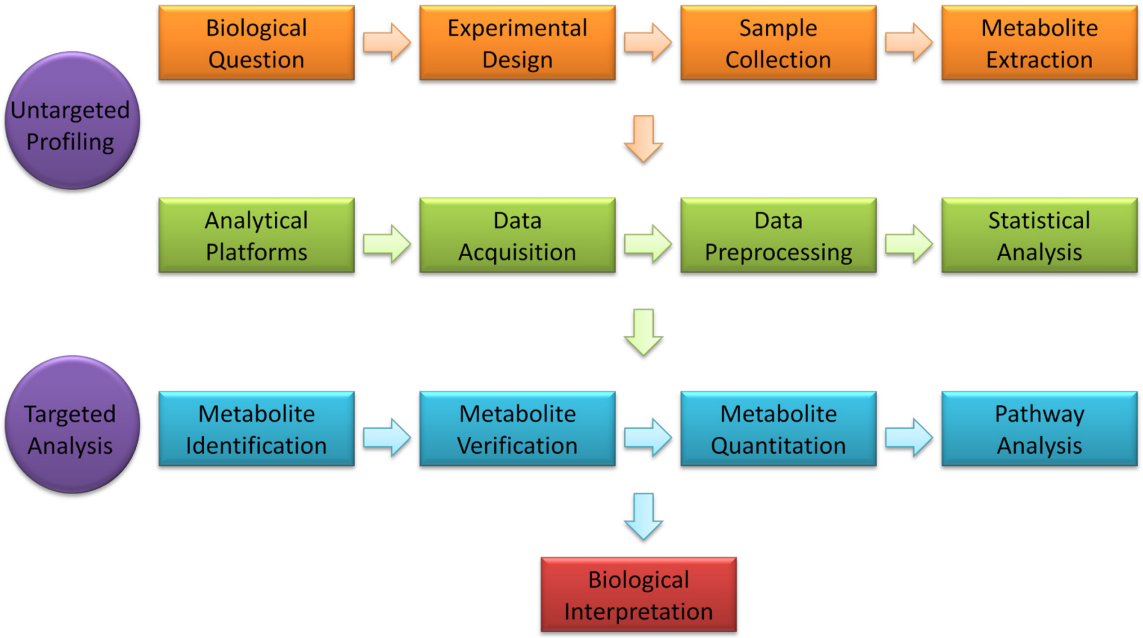


Figure 1.2 A typical metabolomics workflow. Adapted from Reference 21.

As an emerging field, there are still several analytical challenges associated with metabolomics. First of all, there's limited metabolome coverage by current analytical methods and it's not possible to profile all metabolites with a single method. Unlike genes and proteins which are composed of only four nucleotides and twenty amino acids respectively, there's a much greater chemical diversity of metabolites. In addition, there's also a wide metabolite concentration dynamic range. For example, in plasma/serum samples the metabolite concentration range is estimated to be 9 orders of magnitude.[23] These problems may be alleviated by the 'divide-and-conquer' strategy as applied in

proteomics studies.[24] In this context, it means that metabolites may be divided into groups according to their structures or properties, and to each group an optimal analytical method is applied to address chemical variations such as solubility, reactivity and detectability.[25] Secondy, the high complexity of biological samples makes it difficult to quantitatively extract all metabolites and also poses a challenge for the detection of low abundance metabolites. This requires adequate sample preparation methods as well as robust and sensitive analytical methods to be performed. Moreover, metabolite identification is often considered as the most challenging part, particularly in liquid chromatography-mass spectrometry (LC-MS)-based metabolomics studies. It is not feasible in practice to obtain authentic standards of every single compound for definitive identification. Therefore, putative identifications are often used instead by spectral match with database. There are a number of metabolome databases available such as HMDB, [26] Metlin,[27] and MassBank.[28] However, there are still difficulties associated with this approach. For example, LC-MS and MS/MS spectra vary widely with different instruments and experimental conditions, which may lead to erroneous results if not examined carefully.[29] Run-to-run variations in retention time and the lack of a standardized retention index further complicate spectral match in LC-MS-based analysis. There's also limited metabolite coverage in all databases, and in general only less than 30% of peaks can be identified uniquely through database search.[21] Currently there's great research effort in expanding the metabolome database in order to improve metabolite identification.[30-32]

Because of the highly complex nature of metabolites, a variety of analytical tools have been applied to approach metabolomics applications. A brief comparison of the most commonly used analytical techniques will be discussed in the following section.

1.2 Comparison of Major Metabolomics Platforms

Three of the most predominately used techniques are nuclear magnetic resonance (NMR) spectroscopy, gas chromatography mass spectrometry (GC-MS) and liquid chromatography mass spectrometry (LC-MS). NMR has been widely used in metabolomics for its ability to simultaneously identify and quantify a wide range of small molecules at medium to high concentration levels.[33-35] NMR spectra provide rich structural information and are highly reproducible, which is important for unambiguous identification of metabolites in complex mixtures.[36] It also has the advantages of simple sample preparation, instrument automation, and the non-destructive nature, which makes it well-suited for non-invasive analysis of biofluids.[37] The major drawback of NMR is the relatively poor sensitivity, which limits its use for metabolites higher than the micro-molar range. Unfortunately, many potential metabolic biomarkers have very low concentrations, and therefore cannot be detected by NMR. On the other hand, MS-based metabolomics technique is more sensitive than NMR by several orders of magnitude. It is usually coupled with a high-throughput separation technique such as GC and HPLC to reduce spectrum complexity so as to increase the number of detected metabolites. GC-MS is the method of choice for thermally stable, volatile compounds or metabolites that can be made volatile through derivatization. It has been widely used to analyze organic acids, amino acids, sugars and fatty acids.[37] One advantage of GC-MS is that consistent data can be obtained with different instruments and across laboratories because of the

highly reproducible electron ionization (EI) fragmentation pattern and the indexed retention time.[8,25] This enables reliable compound identification through spectra match with database. Nowadays the EI spectra library is very comprehensive.[38,39] However, on the downside, the applicability of GC-MS is limited by the requirement for volatile or derivatizable analytes. For metabolites that are labile, nonvolatile or cannot be derivatized, include many polar and ionic compounds or large molecules, LC-MS becomes the method of choice. Therefore, LC-MS is considered more comprehensive in terms of metabolome coverage. It also provides the advantages of high sensitivity and versatility. Several separation modes are available, and the most common ones used in metabolomics are reversed-phase (RP) (for moderately polar to non-polar) and hydrophilic interaction liquid chromatography (HILIC) (for highly polar to ionic).[40] However, the linear quantification range may be limited by the ion suppression phenomenon with electrospray ionization (ESI),[25] and the use of an internal standard is highly desirable for quantitative analysis. The ESI spectral library is also not as reproducible and comprehensive as the EI library, as described above. Since all analytical techniques have their strengths and limitations, they are complementary to each other and should be combined for a comprehensive metabolomics analysis.[41,42] Since my research focuses on development of LC-MS methods, the analytical techniques employed in LC-MS-based metabolomics will be discussed in more detail below.

1.3 Analytical Techniques in LC-MS-based Metabolomics

1.3.1 Reversed Phase Liquid Chromatography

Reversed phase liquid chromatography (RPLC) is the most widely used separation means in LC-MS-based metabolomics. In RPLC, the stationary phase is made

non-polar by covalently bonding hydrophobic groups such as C18 and C8 onto microporous silica particles. The mobile phase is a mixture of water and an organic solvent that is miscible with water, most commonly acetonitrile and methanol. Other organic solvents such isopropanol and tetrahydrofuran are used less often. There are several features of RPLC that make it particularly useful for metabolomics analysis. Firstly, the hydrophobic stationary phase allows separation of most moderately polar to non-polar compounds, which covers a wide range of metabolites. Secondly, because the mobile phase contains water, it is compatible with water-based biological samples such as urine and serum, thus minimizes sample preparation. Thirdly, there are a wide range of choices for method development, including column dimensions, particle size, stationary-phase type and mobile phase pH.[43] This versatility is especially important for method optimization in untargeted profiling, as efficient separation is required to resolve as many metabolites as possible. Narrow bore columns with particle sizes of 3-5 μm have been widely used in RPLC. More recently, the use of sub-2 μm columns, known as ultra-high performance liquid chromatography (UHPLC or UPLC), have become popular because they enable faster separation without compromise in efficiency.[44-46] Fast analysis allows for improved sample throughput, which is highly desirable in metabolomics studies where a large number of samples need to be analyzed to produce statistically meaningful data. It should be noted though that the use of smaller particles is at the cost of increased column pressure, which requires special equipment. When running at moderate flow rates, sub-2 μm columns can also be used on traditional HPLC systems to provide high column efficiency. In general, RPLC provides a convenient, versatile, efficient and reproducible

approach for metabolite separation, and couples readily with common detection techniques such as UV and ESI-MS.

1.3.2 Ultraviolet-Visible Absorption Spectroscopy

Ultraviolet-visible (UV-Vis) absorption spectroscopy is the most commonly used detection method for HPLC. The principle of UV-Vis absorption is based on the fact that valence electrons can be excited to higher energy levels.[47] Since all organic compounds contain valence electrons, they all have the ability to absorb electromagnetic radiations. However, for single bond electrons, absorption occur in the vacuum ultraviolet region (λ < 185 nm), which is difficult to measure experimentally because of interferences from the atmosphere. The more useful transitions are associated with compounds that have π orbitals. The absorption energies for n or π electrons to be excited to π^* orbitals correspond to bands in the ultraviolet-visible region (190 – 600 nm). Molecules that are capable of absorbing energy in this region are known as chromophores. Most metabolites are organic compounds with unsaturated functional groups or heteroatoms, and therefore can be detected in UV-Vis absorption spectroscopy.

UV-Vis absorption spectroscopy is particularly useful for quantitative determination of analytes. At low analyte concentrations, the measured absorbance is proportional to the sample concentration, as described by Beer's law:

$$A = \log\frac{I_0}{I} = \varepsilon bc \qquad (1.1)$$

Where A is the absorbance, I_0 is the incident light intensity, I is the transmitted light intensity, ε is the molar absorptivity of the sample, b is the path length, and c is the sample concentration. When there is more than one component in the sample solution, the absorbance is additive:

$$A_{total} = A_1 + A_2 + \ldots + A_n = \varepsilon_1 bc_1 + \varepsilon_2 bc_2 + \ldots + \varepsilon_n bc_n \quad (1.2)$$

Although the Beer's law is useful for quantitative measurement of single or a mixture of analytes, it should be noted that it also has certain limitations. For example, at high sample concentrations (> 0.01 M), solute-solvent or solute-solute interactions can affect the analyte environment and may alter the absorptivity. Other deviations include changes in refractive index, compound dissociation or association, interferences from polychromatic radiation or stray radiation.[47] Therefore, it is important to determine the linear range for a particular detection system and to make sure the concentration of measured samples falls in the linear range. Nevertheless, today's UV-Vis detectors can easily achieve linear ranges as wide as 10^5, making it a versatile method for analysis of both trace and major components in a sample solution.[43]

When coupled with LC, UV-Vis detectors typically adapt a Z-shaped flow cell design (Figure 1.3). The light interacts with the liquid stream inside the flow cell, and when a sample passes through the tube, any absorption by the sample will reduce the light intensity compared to the mobile phase alone. The reduction is then converted to absorbance values.

UV-Vis spectroscopy is not a good method for qualitative identification, and therefore is usually not used by itself in metabolomics studies, although there are exceptions.[48] Nevertheless, because of its ease of operation and excellent quantitation capability, it can serve as a useful technique for development of LC-MS-based metabolomics methods. For example, it can be used to compare intensity of peaks from different sample preparation methods, check reagent or product purity,[49] optimize LC gradients, quantify sample amounts prior to MS detection,[50] and guide 1st dimension

fraction collection in offline two-dimensional separations.[50,51] In my research LC-UV has been used as an important tool for metabolite quantification and biological sample normalization prior to LC-MS analysis.

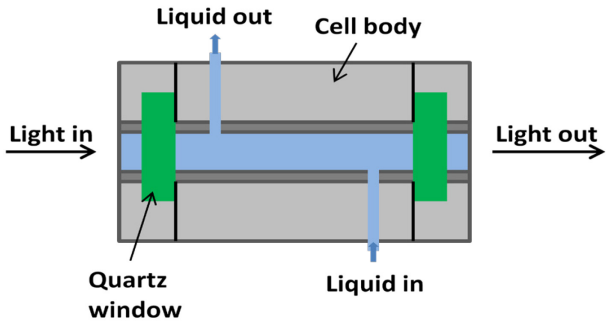


Figure 1.3 Schematic representation of a typical z-shaped flow cell. Adapted from reference 43.

1.3.3 Electrospray Ionization

Atmospheric pressure ionization (API) interfaces play a key role in the success of coupling liquid chromatography to mass spectrometry. The API interface is a specially designed spray chamber that takes liquid elutes from LC, performs aerosol generation, ionization and solvent evaporation under atmospheric pressure, and directs desolvated ions into the vacuum region of the mass spectrometer. Several API techniques have been developed, including electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), and atmospheric pressure photoionization (APPI). Among them, ESI has been most widely used for metabolomics analysis and is the technique I have used in my research. The use of electrospray as ionization source for mass spectrometry was first introduced by Fenn in 1984.[52] In general, the ESI process involves three major steps:[53,54] 1) production of charged droplets at the capillary tip; 2) shrinkage of charged droplets

and repeated droplet disintegration; 3) generation of gas-phase ions. Each step will be elaborated in more detail below.

The schematic representation of the ESI process is illustrated in Figure 1.4. In the first step, a high voltage (2-5 kV) is applied between the metal capillary and the counter electrode. At the narrow capillary tip a strong electric field is created ($E \approx 10^6$ V/m), which will partially penetrate the liquid at the tip and cause the polarized solvent to separate into positive and negative ions. When a positive potential is applied to the capillary, positive ions in the liquid will drift towards the liquid surface. Under the electric field positive ions are drawn downfield, but they cannot escape from the liquid because of surface tension. These two competing forces results in the formation of a stable liquid cone, which is known as Taylor cone (named after Geoffery Ingram Taylor who first investigated the phenomenon).[55] When the field is sufficiently high, the Taylor cone becomes unstable and a liquid filament is emitted from the cone. At some distance downstream excessive positive charges break the filament into a mist of droplets. At high flow rates (>10 μL/min), this process is also facilitated by the strong shear forces created by the nebulizing gas, which is introduced into the spray chamber coaxially through a tube that surrounds the capillary needle.

In the second step, a counter flow of neutral heated drying gas (usually nitrogen) is used to assist evaporation of solvents from the droplets. The evaporation process decreases the droplet diameter, thus increasing the charge density on each droplet. As droplets continue to shrink, they would eventually meet the Rayleigh limit in which the columbic repulsion just becomes sufficient to overcome surface tension. The parent droplets then become unstable and undergo fission to emit a tail of much smaller

offspring droplets. The volume of offspring droplets is only about 2% of parent droplets, but they inherit ~15% of the charges.[54] This much higher charge-to-mass ratio forces the offspring droplets to undergo further disintegrations. The shrink-and-fission process repeats until the resulting charged droplets become sufficiently small to produce gas-phase ions directly.

The last step is generation of gas-phase ions from highly charged nanometer-sized droplets. Two mechanisms have been proposed for this step: the charged residue model (CRM) and the ion evaporation model (IEM). The charged residue model was first proposed by Dole.[56] This model suggests that the repeated shrink-and-fission process eventually results in the formation of extremely small droplets that contain only one analyte molecule. Consecutive solvent evaporation leads to conversion to gas-phase ions with charges originates from the vanished droplet. This mechanism is best used to describe gas-phase ion formation for macromolecules such as proteins, and the resulting ion is multiply charged. The ion evaporation model was proposed by Iribarne and Thomson,[57] which states that when the radii of the droplets decrease to less than 10 nm, direct ion emission will occur instead of Coulomb fission. This mechanism was supported experimentally by the fact that Na^+ and hydrated $Na(H_2O)_k^+$ (k = 1-3) were observed in large abundance in ESI of NaCl solutions in addition to large ion aggregates. In general the ion evaporation mechanism supports gas-phase ion formation for small molecules.

ESI is a soft ionization method that does not involve localized heating during the ionization process. Since there's very little internal energy transferred to the molecular

ions, the molecular ions often remain stable. Therefore the ESI process is particularly useful for molecular weight determination.

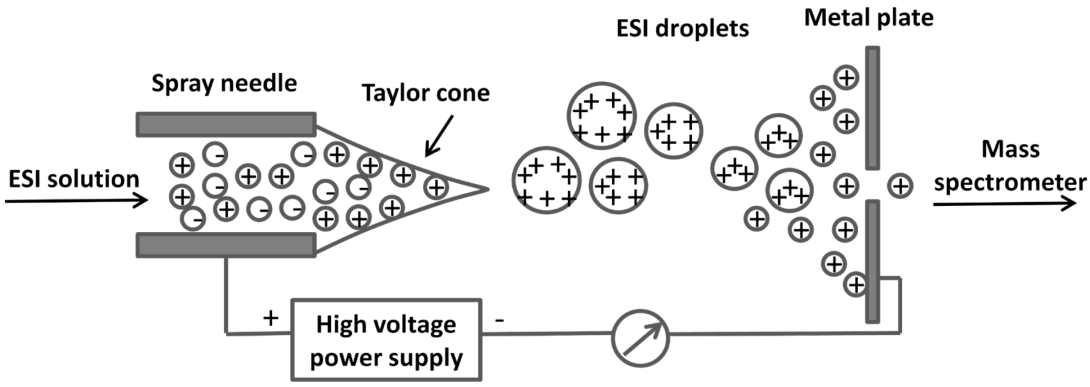


Figure 1.4 Schematic representation of the ESI process. Adapted from Reference 53.

1.3.4 Fourier Transform Ion Cyclotron Resonance Mass Spectrometer

Fourier Transform Ion Cyclotron Resonance Mass Spectrometer (FT-ICR-MS) has received considerable attention in metabolomics because it offers the highest possible resolution among all mass spectrometers.[58,59] High resolution allows distinction between metabolites with very small mass differences and leads to high mass accuracy for molecular formula determination. This is particularly important for comprehensive analysis of complex samples. FT-ICR-MS is fundamentally based on the theory of ion cyclotron resonance developed by Lawrence in the 1930s.[60] The first ICR-based mass spectrometers were developed in the 1950s.[61] However, the popularity of using ICR as the basis for mass detection is largely attributed to the incorporation of Fourier transform into ICR-MS by Comisarow and Marshall. [62,63]

The analyzer cell is the most important component of a FT-ICR-MS instrument, in which ions are stored, analyzed and detected.[64] It consists of three pairs of electric plates. The two electrodes at the ends of the cell are called trapping plates, which are

perpendicular to the magnetic field and create a potential well to confine ions inside the cell. The other two pairs of plates are for excitation and detection, respectively.

As ions enter a unidirectional magnetic field, they experience Lorentz force that causes ions to move circularly in the plane perpendicular to the magnetic field. This is known as the cyclotron motion and is characterized by its cyclotron frequency

$$f_c = \frac{\mathrm{qB}}{2\pi\mathrm{m}} \tag{1.3}$$

Where q and m are the charge and mass of the ion, respectively, and B is the strength of the magnetic field. Since the magnetic field is always held constant in a particular instrument, the *m/z* of an ion determines its cyclotron frequency, and ions with different *m/z* can therefore be resolved by measuring f_c. An important feature of the cyclotron frequency is that it is independent of the velocity (or kinetic energy) of the ion. This is one of the reasons why FT-ICR-MS can achieve such high resolutions.

The FT-ICR-MS operation involves four major steps. In the following discussion, we will assume operation in the positive mode. The first step is quenching, which is used to eliminate any ions remained from a previous experiment. This is achieved by creating a potential difference between the trapping plates, with the exit plate being more negative. After quenching, a new set of ions needs to be introduced into the cell. When coupled to ESI, ions are generated outside the high vacuum region and need to be transported into the cell through a series of ion transfer optics. The next step is ion excitation. The radius of the initial ion cyclotron orbital is usually very small (in the order of sub mm). Therefore, ions need to be excited into larger orbital in order to produce a measurable signal. This is accomplished by applying an RF voltage to the excitation plates with a frequency that is in resonance with the ion cyclotron frequency. The resonance energy

transfer activates ions and allows them to spiral outward and expand their orbital. Ions with the same *m/z* will be excited coherently and form an ion packet. When the ion packet passes the detection plates, an image current is produced as ions attract electrons to the plates.[65] As ions continue to move in circuit a sinusoidal image current will be produced with a frequency equal to the cyclotron frequency of this ion packet (or more precisely the difference between the cyclotron and magnetron frequency). In the ion detection step, ions with different mass-to-charge ratios are excited simultaneously by applying a rapid frequency sweep (RF chirp). This produces a composite of imaging signals with different frequencies and amplitudes. The frequency components is mathematically extracted by Fourier transform to covert the time domain transient to the frequency domain spectrum, which can be readily translated into a mass spectrum by applying a calibration formula derived from eqn (1.3). Since the performance of FT-ICR-MS depends heavily on its resolving power, it is important to understand the factors that affect the resolution. The following equation can be used to describe the resolving power of FT-ICR-MS instrument[66]

$$\frac{\mathrm{m}}{\Delta \mathrm{m}} = \frac{1.274 \times 10^{7}\ \mathrm{Bt}}{m/z} \tag{1.4}$$

Where B is the strength of the magnetic field and t is the data acquisition time. This equation reveals the dependence of mass resolution on the strength of the applied magnetic field and the length of the transient recorded. For this reason, today's FT-ICR-MS instruments usually use superconductive magnets that provide much stronger magnetic fields than permanent magnets and electromagnets. Nowadays, FT-ICR-MS with magnetic field of as high as 21 Tesla is made commercially available. The acquisition time is also an important parameter that affects resolution. In practice the

length of transient is limited by signal decays as a result of collision between ions and neutral molecules in the analyzer cell. Therefore, ultra-high vacuum condition (10^{-9} – 10^{-10} Torr) is required to minimize collisions and to improve the resolution. Other factors may also affect the mass resolution such as the presence of magnetron frequency,[64] and space charge effects.[67]

In my research work I have used the Bruker apex-Qe 9.4 Tesla FT-ICR-MS instrument, which consists of four main sections: the ion source (including both ESI and MALDI), the Qh-interface, ion transfer optics and the analyzer cell (Figure 1.5). Six differential pumping stages, including two roughing pumps and four turbo-molecular pumps were used to achieve ultra-high vacuum environment inside the cell. Under these conditions, the mass resolution is around 50,000 for small molecules with *m/z* of a few hundreds Da, and mass accuracy of 5 ppm can be easily achieved using external calibration.

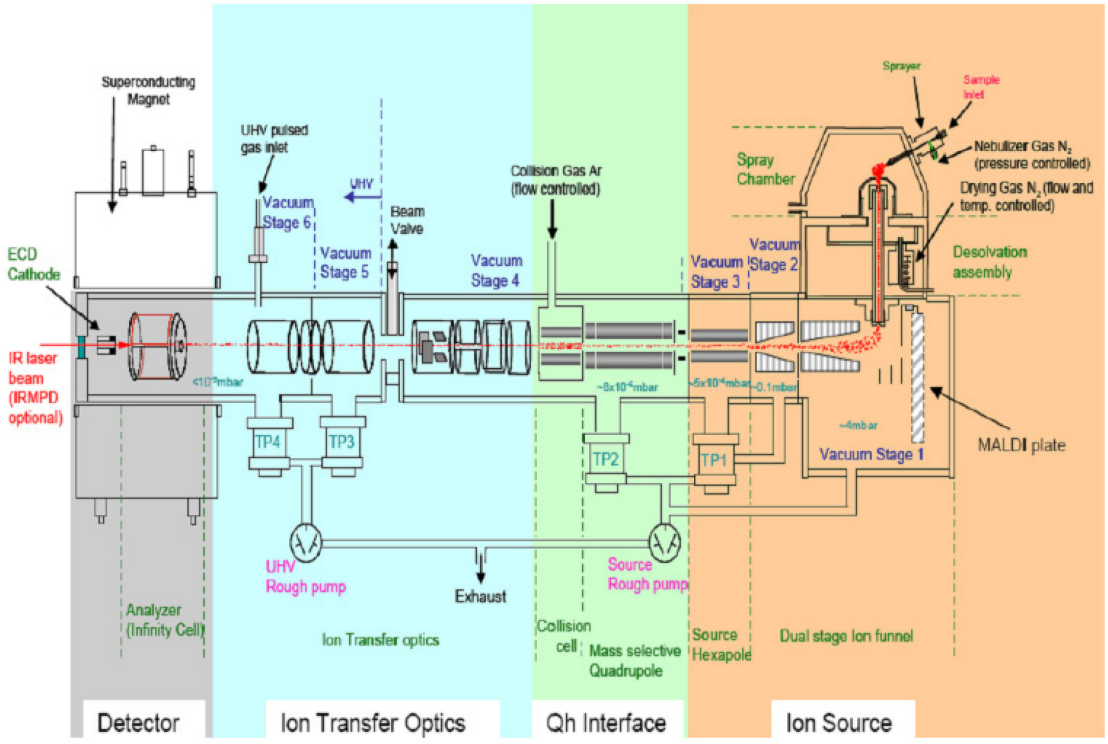


Figure 1.5 Schematic diagram of Bruker apex-Qe 9.4 Tesla FT-ICR-MS. By courtesy of Bruker Daltonics, Bremen.

1.3.5 Quadruple Time-of-flight Mass Spectrometer

Quadruple Time-of-flight Mass Spectrometer (Q-TOF-MS) is a tandem-in-space mass spectrometer that combines the high performance of time-of-flight analysis with MS/MS capability[68] The Q-TOF-MS configuration can be considered as addition of a linear quadrupole mass analyzer to an orthogonal acceleration time-of-flight instrument. Q-TOF-MS is a powerful instrument for metabolomics analysis, as it offers high sensitivity, resolution and mass accuracy in both MS and MS/MS modes, which greatly facilitates metabolic profiling as well as metabolite identification. The Q-TOF instrument is also commonly known as QqTOF, where the q refers to a quadruple (or a hexapole in some commercial instruments) located between the mass-resolving quadrupole and the time-of-flight analyzer. In the MS mode, both Q and q are operated in the RF-only mode to serve as ion guides for transmission of a broad mass range. The transmitted ions are then re-accelerated in orthogonal direction to the required energies and enter the TOF analyzer for spectra recording. In the MS/MS mode, Q acts as a mass filter that only allows selected precursor masses to be transmitted, while q serves as the collision cell where the energized precursor ions collide with neutral gas molecules and undergo collision-induced dissociation (CID). The resulting fragment ions, as well as any remaining precursor ions will then be analyzed in the second MS stage (i.e. the TOF analyzer). In my research I have used the Bruker maXis impact Q-TOF-MS, and a schematic diagram of this instrument is shown in Figure 1.6. In the following text the

three major components (Q, q and TOF) of this instrument will be discussed in more detail.

1.3.5.1 Quadrupole Mass Spectrometer

A quadrupole mass analyzer consists of four parallel rod electrodes. To each pair of opposite electrodes is applied a superimposed potential consists of a time-independent DC component (U) and a time-dependent AC component (Vcos(ωt)). The potentials applied on the two pair of rods are the same in magnitude but opposite in sign. Only ions within a certain mass region can have stable trajectory inside the quadrupole and get detected. The ion trajectory can be described quantitatively using the Mathieu equation,[69] and the solution to the Mathieu equation depends on the reduced Mathieu parameters a and q:

$$a = \frac{4eU}{\omega^2 r_0^2 (m/z)}, \quad q = \frac{2eV}{\omega^2 r_0^2 (m/z)} \tag{1.5}$$

where e is the charge of an electron, U is the value of the DC voltage, V is the amplitude and ω is the angular frequency of the AC component, and r_0 is the radius of the quadrupole. The solution can be represented graphically by the a-q stability diagram (Figure 1.7). Ion with a particular *m/z* corresponds to a specific point in the a-q diagram, and only ions that fall into the stable region will be detected. When the quadrupole is operated as a narrow bandpass mass filter, the ratio of U/V is held constant so that all ions will lie on one straight line (the mass scan line) with slope equals to a/q or 2U/V. The narrow bandpass mass filter can be created by adjusting the slope so that the mass scan line intersects at the sharp tip of the a-q diagram (usually unit resolution). The bandpass region is scanned by simultaneously increasing U and V to allow ions with increasing *m/z* move to the tip of the a-q diagram and become detected. When the

quadrupole is operated in the RF-only mode, the DC potential is removed so that the mass scan line has a slope of zero. This means that all ions with *m/z* higher than a given limit fall within the stability region of the a-q diagram, and the quadrupole serves as a broad bandpass filter for efficient ion transfer. In general, quadrupole mass analyzers are characterized by its mechanical simplicity, low cost, reasonably high scan rates, and ease to couple with chromatography.

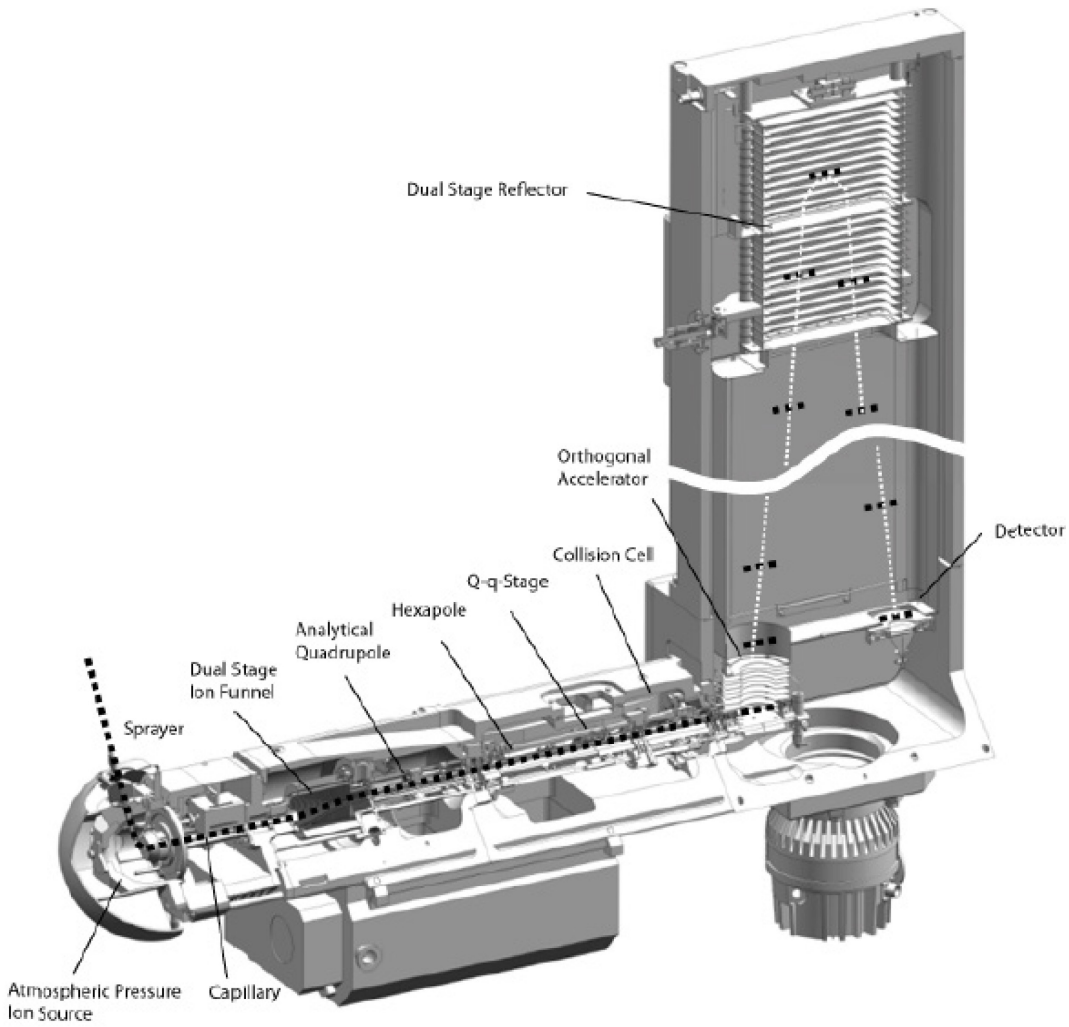


Figure 1.6 Schematic diagram of Bruker maXis impact Q-TOF-MS. By courtesy of Bruker Daltonics, Bremen.

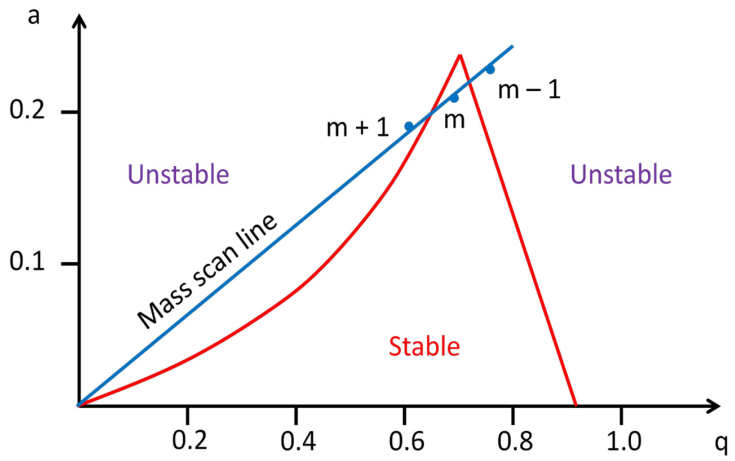


Figure 1.7 The a-q stability diagram.

1.3.5.2 Collision Cell

The collision cell contains a quadrupole enclosed in a chamber with small entrance and exit apertures. A neutral gas molecule, usually nitrogen or argon, is introduced into the chamber at about 10^{-2} mbar. In MS/MS mode, ions are accelerated to an energy of 20-200 eV by increasing the DC voltages in front of the collision cell. During fragmentation the lens voltages are set to block ion transmission. Inside the reaction chamber, the energized parent ion undergoes collision with gas molecules to convert translational energy of ions into internal excitation. When the acquired internal energy exceeds the dissociation energy of chemical bonds, ions may dissociate into fragments. The efficiency of conversion from translational energy into internal energy is a function of several factors, including the magnitude of injection voltage (collision energy), molecular structure and mass of the parent ion, as well as the molecular mass of the collision gas. In particular, the collision energy is an important user-specified parameter which can be optimized to obtain the most abundant fragment information. If

the collision energy is too low, only a few product ions can be observed. On the other hand, too much collision energy may result in loss of some useful fragment peaks. It is also possible to use more than one collision energy in one scan and combine the results to provide maximum information in the fragment spectra. Finally, the quadrupole in the collision cell serves as an ion guide to focus both parent and product ions onto the axis of the collision cell and allow efficient ion transfer into the TOF analyzer.

1.3.5.3 Time-of-flight Mass Spectrometer

In contrast to a quadrupole mass spectrometer where ions are continuously introduced into the analyzer, the TOF instrument works in pulsed mode, which makes it challenge to couple with a continuous ion source such as ESI. This problem is solved using orthogonal acceleration. After the ions leave the collision cell, there are two steps inside the orthogonal accelerator. The first step is the fill phase, in which the collision cell exit lens is open to allow ions enter the extraction volume in their original direction as a parallel beam. At this stage the accelerator voltages are switched off. The time slot for the transfer process is an adjustable parameter that limits the transferred mass range. The next step is the extraction phase, in which the accelerator voltages are switched on to push ions in the orthogonal direction into the flight tube. During this process the accelerated ions gain electric potential energy $E_p = zeU_{acc}$, where z is the number of charges of the ion, e is the charge of an electron and U_{acc} is the accelerator voltage. This potential energy is converted to kinetic energy $E_k = ½ mv^2$ before ions enter flight tube.

The flight tube is a vacuum enclosure free of electrical fields. Inside the flight tube, ions fly uniformly with the velocity v until they hit the detector. If the length of the flight tube is L, then the flight time t can be expressed as

$$t = \frac{L}{v} = L\sqrt{\frac{m/z}{2eU_{acc}}} \quad (1.6)$$

The flight time is thus a function of *m/z* of the ion and is used to resolve ions with different masses.

The ion beam introduced into the accelerator is not a single line but has a radial dimension. During extraction, ions located closer to the repeller plate gain more energy than those further away. This situation creates initial kinetic energy and spatial distributions in the direction of flight tube, which limits the mass resolution. For this reason, a reflector is often used in the TOF instrument to correct for this difference. A reflector consists of a series of electric plates, each with a slightly more positive voltage (for positive ions) than the former. Ions entering the reflector experience a decelerating field, slow down until the velocity reduces to zero and finally get reaccelerated towards the detector. More energetic ions enter the reflector earlier, but penetrate deeper and spend more time inside the reflector. The longer path length of faster ions compensates for their shorter flight time in the flight tube so that they arrive at the detector simultaneously with slower ions. The use of a reflector effectively corrects for the spatial and energy spread and leads to substantially improved mass resolution. Nowadays the mass resolution of TOF instruments can easily achieve 50,000. Other advantages of TOF-MS include its high mass accuracy (with calibration), high mass range, excellent sensitivity, and fast spectra acquisition rate.

1.4 Sample Preparation in Metabolomics

The selection of a proper sample handling procedure for metabolomics studies is of critical influence on the detected metabolite composition.[70,71] It involves all upstream steps prior to LC-MS analysis, including biological medium selection, sample collection,

transport and storage, metabolite quenching and extraction, as well as any sample cleanup, concentration and derivatization procedures. Although there are some discussions on standardization of metabolomics workflows,[72,73] to date there's still lack of standardized sample preparation protocol for biofluids or cells, and the choice of sample preparation methods is often made depending on the particular biological question to be investigated and the analytical platform to be used. Regardless of the protocol selected, it is important to keep consistent for handling the samples within the same study in order to minimize sample variability.

The first step in designing a metabolomics experiment is to select the appropriate biological medium. Several biofluids can be used, depending on the analytical problem under study. Urine and blood are the two most widely used biofluids because of their ease of collection and the fact that they usually reflect the state of the whole organism. Urine contains relatively polar metabolites that are cleared from the body and represents the result of the catabolic process.[74] Blood samples are usually centrifuged first to remove blood cells.[75] The resulting fluid is either plasma (without clotting) or serum (with clotting), which may be used interchangeably as they give very similar metabolic profiles in some studies[76,77], although in most studies it is preferred to use the same type. Plasma or serum reflects the instantaneous metabolic status of the whole organism at the time of sampling, which includes both anabolic and catabolic processes. Other biofluids are more specific for studies of certain diseases. For example, cerebrospinal fluid is the preferred medium when neurological disorders are investigated,[78,79] and saliva is of interest when studying oral diseases.[80] In addition to biofluids, metabolome analysis of tissue or cell culture is also common, which provide more localized description of the metabolite

distribution. They are also less affected by biological variations, such as differences in age, gender and food intake. However, the sample preparation process is often more challenging and is therefore more prone to cause experimental variations.

The next step is sampling and storage of collected samples. Rapid sample collection and handling is desirable, especially when metabolites with fast turnover rates are of interest. To this end, sampling is often performed at reduced temperatures. Occasionally, rapid inactivation of metabolism at the time of collection (metabolic quenching) is also required. Samples are usually stored at -80°C so as to minimize metabolite degradation. However, it should be noted that frequent freeze/thaw cycles may have an influence on the metabolic profile. It is therefore recommended to aliquot the samples at the time of collection in order to minimize the number of freeze/thaw cycles.

The metabolite extraction step is perhaps the most critical part in the entire sample preparation process. For metabolomics studies, it is desirable to detect as many metabolites as possible; however, this is often achieved with an increased risk of lower extract purity. Therefore a good balance should be made depending on the application. The metabolite extraction protocol varies with the biological medium, and herein the most common ones will be discussed. Urine samples are mostly filtered to remove bacterial contamination, followed by the dilute-and-shoot strategy, in which the sample is simply diluted prior to LC-MS analysis. Plasma and serum samples contain a large amount of proteins, and organic solvent-based protein precipitation procedures are often required. Methanol is one of the most commonly organic solvents that provides a good metabolite coverage.[81,82] For liquid samples classical extraction techniques such as liquid-liquid extraction (LLE) and solid phase extraction (SPE) are also often employed

to improve sample quality. For example, LLE separates the liquid sample into polar and non-polar portions, which can be analyzed independently to expand the metabolite coverage. LLE is used more often in lipidomic studies, using chloroform and methanol as the extraction solvents.[83] SPE is an effective method for selective extraction and enrichment of target compounds or for sample cleanup (e.g. desalting), thus enabling higher detection sensitivity for targeted analysis.[84] Because of its selectivity SPE may seem less appropriate for untargeted profiling. Nevertheless, it is possible to use different SPE cartridges for the same sample (e.g., cation exchange and anion exchange), analyze each eluent separately and then combine the results to obtain the comprehensive metabolic profile. For cell cultures, disruption of cell membranes is needed to release intracellular metabolites. Several methods can be used for this purpose, such as freeze-thaw cycle, ultrasonication, or direct addition of a cold or hot extractant. Commonly used extractants include organic solvents (e.g., methanol, ethanol or acetonitrile), acidic or basic solutions (e.g., perchloric acid or KOH solutions).[85] For untargeted profiling, organic solvents are preferred because they are less specific, easy to evaporate, and able to precipitate proteins. The exact extraction protocol varies for different studies and there's no consensus on which method works the best. Therefore, it is recommended that for a specific study, several extraction methods should be compared to determine the most appropriate method.[21] Finally, after extraction, metabolites may be directly analyzed or undergo treatments such as dilution, concentration or derivatization, with the aim of obtaining optimal detectability in LC-MS analysis.

1.5 Biological Sample Normalization

Metabolomics studies usually involve parallel assessment of the metabolite levels of individual samples from two or more biological groups to reveal the metabolic differences. In order to extract meaningful biological information from the quantitative comparisons, it is important to ensure that unwanted concentration variations in biological samples are minimized. In statistical analysis, samples within the same demographic groups or the same phenotypes may be artificially separated according to their concentrations. These intra-group variations often obscure the true inter-group metabolic changes that are of interest and complicate identification of differentially expressed metabolites.[86,87] Therefore, an effective sample normalization strategy is required before data analysis in order to compensate for the total concentration variations. This step is particularly critical for urinary metabolome profiling, because urine volumes can vary by up to 15-fold upon water consumption or other physiological factors.[88] The need for normalization has also been widely recognized in metabolic profiling of various types of cell extracts with unknown cell numbers. Although less investigated, there are many other biological media such as sweat, bronchial lavage fluid and fecal samples, in which the metabolite levels are variable and therefore normalization of the sample concentration is highly desirable for improved quantitative analysis.

Compared to genomics and proteomics, normalization in metabolomics is more challenging because of the greater diversity of metabolite structures. To date, there's no standard method for measuring the total amount of metabolites directly, and many approaches have been proposed to determine the total sample amount as the normalization factor. The performance of a normalization strategy in metabolomics is often evaluated by its ability to correct for unwanted sample amount variations and

subsequent improvements in statistical results. In addition, there are a few other factors to be considered before selecting an appropriate normalization method for a particular application. Firstly, it is important to consider whether the normalization step is performed pre-acquisition or post-acquisition.[89] In pre-acquisition methods, the volumes of biological samples are adjusted so that the same amounts (or concentrations) of sample are taken for LC-MS analysis. One advantage of pre-acquisition normalization is that similar instrumental responses can be obtained for all samples. In electrospray ionization, responses of individual metabolites are often nonuniform at different concentrations because of charge competition and ion suppression. As a result, the analyte signals may not necessarily be linearly scaled with the metabolite concentration in a complex biological matrix. Thus, by adjusting all comparative samples to the same concentration, more accurate quantitative results will be produced. In addition, pre-acquisition methods also allow control of the sample injection amount. It is important to keep the injection amount optimal for detecting low concentration metabolites in order to reduce missing values,[90,91] while avoiding problems brought by overinjection such as signal saturation and sample carryover. Comparisons between pre-acquisition and post-acquisition methods have been investigated. For example, Chen et al. showed that for five serially diluted urine samples, post-acquisition normalization methods failed to overcome urine variability because of the nonlinear response to sample dilution caused by ion suppression or saturated metabolites. On the other hand, pre-acquisition correction was effective for reducing variations introduced by different urine concentrations. They also demonstrated that pre-acquisition injection volume calibration is superior in reducing intra-group bias in the presence of biological variations.[92] Similarly, the study performed

by Edmands et al. suggested that pre-acquisition normalization is a better choice for biological information recovery as it identified the largest number of discriminant MS features when compared to three post-acquisition normalization methods.[89] However, the downside of pre-acquisition normalization is that additional experimental procedures are required. For this reason, post-acquisition approach has been used as an alternative because of its convenience, and it is preferred when high-throughput is of primary concern. Nevertheless, it should be kept in mind that the performance of post-acquisition may be compromised by variations in detection responses, as described above.

The second factor to be considered is the ease of operation, particularly for pre-acquisition normalization, since an extra step is required. In metabolomics larger sample sizes are preferred in order to create valid statistic models and to obtain accurate biological information.[93] However, the sample size is often limited by availability of biological samples and analytical instruments, complexity of sample preparation protocols, as well as length of analysis time per sample. In this regard, it is desirable to keep the normalization method simple, quick and convenient to perform, so that it would not restrain the use of larger sample sets.

Moreover, the selection of normalization methods also depends on the type of biological sample to be analyzed. Some approaches are only applicable to specific biological media while others are more generic. For example, normalization to creatinine is based on the relatively constant excretion rate of creatinine through glomerular filtration, and is therefore only applicable to urine specimens. On the other hand, normalization to MS "total useful signal" (MSTUS) uses the total intensity of peaks that are present in all samples under study as the normalization factor, and it is thus more

universally applicable. While specific approach may be more efficient on occasions for its targeted biological medium, a generic normalization method is usually more desirable because of its adaptability to sample types in which a specific normalization method is not available, and the ease of method transfer from one type of biological matrix to another. In the following sections, we will review some recently reported normalization strategies for MS-based metabolomics studies (Table 1.1), and discuss their performance in terms of concentration variation reduction and biological information recovery.

Table 1.1 Overview of normalization strategies in MS-based metabolomics

Sample Type	Normalization Methods	Pre- or Post-acquisition	Recommended method	Reference No.
Cattle urine	Specific gravity Freeze-drying MSTUS[1]	Pre Pre Post	Similar performance for all three methods	116
Human urine	Specific gravity Specific gravity Median fold change Urine volume	Pre Post Post Post	Pre-acquisition normalization to specific gravity	89
Rat urine	Urine volume Osmolality Creatinine MSTUS	Post Post Post Post	Osmolality and MSTUS	88
Rat urine	Total intensity Median fold change Quantile LOESS[2]	Post Post Post Post	Median fold change	86
Rat urine	All MS signals MSTUS Creatinine value Creatinine peak area Creatinine value Creatinine value + all MS signals Creatinine value + MSTUS	Post Post Post Post Pre Pre + Post Pre + Post	Creatinine value + all MS signals and Creatinine value + MSTUS	92
Human urine	MSTUS	Pre	-	90
Human urine	Conductivity LOESS	Post Post	LOESS	120
Human urine	UV absorbance	Pre	-	125

Adherent cell line (OVCAR-8)	Total protein Cell count Total DNA	Post Post Post	Total DNA	136
Adherent cell line (MDCK)	Protein content Metabolic markers	Post Post	Similar performance for both methods	128
Adherent cell line (MCF-7)	Cell count Sum of peak areas	Post Post	Similar performance for both methods	130
Cyanobacterial strains	Chlorophyll *a* Total protein Glycogen sIC[5] sIC.90[6] sIC.AA[7]	Post Post Post Post Post Post	Similar performance for all six methods	133
E. coli	UV absorbance	Pre	-	137
Human fecal samples	UV absorbance	Pre	-	143

1. MSTUS: MS total useful signal; 2. LOESS: Locally weighted scatter plot smoothing; 3. MSTS: MS total signal; 4. MSGUS: MS group useful signal; 5. sIC: selected ion count for all metabolites; 6. sIC.90: selected ion count for 90% of the non-extreme metabolite pools; 7. sIC.AA: selected ion count for the total amino acid content.

1.5.1 Normalization of urine samples

Urine is one of the most commonly investigated biofluids in metabolomics because it can be easily and non-invasively collected in large quantities, and the sample is relatively clean which requires simple pre-treatment procedures.[94] Unfortunately, urinary solute concentrations often vary widely depending on hydration status, time since previous urination, dietary intake or other physiological factors.[95,96] For example, a recent study carried out in our group (Canadian and Chinese Metabolome Database) has revealed up to 18-fold difference in urinary concentration collected from 100 healthy people. In fact, the majority of research efforts on normalization method development target specifically at urine samples. The most widely accepted approach for urine volume

correction is to express metabolite levels relative to creatinine concentration,[97-99] because the rate of creatinine excretion through glomerular filtration is relatively constant under normal conditions within or across individuals.[100] However, the assumption of constant creatinine excretion is often invalid, as creatinine excretion was found to vary across individuals due to age, gender, race and muscle mass differences or disease states.[101-104] Even within the same individual, the urine creatinine level may change depending on diet, time of day, level of exercise, and physiological conditions.[105-107] As a result, the validity of using creatinine concentration as the normalization factor is often challenged.[108] As an example, Burton et al. showed that normalization of urinary pteridines to creatinine did not improve differentiation between benign and malignant breast cancer samples, and the authors suggested that alternative renal dilution factors are needed.[109] In addition, it is also questionable whether it is sufficient to normalize the wide range of metabolites based on a single compound.[110] Nevertheless, in a recent study it was demonstrated that injection volume calibration based on creatinine value prior to LC-MS analysis is effective in adjusting urinary solute concentrations to similar levels. This pre-acquisition normalization method significantly reduces intra-group variations as indicated by better clustering in PCA score plots and reduced peak area RSDs in intra-group comparisons.[92]

Urine osmolality is a direct measure of the total urinary solute concentration that is only affected by the number of dissociated particles in urine. Therefore, it is often considered as a gold standard for estimating urinary concentration,[111] and has been used as a valid scaling factor for urinary solutes.[112,113] Application of osmolality normalization in urinary metabolomics analysis has been reported, which showed better separation

between different biological groups as well as reduced variations within biological replicates when compared to no normalization or normalization to urine volume and creatinine.[88] However, the procedure of measuring osmolality is often not practically available, and specific gravity is used usually measured instead as a fair estimation of osmolality.[111] Urine specific gravity is the ratio between the density of urine and that of pure water at constant a constant temperature, which can be measured either directly by gravimetry or indirectly by refractometry. Specific gravity has been used as a normalization method for urinary metabolites in many applications.[95,114,115] Recently, the feasibility of using specific gravity as a normalization strategy in urine metabolomics has been assessed by Jacob et al.,[116] in comparison with freeze-drying, a valid normalization strategy for anabolic practices in cattle.[117,118] The authors have showed that normalization by specific gravity improved separation between two study groups and revealed the same differentiating ions as the freeze-drying method, and they thus proposed that specific gravity can be used as an alternative to the time-consuming freeze-drying for urine metabolome normalization. Besides creatinine, osmolality and specific gravity, other conventional urine normalization methods include normalization to 24-hour urine volume, [88,119] conductivity[120] and flow rate correction,[121] but these methods are used less often and will not be further elaborated here.

Data-driven normalization approaches have been used for some metabolomics studies in recent years. Warrack et al. have proposed the idea of using the total intensity of peaks that are common to all samples, known as MS "total useful signal" (MSTUS), as the scaling factor.[88] This concept is similar to the use of total integrated proton signal for normalization in proton NMR-based metabolomics analysis.[122] Incorporation of only

"useful signals" ensures that contributions from xenobiotics and artefacts are minimized. The authors have compared MSTUS normalization to other common normalization approaches in the non-targeted metabolomic profiling of rat urine from different dose groups. They recommended both MSTUS and osmolality for detection of significant metabolic changes as these methods are most efficient in differentiation between high and low dose groups. Since reported in 2009, MSTUS normalization has been employed in several applications.[90,123] In addition to MSTUS, more sophisticated statistical strategies have been introduced. For example, Veselkov et al. have compared four normalization techniques.[86] The first two methods, total intensity and median fold change, assume that metabolite peak intensities vary linearly with concentration, while the other two approaches, quantile and locally weighted scatter plot smoothing (LOESS), consider peak-intensity-dependent scaling factors (i.e. in the presence of ion suppression or saturation). They found that for the majority of urinary metabolites, the peak intensities did not respond differentially to dilution, which supports validity of the first two methods. In terms of normalization performance, all four methods are equally well in the absence of biological variation. However, when biological variation is considered, the performance of total intensity normalization is slightly inferior due to variations in total metabolite output between samples. Despite the comparative effectiveness of the other three methods, the authors recommended the median fold change approach because of its relaxed assumption with regard to the proportion of asymmetrical metabolite changes. More comprehensive evaluations of statistical treatments have been discussed by Ejigu et al..[124] In general, the advantage of data-driven approaches over conventional urine normalization methods is that the methodology is not restricted to urine samples and can

be readily adapted to other types of biological matrices. However, since the scaling factor is based on MS data, the normalization processed has to be performed after LC-MS analysis. As discussed before, post-acquisition normalization cannot control the amount of sample injected into the mass spectrometer, and therefore cannot be used to alleviate problems associated with varying injection amount such as signal saturation and missing values. Although pre-acquisition MSTUS has been proposed by Mattarucchi et al. to improve the quality of extracted LC-MS data,[90] it requires LC-MS acquisitions to be performed for all samples prior to metabolomics analysis in order to obtain MSTUS values. This would increase the workload considerably and is not practical for large sample sizes.

Another concept for sample normalization is to determine the UV absorbance of the sample solution as a measure of the total concentration of solutes that absorb at the specific wavelength. This type of normalization approach is more representative of the overall sample composition compared to the use of a single compound such as creatinine, and is independent of the biological medium. In addition, UV measurement can be performed prior to LC-MS analysis to allow injection amount adjustment. Kemperman et al. has reported normalization of urine samples to the area under the curve at 214 nm (AUC_{214}) and showed that this method is preferred to creatinine normalization for minimizing peak area and intensity variations of peptides.[110] More recently, we have developed a general approach of determining the total metabolite concentration based on the use of chemical labeling to attach a UV absorbing dansyl moiety to amines and phenols, followed by a rapid step-gradient LC-UV detection of the labeled metabolites at 338 nm (Chapter 2).[125] This method was incorporated into the differential isotope

labeling LC-MS metabolome profiling workflow to ensure similar and optimal amounts of sample are injected. We showed that this normalization strategy effectively corrects for the dilution effect in intra-day urine samples and minimizes artificial separation caused by the variations in the original urine concentrations. Moreover, although parallel comparison with other normalization methods was not performed, we did observe a good correlation between the UV absorbance and creatinine or osmolality values. However, as pointed out by Kemperman et al., one potential drawback of UV-based normalization is that the peak area may be dominated by a single compound.[110] Nevertheless, the main purpose of pre-acquisition normalization is to ensure similar amounts of sample are injected into LC-MS, while the data quality can always be further improved by post-acquisition curative. For example, Chen et al. have illustrated that the combination of pre-acquisition injection amount calibration based on creatinine values and post-acquisition MSTUS normalization provides the best results in overcoming urine sample variability.[92]

1.5.2 Normalization of cell culture

Cell metabolomics plays an important role in systems biology and has found applications in many areas such as toxicology and preclinical drug testing, in which *ex vivo* models are required.[126,127] In cell metabolomics, it is often of interest to investigate quantitative metabolic changes in response to different environmental stimuli. Unfortunately, it is difficult to control the amount of cells harvested from different culture plates due to variations in seeding density and/or treatment conditions.[128] For example, exposure to toxic compounds such as 2,3,7,8-tetrabenzodi-p-dioxin (TCDD) may lead to decreased cell proliferation rate and subsequently lowered overall metabolite level.[129] The same problem is also encountered in microbial metabolomics, for

comparative analysis of bacterial species under different environmental conditions or for microorganism differentiation and identification. Therefore, normalization of mammalian and microbial cells will be discussed together in this section. For mammalian cells, cell counting using a hemocytometer is commonly used for cell amount normalization. However, the normalization performance of this method is often impaired by direct scraping of adherently growing cells, as high variability can be introduced during cell aliquot and transfer.[127,130] This procedure is also difficult to implement for bacterial cells because of their small sizes. Alternatively, optical density at 600 nm (OD_{600}) is often determined as a measure of the light scattered by a culture, and the OD_{600} value is then correlated to a known colony number to estimate the amount of cells. As an example, Marcinowska et al. have showed that normalization to OD_{600} values provides a robust basis for quantitative analysis and differentiation of clinically relevant bacterial cells.[131] While this method is applicable to homogeneous cell suspensions, it may not be as convenient for adherent cell cultures. Also, since this method is usually performed at the time of harvest, it would delay subsequent quenching procedure and may result in alterations in the metabolic profile. Other conventional approaches for cell amount determination include measurement of the dry weight of cell debris and quantification of the total protein content.[132] The dry weight method is time-consuming and is not preferred for metabolomics studies in which a large number of samples need to be processed. Also, relatively large errors may be introduced in dry weight measurement when the sample amount is small.[133] The protein amount can be readily determined using well-established colorimetric methods such as Bicinchoninic acid (BCA) assay and Bradford assay, and has been widely used in several applications for study of metabolic

changes.[134,135] However, Silva et al. have showed that the protein concentrations of both the metabolite extraction solution and the remaining cell pellet failed to produce expected correlations with seeded cell number, due to poor protein recovery in metabolomics-compatible solvents and incomplete protein resolubilization from the pellet.[136] Therefore, assaying the protein amount would require a separate experiment to be performed in order to obtain accurate quantitation results, which is not desirable as it consumes part of the samples and prolongs the sample preparation process. Instead, the authors have proposed the use of DNA concentration as the normalization factor for metabolomic data, as they have demonstrated good correlation between DNA concentration of the cell pellet and seeded cell number for four adherent cell lines. Despite the reported efficiency and robustness of this method, examples of how normalization to DNA concentration could improve statistical analysis in metabolomics were not given in this work.

In recent years, several data-based normalization strategies have also been introduced for cellular metabolomics. One approach involves the use of specific metabolic markers as the cell amount indicator,[128,129,133] with the assumption that the concentrations of these selected metabolites are directly proportional to the cell number and are independent of the treatment conditions under study. For example, Cao et al. has identified pantothenate and inositol as the best candidate markers for normalization of Madin–Darby canine kidney (MDCK) cells based on three criteria: good linearity between the metabolite signal intensity and the cell amount for serially diluted cell suspensions; high linear correlation between the metabolite abundance and the protein content for cells seeded at different concentrations; improved separation between two cell lines and closer clustering within each cell line. In general, the use of single or a few

metabolic markers is simple, fast and convenient. However, it should be noted that the same metabolite markers are not necessarily applicable in other cellular systems, as their cell line and treatment independence may not always hold true. Therefore, the validity of these metabolic markers for other studies requires further assessment, which poses a limitation to the utility of this normalization method. Alternatively, the use of total metabolite intensity is often considered as a more robust way of estimating the cell number. Huege et al. have evaluated the normalization effects of three data-based parameters, namely the intensity sum of all metabolites, the intensity sum of 90% non-extreme metabolites and the intensity sum of amino acids and their conjugates, in cyanobacterial metabolomics.[133] They compared these intrinsic parameters with three cellular constituents (chlorophyll *a*, total protein and glycogen), and it was concluded that these six normalization factors are essentially equivalent in terms of their influence on sample cohesion within strain groups and separation among different strains. Hutschenreuther et al. have also evaluated the performance of peak area sum normalization in comparison to cell count.[130] They observed a good linear correlation between sum of peak areas and cell count within a specified linear range. However, they noted that this normalization method should only be applied when the cell extract concentration in two comparative sample sets differ by less than two-fold, as otherwise the number of "false significants" would increase to over 10%. This is likely attributed to the fact that not all metabolites exhibit linear response with concentration (e.g. presence of borderline metabolites and ion suppression effect). Therefore, the authors suggested that similar extract concentrations should be used for comparison in cellular

metabolomics. This conclusion is in accordance with our earlier discussion on the importance of pre-acquisition normalization.

As we discussed in Section 1.5.1, normalization to UV absorbance of the sample solution is independent of the biological matrix and is performed prior to MS acquisition, which makes it a promising approach for cell amount adjustment. In addition, when incorporated into the LC-MS metabolome profiling workflow, this normalization step is usually carried out immediately before LC-MS (i.e., after all initial sample preparation steps). This is particularly advantageous to cell metabolomics for two reasons. Firstly, there's no extra procedure during the cell harvest stage, so that metabolite quenching can be performed without any delay. Secondly, metabolite extraction from cells often involves more steps compared to biofluids, which makes it more prone to experimental errors. The UV normalization can also be used to correct for concentration variations introduced during the sample workup process. Application of the chemical labeling UV normalization method on microbial metabolomics has been demonstrated on bacterial differentiation (Chapter 4) and the study of butanol tolerance in *Staphylococcus warneri* SG1.[137,138] More recently, we have adapted this UV absorption method into a dansylation metabolite assay (DMA), which measures absorbance of labeled metabolites using a microwell plate reader instead of the more expensive LC-UV system (Chapter 3).[139] The DMA improves throughput by allowing simultaneous measurement of multiple samples within a short period, and can be readily implemented because of the low cost and simple procedures. We showed good linear relationships between the UV absorbance values and the cell suspension volume or the protein amount. The validity of this normalization strategy for metabolomics has been demonstrated by the improved separation between

two *E. coli* strains as well as decreased %RSD values within each strain. Although this UV absorbance normalization approach has so far only been applied to microbial metabolomics, we envisage that it should in principle be equally applicable to mammalian cells.

1.5.3 Normalization of other biological samples

Compared to urine and cellular samples, the need for sample amount normalization in other biological media is far less explored. While most biofluids, such as blood and cerebrospinal fluid, are considered homeostatically regulated, there are still a number of sample types in which proper control of the solute concentration is absent. For example, the sample volume of sweat can often vary depending on the water content, which can be affected by several factors such as ambient temperature and relative humidity. There's also large variation in sweat production for different individuals.[140] As commented in a recent review on sweat metabolomics, the absence of proper normalization methods to account for the sample volume variations presents a major drawback in the quantitative analysis of sweat.[141] To address this problem, Appenzeller et al. has proposed the use of sodium and potassium concentrations as normalization factors for sweat, which is similar to normalization of urine samples to their creatinine content.[142] They found that the potassium concentration was highly variable within both females and males, making this species unsuitable for determination of the sweat volume. In contrast, the authors recommended the use of sodium as an internal standard for sweat volume correction because of its low inter-individual variability. However, similar to the limitations of creatinine normalization, this method is also subject to disease states, and the use of a single species for normalization may not be appropriate in metabolomics.

Fecal samples represent another biological medium in which proper normalization is required, because they contain both solid and liquid materials with varying proportions.[143] Even with the same sample weight or volume, the metabolite concentration can be different depending on the liquid content and solid density. Unfortunately, there's few discussion on normalization of fecal samples. In this case, normalization in fecal metabolomics can only be performed with universal approaches such as post-acquisition statistical methods or pre-acquisition UV absorbance measurement. As an example, Xu et al. has applied the LC-UV-based normalization method described in Chapter 2 for the profiling of human fecal metabolome.[143] They observed a wide total concentration distribution from 0.60 to 6.37 mM. Even within the same individual at three different days, the total metabolite concentration can vary by more than 3-fold. This result highlights the importance of sample amount normalization in quantitative fecal metabolomics.

1.6 Differential Isotope Labeling in LC-MS-based Metabolomics

Although LC-MS analysis of metabolites can be performed conveniently without chemical derivatization, the addition of a derivatization step is still widely applied for improved metabolite detection and quantification. For MS analysis, the derivatization step typically involves reaction of a labeling reagent and its isotopic counterpart to two comparative samples respectively, which is known as "differential isotope labeling" (DIL). A judiciously chosen differential isotope labeling reagent can benefit LC-MS-based metabolomics analysis in several ways. Firstly, one form of the isotope reagent (usually the heavier isotope) can be used to label a reference sample that has the same composition as the samples to be analyzed (e.g. a pooled sample) to serve as the internal

standard.[144] The labeling procedure thus creates an internal standard for every labeled metabolite, which effectively accounts for instrumental variations such as ion suppression in ESI and improves quantification accuracy. Secondly, by incorporating a non-polar moiety into the labeling reagent, the polar or ionic compounds in biological samples can be made more hydrophobic, resulting in better separation by RPLC columns and reduced ion suppression caused by co-elution. Thirdly, derivatization can provide enhanced ionization efficiency if the labeling reagent has a chargeable functional group (e.g. amines in positive mode) and/or good surface activity. In general, compounds with good hydrophobicity are associated with higher ionization efficiency because non-polar ions prefer the droplet air interface (better surface activity).[145] In addition, hydrophobic compounds are usually eluted out at a relatively high percentage organic mobile phase, which promotes the ESI desolvation process and thereby improves ionization efficiency.[146] Finally, the addition of a labeling reagent shifts low-mass metabolites to higher mass region. Low-mass region is usually associated with high background noises caused by solvent clusters or other contaminants.[147] By shifting molecules to the high-mass region, the background spectrum is cleaner and as a result signal-to-noise ratio is improved. Despite the many great features brought by chemical derivatization, it inevitably requires an extra step in sample preparation in which analytical errors may be introduced. Therefore, it is important to evaluate the performance of the selected labeling reagent and the chemical reaction beforehand, including the reagent purity, reaction conditions, labeling efficiency and repeatability, in order to obtain high quality DIL results.

There are many isotope labeling reagents developed for metabolomics research, targeting at different metabolite groups. The most frequently labeled functional groups are amines,[146,148,149] carboxylic acids,[150,151] fatty acids[152] and carbonyls,[153] which cover the majority of metabolites. In my research I mainly work with dansylation reaction, which is a classical derivatization technique for primary amines, secondary amines and phenols (Figure 1.8).[146] In this derivatization method, two ^{13}C isotopes are introduced into the dimethylamino group of dansyl chloride, which results in mass shift of 2 Da from the natural ^{12}C isotope. A peak pair picking algorithm has been developed to selectively pick the differential-isotope-labeled peak pairs with high accuracy (low false positive rate).[154] This derivatization strategy provides 10-1000 fold increase in sensitivity as well as good quantification precision. Application of this labeling approach for metabolomics analysis has been successfully demonstrated in a variety of biological samples, such as urine, cerebrospinal fluid and saliva.[49,155,156]

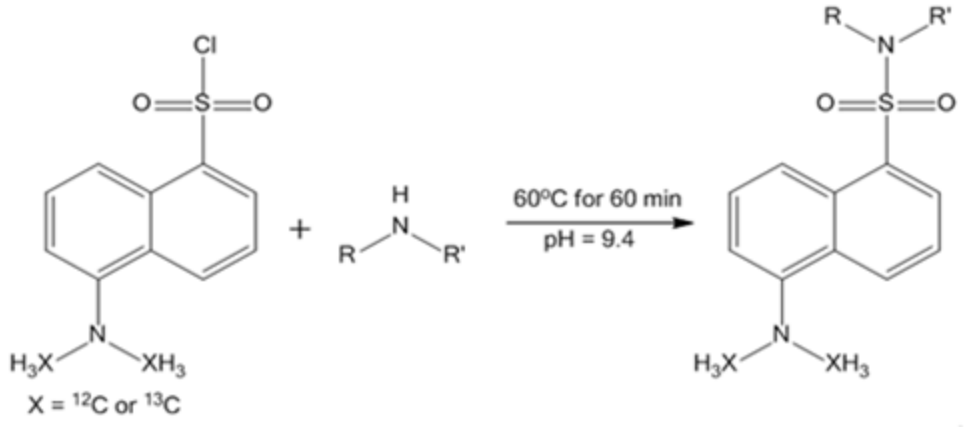


Figure 1.8 Reaction scheme for chemical derivatization of primary/secondary amines using dansyl chloride as the labeling reagent.

1.7 Overview of Thesis

Based on my research objective, this thesis can be divided into two parts. The first part (Chapters 2 and 3) aims to develop a convenient, robust and universally applicable pre-acquisition normalization approach to account for sample concentration variations in comparative metabolomics. The second part (Chapters 4-6) focuses on development of differential isotope dansylation labeling metabolomic workflows for various biological matrices including bacterial cells, plasma and cerebrospinal fluid (CSF), as well as applications of the optimized protocols to bacterial differentiation and discovery of disease-related biomarkers in plasma and CSF.

Specifically, Chapter 2 describes a general approach of determining the total concentration of metabolites based on the use of chemical labeling to attach a UV absorbent to the metabolites to be analyzed, followed by rapid step-gradient LC-UV detection of the labeled metabolites. Chapter 3 presents a modified total metabolite quantification method that involves solvent extraction of the labeled metabolites followed by UV absorbance measurement with a microplate reader instead of using the LC-UV system, in order to enhance the cost- and time-effectiveness. Both methods were shown to be effective in generating reliable metabolome profiles for comparison between different sample groups. In Chapter 4 a robust and sensitive differential isotope labeling LC-MS workflow for comprehensive profiling the amine- and phenol-containing sub-metabolome of bacteria cells is reported, which allows over a thousand peak pairs or putative metabolites to be detected from bacterial cells in one LC-MS run. This workflow is then applied to the differentiation of three bacterial species in cultured media and spiked human urine samples. Chapter 5 focuses on application of the differential isotopic labeling LC-MS technique for the discovery and validation of metabolic signatures in

autism spectrum disorder (ASD). Significant differences were observed in the metabolic profiles of ASD patients and healthy controls, and a set of metabolic signatures were identified with high prediction accuracy. Chapter 6 demonstrates the high analytical performance of the differential isotopic labeling LC-MS platform in parallel metabolomic profiling of CSF and serum. This workflow is then applied to the study of metabolic changes induced following spinal cord injury in both biofluids. Finally, Chapter 7 provides a conclusion of the thesis as well as a brief discussion on the future research directions.

Chapter 2

Determination of Total Concentration of Chemically Labeled Metabolites as a Means of Metabolome Sample Normalization and Sample Loading Optimization in Mass Spectrometry-Based Metabolomics

2.1 Introduction

Liquid chromatography mass spectrometry (LC-MS)-based metabolomic techniques have been widely applied to the qualitative and quantitative analysis of biofluids, cells or tissue extracts for biological studies and biomarker discovery with high sensitivity, high resolution and wide metabolite coverage.[157-160] However, for quantitative studies, variation in total metabolite concentration among different samples can complicate the relative quantification of the metabolome changes in comparative metabolomics.[86] This is particularly true for urinary metabolome profiling. Urinary metabolite concentrations can be governed by many factors, such as kidney filtration and water consumption, and up to 15-fold variations in urine volume can be observed for normal individuals.[161] This variation can be even greater due to disease or drug effects. The changes of the overall concentration can often obscure specific changes of metabolites that are of interest in metabolomic studies.[87] Although the collection of urine samples from an individual over a long period (e.g., 24-hr) may account for the variation of total concentration, the collection and storage process is often inconvenient and cumbersome in practice. Therefore, a good sample normalization strategy is required to compensate for variations in the overall urine concentration. For other biofluids, such as

bronchial lavage fluid and saliva, as well as metabolome extracts of various types of cells with unknown cell numbers, it is also desirable to normalize the sample concentration prior to sample workup and quantitative analysis.

For urine samples, various normalization methods have been reported in the literature. The most common one is creatinine normalization,[162-164 95]because the rate of creatinine formation and excretion is relatively constant under normal conditions. In this method, the concentrations of metabolite analytes are normalized relative to the concentration of creatinine. However, the assumption of constant creatinine excretion is often not valid because creatinine excretion does vary among individuals due to age, gender, lean body mass differences, and even within the same individual, the urine creatinine level can change due to variables such as creatinine intake, time of day, level of exercise, and disease states.[94] Normalization to osmolality or specific gravity has also been used.[111,112,163] This method measures the total solute concentration of urine samples for normalization. However, the solute concentration includes contributions of all dissolving solids and may not directly reflect the total urinary metabolite output. Statistical normalization strategies have also been routinely used for NMR data[87,165] and in LC-MS studies a similar normalization method based on the use of MS "total useful signal" (MSTUS) has been recently reported by Warrack et al,[88] which uses the total intensity of peaks that are present in all samples as the normalization factor. This method avoids contribution from xenobiotics and artifacts and has been demonstrated to be useful in detecting statistically significant changes in the endogenous metabolite profile of urine samples and reduces variation between biological replicates. However, this method does not allow the adjustment of relative sample amounts for mixing in the cases where two

comparative samples (e.g., sample vs. control) need to be mixed prior to MS analysis. Furthermore, this method does not offer a means of controlling the amount of samples to be injected into a mass spectrometer for analysis. The amount of sample injected can be very important for metabolite detection. If the injection amount is small, the low concentration metabolites will not be detectable because their concentration level becomes lower than the detection threshold of the instrument.[90] On the other hand, if a large injection amount is used, the electrospray ionization (ESI) source and the detector will be easily saturated with ions, and the high abundance peaks can obscure small peaks and make them undetectable. Because each normalization method has its strengths and limitations, there is no consensus on which normalization method works the best for LC-MS metabolome studies.

In this work, we present an alternative strategy for sample normalization while offering the possibility of controlling and optimizing the sample injection for optimal mass spectrometric metabolome detection. This method is based on the use of LC-UV for quantifying the total concentration of the chemically labeled metabolites to be analyzed in any type of biological samples including urine. We note that the use of LC-UV for quantification of the total peptides generated from a proteome digest has been reported earlier, using a probing wavelength of 214 nm that corresponds to the carbonyl group in the peptide backbone.[50] Unlike proteins and peptides, which have a relatively uniform backbone structure, metabolites have a wide variety of structures and thus very different UV absorptivity. As a result, it is very difficult to choose a single wavelength for detection. Most of the studies using LC-UV quantification of metabolites are focused on the analysis of a certain class of compounds.[166-168] To the best of our knowledge, there

is no report of using LC-UV as a general tool for the quantification of a metabolome or a subset of the metabolome (e.g., all the metabolites with each containing an amine group, i.e., the amine-containing metabolome).

Although metabolites have very different structures, the use of labeling chemistry can somehow 'unify' the metabolites by attaching the same functional group to each molecule. If this labeling group has a very specific absorption wavelength, then it is possible to quantify the labeled metabolites based on absorption at this wavelength. Recently, our group reported a $^{12}C/^{13}C$-dansylation labeling technique for absolute and relative quantification of the amine- and phenol-containing metabolome by LC-MS.[146] This labeling strategy allows separation of polar or ionic metabolites on a reversed phase (RP) column, as well as provides signal enhancement of 10 to 1000-fold over the unlabeled counterparts. Experimental variation can also be compensated for by using a pooled ^{13}C-labeled sample as the internal standard. Another advantage of this technique is that the aromatic ring structure of the dansyl group can also act as a good chromophore to facilitate UV quantification. In this work, we report a LC-UV method to quantify all of the labeled metabolites in urine samples using a fast step-gradient elution. The quantification results were then used to normalize the urine samples, and to optimize the sample injection amount. We describe the procedures and rationales for selection of detection wavelength in LC-UV, appearance of chromatographic peak profiles and method of peak area integration in step-gradient LC, selection of calibration standards and calibration method for relative and absolute quantification of the total labeled metabolites, and strategy of optimizing the sample injection amount in LC-MS. Finally,

integration of the LC-UV method into a differential isotope labeling LC-MS workflow for improved metabolome quantification is discussed.

2.2 Experimental

2.2.1 Chemicals and Reagents.

^{12}C-dansyl chloride (DnsCl) and amino acid standards were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). The isotopic compound used to synthesize ^{13}C-dansyl chloride was purchased from Cambridge Isotope Laboratories (Cambridge, MA, USA). ^{13}C-dansyl chloride was synthesized in our lab as described previously,[146] and the other chemicals used to synthesize this isotope reagent were also purchased from Sigma-Aldrich. LC-MS grade water, acetonitrile (ACN) and formic acid were purchased from Thermo Fisher Scientific (Edmonton, AB, Canada).

2.2.2 Urine Sample Collection.

Urine samples were collected from two individuals of each gender in three consecutive days. An informed consent was obtained from individual volunteers and ethics approval was obtained from the University of Alberta in compliance with the University of Alberta Health Information policy. On each day three samples were taken with a collection interval of 1 hr, denoted as A, B and C. Between collection of A and B the individuals were instructed to refrain from drinking water, while between collection of B and C the individuals were asked to drink a large amount of water (1 L for individual 1 and 0.5 L for individual 2). The samples were stored at 4°C immediately after collection. The urine samples were centrifuged at 4,000 rpm for 10 min, and the supernatant was filtered twice through a 0.2 μm filter. The filtered urine was aliquoted and stored at -80°C until further use.

2.2.3 Dansylation Labeling Reaction.

The frozen urine samples were thawed in ice-bath and then diluted two-fold prior to the labeling reaction. Fifty μL of urine or amino acid standard solution was mixed with sodium carbonate/sodium bicarbonate buffer (pH 10,3) and ACN. The solutions were vortexed, spun down and mixed with 50 μL freshly prepared ^{12}C-dansyl chloride solution (18 mg/mL) (for light labeling) or ^{13}C-dansyl chloride solution (18 mg/mL) (for heavy labeling). The reaction was allowed to proceed for 1 hr at 60°C. After 1 hr, NaOH was added to the reaction mixture to quench the excess dansyl chloride. The solution was then incubated at 60°C for another 10 min. Finally, formic acid in 50/50 ACN/H_2O was added to consume excess NaOH and to make the solution acidic. The ^{12}C- or ^{13}C-labeled mixtures were centrifuged at 14,000 rpm for 10 min before injecting onto the ultra-high performance liquid chromatography (UPLC) column for UV quantification. For MS analysis, the ^{12}C- and ^{13}C-labeled mixtures were combined in a ratio determined by the quantification results.

2.2.4 LC-UV Quantification.

A Waters ACQUITY UPLC system with a PDA detector was used for the quantification step. Two μL of the labeled urine or amino acid solution was injected onto a Waters ACQUITY BEH C18 column (2.1 mm × 5 cm, 1.7 μm particle size, 130 Å pore size) for a fast step-gradient run. Solvent A was 0.1% (v/v) formic acid in 10% (v/v) acetonitrile, and solvent B was 0.1% (v/v) formic acid in acetonitrile. The gradient started with 0% B for 1 min and was increased to 95% within 0.01 min and hold at 95% B for 1 min to ensure complete elution of all labeled metabolites. The gradient was restored to

0% B in 0.5 min and hold at this condition for 3.5 min to re-equilibrate the column. The flow rate used was 0.45 mL/min.

2.2.5 LC-FTICR-MS Analysis.

The labeled urine samples were analyzed using a Bruker 9.4 Tesla Apex-Qe Fourier transform ion-cyclotron resonance (FTICR) mass spectrometer (Bruker, Billerica, MA) linked to an Agilent 1100 series binary HPLC system (Agilent, Palo Alto, CA). The samples were injected onto an Agilent reversed phase Eclipse Plus C18 column (2.1 mm × 10 cm, 1.8 μm particle size, 95 Å pore size) for separation. Solvent A was 0.1% (v/v) formic acid in 5% (v/v) acetonitrile, and solvent B was 0.1% (v/v) formic acid in acetonitrile. The chromatographic conditions were: t = 0 min, 20% B; t = 3.5 min, 35% B; t = 18 min, 65% B; t = 21 min, 95% B; t = 21.5 min, 95% B; t = 23 min, 98% B; t = 24 min, 98% B; t = 26.5 min, 99% B; t = 28.5 min, 99% B; t = 29.5 min, 20% B. The flow rate was 180 μL/min and the flow from LC was split 1:3 before entering the electrospray ionization (ESI) source. All MS spectra were obtained in the positive ion mode. The resulting MS data were processed using R language program based on XCMS[24] written specifically for ^{12}C-/^{13}C-peak pair picking.[51] The program eliminated many false positive peaks, such as isotopic peaks, common adduct ions, and multiply charged ions, and only the protonated ion pairs were exported for further analysis.

2.2.6 Statistical Analysis.

The extracted peak pair data for the two individuals' three-day urine samples was aligned by retention time and accurate mass, and only those peak-pair features shared by no less than 50% of the samples were retained for multivariate analysis. The resulting multivariate dataset contains 108 observations (individual urine samples) and 467

variables (peak-pair features). The multivariate analysis was performed by Metaboanalyst[169] (www.metaboanalyst.ca) and SIMCA P+12 (Umetrics, Umeå, Sweden) and the data were mean-centered and auto-scaled (unit variance) prior to analysis. Principal component analysis (PCA) was applied first to evaluate general clustering of normalized and un-normalized data for two different individuals. Supervised partial least square-discriminate analysis (PLS-DA) was then used to reveal subgroups within an individual.

2.3 Results and Discussion

2.3.1 Wavelength Selection.

Dansylation is a well-studied labeling chemistry that works for primary amines, secondary amines and phenols. The aromatic ring structure of the dansyl group makes it a good chromophore with very characteristic absorptions. Figure 2.1 shows the absorption spectra of four standards, dansyl-tryptophan, dansyl-alanine, dansyl-putrescine and dansyl-threonine, from 210 – 400 nm. The spectra features are very similar, suggesting that the dansyl group plays a major role for the absorption. As expected for aromatic hydrocarbons, three sets of bands were observed that originate from $\pi \rightarrow \pi^*$ transitions: one strong absorption band centered at ~220 nm, one weaker band at ~252 nm, and a weakest one at ~338 nm.[47] The absorption spectra of other 17 labeled amino acid standards as well as several labeled urine samples were examined (data not shown). It was found that the peak wavelength could shift up to about 20 nm for more complex mixtures (urine), indicating that the presence of other functional groups can have minor effects on the overall absorption. For example, for the weakest band, the peak wavelength

can vary from 326 to 349 nm. Since only one wavelength should be chosen for quantification, the median value of each set of bands was used for comparison.

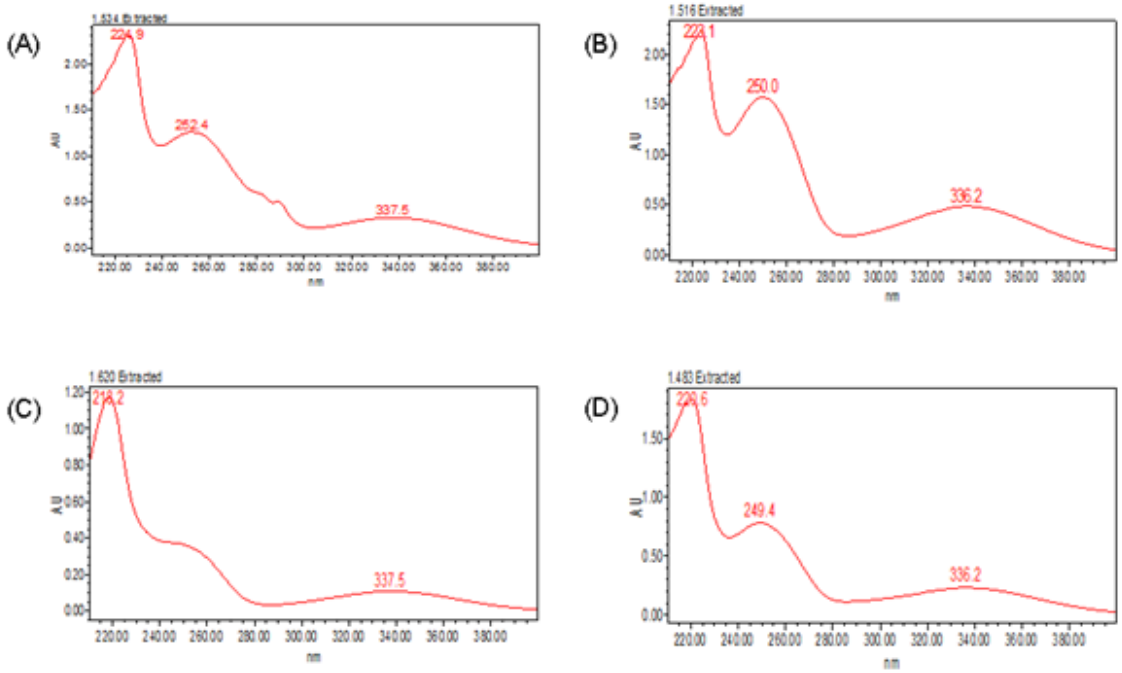


Figure 2.1 Absorption spectra of (A) dansyl-tryptophan (0.25 mM), (B) dansyl-alanine (0.5 mM), (C) dansyl-putrescine (0.25 mM), and (D) dansyl-threonine (0.5 mM).

There are several considerations when selecting a proper wavelength for quantification. Firstly, the absorption should be specific to the dansyl group. Because UV absorbance is additive, absorption of other functional groups in a metabolite would also contribute to the measurement, which can affect accuracy of quantification. Many common organic chromophores, such as carbonyl, carboxyl and phenyl, have absorption peaks under 300 nm, but very few functional groups absorb at higher wavelengths. This can be illustrated by the overlaid chromatograms of 17 unlabeled amino acids at 220, 252 and 338 nm (Figure 2.2). All amino acids elute out between 0.2 to 0.6 min. It is clear that significant absorption was observed at 220 and 252 nm, but not at 338 nm. Thus, using a

wavelength of 338 nm for detection can provide the least interference from other chromophores. Secondly, the high absorbance at 220 and 252 nm can easily saturate the UV signal at practically useful concentrations, such as for a two-fold diluted urine sample, as shown in Figure 2.3. Finally, the rapid solvent change can lead to a non-flat background because of the differences in absorption coefficients and refractive indexes of different solvents. The background change would affect the accuracy for peak integration and should be kept as small as possible. Comparing the background change at these three wavelengths in Figure 2.2, it is apparent that the change in background is smallest at 338 nm. Taken together, 338 nm was chosen as the probing wavelength for all of the following quantification work.

2.3.2 Calibration of Labeled Amino Acid Standards.

Since the main purpose of this method is to quantify all the labeled metabolites in a sample, a step-gradient was used to elute all compounds together. By using the UPLC system, it is possible to run at a high flow rate to increase the throughput. Figure 2.4A shows the overlaid elution profiles of the mixtures of 17 labeled amino acid standards (17-aa-std) at different concentrations. The early eluting peak corresponds to the quenched dansyl chloride (Dns-OH), which does not retain on the column well. The peaks between 1.4 min to 2.0 min are from the labeled amino acids. As shown in Figure 2.4A, the step-gradient allows fast elution of all labeled compounds in 2 min while separation of these compounds from the quenched DnsCl is achieved to avoid reagent interference in the UV measurement. The peak area increases accordingly with increasing amino acid concentration, and the elution profile is very similar for different concentrations. Figure 2.4A (inset) also shows that the peak area of the quenched DnsCl

decreases with increasing amino acid concentration, because there is less excess DnsCl at higher concentrations of analytes. Even with a sharp solvent change in 0.01 min, there is still a slight separation of the amino acids, as evident from several peaks shown in each elution profile (see Figure 2.4A between 1.4 and 2.0 min). This is due to the wide variety of side-chains in amino acids that can interact differently with the column. Nevertheless, the total area of the labeled amino acid peaks can be readily integrated using the Empower software of the LC-UV instrument.

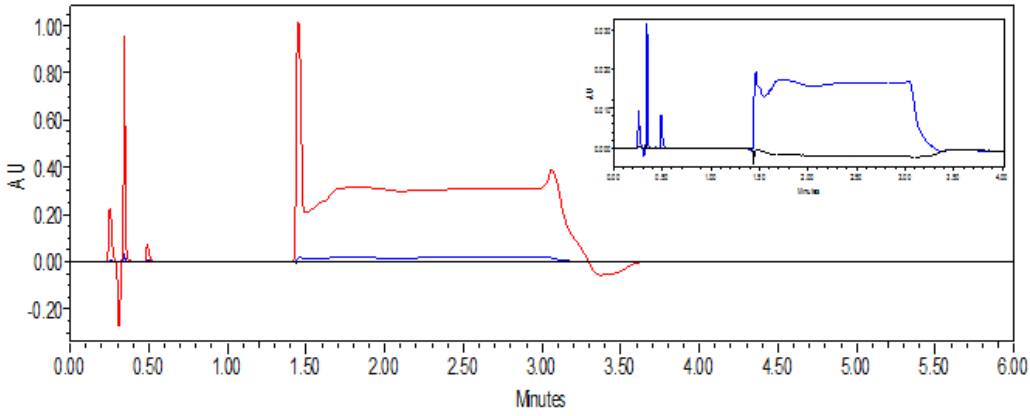


Figure 2.2 Overlaid chromatograms of unlabeled 17 amino acid standards at 220 nm (red), 252 nm (blue) and 338 nm (black). Inset: Comparison between 252 nm and 338 nm only.

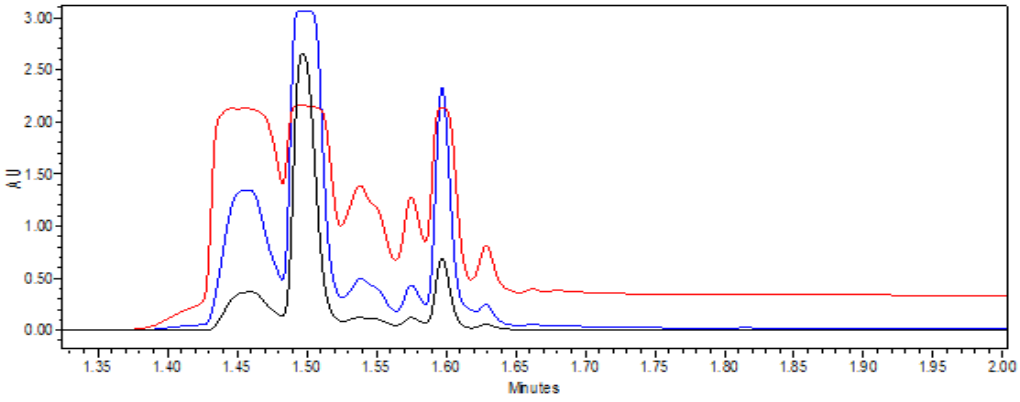


Figure 2.3 UV signal saturation at 220 nm (red) and 252 nm (blue) for two-fold diluted urine; saturation does not occur at 338 nm (black).

It should be noted that although the baseline of the chromatograms is relatively flat at 338 nm, there is a very small negative peak that appears at 1.43 min due to the rapid solvent change. This system peak is very reproducible, with a retention time shift of less than 0.06% and a peak area variation of 2.32% in five replicate runs. Therefore, the peak area difference was used for quantification of the metabolites, which was calculated from the peak area measured for a given sample minus the system peak area measured in a blank run. Here the system peak area has a negative value so the peak area difference is actually the sum of the sample peak area and the absolute value of the system peak. Integration was made from 1.43 to 2.0 min to ensure every peak from the sample has been included. While not tested in this work, other manufacturers' LC-UV instruments and LC columns may likely give different profiles. However, the strategy of integrating the entire elution peak including the system peak as described above should be applicable to other systems.

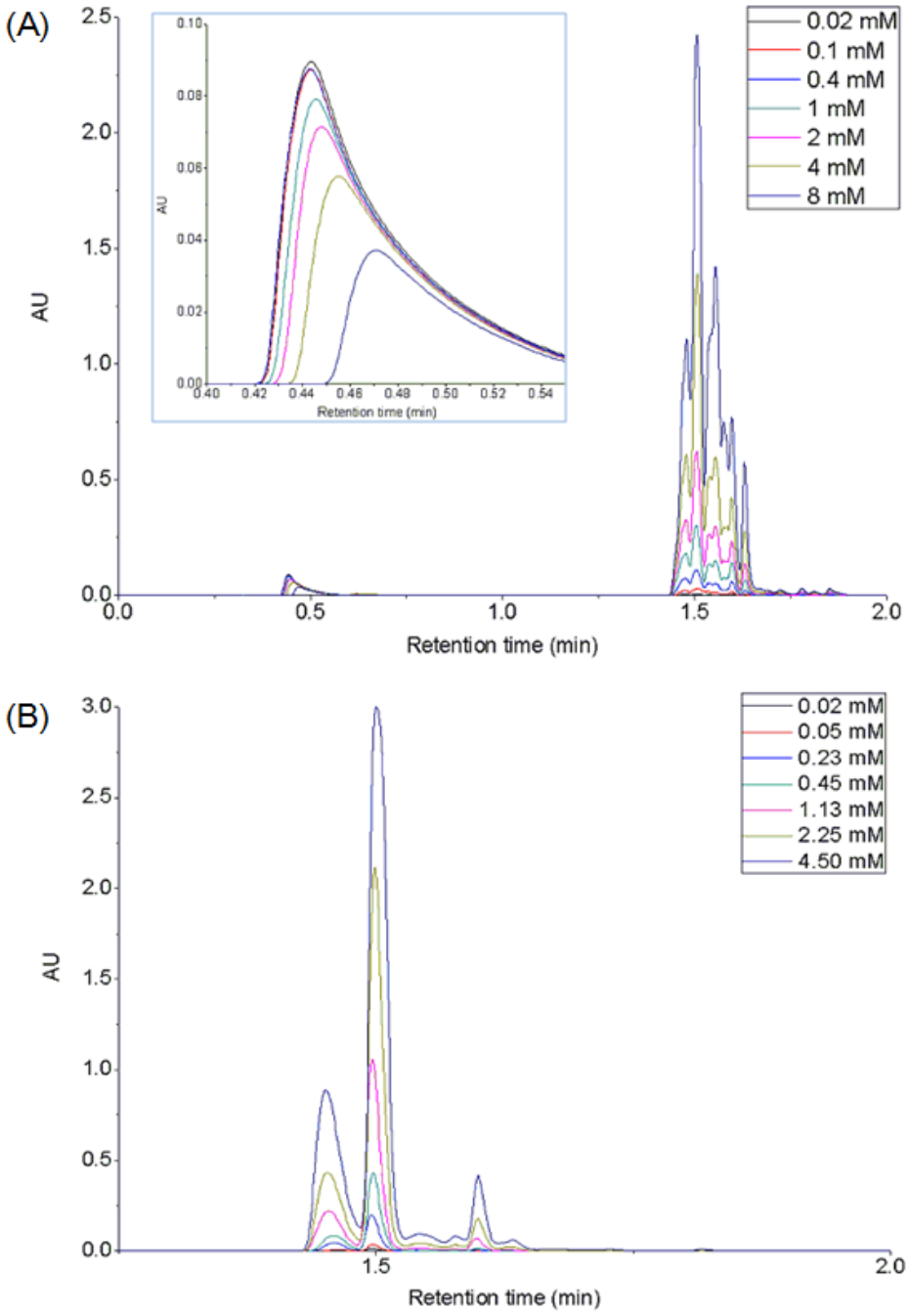


Figure 2.4 (A) Overlaid UV chromatograms of a mixture of 17 labeled amino acid standards (17-aa-std) at different concentrations. Inset: zoom-in region of 0.4 – 0.55 min. (B) Overlaid UV chromatograms of a labeled pooled urine sample at different concentrations.

Figure 2.5A shows the calibration curve of peak area versus the concentration of 17-aa-std. The curve is linear from 0.02 mM to 6.25 mM amino acids with very good correlation (R^2 = 0.9992). In this case, a series of diluted 17-aa-std solutions were prepared and labeled individually. A similar curve was obtained when 6.25 mM standard solution was diluted after labeling, indicating that one can establish the calibration curve by preparing diluted solutions either before or after labeling. As dilution after dansylation is more convenient and consumes fewer reagents, this strategy was used for all the subsequent works. However, it should be noted that, if too high concentrations of analytes (>6.25 mM standard solution) were used to prepare the stock solution, a non-linear response was observed due to the decreased dansylation efficiency as the relative amount of the dansyl chloride reagent was not sufficient. The dansyl chloride used for the reaction was 3.35 μmol, while the deviation from linearity became significant from 1.75 μmol of amino acids and no UV signal saturation was observed at these concentrations. Therefore, in order for the dansylation reaction to be complete, the amount of dansyl chloride needs to be more than ~2-fold in excess. The low end of this linear range was limited by the formation of side products, such as Dns-NH_2 and Dns-$N(CH_3)_2$.[170] However, the acquired linear range of the calibration curve was adequate for quantification of most urine samples, as it will be discussed later.

Because the analyte composition of biological samples can be very different, it is important to investigate how well a calibration curve established from one sample can be used to quantify the amount of metabolites in another sample. To do this, we compared the calibration curves of the 17-aa-std and a mixture of 15 other labeled amine and phenol standards (15-std-mix) with varying structures (see Figure 2.6). These curves

were obtained independently based on the actual concentration of standards used. Similarities in UV absorptivity were evaluated by comparing the slopes (sensitivity) of the two calibration curves using a modified student t-test at 95% confidence level[171] and the result showed that there was no statistical difference between these two slopes. In addition, if we use the linear regression equation obtained from the 17-aa-std to calculate the concentration of the 15-std-mix, the error was less than 4%. These results indicate that although the absorptivity of individual dansyl metabolites at 338 nm can be different (some evidences are shown in Figure 2.1), the average absorptivity of a mixture of many dansyl labeled metabolites can be very similar, because the high absorption of some metabolites can be averaged out by other low absorption compounds. Finally, as it is shown below, there is no significant difference between the slopes of the 17-aa-std and the labeled urine curve. Therefore, we can use the calibration curve of the 17-aa-std to determine the absolute concentration of the total labeled metabolites in biological samples. We note that, mixtures of amino acids can be purchased from chemical suppliers and readily prepared for dansyl labeling, providing a convenient means of establishing a calibration curve.

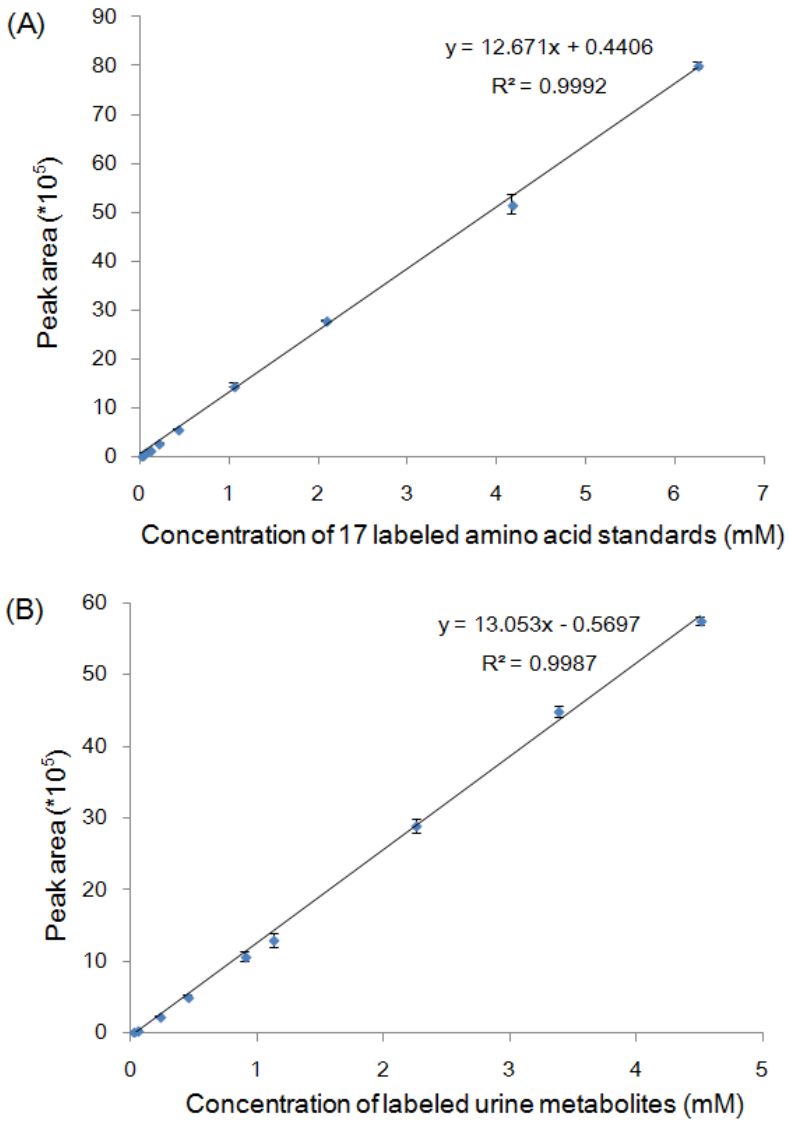


Figure 2.5 (A) Calibration curve of the mixture of 17 labeled amino acid standards (17-aa-std) from triplicate labeling experiments at each concentration. The labeled amino acid mixtures were diluted before dansylation. (B) Calibration curve of a labeled urine sample from a series of dilution of the highest urine concentration which was labeled in triplicate experiments. The concentration of the labeled metabolites in each diluted sample of the labeled urine was calculated from the undiluted labeled urine concentration determined from Figure 2.5A multiplying by the dilution factor at each data point.

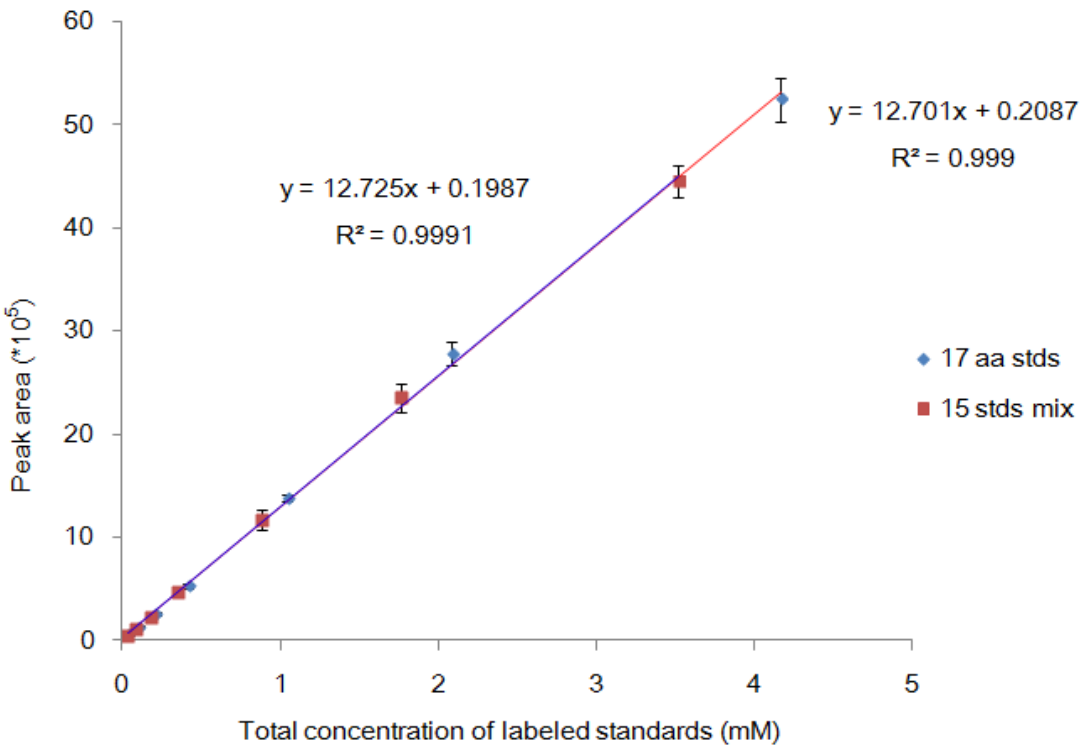


Figure 2.6 Slope comparison for two standard mixture solutions. '17 aa stds' refers to the amino acid standards solution containing L-alanine, L-arginine, L-aspartic acid, L-cystine, L-glutamic acid, glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-serine, L-threonine, L-tyrosine and L-valine. The '15 stds mix' solution contains L-asparagine, cystathionine, ethylamine, *trans*-ferulic acid, L-glutamine, γ-glutamylcysteine, glycylproline, homocitrulline, homovanillic acid, 4-hydroxy-3-methoxyphenyllactic acid, ornithine, phenol, pantothenic acid, L-tryptophan and xanthurenic acid. The two linear regression lines were obtained independently. The slopes were not statistically different using the modified student t-test at 95% confidence level ($t = 0.07 < t_{crit} = 4.30$).

2.3.3 Quantification of Labeled Metabolites in Urine.

Figure 2.4B shows the elution profile of the labeled urine, which is different from that of the labeled amino acids, because there are so many metabolites in the urine sample that elute out closely together, which tends to smooth the chromatographic peaks. As a result, fewer peaks were observed for urine. The calibration curve for urine samples was established from a pooled urine sample, with dilution from 1.3 to 200 fold (see Figure 2.5B). In this case, a pooled urine sample was labeled with dansyl chloride and the peak area of the labeled urine was measured by LC-UV. The total concentration of the dansylated metabolites in the labeled urine was determined using the calibration curve shown in Figure 2.5A, assuming that the absorptivity of the labeled urine is the same as that of the labeled 17 amino acid standards. The labeled urine was then diluted to produce a series of diluted samples for LC-UV measurements. The concentration of each diluted sample was calculated by the concentration of the undiluted sample multiplying by the dilution factor or 1/dilution-fold.

As it is shown in Figure 2.5B, a linear relationship was observed between peak area and the labeled urine concentration. In this case the original urine concentration lies well within the linear range of the labeled amino acid standards. However, we note that, in using this method for quantification of labeled urine metabolites, occasionally the original urine concentration of an individual sample could be slightly higher than 6.25 mM (the upper limit of the calibration curve shown in Figure 2.5A), which can result in incomplete labeling. Therefore, in our work, all of the original urine samples were diluted 2-fold prior to dansylation reaction to ensure the labeling was complete. For dansylation reaction, there was no matrix effect from the urine sample. If the matrix of urine (or any

biofluid) affected the dansylation reaction, the extent of matrix effect would be dependent on the urine concentration. The higher the urine concentration the larger the matrix effect would be. A simple procedure was applied to check the presence of any matrix effect on dansylation. In this procedure, a series of dilutions of a urine sample (undiluted, 2.5×, 5×, and 10×) are prepared and individually labeled by dansylation, the resultant labeled samples are analyzed by LC-UV. The peak area of the labeled samples is plotted as a function of the dilution factor or 1/dilution-fold (Figure 2.7A). In a control experiment, the undiluted urine sample is labeled by dansylation and then the labeled concentrated sample is diluted by 2.5-, 5- and 10-fold. The peak areas measured from the diluted samples are plotted against the dilution factor (Figure 2.7B). If the dilution calibration curves obtained from the two experiments are found to be statistically the same, then we can conclude that there is no matrix effect from the urine sample on the dansylation reaction. In this particular case, we found that the calibration curves shown in Figures 2.7A and 2.7B were statistically the same and thus there was no matrix effect of urine on dansylation reaction.

It should be noted that chemical labeling to "unify" the absorptivity of a known metabolite mixture (used as the calibration standard) and the urine samples is critical for determining the total concentration of the urinary metabolites. Without labeling, total metabolite quantification by UV absorbance measurement is not possible. For example, it is anticipated that some metabolites in urine would have functional groups that can absorb at 338 nm. Indeed, if the chromatogram at 338 nm of unlabeled urine was examined, there was a small peak that has a similar retention time as the dansyl labeled peaks. Since this peak area was also proportional to urine concentration, we investigated

the ability of using unlabeled urine for quantification. We compared the slopes of the calibration curves established at different wavelengths and retention time windows for three individuals' urine samples (Figure 2.8). We found there were several problems associated with direct quantification of the unlabeled urine. Firstly, the slopes of curves depend heavily on the wavelength used. For example, for the peak at 1.43 – 1.5 min, the slopes at 338 nm were all similar for three individuals (Figure 2.8C). However, at 280 nm the slope for individual B becomes different from individual A and C (Figure 2.8B), and at 254 nm all three slopes are significantly different from each other (Figure 2.8A). Likewise, for the peak at 0.22 − 1.3 min, the slopes at 254, 280 and 338 nm are also very different (Figure 2.8D-F). As a result, it is very difficult to choose a wavelength at which the absorptivity is similar for all different individuals. Secondly, although the slopes at 338 nm and 1.43 – 1.5 min were similar for all three individuals, the peak area was much smaller compared to the dansyl peaks (<5%), therefore the linear range was significantly reduced at the lower end. Moreover, since there is no good standard for calibration, it would be difficult to determine the total concentration of the metabolites. Thus, determination of total metabolite concentration cannot be done by direct UV analysis of urine samples.

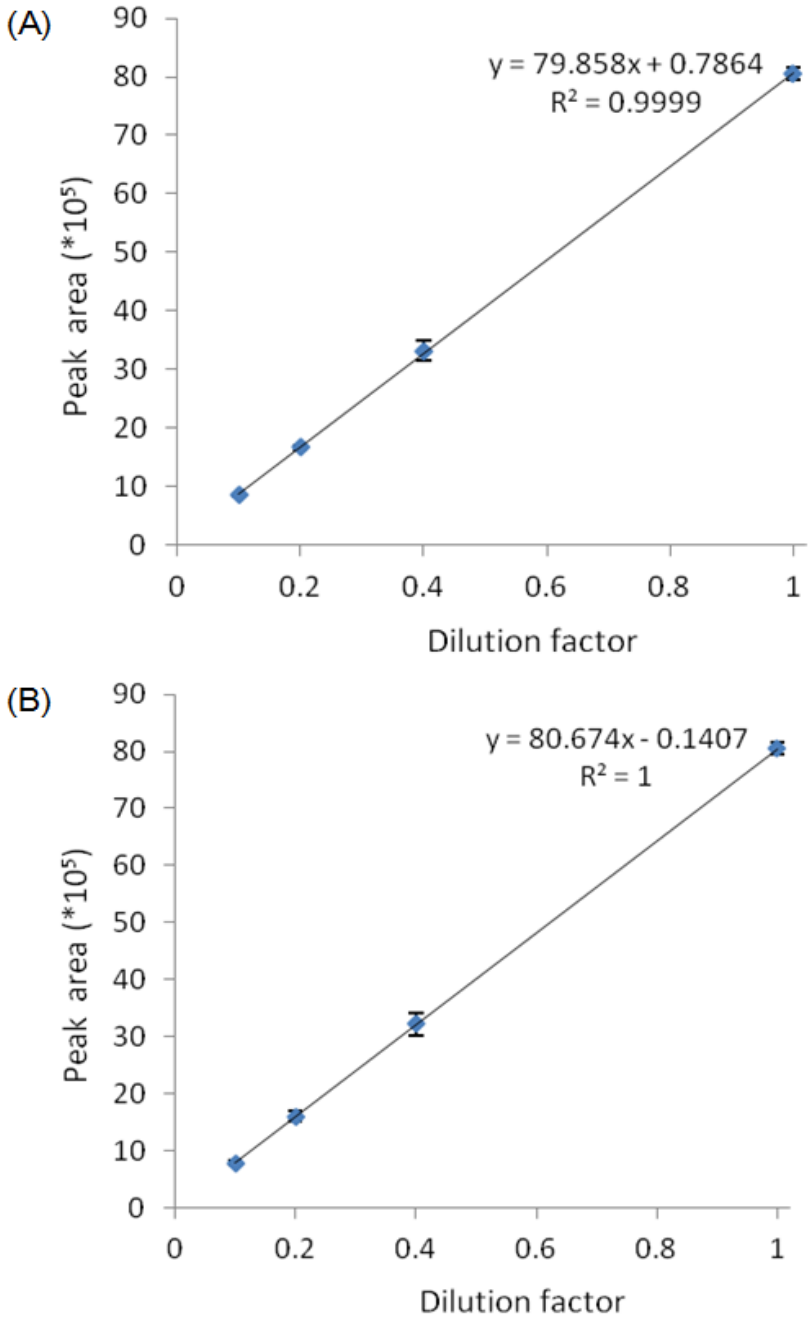


Figure 2.7 (A) Plot of peak area as a function of dilution factor for individually labeled urine samples. (B) Plot of peak area as a function of dilution factor for urine samples diluted after labeling.

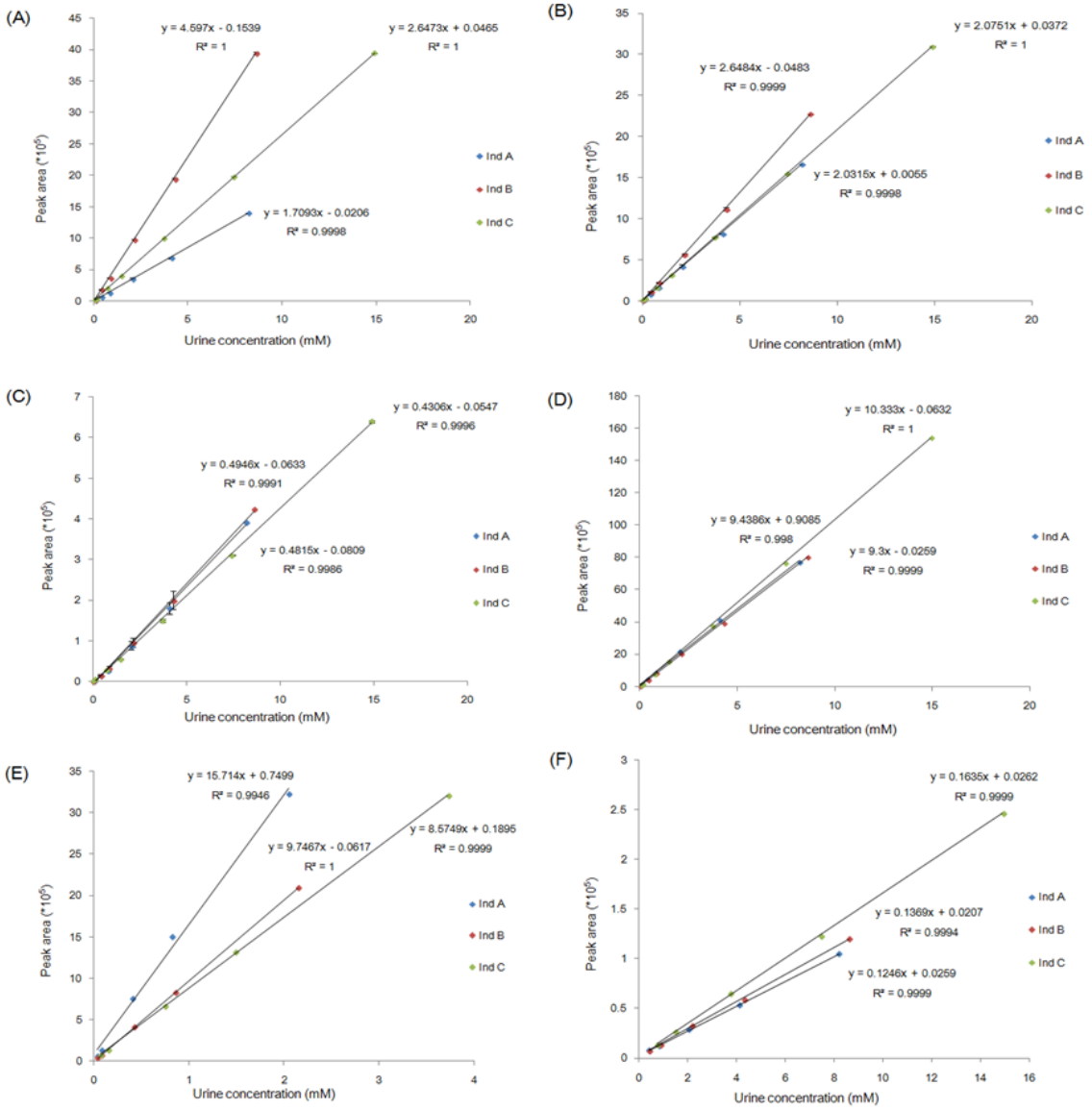


Figure 2.8 Slope comparison between three individuals' unlabeled urine samples: (A) 254 nm, 1.43-1.5 min; (B) 280 nm, 1.43-1.5 min; (C) 338 nm, 1.43-1.5 min; (D) 254 nm, 0.22-1.3 min; (E) 280 nm, 0.22-1.3 min; (F) 338 nm, 0.22-1.3 min. The majority of peaks were eluted out between 0.22 to 1.3 min under isocratic condition. The compounds that elute out at high organic condition were eluted out as a single peak from 1.43 to 1.5 min. The urine concentration was determined by labeling the urine sample and then calculating the concentration from the calibration curve shown in Figure 2.5B.

2.3.4 Optimization of Sample Loading to LC-FTICR-MS.

The ability of the LC-UV method for quantification of labeled metabolites allows us to control the amount of sample to be injected into the LC-ESI-FTMS instrument. In order to investigate the effect of sample injection amount on the FTMS results, the same sets of pooled urine samples (in triplicate) used to establish the calibration curve were injected into the instrument with an injection volume of 2 μL. The urine samples were prepared by mixing equal amount of ^{12}C- and ^{13}C-labeled solutions, which will give peak pairs with a mass difference of 2.0067 in the FTMS run. The number of peak pairs obtained was plotted against the labeled urine concentration as shown in Figure 2.9. It is clear that the number of peak pairs increases as the labeled urine concentration increases, because at lower concentrations, the low abundance peak pairs would be buried in the background and thus either becomes undetectable or be filtered out during the data processing step (i.e., S/N<10).

As Figure 2.9 shows, when the labeled urine concentration is sufficiently high, the number of peak pairs levels off. This can be explained by considering the dynamic range of FTICR and the ion suppression effect of the ESI source. Compared to some other mass spectrometers, such as quadrupole MS, FTICR-MS has relatively small ion detection dynamic range.[172] If the sample concentration is too high, it is possible that the detector cell will be overloaded with ions, which can obscure the detection of small peaks from low abundance or less ionizable metabolites eluted out close together. The ESI source also poses an upper limit on the number of ions, because at high concentrations competition for either space or charge becomes important. In this case, the number of peak pairs comes to a plateau at a labeled urine concentration of 3.4 mM with 2 μL

injection. Therefore, in the subsequent metabolome profiling work, the urine amount was adjusted to be equal to this value for optimal sample injection.

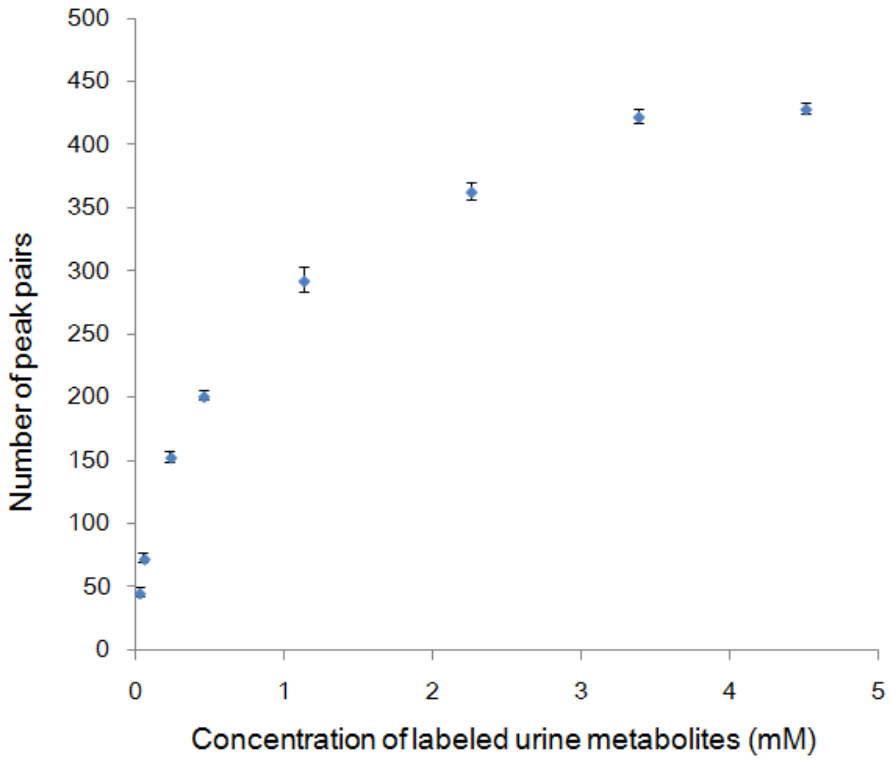


Figure 2.9 Number of peak pairs detected by LC-FTICR-MS vs. labeled urine metabolite concentration (n=3). The MS data was pre-processed to exclude peak pairs with ratios of larger than 1.5 or smaller than 0.67, as well as peaks with intensity of less than 25000 counts (i.e., S/N<10).

2.3.5 Normalization of urine sample concentrations for differential isotope labeling LC-MS.

The LC-UV quantification method was applied to the normalization of two individuals' three-day urine samples. For each day, three urine samples were collected and denoted as

A, B and C. Sample B was collected 1 hr after sample A without drinking water. Sample C was collected 1 hr after sample B, but a large amount of water was taken during this interval. It is therefore expected that the urine concentration of sample A should be similar to sample B, and sample C would be much diluted. Indeed, the chromatographic peak area of sample C was significantly smaller than that of A and B for all three-day samples of the two individuals (data not shown). Triplicate experiments of dansylation were done for each sample and the peak area variation was found to be in the range of 0.2-6.6%, which indicates good reproducibility for the labeling reaction. The peak areas obtained from the LC-UV measurement were compared to creatinine assay and osmolality measurement results of the same set of urine samples (Figure 2.10). Since this is a relatively simple set of samples that only involve two healthy individuals in three consecutive days, it is unlikely to have a large variation in creatinine excretion. As Figure 2.10 shows, a good correlation was obtained in both cases, which is quite reassuring on the validity of the LC-UV method for sample normalization. However, LC-UV quantification is more reproducible than creatinine assay and osmolality measurement, as the relative standard deviation for those two methods can be as high as 20%. Moreover, the LC-UV method can potentially be applied to many different biofluids. As indicated in the Introduction, the creatinine normalization method cannot be applied to samples where the concentration of creatinine itself varies due to biological processes or creatinine is totally absent in a sample (e.g., in cell extracts). The osmolality method also has shortcomings, such as inaccuracy due to salt content variations in samples.

We have incorporated this LC-UV sample normalization method into the differential isotope labeling LC-MS metabolome profiling workflow. The isotope

labeling strategy allows us to use a ^{13}C-labeled pooled urine as the internal standard, while each individual sample is ^{12}C-labeled. For quantitative analysis, we can compare the metabolite concentration in two samples by comparing their ^{12}C-/^{13}C-peak ratio, if the same amount of ^{13}C-labeled pooled urine was used.[146] In a previous work the pooled urine was prepared by adding equal volume of each individual sample. However, the problem of this pooling strategy is that the contribution of each sample would be different due to different concentrations, and some of the low abundance metabolites in low concentration samples may be lost. By taking advantage of the quantification method described here, we can prepare a pooled urine sample by adjusting the volume of individual samples so that an equal amount of each sample was aliquoted for pooling. This procedure can alleviate the bias towards high concentration samples. More significantly, when a ^{12}C-labeled individual sample is taken to mix with the ^{13}C-labeled pooled urine, the volume or concentration of the individual sample can be normalized based on its labeled urine concentration to ensure an equal amount of an individual sample and the pooled urine is mixed.

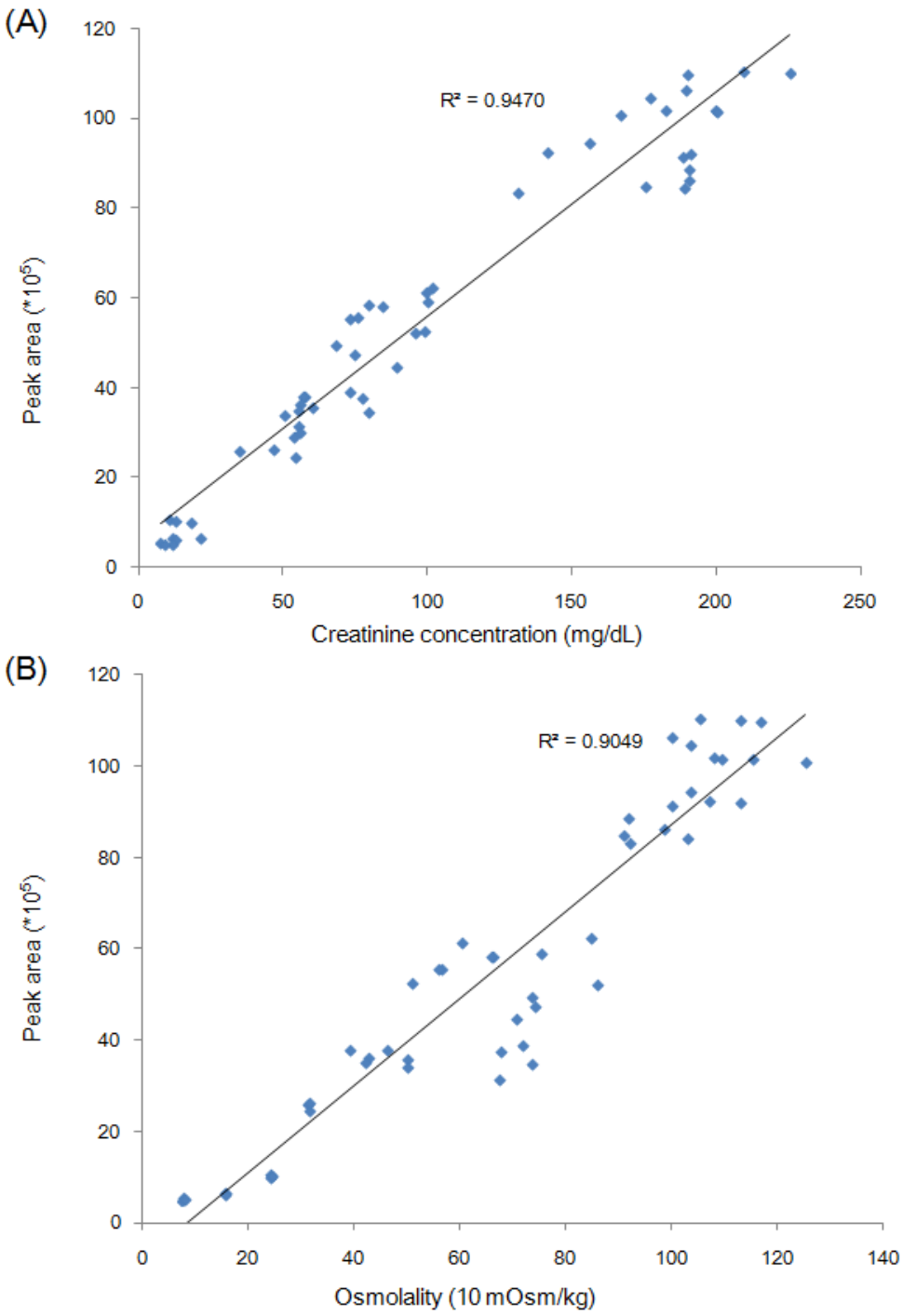


Figure 2.10 Correlation between (A) UV peak area and creatinine concentration, and (B) UV peak area and osmolality.

In order to examine the effect of normalization, two sets of samples were prepared. The first set of samples were un-normalized (denoted as UN). These samples were prepared by mixing equal volume of each ^{12}C-labeled urine sample with the ^{13}C-labeled pooled urine. The second set of samples was normalized (denoted as NOR) by measuring the concentration of the labeled urine metabolites in each sample with the use of the calibration curve shown in Figure 2.5B and the peak area of the eluted labeled metabolites in LC-UV and then adjusting the volume of each sample to ensure an equal sample amount was mixed with the pooled sample. At the same time, the volume was calculated to ensure the total amount of the ^{12}C- and ^{13}C-labeled samples was the same before mixing. In both cases, the same pooled urine sample was used with a concentration of 3.4 mM. The injection volume was adjusted to account for the volume variation after mixing so that the injection amount in each case remained optimal.

Figure 2.11 shows the representative mass spectra of a selected peak pair and the calculated $^{12}C/^{13}C$ ratios. This peak pair was identified to be Dns-alanine by matching the accurate mass and retention time with the amino acid standard solution in our standard library. For the un-normalized samples, the ratio for sample C was much smaller than samples A and B, while the ratios were all similar for the normalized samples. Since the ^{13}C-labeled pooled urine amount was the same in each sample, we can calculate the ratios of A/B, B/C and A/C from their $^{12}C/^{13}C$ ratios. For the un-normalized samples, the ratios were A/B = 1.57, B/C = 2.34, A/C = 3.68, and the ratios for the normalized samples were A/B = 0.92, B/C = 1.06, A/C = 0.97. Because the samples were collected in 2 hr, one would expect that the amount of most metabolites would not change significantly, and therefore the ratios should be close to 1. This was the case for the normalized samples.

But for the un-normalized samples the B/C and A/C ratios were clearly skewed due to the dilution of sample C.

Similar observations were obtained when we examined the intra-day sample peak ratio distributions using box plots, as shown in Figure 2.12. The log ratios between the intra-day samples should be close to 0 for most metabolites under the reasonable assumption that there are little biological variations during the 2-hr collection interval, which should result in a box with a small spread around 0. This is illustrated to be the case in the box plot of the normalized samples (Figure 2.12A) for both individuals. For the un-normalized samples (Figure 2.12B), the log A/B ratio was still close to 0 because the concentrations of samples A and B were similar. However, the log ratios for A/C and B/C show significant deviation from 0, due to the dilution effect in sample C. Thus, there would be a large error, if we used these ratios for metabolite quantification.

The whole set of data were analyzed by multivariate analysis. The unsupervised PCA was first applied to the normalized and un-normalized samples to generate an overview on how the data was scattered, as illustrated in panels A and B of Figure 2.13, respectively. It can be seen that separation between the two individuals was observed in both cases. However, for the normalized samples, the data was scattered more randomly, and the difference between the two individuals was mostly reflected by the first principal component. On the other hand, the separation between individuals of un-normalized samples was mainly attributed to the second principal component, and subgroups within each individual were observed due to different sample concentrations. In particular, the low concentration samples from the two individuals tend to gather together to form a third group, as shown on the middle right of Figure 2.13B. In order to better visualize

these subgroups within each individual, supervised PLS-DA was utilized to make the classification of the two individuals. Figure 2.13C shows the 3D plot established by using this model. The index number (Num) was used as one axis in order to separate the different sets for easier visualization. The PLS-DA fit criteria for the normalized samples were found to be $R^2 = 0.978$ and $Q^2 = 0.941$, indicating an excellent model, and over-fitting is not a main concern here because this model is not forced to show separation of the subgroups.[173] It is clear from the plot that without normalization (blue and yellow), different subgroups were artificially created based on sample concentrations. On the other hand, the normalized samples (red and green) only show separation of two individuals, with no further division into sub-groups. These results illustrate that the LC-UV sample normalization strategy can overcome the problem of artificial separation caused by the variations of the original urine concentrations in relative quantification of urine metabolomes using differential isotope labeling LC-MS.

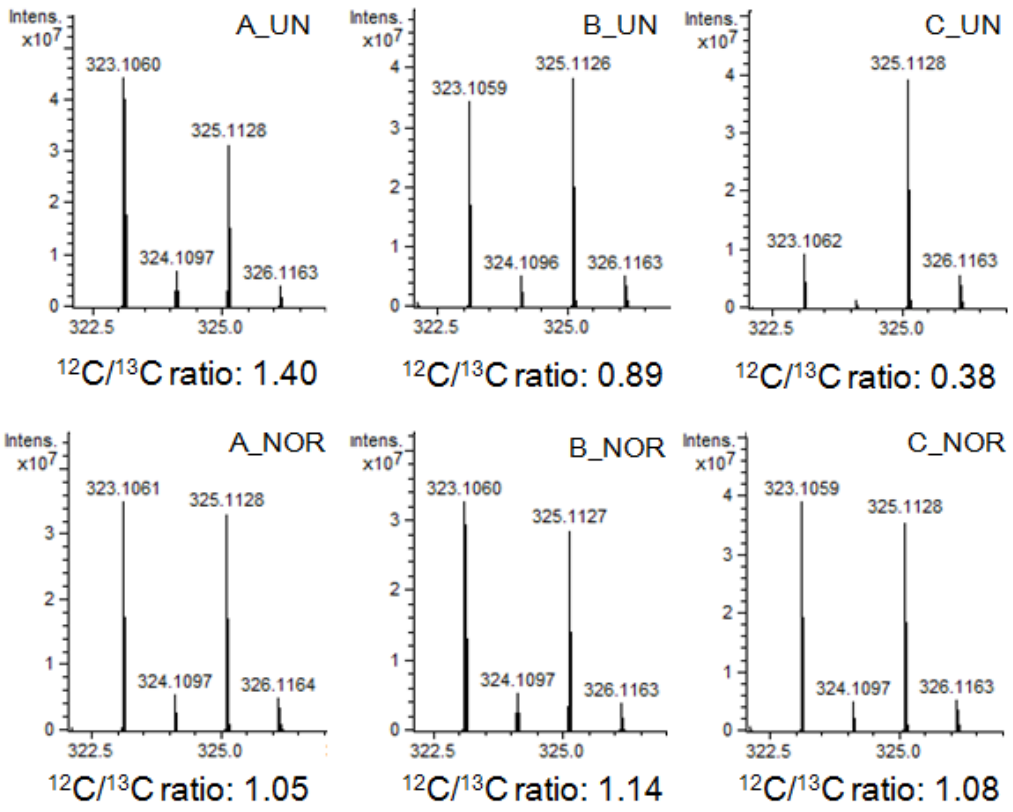


Figure 2.11 Representative mass spectra showing the ratios of a selected peak pair belonging to Dns-alanine.

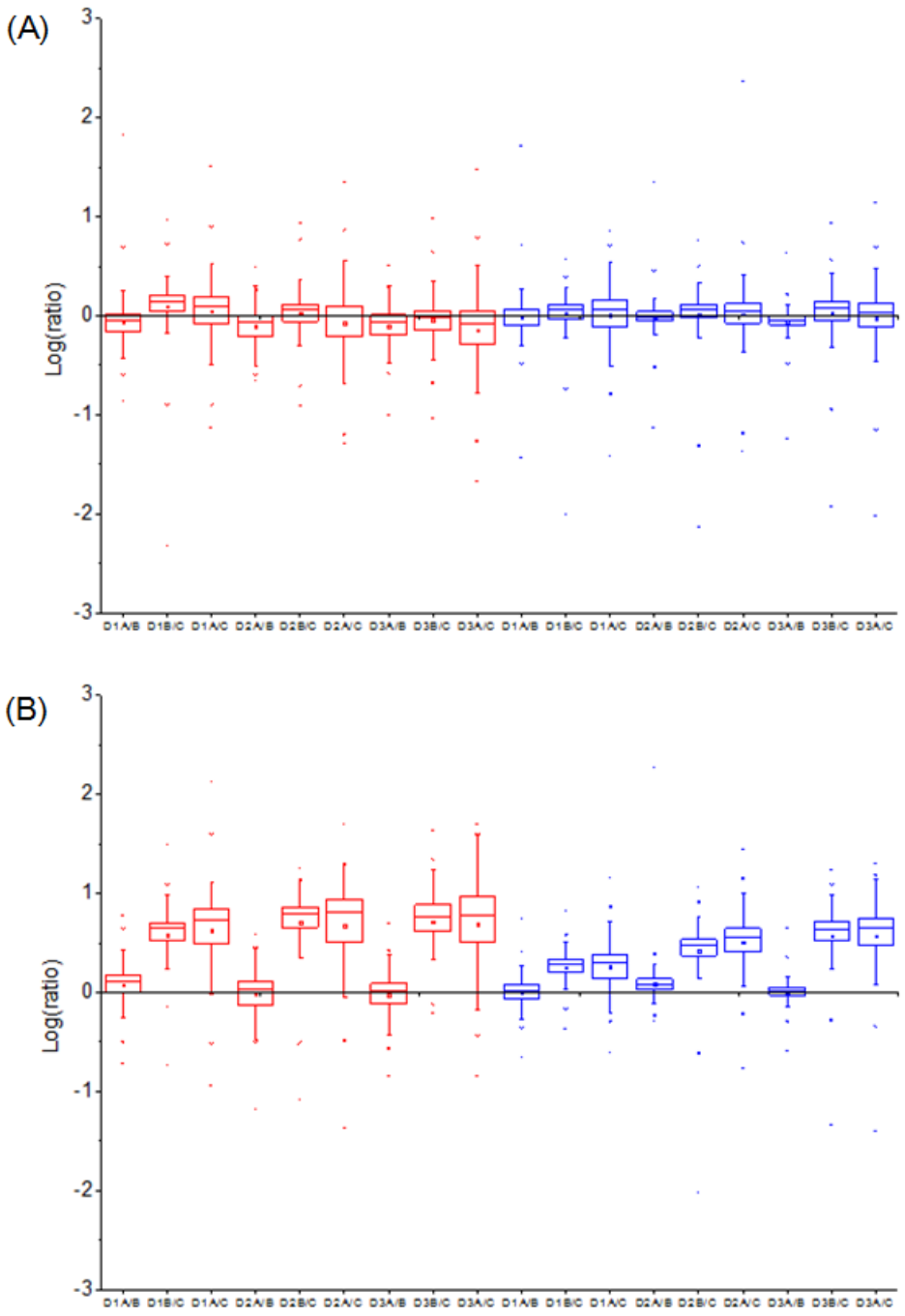


Figure 2.12 Box plots of the log intra-day ratio for individuals 1 (in red) and 2 (in blue): (A) normalized data and (B) un-normalized data. The range of the box is 25 to 75 percentile. The line in the box represents the median value and the mean value is shown as a dot in the box.

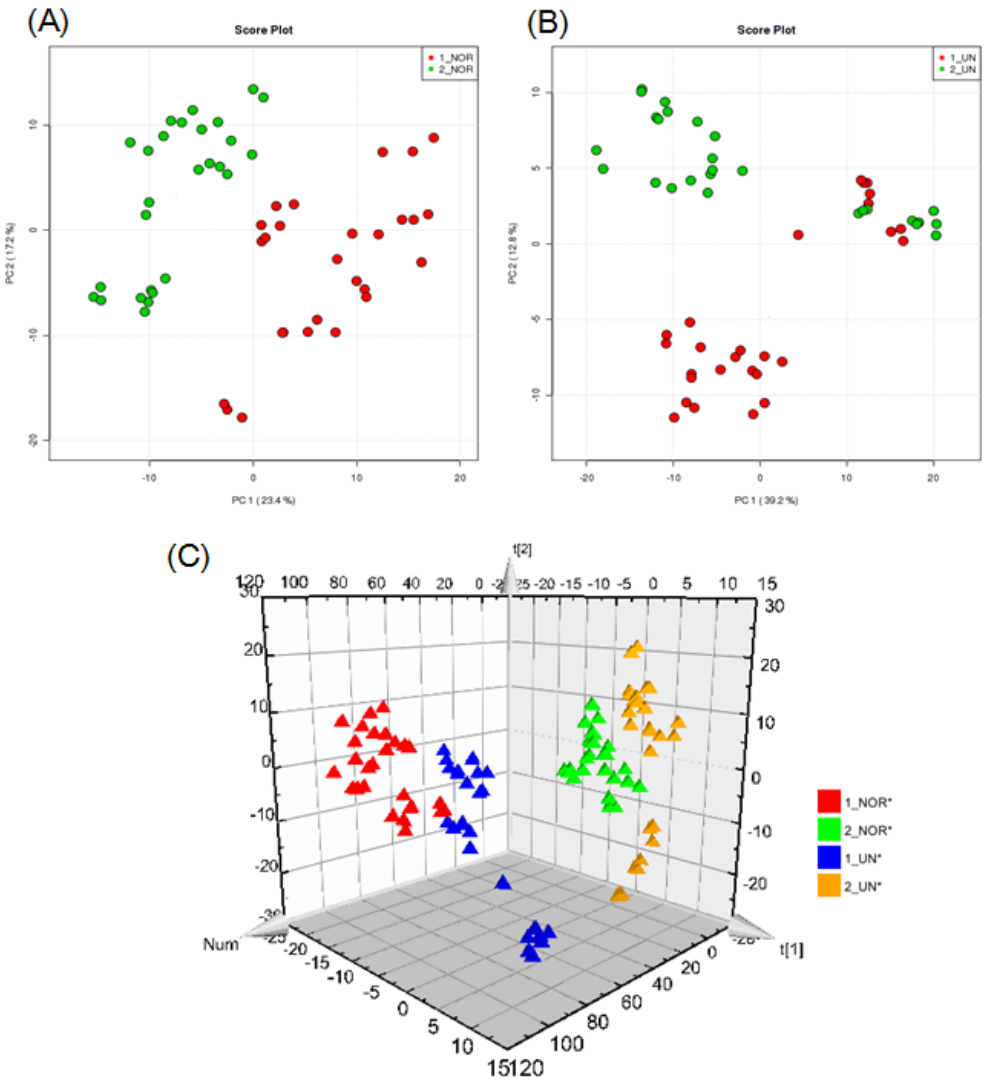


Figure 2.13 PCA plots for (A) normalized urine samples and (B) un-normalized urine samples: 1_NOR (in red), normalized urine samples from individual 1; 2_NOR (in green), normalized urine samples from individual 2; 1_UN (in red), un-normalized urine samples from individual 1; 2_UN (in green), un-normalized urine samples from individual 2. (C) PLS-DA plots of normalized and un-normalized urine samples: 1_NOR (in red), normalized urine samples from individual 1; 2_NOR (in green), normalized urine samples from individual 2; 1_UN (in blue), un-normalized urine samples from individual 1; 2_UN (in yellow), un-normalized urine samples from individual 2.

2.4 Conclusions

A strategy of determining the total concentration of chemically labeled metabolites and its incorporation into a differential isotope labeling LC-MS workflow have been developed and demonstrated for improved relative quantification of urine metabolomes. Quantification of total concentration of labeled metabolites can be done by using a fast step-gradient LC-UV method, in combination with chemical labeling to "unify" the UV absorptivity of diverse metabolites in a metabolome sample. In the present work, dansylation chemistry was used to label the metabolites containing amine and phenol groups, followed by LC-UV detection at 338 nm of the labeled sample in about 2 min. Knowing the total concentration of the labeled metabolite analytes, sample volume or concentration during the metabolome sample workup can be normalized to account for concentration variations in different metabolome samples. In addition, the amount of the sample injected into the mass spectrometric detection system can be optimized and ultimately controlled to maximize the metabolite detectability to improve metabolome coverage. Detailed workflow incorporating this LC-UV metabolite quantification strategy into the LC-MS metabolomics analysis of human urine samples has been discussed. It was demonstrated that concentration normalization among different samples with varying total metabolite concentrations was critical for generating reliable metabolome profiles for comparison. While this present work focuses on urinary metabolome profiling using isotope labeling LC-MS, this dansyl labeling LC-UV method, in principle, should be applicable to any other biological samples and MS platforms where knowing the total concentration of metabolites is desirable for optimal metabolome analysis.

Chapter 3

Dansylation Metabolite Assay: a Simple and Rapid Method for Sample Amount Normalization in Metabolomics

3.1 Introduction

In recent years, liquid chromatography mass spectrometry (LC-MS)-based metabolomics techniques have become a popular choice for the study of biological processes and biomarker discovery.[21,174] In LC-MS-based metabolomics, individual samples from two or more groups are analyzed to study the metabolome profile differences among these samples. Because the total concentration of metabolites can vary significantly from sample to sample,[86,87,128] sample amount normalization to equalize the amounts of individual samples prior to quantitative analysis is required in order to generate accurate and precise results. In this paper, the term "normalization" refers to adjust either the volume or concentration of an individual sample so that the same amount is taken from all the individual samples used in a metabolomics study.

Ideally a good normalization strategy for metabolome analysis should have the following features. Firstly, it should be convenient to perform and should not add too many extra steps or cost to the overall sample processing procedure. Secondly, it is desirable to carry out normalization after the initial sample preparation steps so that any variations during the sample work-up process can be accounted for. Thirdly, it is preferable to perform sample normalization before LC-MS analysis to ensure similar instrumental responses are obtained for all samples. Because of non-uniform responses of individual metabolites, analyte signals obtained from different concentrations of samples cannot be linearly scaled. Thus, using the same concentration of samples for LC-MS will

produce more accurate results. Fourthly, normalization can provide information on the absolute concentration of the samples, relative to a standard.[125] This would allow a user to control the sample injection amount into LC-MS precisely to ensure that an optimal amount is injected. This is important for detecting low concentration metabolites in a sample and avoiding over-injection that can cause problems such as column saturation and sample carry-over from one run to another. Finally, the normalization method should be universally applicable to all biological media.

There are several normalization methods reported in the literature for metabolomics. Normalization to creatinine or osmolality has been used for urine samples.[92,112,162,163] However, in some cases, creatinine itself may vary according to the disease state.[109] Normalization to cell counts, total protein concentration or DNA concentration has been described for cultured cells;[130,134,136] the cell counts or protein/DNA amounts were shown to be useful as sample amount indicators. However, these methods were targeted at specific biological media and cannot be readily extended to other biological samples such as saliva, cerebrospinal fluid (CSF) and fecal samples. For these types of samples, using the same volume or weight does not guarantee that the same total amount of metabolites is taken from each sample. Post-analysis data normalization strategy has also been reported in recent years. The advantages of this strategy are that it is convenient to perform (i.e., no extra experimental procedures required) and widely applicable. Various forms of this strategy has been reported, including normalization to the sum of all metabolite abundance,[175] normalization to the MS "total useful signal",[88] as well as normalization to specific metabolic markers.[128,133] However, the major disadvantage of this strategy is the lack of control of the sample amount injected into a mass spectrometer.

As a result, uneven LC-MS responses may be generated from samples of different concentrations, which can compromise the accuracy and precision of metabolite quantification as well as the metabolome coverage.

We have recently reported a sample normalization method based on the use of LC-UV for quantifying the total concentration of chemically labeled metabolites.[125] In this method, metabolites were first labeled with ^{12}C-dansyl chloride and the absorbance was measured at 338 nm targeting the dansyl chromophore; dansylation labeling is a robust and proven method used in techniques such as LC-UV, LC-fluorescence, and MS.[176-180] A fast step-gradient was applied to allow co-elution of all labeled metabolites, and the total absorption was measured to determine the sample concentration. This method can be readily applied to any type of biological samples and has been demonstrated to be useful as a sample normalization strategy in various applications.[137,138,156,181] However, one major drawback of this method is that it requires an expensive LC-UV system to perform the analysis. The cost per analysis can be relatively high, considering the high consumption of HPLC grade solvents (and columns). Another drawback is that sample throughput is not high. For a large scale metabolomics study, this can be a concern.

In this work, we report a dansylation metabolite assay sample normalization method that measures absorbance of labeled metabolites using a microwell plate reader, instead of an LC-UV system. Microwell plate reader is relatively inexpensive and commonly used in biological laboratories for measuring total concentration of proteins or DNA. The dansylation metabolite assay allows simultaneous measurement of multiple samples within a very short period, which greatly increases the throughput. It also requires less lab consumables and instrument maintenance. Herein we describe the

workflow of this assay and discuss the rationale and procedure of each step involved. We report the performance of the assay in terms of quantification linearity, linear range and reproducibility. We demonstrate an application of this method to quantify the total labeled metabolites in cultured *Escherichia coli (E. coli)* cells for a metabolomics study using a chemical isotope labeling LC-MS platform.

3.2 Experimental

3.2.1 Chemicals and reagents.

Chemicals were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). The isotopic compound used to synthesize ^{13}C-dansyl chloride was purchased from Cambridge Isotope Laboratories (Cambridge, MA, USA). ^{13}C-dansyl chloride was synthesized in our lab as described previously.[146] The BCA assay kit was from Pierce. (Rockford, IL, USA).

3.2.2 Cell culture and harvest conditions.

For quantification linearity test, *E. coli* (ATCC 47076) cells were grown in 75 mL Luria-Bertani (LB) medium (0.5% yeast extract, 1% tryptone, 1% NaCl) at 37°C and 225 rpm in a shaking incubator for ~18 h. The cell density was estimated to be 4.4×10^9 cells/mL by measurement at OD 600. Three portions of 20 mL cell culture were taken and spun at 3900 ×*g* for 10 min at 4°C. For each 20 mL portion, the supernatants were removed and the pellets were washed twice with 20 mL of ice-cold 0.9% NaCl. The washed cell pellets were resuspended in 20 mL of ice-cold 0.9% NaCl and were divided into 8, 4, 2, 1, 0.5 and 0.25 mL aliquots. Each aliquot was spun down at 3900 ×*g* for 10 min at 4°C and the pellets were stored at -80 °C until further use.

For method validation test, 5 individual colonies of *E. coli* (ATCC 47076) were streaked from one LB plate (0.5% yeast extract, 1% tryptone, 1% NaCl, 1.8% agar) onto 5 LB plates with one colony per plate. Similarly, 5 individual colonies of *E. coli* (ATCC 9637) were streaked from one nutrient agar plate (0.2% yeast extract, 0.1% beef extract, 0.5% peptone, 0.5% NaCl, 1.8% agar) onto 5 nutrient agar plates with one colony per plate. All plates were incubated at 37°C for ~18 h. During harvest, different amounts of cells were scraped from the plates and washed twice with 1 mL 0.9% NaCl. The cell pellets were resuspended in 1 mL 0.9% NaCl and were aliquoted into two 350 μL (for absorbance measurement in duplicate) and one 100 μL (for BCA assay) portions. After centrifuging for 10 min at 3900 ×*g* and 4°C, the supernatants were discarded and the pellets were stored at -80°C until further use.

3.2.3 Determination of protein amount.

The protein concentration was determined using a BCA assay kit. 500 μL of 1% SDS in PBS buffer were added into the cell pellets from the 100 μL portion, followed by sonication on ice (2 × 10 s pulses) to disrupt the cells. 25 μL of the lysate was mixed with 200 μL of the working reagent and incubated at 37°C for 30 min. The absorbance measurement was made at 562 nm. A series of diluted bovine serum albumin (BSA) standards were used to establish the calibration curve. The protein amount calculated for this 100 μL portion was then multiplied by a factor of 3.5 to obtain the protein amount in the 350 μL portions.

3.2.4 Metabolite extraction.

Metabolites were extracted by adding 1 mL of 50% MeOH (-20°C) into the cell pellets, followed by ultrasonication-assisted disruption in a Branson ultrasonic cleaner

1510-MT (Branson Ultrasonics Corporation, Danbury, CT, USA) with ice bath for 10 min. The resulting suspensions were centrifuged at 20817 ×*g* for 10 min at 4°C. The extraction procedure was repeated with another 1 mL 50% MeOH and the two supernatants were combined, dried using a SpeedVac and resuspended in 150 μL water. The resulting solutions were used for the labeling step.

3.2.5 Dansylation labeling reaction.

Dansylation reaction was done according to the reported protocol.[28,29] Fifty μL of the extracted solution was mixed with sodium carbonate/sodium bicarbonate buffer and ACN. The solutions were vortexed, spun down and mixed with 50 μL freshly prepared ^{12}C-dansyl chloride solution (18 mg/mL) (for light labeling) or ^{13}C-dansyl chloride solution (18 mg/mL) (for heavy labeling). The reaction was allowed to proceed for 1 hr at 60°C. After 1 hr, NaOH was added to the reaction mixture to quench the excess dansyl chloride. The solution was then incubated at 60°C for another 10 min. Finally, formic acid in 50/50 ACN/H_2O was added to consume excess NaOH and make the solution acidic. For method validation test, 20 μL of each extracted solution was pooled and labeled with ^{13}C-dansyl chloride to serve as the reference sample.

3.2.6 LC-UV quantification of extracted solutions.

The LC-UV method used for determining extraction efficiency was the same as that described previously.[125] Briefly, two μL of the extracted solution was injected onto a Phenomenex Kinetex C18 column (2.1 mm × 5 cm, 1.7 μm particle size, 100 Å pore size). Solvent A was 0.1% (v/v) formic acid in 5% (v/v) acetonitrile, and solvent B was 0.1% (v/v) formic acid in acetonitrile. The gradient started with 0% B for 1 min and was increased to 95% within 0.01 min and hold at 95% B for 1 min. The gradient was

restored to 0% B in 0.5 min and hold for 3.5 min before the next run. The flow rate was 450 μL/min. The wavelength used for the absorbance measurement was 338 nm.

3.2.7 Microplate quantification of labeled metabolites.

To extract the labeled metabolites into the organic layer, three volumes of ethyl acetate were added to the labeled solution followed by vortex for 30 s and centrifugation at 5220 ×*g* for 2 min. For quantification of labeled amino acid standards, 25 μL of the organic layer was pipetted into a Greiner UV-Star 384-well microplate (Monroe, NC, USA) and absorbance measurement was made at 340 nm on a SpectraMax 340PC plate reader from Molecular Devices (Sunnyvale, CA, USA), while 50 μL of the organic layer was added into the plate for quantification of labeled metabolites in *E. coli*.

3.2.8 LC-MS analysis and data processing.

For un-normalized analysis, the ^{12}C- and ^{13}C-dansyl labeled solutions were combined in 1:1 volume ratio. For normalized analysis, the ^{12}C- and ^{13}C-labeled metabolites were mixed in a ratio determined by the quantification results. The combined mixture was analyzed using a Bruker Maxis Impact QTOF mass spectrometer (Billerica, MA, USA) linked to an Agilent 1100 series binary HPLC system (Palo Alto, CA, USA). The LC-MS conditions were the same as those reported.[154] For each LC-MS run, masses were calibrated to the spectrum that contained the dansyl-amine peaks at m/z 242.57160 (two tags two charges), m/z 484.13592 (two tags one charge) and m/z 971.27799 (dimer) using the Data Analysis software and the calibration result was applied to all the other spectra in the same LC-MS run. The resulting MS data were processed using a peak-pair picking software, IsoMS.[154] The level 1 peak pairs were aligned from multiple runs by retention time within 30 s and accurate mass within 5 ppm. Only the common peak-pairs

were retained for statistical analysis. Principle component analysis (PCA) and volcano plot analysis were performed by Metaboanalyst (www.metaboanalyst.ca).[169] The data were mean-centered and auto-scaled (unit variance) prior to analysis.

3.3 Results and Discussion

Figure 3.1 shows the overall workflow of the dansylation metabolite assay (DMA) for normalization of biological samples using a microplate reader. The assay involves three key steps: dansylation labeling of amines and phenols in a sample, extraction of the labeled metabolites using ethyl acetate, and UV absorbance measurement of the organic extract. Based on the measured total concentration of the labeled metabolites, same sample amount is taken from all the samples. To measure the absolute concentration of labeled metabolites relative to a standard in a sample, a calibration curve of a standard (e.g., a mixture of 17 dansyl labeled amino acid standards) can be used. In developing this assay, several experimental parameters and procedures were considered, which are described below. The assay was then applied to a cellular metabolomics study to evaluate its performance.

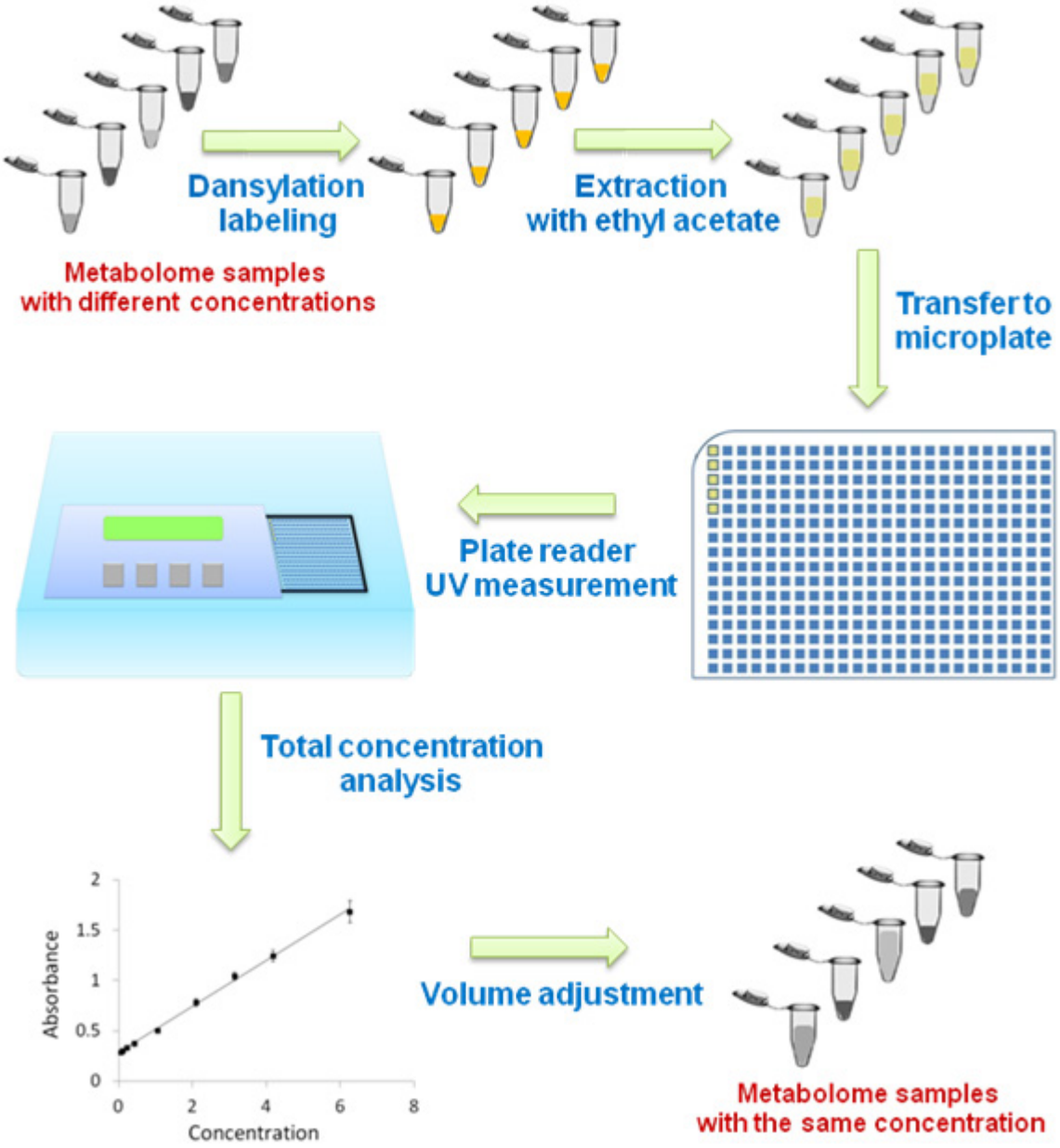


Figure 3.1 Dansylation metabolite assay (DMA) workflow for sample amount normalization.

3.3.1 Extraction of labeled metabolites.

After the dansylation labeling reaction, excess dansyl chloride is quenched with sodium hydroxide to form the hydrolyzed product (Dns-OH). To quantify the labeled

metabolites, Dns-OH has to be removed from the labeled solution prior to quantification. In LC-UV[125], Dns-OH elutes at the high aqueous phase and the labeled metabolites elute at the high organic phase and thus a step-gradient elution can be used to separate them. This LC retention behaviour also indicates that Dns-OH is much more hydrophilic compared to other labeled metabolites, suggesting the possibility of their separation by using a simple liquid-liquid extraction (LLE) method. We evaluated the performance of LLE with ethyl acetate, a commonly used extraction solvent with moderate insolubility, relatively high boiling point, and low toxicity.

Figure 3.2A compares the step-gradient LC-UV chromatograms of a solution of 17 dansyl labeled amino acid standards (17-Dns-aas) before and after extraction. The early eluting peak corresponds to Dns-OH and the peaks between 1.4 min to 2.0 min are from the labeled amino acids. Comparing the areas of the 17-Dns-aas peaks, it can be seen that after extraction the majority of Dns-aas has been removed from the aqueous solution, indicating a successful separation of the quenched reagent from the labeled metabolites by LLE. The Dns-OH peak is higher after extraction, likely due to the fact that the final labeled solution was in 50/50 ACN/H_2O, and during the extraction some of the ACN was distributed into the organic layers, making the remaining aqueous solution more concentrated. The LC-UV chromatogram of the organic extract is shown in Figure 3.2B. This chromatogram confirms the presence of the labeled metabolites with very little Dns-OH in the organic extract. The extraction efficiency of the labeled metabolites was calculated as the % peak area difference before and after extraction in the aqueous layer. In the example shown in Figure 3.2A, the extraction efficiency was found to be 91%.

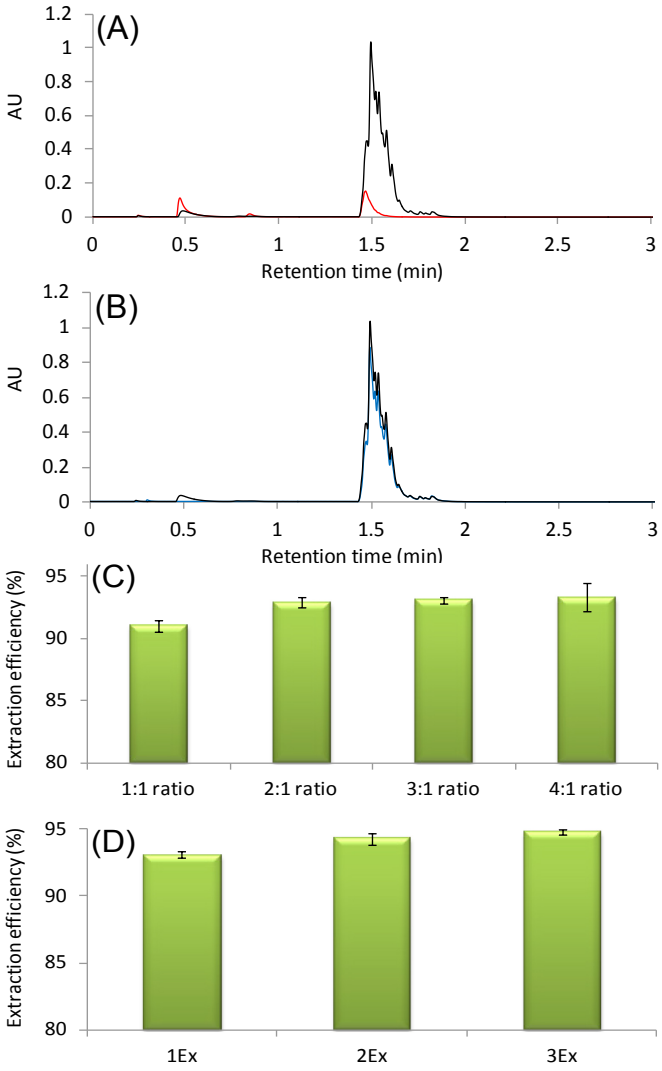


Figure 3.2. Step-gradient LC-UV chromatograms of a mixture of 17 dansyl labeled amino acid standards (17-Dns-aas) before and after ethyl acetate extraction: (A) comparison between un-extracted solution and the aqueous layer after extraction and (B) comparison between un-extracted solution and the ethyl acetate layer after extraction. Comparison of extraction efficiencies under different extraction conditions: (C) with different ethyl acetate to aqueous solution ratios and (D) with different numbers of extractions while the total volume of ethyl acetate was held constant (1Ex, 2Ex and 3Ex refer to extraction once, twice and three times, respectively).

The extraction condition was then optimized using different organic-to-aqueous solvent ratios and different numbers of extractions. We compared extraction with four different organic-to-aqueous solvent ratios: 1:1, 2:1, 3:1 and 4:1 and the results are shown in Figure 3.2C. The extraction efficiency increases from the 1:1 solvent to the 2:1 solvent, but there is no significant difference for the 2:1, 3:1 and 4:1 solvents. In the 4:1 solvent, the standard deviation is larger than the other solvents for an unknown reason, but is still less than ±2%. For a microplate, a larger working volume is desirable to increase the absorption signals and generate more reproducible results. Therefore, a higher organic-to-aqueous ratio is preferred. However, if the ratio is too high, the metabolites may be too diluted and organic solvent consumption is also increased. We chose the 3:1 solvent to balance these two factors. The extraction efficiency also increases with the number of extraction if the total organic volume is held constant, as shown in Figure 3.2D. However, the sample throughput would be reduced considerably with each additional extraction. As there was only about 1-2% increase in extraction efficiency from 1 to 3 extractions, we decided to perform only one extraction in order to maintain a high throughput. As a result, the final extraction protocol involves a one-time extraction with 3 volumes of ethyl acetate and 1 volume of labeled solution, which gives an extraction efficiency of ~93% for the labeled amino acid standards. It should be noted that in the dansyl metabolite assay the labeled metabolites are dissolved in ethyl acetate, while in LC-UV they are present in the elution solvent composed of mainly ACN, water and FA. These solvent systems do not absorb at 338 nm (for LC-UV) or 340 nm (for microplate UV measurement) so there is no solvent interference.

3.3.2 Absorbance measurement.

The absorbance of dansyl labeled metabolites was measured at 340 nm, which is the smallest detection wavelength available on the SpectraMax 340PC plate reader. This wavelength is close to 338 nm used in the LC-UV sample normalization method.[125] Figure 3.3 shows the absorption spectrum of 17-Dns-aas; the absorbance starts to decrease at higher wavelengths than 340 nm. The measurement was done using a UV-transparent 384-well plate with low background absorbance at 340 nm. This plate also shows good resistance to ethyl acetate with an optimal working volume of 15-110 μL, which matches with our applications.

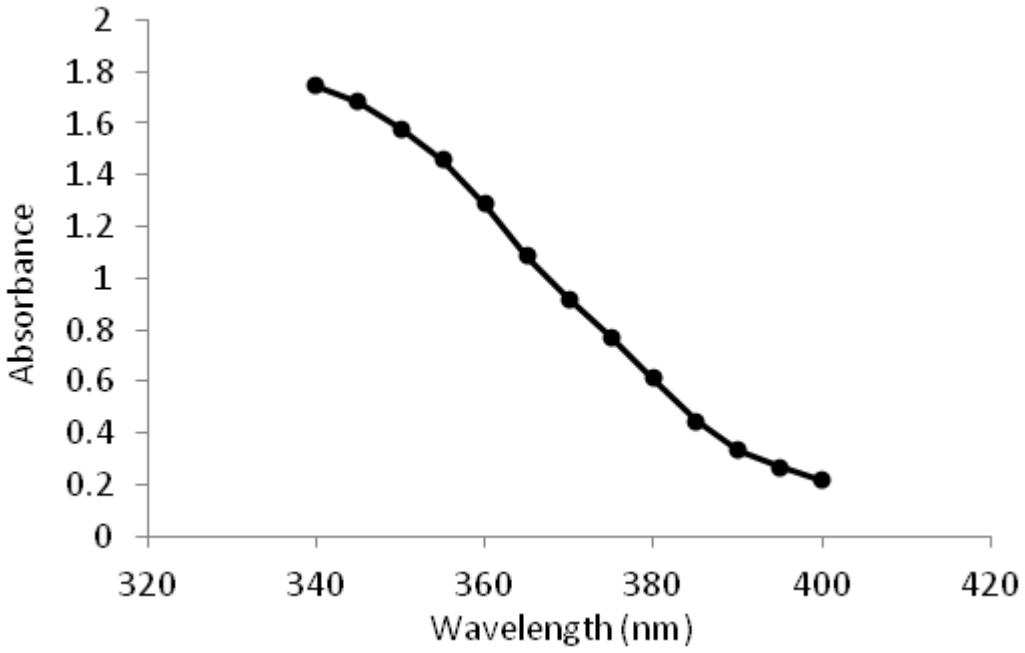


Figure 3.3. Absorption spectrum of a mixture of 17 dansyl labeled amino acid standards (17-Dns-aas) (6.3 mM) from 340 to 400 nm.

We prepared a series of diluted 17-aas solutions and labeled them separately to evaluate the linearity and linear range of this quantification method. In this case, 10 μL of labeled amino acid solutions were extracted with 30 μL of ethyl acetate, and 25 μL of each extracted solution was added into the 384-well plate. Figure 3.4 shows the

calibration curve generated which is linear from 0.04 mM to 6.25 mM with good correlation (R^2 = 0.9981); note that the three points at the lower end departs from this linear curve to some extent. As in the case of our LC-UV work, the upper end of this linear curve was limited by the decreased dansylation labeling efficiency at higher analyte concentrations. Increasing the reagent amount relative to the analyte amount can extend the upper limit, but this is not needed as the upper limit shown in Figure 3.4 is already sufficient for normalization of metabolomic samples. The lower end was limited by the formation of side products that produced the background signals. We note that the lower end of this linear curve (0.04 mM) was higher than that could be achieved with LC-UV (0.02 mM), since the sensitivity (slope) of the current method was lower; the absorbance of 0.02 mM standard was close to the absorbance of a labeled blank solution. Nevertheless, the linear range of this calibration curve should be sufficient for quantification of most real biological samples. Decreasing the detection limit of the assay is not necessary, as the total concentration of labeled metabolites in most metabolomic samples would be in the range of high μM to mM. In addition, the sensitivity can be improved by increasing the path length (i.e., increasing the volume of solution added into a well). From our working experience with different types of samples, we suggest that the volume of a solution used for the microplate reader measurement be adjusted, depending on the type of biological samples analyzed. For example, we found that 25 μL was generally sufficient for quantification of human urine samples, while a larger volume (e.g., 50 μL) was optimal for analyzing the extracts of bacterial cells in which the total metabolite concentration is lower than urine.

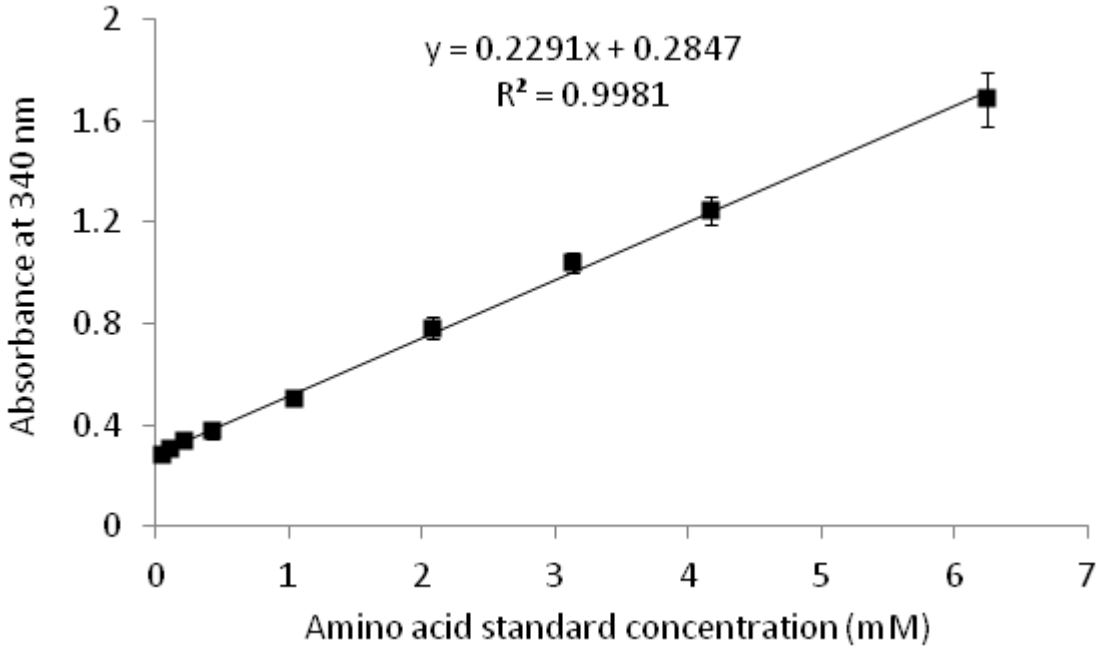


Figure 3.4 Calibration curve of a mixture of 17 dansyl labeled amino acid standards (17-Dns-aas) from triplicate labeling experiments at each concentration. Labeled metabolites were extracted with three volumes of ethyl acetate. Absorbance was measured at 340 nm with 25 µL of the organic extract placed in a microplate well.

Method reproducibility was evaluated in terms of experimental reproducibility and run-to-run reproducibility. Table 3.1 shows a summary of the results obtained. In this case, experimental reproducibility was determined from three experimental replicates that should account for variations during labeling, extraction, solution transfer and absorbance reading. As Table 3.1 shows, the %RSD values at different 17-Dns-aas concentrations are all below 10%, indicating a good experimental reproducibility. Run-to-run reproducibility was determined by measuring absorbance at 0, 5 and 10 min after the samples were added into the plate. In practice, there will be a time interval between the first added sample and the last added one. Although ethyl acetate has a relatively high boiling point, compared to other organic solvents, some evaporation might still occur

during this time period. Therefore, we evaluated the run-to-run reproducibility within a 10-min period to see whether solvent evaporation would affect the measurement. As shown in Table 3.1, the %RSD values within a 10-min period were only 1-3%, indicating that solvent evaporation would not cause a problem for absorbance measurement of samples added within a 10-min interval. This usually allows approximately 30 samples to be processed and measured at the same time. Using a multichannel pipette should increase the number of samples handled in 10-min interval.

Table 3.1. Summary of reproducibility of absorbance measurements.

17-dns-aas concentration (mM)	Run-to-run reproducibility (%RSD)	Experimental reproducibility (%RSD)
0.042	2.5	5.4
0.10	3.5	1.9
0.21	2.4	3.5
0.42	3.0	6.7
1.0	2.0	1.1
2.1	1.5	5.2
3.1	2.4	3.9
4.2	1.2	4.4
6.3	1.3	6.4

3.3.3 Quantification of labeled metabolites in *E. coli*.

As an example of applications of DMS for sample amount normalization, we applied this method for metabolomic profiling of *E. coli* cells. We first determined the linearity between the measured absorbance and the cell amount. *E. coli* cells from the same culture medium were aliquoted into 0.25, 0.5, 1, 2, 4 and 8 mL portions, and metabolites in each portion were extracted and labeled under the same conditions. A good linear relationship ($R^2 > 0.99$) was observed between the absorbance of the labeled metabolites and the volume of the cell suspensions (Figure 3.5) (note that the last point at

the lower end departs from the linear curve to some extent), indicating that the total amount of extracted metabolites correlates very well with the cell amount. Thus, in real world applications where the number of cells is not known, the DMA quantification results can serve as a surrogate of the cell amount. To validate this, we compared the metabolite measurement results obtained by our assay to the protein amounts measured using a well-established BCA quantitative method.

In this work, two different *E. coli* strains (ATCC 47076 and ATCC 9637) were each grown on five agar plates. During harvest, different amounts of cells were collected from each plate. Because of the small size of *E. coli* cells, cell counting using a microscope is difficult. Instead, we measured the protein concentrations as the cell amount indicator and then compared the absorbance values of the labeled metabolites measured by the dansylation assay with the measured protein amounts. The results are shown in Figure 3.6A, B for *E. Coli* ATCC 47076 and ATCC 9637, respectively. Figure 3.6 shows that there is a good linear correlation between the absorbance and the protein amount for both strains. It is interesting to note that the linear regression equations for these two strains were very close to each other, indicating that the two *E. coli* strains have similar amounts of labeled metabolites at the same protein levels.

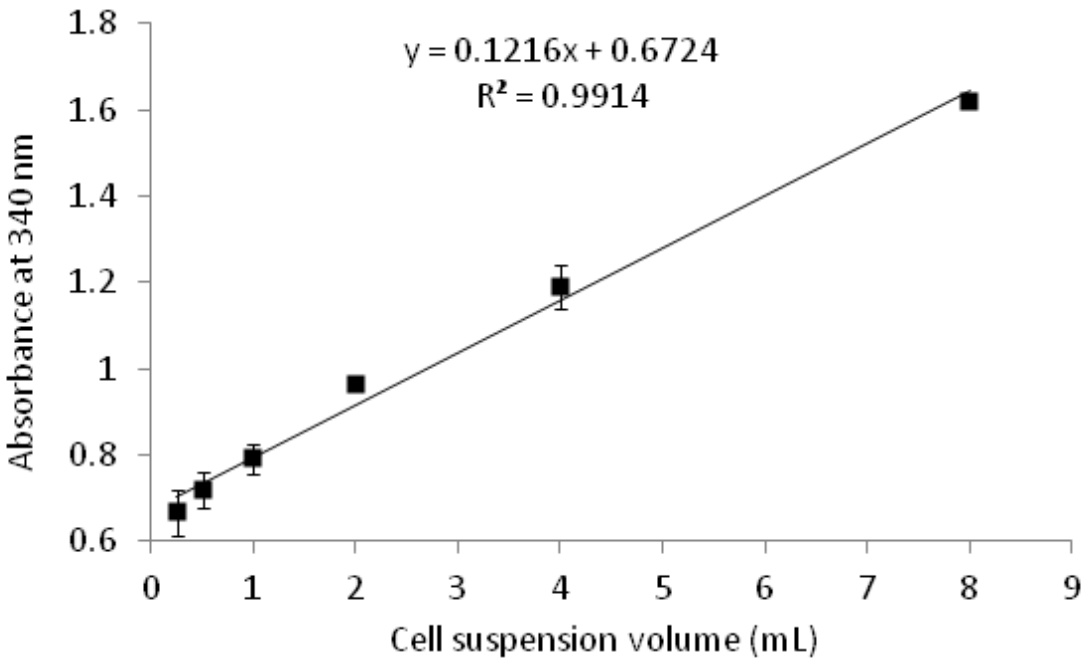


Figure 3.5 Determination of linearity between absorbance of labeled metabolites and *E. coli* cell amount. Different cell suspension volumes were taken from the same culture and thus the cell amount should be proportional to the suspension volume. Labeled metabolites were extracted with three volumes of ethyl acetate. Absorbance was measured at 340 nm with 50 μL of the organic extract placed in a microplate well. The blank absorbance was 0.6201, which was not subtracted out in the total absorbance shown in x-axis.

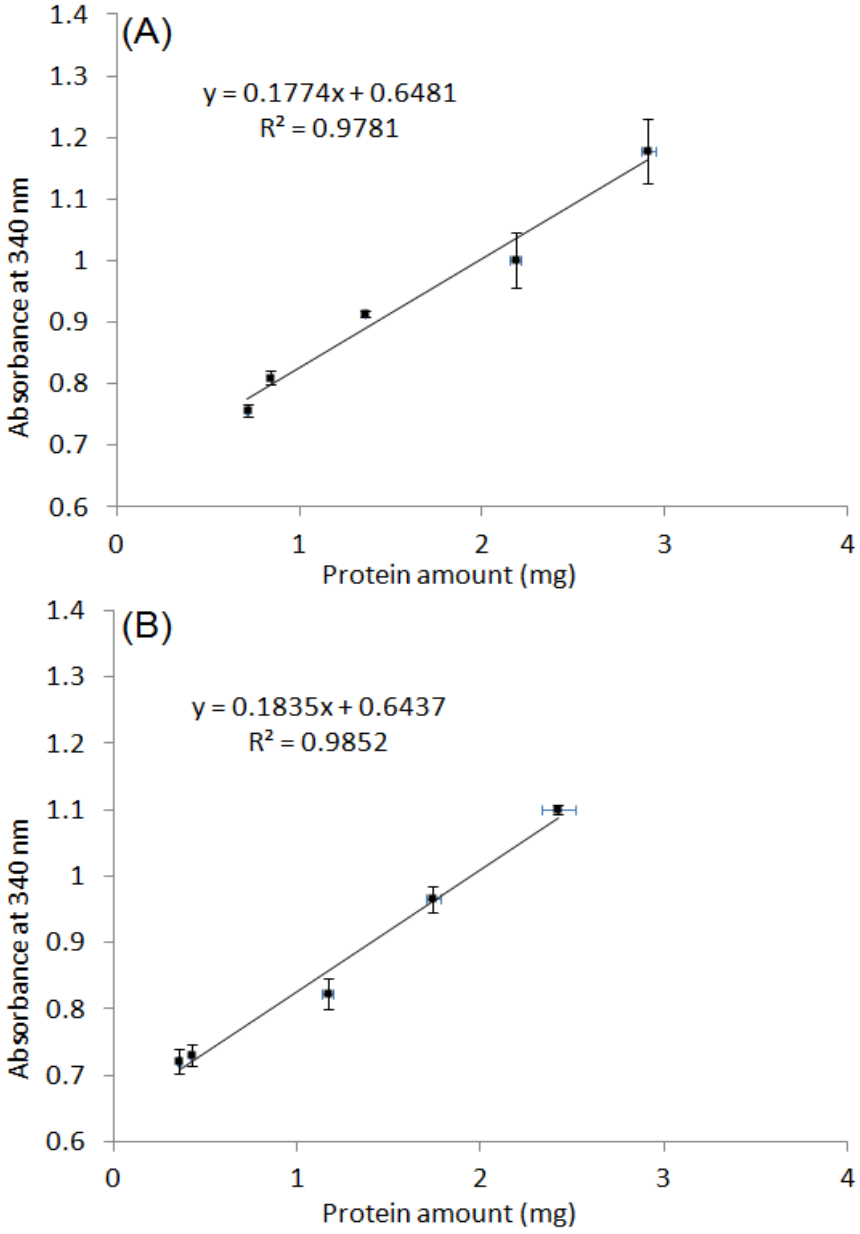


Figure 3.6. Determination of linear correlation between absorbance of labeled metabolites and protein amounts for (A) *E. coli* strain ATCC 47076 and (B) *E. coli* strain ATCC 9637. Labeled metabolites were extracted with three volumes of ethyl acetate. Absorbance was measured at 340 nm with 50 µL of the organic extract placed in a microplate well. The

blank absorbance in (B) was 0.6201, which was not subtracted out in the total absorbance shown in x-axis.

3.3.4 Sample amount normalization in *E. coli*.

To evaluate whether the labeled metabolite absorbance measurement could serve as a valid normalization strategy for metabolomics, we compared the metabolomic profiles of the two *E. coli* strains with and without normalization. A differential isotope dansyl labeling LC-MS platform[137,146,154] was used to profile the amine/phenol submetabolome differences of the two strains. The un-normalized set of samples was prepared by mixing equal volume of each ^{12}C-labeled sample with a ^{13}C-labeled pool, while the second set of samples was normalized based on the absorbance values of ^{12}C-labeled individual samples. It is important to note that there is background absorbance from a labeled solution, which is likely caused by the presence of by-products of the labeling reagent (e.g., dimerization). Therefore, a blank subtraction step must be carried out first, which can be done using a pre-determined calibration curve, where the y-intercept reflects the background absorption (see Table 3.2). Column 2 in Table 3.2 shows the results of the absorbance measurement from the ^{12}C-dansylation metabolite assay. The absorbance values after blank subtraction is shown in Column 3. To mix an equal amount of ^{12}C-labeled sample with the ^{13}C-labeled pooled sample, the volume of an individual sample was determined by the absorbance value after blank subtraction. These volumes are shown in Column 4.

Table 3.2. Normalization of the sample amounts from two *E. coli* strains

E. coli* sample	**Before blank subtraction**	**After blank subtraction****	**Normalized volume (µL)**
47076_1	0.914	0.294	14.5
47076_2	0.810	0.190	22.5
47076_3	0.757	0.137	31.2
47076_4	1.002	0.382	11.2
47076_5	1.179	0.559	7.64
9637_1	0.721	0.101	42.3
9637_2	0.730	0.110	38.8
9637_3	0.823	0.203	21.0
9637_4	0.965	0.345	12.4
9637_5	1.100	0.480	8.88
Pooled sample	0.904	0.284	15.0

*The number after the strain number indicates the culture plate number.

**Blank absorbance was determined from the y-intercept of a pre-determined calibration curve: $y = 2.154x + 0.6201$.

3.3.5 Metabolome comparison of two *E. coli* strains.

In comparative metabolomics studies, it is common to perform statistical analysis of the metabolome data to identify metabolites that are differentially expressed in two or more groups. For example, it is often of interest to investigate cellular metabolome changes under different culture conditions or specific treatments. One prerequisite for a fair comparison between different sample groups is that the variation within each group should be small.[128] In this example of comparing the metabolome profiles of two *E. coli* strains, we first applied a PCA model to the two metabolome datasets (Figure 3.7). As Figure 3.7A shows, for the un-normalized sample dataset, separation between the two strains is attributed to the second principal component (19.7% of the total variation), while the most important variation reflected by the first principal component is the cell amount (65.8% of the total variation). In contract, for the normalized sample dataset (Figure 3.7B), the two strains are clearly separated on the first principal component

which represents 43.2% of the total variation, indicating that metabolic difference between the two strains is the major variation in the dataset. These PCA score plots illustrate that our sample normalization strategy is effective in reducing the artificial variation caused by different sample amounts used in the un-normalized dataset.

We further analyzed the two datasets using the volcano plot statistical analysis. Figure 3.8 shows the volcano plots generated by examining the differentiating metabolites with the criteria of fold change (FC) > 2 or FC < 0.5, and $p < 0.01$. Only 29 metabolites were found to be differentially expressed using these criteria in the un-normalized samples, while there were 145 metabolites at significantly different levels in the normalized dataset. Much lower identification rate in the un-normalized dataset is mainly due to the large variations within each strain caused by the sample amount differences. Figure 3.9 shows the %RSD values of relative metabolite quantities measured from multiple samples within each strain. In the normalized data, the %RSD values were lowered by almost 50%. These results again confirm that the variations within each strain have been reduced through this normalization process to allow identification of a larger number of differentiating metabolites between the two strains.

In Figure 3.8A, there are fewer metabolites on the left-hand-side of the volcano plot (FC < 0.5) in the un-normalized set. This is probably because the average cell amount is higher in the ATCC 47076 strains (Figure 3.6), which could disguise metabolites expressed at lower levels in this strain. Therefore, in addition to perform normalization within one group, it is also important to adjust the sample amount of all comparative groups to the same level in order to avoid any bias created by the amount differences between the sample groups. In our normalization strategy, the total metabolite

amounts in all samples were adjusted to be equal to that of the pooled reference sample. As a result, a more even distribution of differentiating metabolites was observed in the normalized set (Figure 3.8B).

The above example demonstrates that sample amount normalization is very important for comparative metabolomics and the dansylation metabolite assay can be used as a simple and rapid normalization method. It should also be noted that this assay is not destructive if the dansyl labeled sample is used for metabolome profiling, as in the case of using differential dansyl labeling for profiling the amine/phenol submetabolome. After metabolite quantification is finished, the solution can be re-collected for further analysis. Although this assay only measures the total concentration of labeled amine/phenol submetabolome, the large diversity of amines and phenols in a metabolome sample ensures that the measured concentration is a good representation of the total metabolome concentration. This assay should be applicable to many types of biological samples including biofluids. One potential limitation of this assay is that the UV measurement is done at 340 nm and thus if a sample contains high concentrations of chemicals that absorb at 340 nm, interference from these chemicals may cause errors in the measurement of the labeled metabolites. However, considering that the total concentration of labeled metabolites is in the mM range for many samples, lower concentrations of 340-nm absorbing compounds present in a sample should not affect the quantitative results significantly.

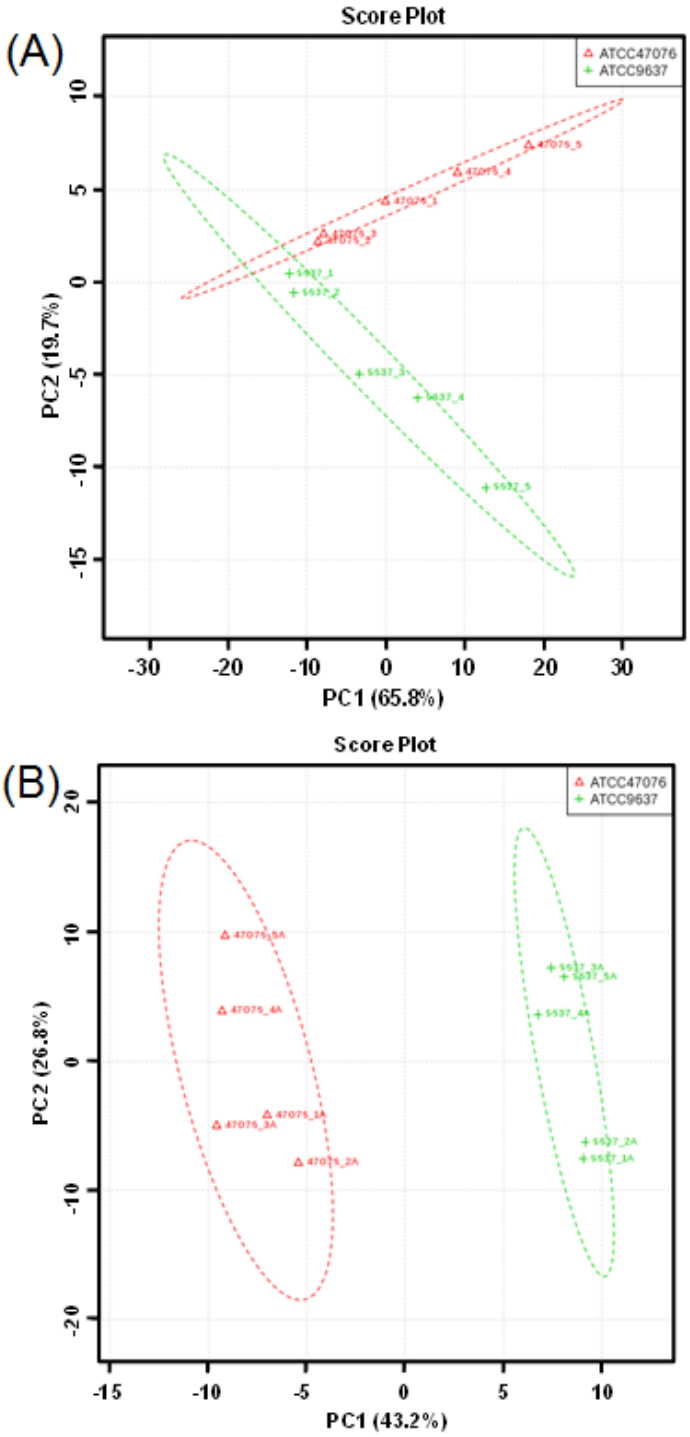


Figure 3.7 PCA score plots for (A) un-normalized dataset and (B) normalized dataset. Red triangles and green crosses represent *E. coli* strain ATCC 47076 and 9637 harvested from 5 different agar plates, respectively.

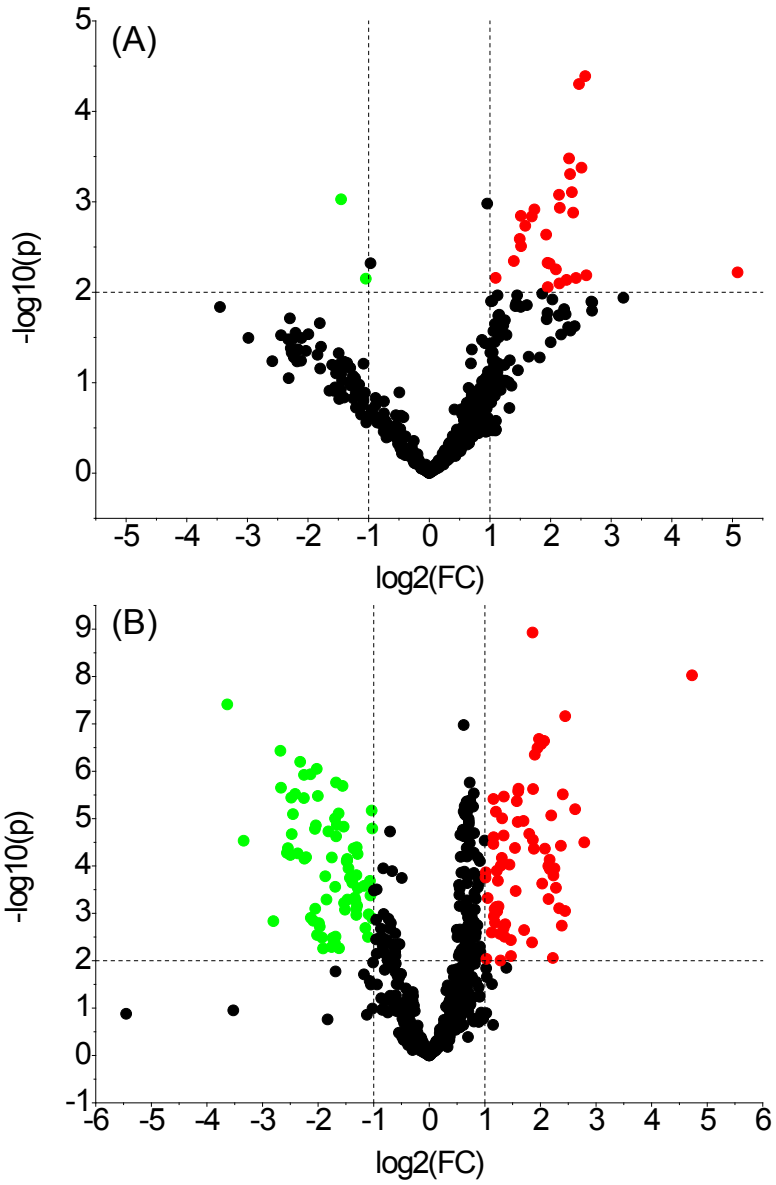


Figure 3.8. Volcano plots for metabolome comparison between two strains from (A) un-normalized dataset and (B) normalized dataset. Fold change (FC) was expressed as the average peak pair ratio in strain ATCC 47076 over that in strain ATCC 9637. The horizontal dotted line represents cutoff at p = 0.01 and the vertical dotted line represents cutoff at FC = 2 and FC = 0.5.

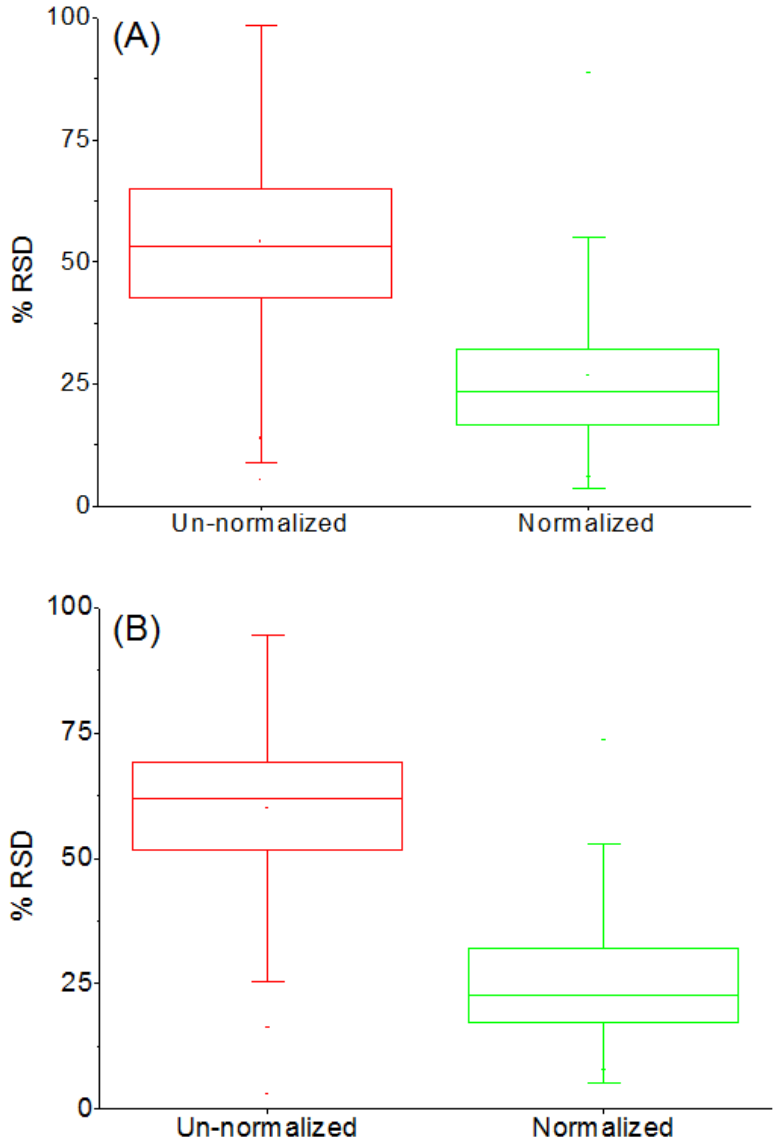


Figure 3.9 Box plot showing the %RSD of un-normalized and normalized data in (A) ATCC 47076 and (B) ATCC 9637. The range of the box is 25 to 75 percentile. The line in the box represents the median value and the mean value is shown as a dot in the box.

3.4 Conclusions

We have developed a dansylation metabolite assay for sample amount normalization in quantitative metabolomics. It uses a microplate reader to measure the absorbance of labeled metabolites at 340 nm in a sample after dansylation labeling of the metabolites and ethyl acetate extraction to remove the quenched excess dansyl reagent. This method is simple, rapid and easy to implement. We envisage the application of this dansylation metabolite assay, analogy to the widely used BCA assay in quantitative proteomics, as a robust sample normalization method in metabolomics.

Chapter 4

Development of Isotope Labeling LC-MS for Metabolic Profiling of Bacterial Cells and Its Application for Bacterial Differentiation

4.1 Introduction

Metabolomics refers to the comprehensive detection and quantification of all metabolites present in a given biological sample, such as urine, plasma, tissues and cells.[2] Since metabolites are the end product of all regulations, the study of the metabolite levels can provide valuable and complementary information to the genomics, transcriptomics and proteomics data.[182-185] In recent years, the study of microbial metabolomics has received growing research interest because of its potential applications in a wide range of microbial research fields including metabolic engineering.[186,187] There are at least two major research areas in microbial metabolomics.[131] The first one is to investigate changes of the metabolic profile under different environmental conditions or look for the key metabolic changes in a mutant strain.[188] The second one is to generate a metabolic fingerprint of various bacterial species for the purpose of microorganism identification or differentiation.[131,189-191] In both cases, a robust and sensitive method capable of detecting and quantifying a large number of metabolites is desirable.

The major analytical platforms currently used for metabolomics studies include nuclear magnetic resonance (NMR),[192-194] and mass spectrometry (MS) coupled with gas chromatography (GC),[195,196] liquid chromatography (LC)[197-200] and capillary electrophoresis (CE).[201-203] Among these, LC-MS, particularly reversed phase LC-MS (RPLC-MS) has been widely used because of its high sensitivity, high resolution, and wide metabolite coverage. However, the performance of RPLC-MS for detecting very

polar and ionic metabolites, which are highly abundant in biological samples, is not as good. To address this problem, our group has reported a $^{12}C/^{13}C$-dansylation labeling technique for the analysis of the amine- and phenol-containing sub-metabolome; dansyl labeling allows separation of polar and ionic metabolites on a RP column while providing a signal enhancement of 10- to 1000-fold.[146] This method enables detection of hundreds to over a thousand of metabolites using one dimensional LC-MS and has been shown to be useful for metabolic profiling of biofluids such as urine,[146] cerebrospinal fluid[49] and saliva.[156]

However, differing from biofluid analysis, cellular metabolome profiling is a much more challenging task. Prior to metabolite analysis by NMR or MS, several sample preparation processes are needed, including (1) separation of cells from the growth medium; (2) washing procedure to remove any interfering compounds from the cell surface; (3) rapid quenching to stop cellular activity; (4) extraction of intracellular metabolites and (5) disruption procedure to enhance the extraction process.[186] A number of studies have been reported with a focus on evaluating the sample preparation methods for microbial metabolome analysis.[131,200,204-208] The cells are usually separated from the growth medium by centrifugation or filtration, or scraped off the plates if grown in agar medium. The choice of an improper washing solution (e.g., distilled water) might result in leakage and loss of intracellular metabolites, and it was recommended to use a washing solution with salt content matching the ionic strength of the medium.[208,209] The quenching step is usually applied to rapidly arrest metabolic activity and is particularly important for detecting metabolites with high turnover rates and in flux analysis.[210] The most commonly used quenching agent is -40°C 60% (v/v) MeOH/H_2O,[205] either alone or with

the addition of a buffer such as HEPES to reduce leakage.[209,211] Following quenching of the metabolism, the intracellular metabolites need to be extracted. Various extraction solvents have been evaluated, including cold and hot organic solvents with different percentages of water, acid (e.g., perchloric acid), and base (e.g., potassium hydroxide).[132,182,204-206,212,213] It was found that cold organic solvents, such as methanol and acetonitrile (50-100% v/v in H_2O), generally show less bias and less degradation products. Finally, a disruption procedure can be applied after addition of the extraction solvent to facilitate cell breakage and metabolite extraction, although this step is often omitted in many works. The most commonly employed disruption method is freeze-thaw cycle,[132,204] but other mechanical methods, such as mill/beads, microwave and ultra-sound, have also been used with varying degree of success.[131,214]

It is clear that a variety of sample preparation methods may be used for microbial metabolome analysis. However, the analytical performance of each method is very likely dependent on the type of metabolites analyzed and the detection method used. For example, evaluating the performance of a method by analyzing a few metabolites does not provide a complete assessment of the method for profiling a large number of metabolites present in a metabolome sample.

In this work, we describe a method based on differential isotope dansylation labeling LC-MS, in combination with a fast step-gradient LC-UV quantification method for sample amount normalization,[125] for microbial metabolome profiling. This method allows the detection and quantification of thousands of putative metabolites from a bacterium with high precision and accuracy, compared to current methods of detecting less than 300 putative metabolites in general. This high level of detectability allows us to

evaluate various sample preparation methods more completely in order to determine the optimal conditions for microbial metabolome profiling. As an example of potential applications of this method for microbial metabolomics, we demonstrate the ability of this method for differentiation of three model organisms, *Escherichia coli* (*E. coli*), *Bacillus subtilis* (*B. subtilis*) and *Bacillus megaterium* (*B. megaterium*), as well as identification of the important metabolites that contribute to the differentiation. The possibility of employing this isotope labeling LC-MS method for identifying bacteria in clinical samples is assessed by analyzing bacteria spiked in human urine samples. To our knowledge, most reported studies of using metabolomics for bacterial identification were based on the use of GC-MS.[190,191,215]

4.2 Experimental

4.2.1 Chemicals and Reagents.

^{12}C-dansyl chloride (DnsCl) and formic acid were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). The isotopic compound used to synthesize ^{13}C-dansyl chloride was purchased from Cambridge Isotope Laboratories (Cambridge, MA, USA). ^{13}C-dansyl chloride was synthesized as described previously[146] and the other chemicals used to synthesize this isotope reagent were purchased from Sigma-Aldrich. LC-MS grade water, acetonitrile (ACN) and methanol (MeOH) were purchased from Thermo Fisher Scientific (Edmonton, AB, Canada).

4.2.2 Cell Culture and Harvest Conditions.

For the method optimization work, *E. coli* (ATCC 47076) cells were grown in nutrient broth (0.3% beef extract, 0.5% peptone) at 37°C and 225 rpm in a shaking incubator for ~24 h. All cultures were harvested at OD600 of 1.5 and were spun at

4640×*g* for 10 min at 4°C. The pellets were resuspended in 1 mL of ice-cold 0.9% NaCl, and spun in an Eppendorf 5415C microcentrifuge at 16 000×*g* for 1 min at 4°C. The cell pellets were flash frozen in a dry ice/ethanol bath and stored at -80 °C until further use. For the cell differentiation study, *E. coli* (ATCC 47076) cells were grown at 37°C on nutrient agar plates (0.3% beef extract, 0.5% peptone, 1.8% agar) for ~24 h. *B. subtilis* (ATCC 9372) and *B. megaterium* (ATCC 14581) cells were grown at 30°C on the same nutrient agar plates for ~24 h. For the urine experiments, *E. coli* was grown overnight at 37°C in LB medium (1% tryptone, 0.5% yeast extract) to an OD600 of ~4 and then diluted to 1×10^5 cells/mL in each urine sample. Ten µL of the spiked urine was diluted to 100 µL in water and then spread onto the nutrient agar plate and incubated overnight at 37°C. Cells from the plates were scraped into 1 mL 0.9% NaCl and rapidly centrifuged in an Eppendorf 5415C microcentrifuge at 16 000×*g* for 1 min. The cell pellets were resuspended in 1 mL 0.9% NaCl and spun again. The final cell pellets were stored at -80°C until further use.

4.2.3 Metabolite Extraction.

The performance of three solvent systems were evaluated in this work: 50/50 MeOH/H_2O (MeOH), 50/50 ACN/H_2O (ACN), and 40/40/20 MeOH/ACN/H_2O (MAW). Each solvent extraction experiment was carried out in triplicate. For each extraction, the cell pellets were resuspended in 1 mL of the corresponding solvent system (0°C for ACN and -20°C for MeOH and MAW), and disrupted using ultrasonication, as described below. The resulting suspensions were centrifuged at 16 000×*g* for 10 min. The supernatants were dried using a SpeedVac and resuspended in 250 mL water. The resulting solutions were used for the labeling step.

Three cell disruption methods were compared, together with a control experiment (i.e., no disruption, CT). They were ultrasonication (SN), microwave (Mic) and freeze-thaw cycle (FT). Each disruption method, as well as the control experiment, was carried out in triplicate. Prior to disruption, the cell pellets were resuspended in 1 mL of 50/50 MeOH/H_2O. For the ultrasonication-assisted extraction, the cell suspensions were placed in a Branson ultrasonic cleaner 1510-MT (Branson Ultrasonics Corporation, Danbury, CT, USA) with ice bath for 10 min. Microwave-assisted extraction was performed for 10 min at 240 W power with a 1200 W microwave oven (Panasonic, Toronto, ON, Canada). For freeze-thaw cycle extraction, the cell suspensions were rapidly dipped into liquid nitrogen bath for 30 sec and thawed on ice for 1 min. This procedure was repeated 3 times. All of the resulting suspensions were centrifuged at 16 000×*g* for 10 min. The supernatants were dried and resuspended in 250 mL water.

4.2.4 Dansylation Labeling Reaction.

Fifty μL of the extracted solution was mixed with sodium carbonate/sodium bicarbonate buffer and ACN. The solutions were vortexed, spun down and mixed with 50 μL freshly prepared ^{12}C-dansyl chloride solution (18 mg/mL) (for light labeling) or ^{13}C-dansyl chloride solution (18 mg/mL) (for heavy labeling). The reaction was allowed to proceed for 1 hr at 60°C. After 1 hr, NaOH was added to the reaction mixture to quench the excess dansyl chloride. The solution was then incubated at 60°C for another 10 min. Finally, formic acid in 50/50 ACN/H_2O was added to consume excess NaOH and to make the solution acidic.

4.2.5 LC-UV Quantification.

For the cell differentiation experiment, an LC-UV quantification step was carried out prior to mass analysis in order to control the amount of sample used for metabolome comparison. The LC-UV method used has been described previously.[125] Briefly, two μL of the labeled solution was injected onto a Waters ACQUITY BEH C18 column (2.1 mm × 5 cm, 1.7 μm particle size, 130 Å pore size). Solvent A was 0.1% (v/v) formic acid in 10% (v/v) acetonitrile, and solvent B was 0.1% (v/v) formic acid in acetonitrile. The gradient started with 0% B for 1 min and was increased to 95% within 0.01 min and hold at 95% B for 1 min. The gradient was restored to 0% B in 0.5 min and hold at this condition for 3.5 min to re-equilibrate the column. The flow rate used was 0.45 mL/min.

4.2.6 LC-MS and Data Analysis.

The labeled metabolites were analyzed using a Bruker 9.4 Tesla Apex-Qe Fourier transform ion-cyclotron resonance (FTICR) mass spectrometer (Bruker, Billerica, MA) linked to an Agilent 1100 series binary HPLC system (Agilent, Palo Alto, CA). The samples were injected onto an Agilent reversed phase Eclipse Plus C18 column (2.1 mm × 10 cm, 1.8 μm particle size, 95 Å pore size) for separation. Solvent A was 0.1% (v/v) formic acid in 5% (v/v) acetonitrile, and solvent B was 0.1% (v/v) formic acid in acetonitrile. The chromatographic conditions were: t = 0 min, 20% B; t = 3.5 min, 35% B; t = 18 min, 65% B; t = 21 min, 95% B; t = 26 min, 95% B. The flow rate was 180 μL/min. All MS spectra were obtained in the positive ion mode. The resulting MS data were processed using our in-house peak-pair picking software, IsoMS, written in R language.[154] This program eliminated the false positive peaks, such as isotopic peaks, common adduct ions, and multiply charged ions. Only the protonated ion pairs were exported for further analysis. The extracted peak-pair data were aligned by retention time

and accurate mass, and only those peak-pair features shared by no less than 50% of the samples were retained for multivariate analysis. Heatmap comparison, ANOVA and multivariate analysis was performed by Metaboanalyst[169] (www.metaboanalyst.ca) and the data were mean-centered and auto-scaled (unit variance) prior to analysis.

4.3 Results and Discussion

The major objective of this work is to develop an analytical method for quantitative profiling of as many microbial metabolites as possible in bacterial cells. One key step for achieving a high metabolome coverage is to optimize the sample preparation process leading to the isotope labeling and LC-MS analysis. In this work, dansylation chemistry is used to label the amine- and phenol-containing metabolites. However, the sample preparation process should be equally applicable to other labeling chemistries targeted at the analysis of other sub-metabolomes, such as acid-containing metabolites.[150] Differential isotope dansylation labeling combined with LC-MS offers three unique attributes for developing an optimal sample preparation process for microbial metabolome profiling. First of all, this detection scheme is much more sensitive than direct LC-MS without labeling. Thus a larger number of metabolites can be profiled to provide a more complete view of the performance differences among different conditions under investigation. Secondly, with differential isotope labeling, differences in metabolite quantities in samples prepared under different conditions can be evaluated. Finally, high detection precision can be obtained with isotope labeling LC-MS and therefore subtle differences in analytical performance among different experimental conditions used can be revealed. In the following section, the optimization and performance of each key step involved in the sample preparation of bacteria cells are described, followed by the

demonstration of an application of the developed method for differentiation of different bacteria in cultured samples as well as spiked urine samples based on the metabolome fingerprints of these microorganisms.

4.3.1 Effect of Cell Washing.

Since we are interested in intracellular metabolite profiling, it is important to ensure that the metabolites detected are from inside of the cells rather than from the growth medium. In this work, after the cells were scraped off the dish, they were washed twice with 0.9% NaCl solution to remove any extracellular compound potentially stuck to the surface of the cells. The isotonic 0.9% NaCl solution was selected because it would not cause significant leakage[131] and was reported to be effective for quenching the cell metabolism.[182]

Panels A-C in Figure 4.1 show a comparison between the base peak chromatograms of the washing solutions and the *E. coli* cell extract. By comparing the first and second wash solutions (panels A and B in Figure 4.1), we can see a significant decrease in signal intensity for most of the peaks, suggesting that the washing step is effective to remove the extracellular metabolites from the cells while not causing observable cell lysis (i.e., no leaking of metabolites from inside of the cells). A comparison between the second washing solution (Figure 4.1B) and the cell extracts (Figure 4.1C) reveals that there are fewer peaks detected in the washing solution with much lower intensities. Figure 4.1C shows that many metabolite peaks are observed and distributed across the entire gradient elution window. At a given retention time, a number of peak pairs from the differentially labeled metabolites can be observed in a mass spectrum. One example is shown in Figure 4.1D where several peak pairs with different

absolute intensities are detected. The expanded mass spectra at different intensity levels (see panels E-G in Figure 4.1) show a wide intensity dynamic range (from 1×10^5 to 3×10^7 counts) from which peak ratio of each peak pair can be determined. In this particular case, 943 peak pairs were detected from the cell extract, while 65 peak pairs were found in the 2nd washing solution. Among them, only 16 ion pairs have overlaps, which is less than 1.7% of the detectable peak pairs from the cell extracts and the intensities of these 16 ion pairs found in the washing solution are much lower than those from the cell extracts. The large chromatographic peaks observed at the front (at ~1.4 min) in both the washing solution and the cell extract were from the quenched reagent (dansyl-OH). The large peak observed at 23.7 min is from a singlet mass peak with m/z 457.2772, which indicates that it is from a background compound, not a metabolite. One possible source of this peak is the sodium adduct of polysorbate 60 based on the mass match; peaks corresponding to its H and NH_4 adducts were also observed in the mass spectrum. The comparison results obtained, as shown in Figure 4.1, give us confidence that compounds from the medium would not interfere and almost all detected metabolites in the ion chromatograms shown in Figure 4.1C should be originated from the cells.

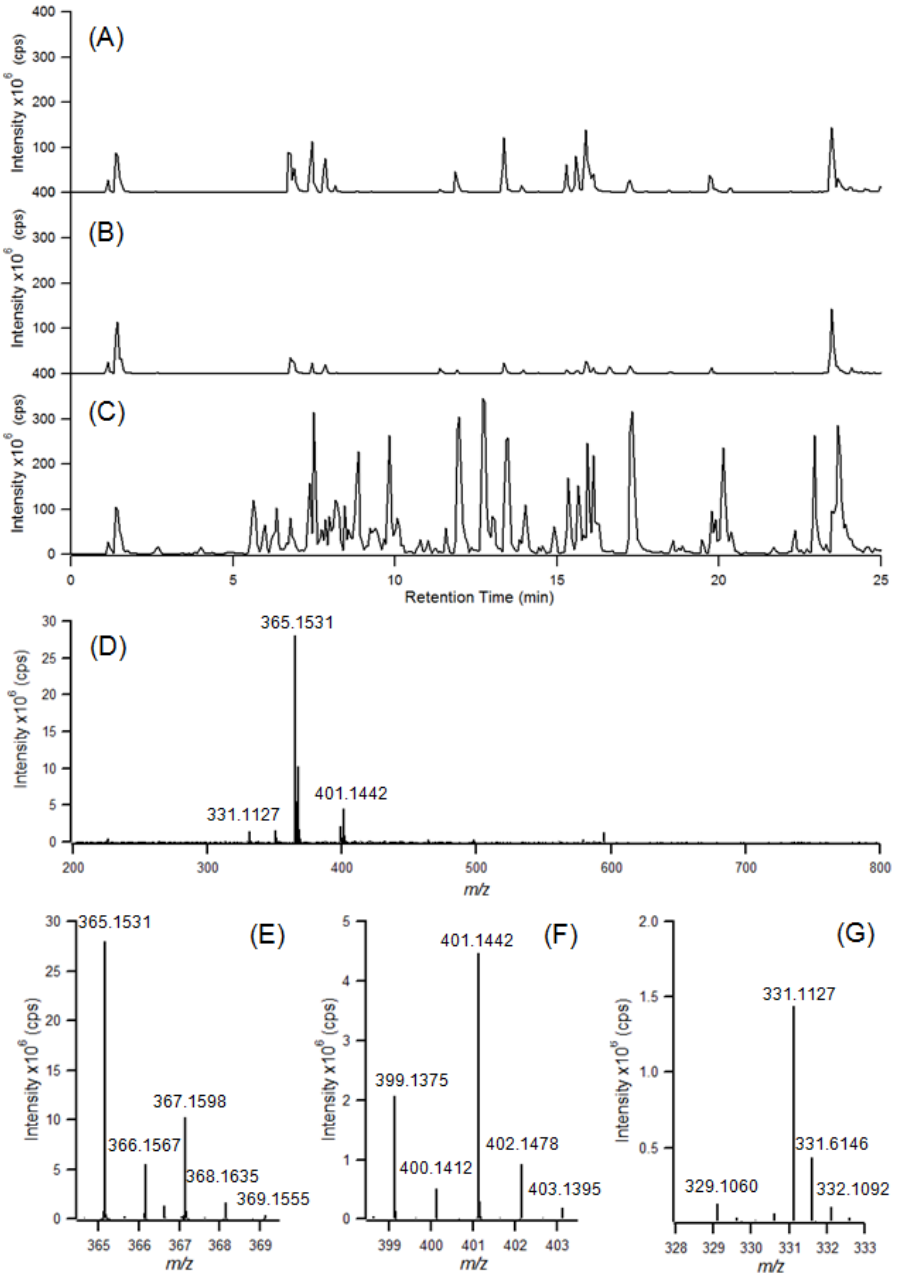


Figure 4.1 Base peak chromatograms of (A) the first wash solution, (B) the second wash solution and (C) the metabolites extracted from *E. Coli*. Mass spectra of (D) full scan at the retention time of 15.6 min in (C) and the expanded regions showing a peak pair with a ^{12}C-dansyl labeled metabolite (E) at m/z 365.1531 identified as isoleucine, (F) at m/z 399.1375 identified as phenylalanine and (G) at m/z 329.1060 identified as diaminopimelic acid.

4.3.2 Metabolism Quenching.

Although a quenching step is usually recommended before carrying out the extraction in order to stop further metabolism, the use of a quenching solution (typically 60% methanol) can often lead to metabolite leakage that can be as large as 60%.[208] The quenching procedure is more critical for studying metabolites with fast turnover rates, such as those involved in the energy metabolism (e.g., ATP) and glycolic pathways (e.g., glucose-6-phosphate).[216] Since our study focuses more on the general metabolic profile of the bacterial cells, rather than studying their metabolite fluxes, we decided not to include any additional quenching solutions in our experiment. Nevertheless, as noted before, we used cold 0.9% NaCl solution to rapidly wash the cells, which could at the same time serve as the quenching step. Although the extent of metabolism quenching during the cell washing step is unknown, as it will be shown below, the relative quantities of most of the putative metabolites detected in biological replicate experiments are reproducible, suggesting that any further metabolism during the sample handling process, if present at all, does not significantly affect the overall metabolic profiles. This may be due to the fact that, prior to metabolite extraction, the amine- and phenol-containing metabolites profiled by dansylation LC-MS do not undergo further metabolism extensively.

4.3.3 Comparison of Extraction Solvents.

As both methanol (MeOH) and acetonitrile (ACN) have been reported to be the optimal solvent for extraction of intracellular metabolites for different types of cells,[182,213] we have compared these two solvents (1:1 organic:water) to see which one works better for extracting the amine- and phenol-containing sub-metabolome from bacterial cells, using *E. coli* as the model system. In addition, a combination of the two solvents in water

(2:2:1 MeOH:ACN:H_2O, MAW) was also compared to see if a mixture of solvents can perform better than the use of a single solvent, with an expectation that a pure solvent and a solvent mixture may have different metabolite extraction and solubility properties. The performance of the extraction solvents was evaluated according to three criteria,[206] namely the number of peak pairs detected, the relative intensity of each peak pair and the reproducibility of each extraction method. In this experiment, each individual sample was labeled with ^{12}C-dansyl chloride and a pooled sample was labeled with ^{13}C-dansyl chloride to serve as the internal standard.

The number of peak pairs detected in each extraction solvent is plotted in Figure 4.2A. More than 1000 peak pairs were detected for each extraction solvent; this is a much larger number, compared to around 300 peaks detectable in similar cellular samples reported in the literature.[204,206] Since only the labeled amine- and phenol-containing compounds present in both the individual samples and the pooled sample can be picked up as peak pairs, the dansylation isotope labeling LC-MS detection scheme eliminates artifacts from the instrument[217] and other interferences (e.g., impurities leached from the plastic container). Moreover, we have also used a built-in function in the peak extraction software to filter the peak pairs found in the method blanks, thereby eliminating contributions from dansyl products of solvents, reagents and any impurities present therein. Therefore, each of the resulting peak pairs extracted by the software should represent a true metabolite. Since many of them were not identified, each peak pair detected is considered to be from a putative metabolite. Figure 4.2A shows that ACN and MeOH extractions gave a similar number of peak pairs. The use of solvent combination

(MAW), on the other hand, was not capable of extracting as many metabolites as the use of ACN or MeOH.

The relative intensity of each peak pair was calculated by taking the $^{12}C/^{13}C$ ratio, i.e., ratio of each individual sample to the pooled internal standard. Only those peaks commonly detected across all three extraction solvents were used for comparing the relative intensity. Figure 4.3 shows the number distributions of the peak pairs detected within a solvent (triplicate) and among the three solvent systems. From the combined results of all 9 runs, 2381 unique peak pairs or putative metabolites were detected. Among them, 851 peak pairs were found in all three extraction solvents. Figure 4.2B shows the relative intensity of 10 selected amino acids with different type of side-chains (i.e., hydrophobic, polar, acidic and basic) and Figure 4.2C compares the relative intensity of 10 other compounds encompassing a variety of classes (e.g., purine and pyrimidine derivatives, amino acid derivatives, dipeptide, amino sugar). These metabolites were identified based on the accurate mass and retention time matches with those of the dansyl labeled standard compounds. It can be seen that in most cases the results from these three solvents were comparable, with MeOH performing slightly better on average. For some compounds, the performance of MeOH extraction was significantly better than ACN or MAW (e.g., pyridoxal 5'-phosphate, adenosine), but it is difficult to observe specific trends on which class of compounds are more favorable in each solvent.

In addition to comparing the relative intensity of representative compounds, we have also carried out a heatmap comparison of all commonly detected peak pairs in order to avoid any bias. The heatmap comparison results are shown in Figure 4.4 where more red colored features indicate higher signal intensities. It was found that MeOH extraction

in general could give higher intensities for most of the peak pairs detected, compared to ACN or MAW extractions. These results indicate that MeOH is a better extraction solvent in terms of the relative intensity of the extracted metabolites.

To compare the reproducibility of different solvents, box plots showing the distributions of relative standard deviations of peak pairs were constructed and they are given in Figure 4.5. For all three solvents, the relative standard deviations for the majority of peak pairs were below 20%, indicating a good reproducibility. The mean values (represented by the dots) were 12.6% 10.2% and 9.8% for MeOH, ACN and MAW, respectively. The results from t-test of these values indicate that ACN and MAW were not significantly different, while MeOH's mean value was different from those of ACN and MAW. Thus, ACN and MAW gave slightly better reproducibility than MeOH. However, the reproducibility of all of these methods should be sufficient for most metabolic profiling applications.

The above results indicate that the three solvent systems studied detected a similar number of peak pair with similar number reproducibility. In terms of signal reproducibility, ACN and MAW were slightly better than MeOH. However, MeOH was more effective in getting higher amounts of compounds. Since the purpose of the work is to do metabolic profiling of bacteria cells, it is desirable to get as high yields as possible from the extraction process. Therefore, MeOH was chosen as the extraction solvent for all the subsequent experiments.

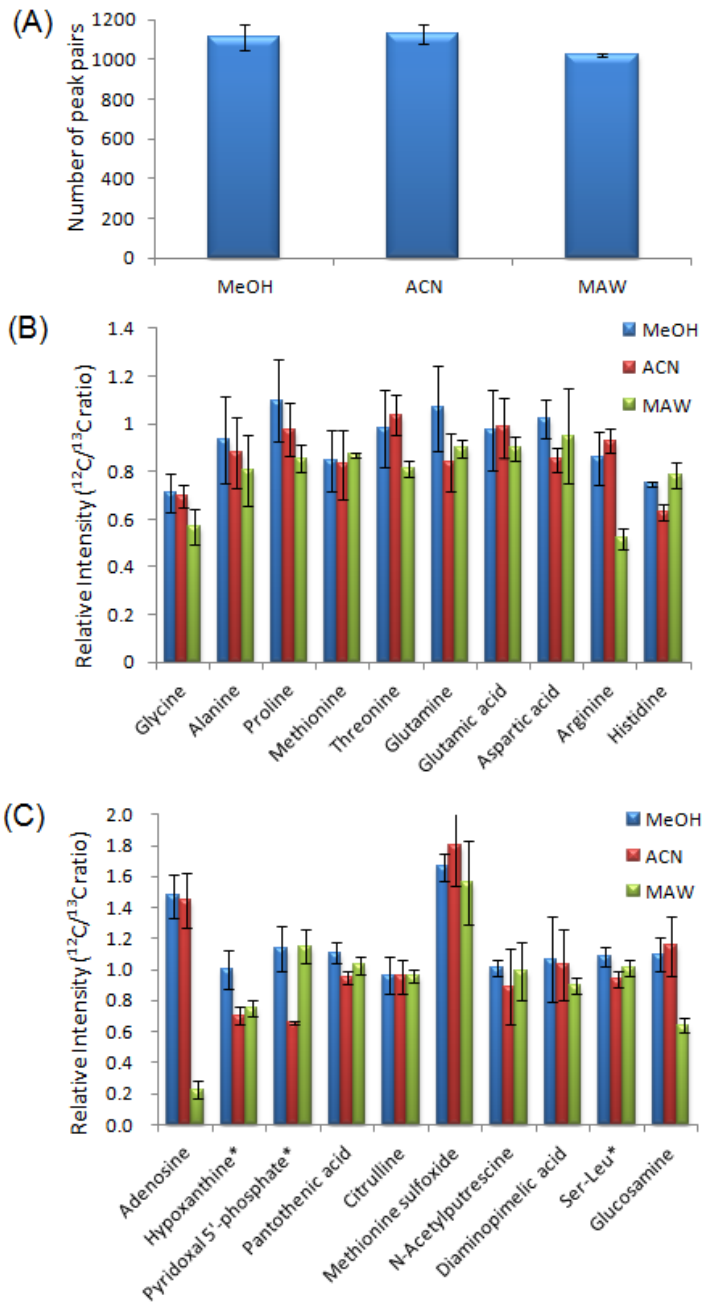


Figure 4.2 Comparison of (A) average number of peak pairs detected, (B) relative intensities of 10 amino acids and (C) relative intensities of 10 other selected compounds extracted by the three solvent systems. Compounds labeled with an asterisk (*) indicates level 2 identification (see Text). All the other compounds were definitively identified (level 1).

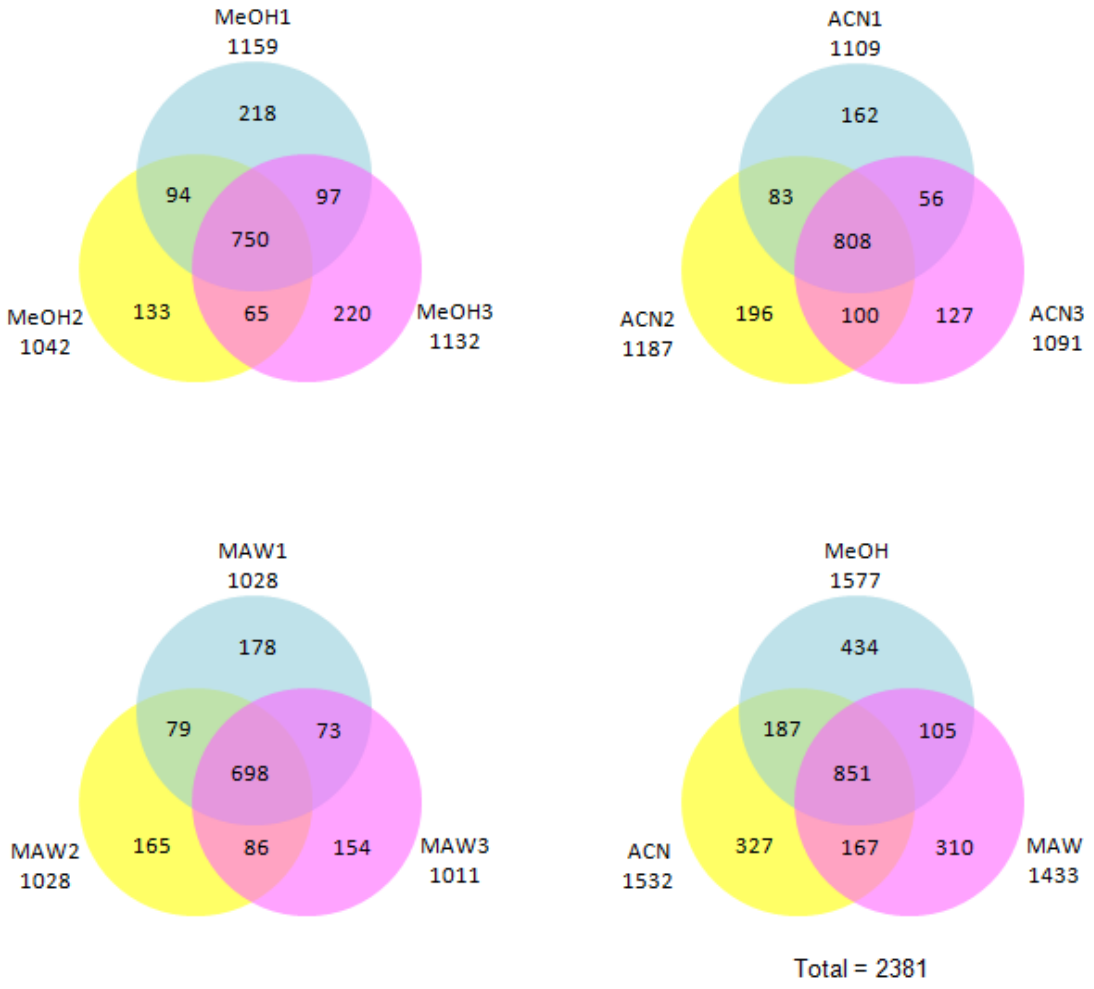


Figure 4.3 Distributions of the number of ion pairs detected in cell extracts prepared using three different solvent systems.

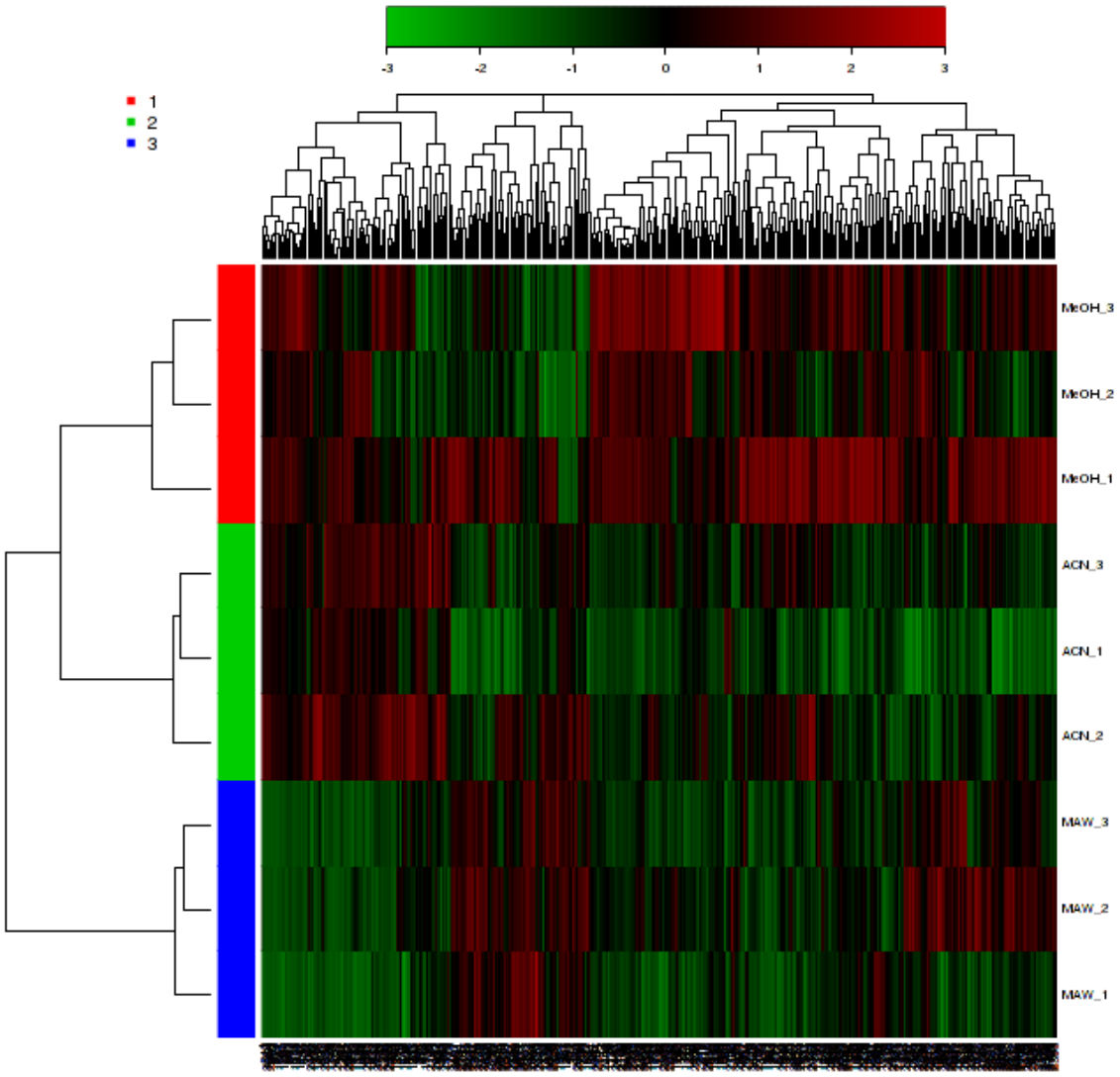


Figure 4.4 Clustered heatmap showing comparison of the relative intensity of each peak pair in three solvent extraction methods. More red colored features indicate higher signal intensities.

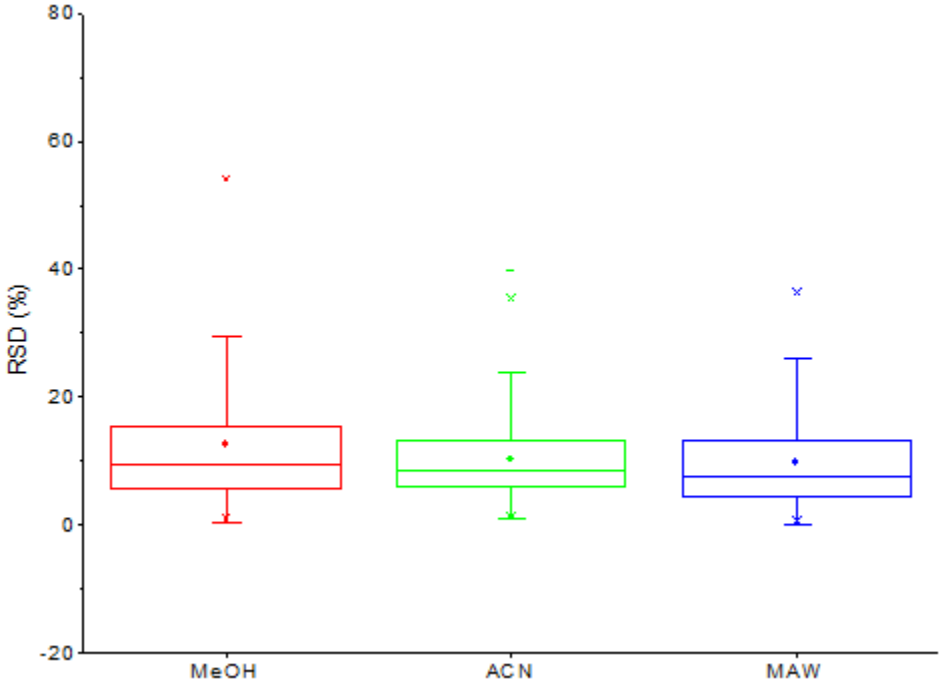


Figure 4.5 Box plot showing distribution of %RSD for three solvent extraction methods. The range of the box is 25 to 75 percentile. The line in the box represents the median value and the mean value is shown as a dot in the box.

4.3.4 Comparison of Sample Disruption Methods.

In addition to the selection of an appropriate extraction solvent, it is also important to use an effective disruption method to facilitate metabolite extraction into the extraction solvent. In this work, four protocols were evaluated: no disruption (or control, CT), freeze-thaw cycle (FT), ultrasonication (SN) and microwave (Mic). In many of the cellular metabolomics studies reported,[204,205] freeze-thaw cycle or no cell disruption was applied. Ultrasonication has also been used, but mostly with a sonicator containing a metal tip. However, this way of ultrasonication is very time-consuming as only one

sample can be processed at a time and may cause cross-contamination if the tip is not washed thoroughly. In our work, we used an ultrasonic cleaner which is capable of handling multiple samples without direct contact with the sample. The use of microwave has also been demonstrated to be efficient for extracting metabolites from biological samples.[218,219]

The same three criteria have been applied to evaluate the performance of the sample disruption methods. Figure 4.6A shows the number of peak pairs detected in each method. Among these four protocols, ultrasonication and microwave gave slightly higher numbers of peak pairs. Comparison of the relative peak intensity was also carried out on the commonly detected peak pairs (i.e., 834 out of a total of 2484 peak pairs from 12 runs, Figure 4.9). The relative intensities of 20 selected compounds are shown in panels B and C of Figure 4.6. The heatmap showing the comparison of relative intensities of all the peak pairs commonly detected in the four disruption methods is shown in Figure 4.7. It is apparent that ultrasonication gave higher intensities for most of the compounds, followed by microwave. The performance of freeze-thaw cycle was very similar to the control, which is not surprising considering that the control samples were flash frozen and stored under -80°C before the extraction, and therefore cell lysis is expected to occur to some extent through this freeze-thaw process. The heatmap shown in Figure 4.7 indicates that the extraction with no disruption or freeze-thaw gave low amounts for the majority of the peak pairs. Applying a disruption procedure appears to be essential in order to achieve high efficiency of extraction. Our results also suggest that the commonly used freeze-thaw cycle procedure is insufficient for metabolite extraction. The performance of

ultrasonication and microwave was similar, indicating that both methods can be used to effectively break the cells and facilitate metabolite extraction into the extracting solvent.

The reproducibility of the four disruption methods was also examined and the results are shown in Figure 4.8. For most of the metabolites the relative standard deviation of peak intensity from triplicate experiments was below 20%. The mean values for CT, FT, SN and Mic were 15.2%, 16.2%, 12.9% and 12.6%, respectively. The larger relative standard deviation values for CT and FT is another indication of insufficient extraction of the metabolites into the solvents; with low extraction efficiency, the level of metabolites can more randomly vary from one experiment to another, compared to a more completed extraction. Therefore, in terms of intensity reproducibility, SN and Mic were found to be better disruption methods than CT and FT. We also compared the reproducibility of the number of peak pairs detected among these methods. Figure 4.9 shows the distributions of the number of peak pairs detected from the four disruption methods. The average number of peak pairs detected was 981±46, 1012±50, 1111±61 and 1058±70 for CT, FT, SN and Mic, respectively. There was no significant difference in terms of peak pair reproducibility. Based on the overall consideration of the three criteria examined, it can be concluded that the use of SN or Mic is a better choice to facilitate the metabolite extraction. Because SN is very convenient to do, compared to Mic, we used ultrasonication with a sonicator to perform solvent extraction for the subsequent experiments.

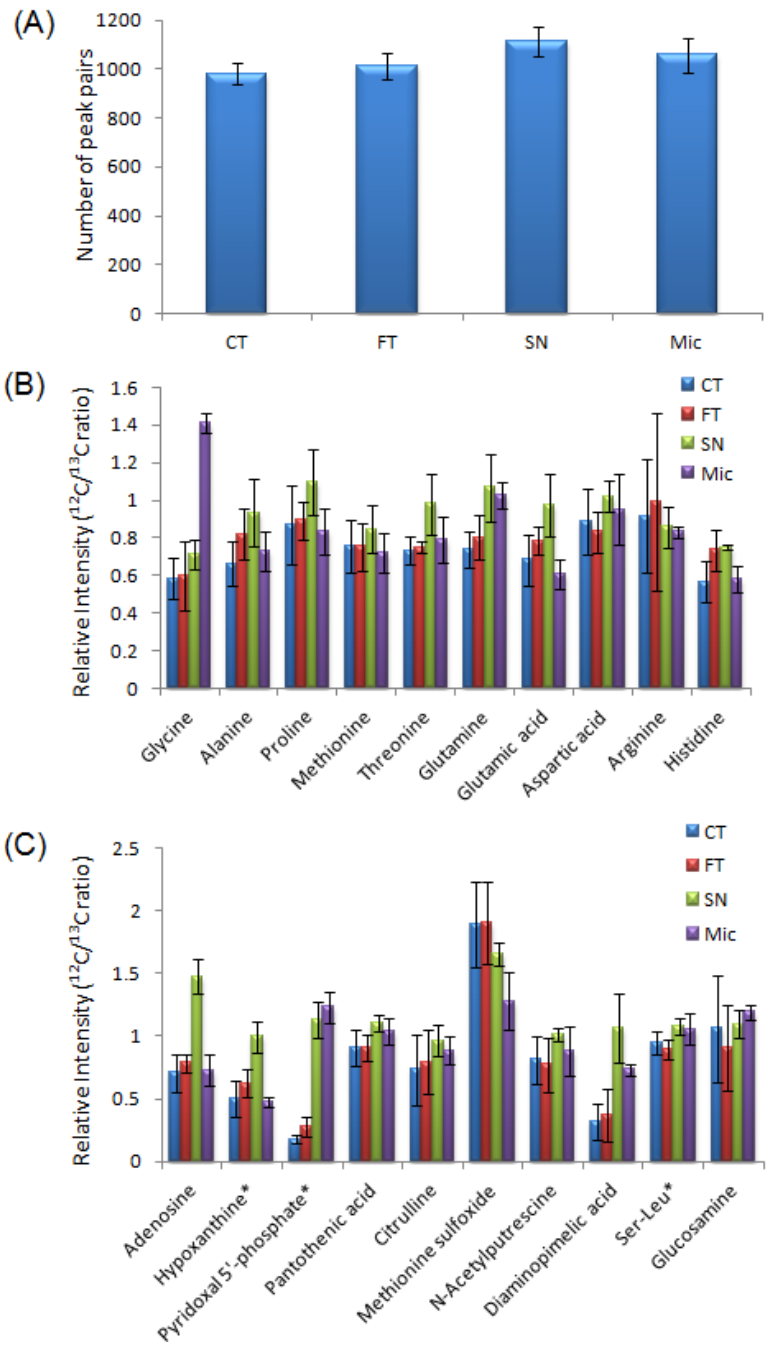


Figure 4.6 Comparison of (A) average number of peak pairs detected, (B) relative intensities of 10 amino acids and (C) relative intensities of 10 other selected compounds extracted in 1:1 MeOH:H_2O in combination with one of the four disruption methods. Compounds labeled with an asterisk (*) indicates level 2 identification (see Text). All the other compounds were definitively identified (level 1).

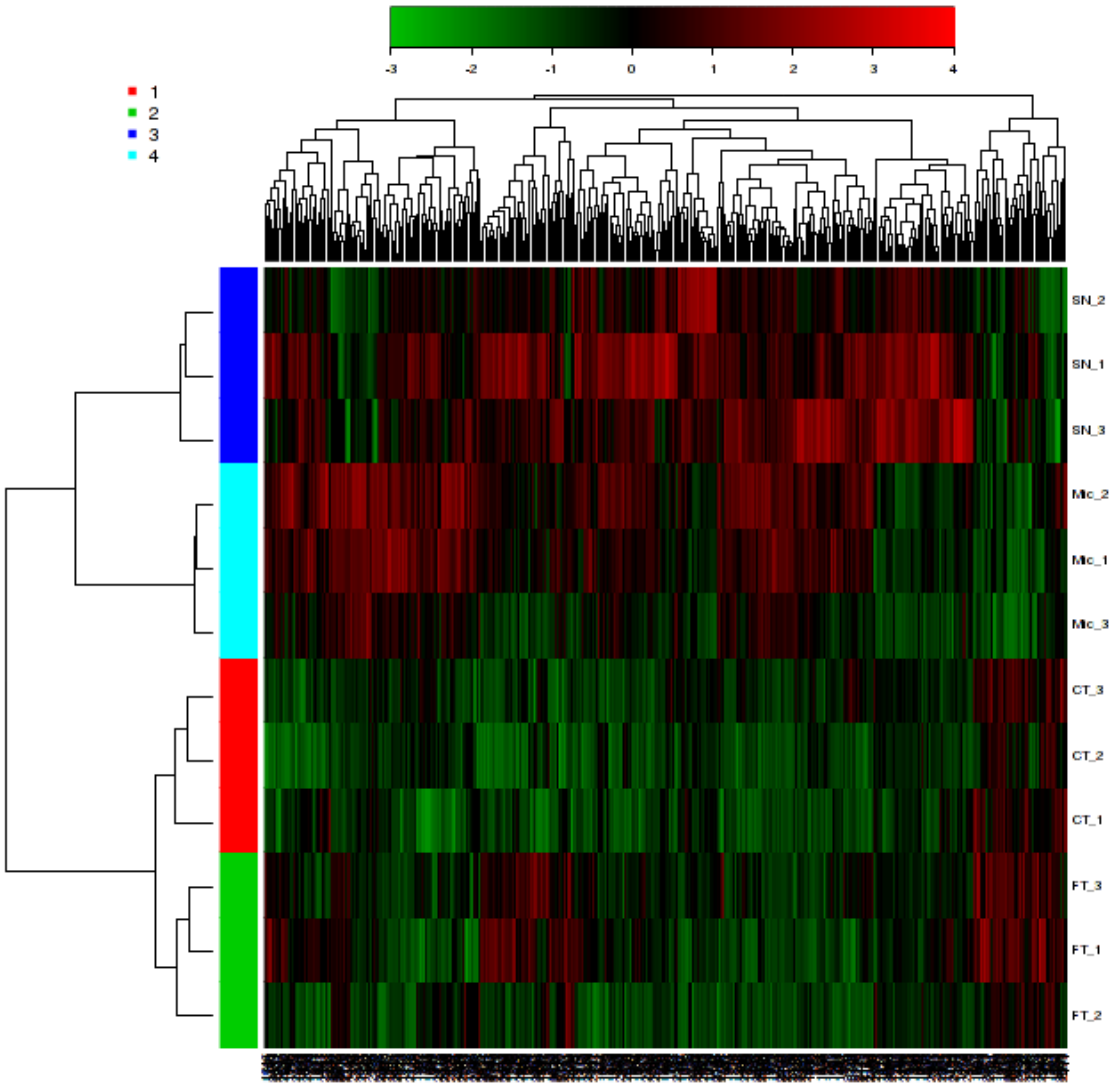


Figure 4.7 Clustered heatmap showing comparison of the relative intensity of each peak pair in four disruption methods. More red colored features indicate higher signal intensities.

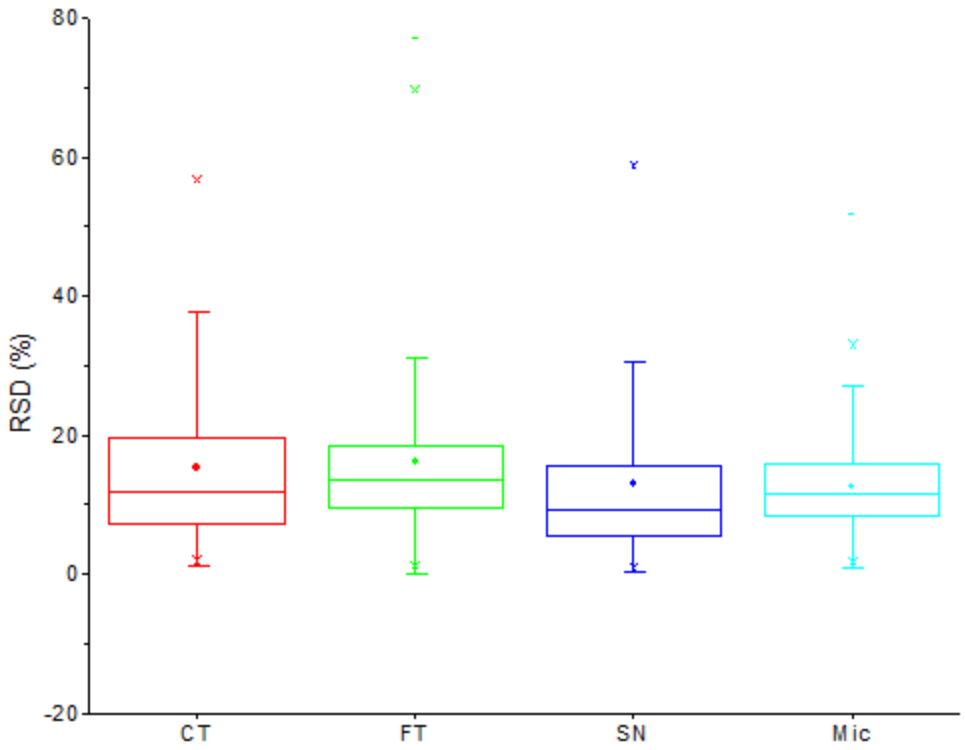


Figure 4.8 Box plot showing distribution of %RSD for four disruption methods. The range of the box is 25 to 75 percentile. The line in the box represents the median value and the mean value is shown as a dot in the box.

It should be noted that the entire process of sample extraction with sonication, dansylation labeling and LC-MS analysis can be accomplished within 5 hours. For a trained researcher, 8 samples can be processed simultaneously for the metabolite extraction and 50 samples can be labeled in one batch. This procedure thus allows routine processing of tens to hundreds of samples within one day and the LC-MS analysis of these samples with automatic running sequences to be done within a few days at a rate of 30 min per sample including column washing in between runs. The sample throughput should be expandable with the use of a multi-channel pipette and a larger or more

sonicators for sample processing and the use of a faster, high resolution mass spectrometer, such as time-of-flight MS, instead of FT-ICR-MS, for collecting the LC-MS data.

4.3.5 Bacterial Differentiation: Sample Amount Normalization.

The application of the extraction-labeling method described above is demonstrated on the differentiation of three different bacteria cells, namely *E. coli* (EC), *B. subtilis* (BS) and *B. megaterium* (BM), based on their metabolic profiles. These bacteria were chosen as the model system because they include both Gram (+) (*B. subtilis* and *B. megaterium*) and Gram (-) (*E. coli*) species. Each cell was analyzed in a total of 9 biological replicates, including 3 inter-day replicates and 3 intra-day replicates on each day. In addition to the biological replicates, one additional experimental replicate was carried out for each of the three bacteria on each day. These replicate experiments were designed to demonstrate the robustness of this method, which is an important prerequisite for studying the differences in metabolic profiles. If the metabolic profiles of each cell type obtained from all these replicates are similar, we will then be confident to apply our method to study the metabolic differences among different cells, i.e., either the same cell type grown under different conditions to investigate the effect of cellular perturbation (e.g., via a stimulus) for biological studies, or different cell types for bacteria differentiation or identification.

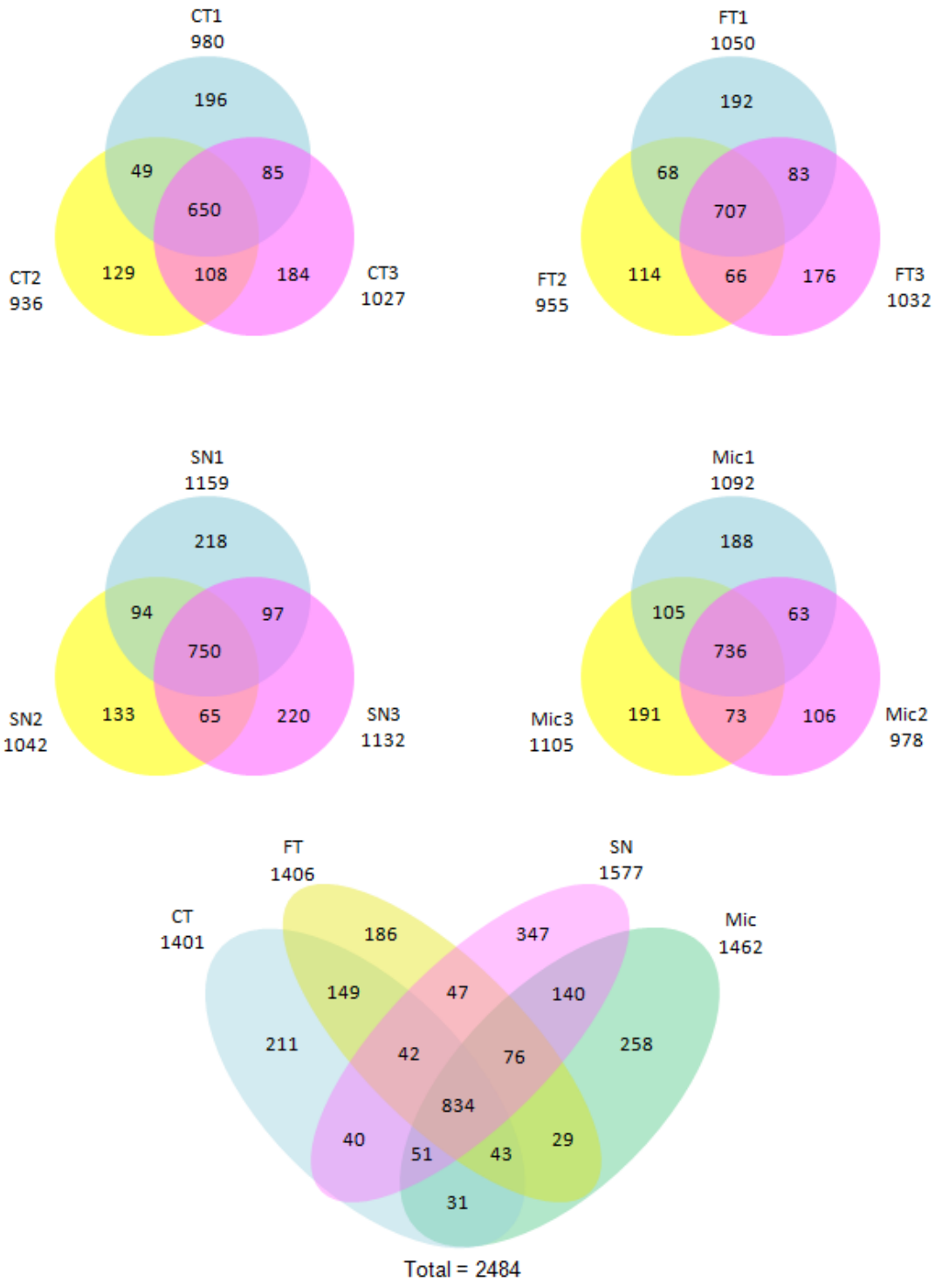


Figure 4.9 Distributions of the number of ion pairs detected in cell extracts prepared using four different sample disruption methods.

Since the cells were scraped off the plates, it is difficult to control as well as count the number of cells harvested. For fair comparison of the metabolic profiles of the same or different cells, sample amount normalization among the different comparative samples is needed. We recently reported a normalization method based on the use of dansylation labeling LC-UV for determining the total amount of the labeled metabolites in a sample, followed by adjusting the sample volume of individual samples to mix with a control sample (i.e., a pooled sample from several individual samples).[125] In this work, we applied this method to normalize the metabolite amounts among the comparative cell samples. After dansylation labeling of the extracted metabolites from each cell sample, a LC-UV quantification step was carried out to determine the total amount of the labeled metabolites.

The quantification results are shown in Table 4.1, which illustrates that although the sample amount within each cell type was similar in most cases, a variation of as large as 2-fold could still be observed. This indicates that a sample amount normalization step is very important in order to minimize the contribution of the sample amount variation to the metabolic profiles of comparative cells. Table 4.1 also shows that the amount of metabolites extracted from *B. subtilis* and *B. megaterium* was considerably larger than that extracted from *E. coli*. This is due to the larger number of cells harvested, because *B. subtilis* and *B. megaterium* grow better in the growth medium (nutrient agar) than *E. coli*. In this work, to compare the metabolic profiles of different cell species, we took the same total amount of the labeled metabolites from each sample and assumed that any difference observed from their metabolic profiles solely came from differences in the

abundances of individual metabolites. As it will be discussed later, this assumption was proved to be valid for the purpose of differentiating cell types.

Table 4.1 Quantification results of bacterial samples by UV absorbance.

	Peak area	Concentration (mM)*	Volume needed (µL)
EC D1_1#	1208390	0.99	70.3
EC D1_2	1377843	1.12	62.2
EC D1_3	887078	0.75	93.3
EC D2_1	1234665	1.01	68.9
EC D2_2	1465343	1.19	58.8
EC D2_3	1385553	1.13	61.9
EC D3_1	626914	0.55	126.9
EC D3_2	1023379	0.85	82.0
EC D3_3	1240886	1.02	68.6
BS D1_1	3091836	2.43	28.8
BS D1_2	3199967	2.51	27.8
BS D1_3	3118720	2.45	28.5
BS D2_1	3230289	2.54	27.6
BS D2_2	3451369	2.70	25.9
BS D2_3	4257265	3.32	21.1
BS D3_1	2477912	1.96	35.6
BS D3_2	3178666	2.50	28.0
BS D3_3	3336383	2.62	26.7
BM D1_1	3364010	2.64	26.5
BM D1_2	4661662	3.63	19.3
BM D1_3	4673824	3.64	19.2
BM D2_1	3519363	2.76	25.4
BM D2_2	2844405	2.24	31.2
BM D2_3	3789028	2.96	23.6
BM D3_1	4192583	3.27	21.4
BM D3_2	4032917	3.15	22.2
BM D3_3	4467401	3.48	20.1
Pooled sample	2962148	2.33	30.0

*The concentrations were calculated using a pre-determined calibration curve

[#]EC: *E. coli*; BS: *B. subtilis*; BM: *B. megaterium*. D1, D2, D3 refer to samples collected on day 1, day 2 and day 3, respectively. The numbers after the underline indicate the intra-day replicates.

4.3.6 Bacterial Differentiation: Data Analysis.

To produce a reference sample from which all the individual samples can be compared to, aliquots of individual extracts from the three different cells were mixed to generate a pooled sample which was subsequently labeled by ^{13}C-dansyl chloride. An equal amount of ^{12}C-dansyl labeled individual sample and ^{13}C-dansyl labeled pooled reference sample was mixed, followed by LC-MS analysis. The peak ratios of individual peak pairs found in the mass spectra were calculated. Since the same reference sample was used, the peak ratios of an individual peak pair obtained from different samples reflected the concentration differences of the putative metabolite in these samples.

Among the comparative samples, there are a total of 704 peak pairs or putative metabolites with ratios detectable in at least 50% of the samples. These ratios were used for the principal component analysis (PCA) and the resulting PCA score plot is shown in Figure 4.10. The score plot clearly demonstrates that the three different bacteria can be well separated. Although *B. subtilis* and *B. megaterium* belong to the same *Bacilli* genus, there are still substantial differences in their metabolic profiles, which are reflected by the second principle component. These results indicate the potential of this method for differentiating bacteria at the species level. Of course, the specificity and applicability of the method for differentiating a great variety of bacteria species and strains, including clinically relevant microorganisms, requires further investigation.

The distribution of the intensity ratio of the peak pairs detected from each sample that were used for the PCA analysis is shown in Figure 4.11 (only the data from biological replicates were plotted). It is interesting to note that the intensity ratio distributions of all 9 *E. coli* samples were larger than those of *B. megaterium* and *B. subtilis*, while the distributions of *B. megaterium* and *B. subtilis* were very similar. The median of the intensity ratios for each sample ranges from 0.67 to 1.56 with a mean value of 1.06. Thus, many of the metabolites have similar ratios, which are expected as they are likely involved in essential metabolisms common to different cell types. The most important variables that contribute to the differentiation were determined by PLS-DA analysis. Table 4.2 lists the top 20 metabolites (i.e., with the highest VIP scores). The metabolites were identified either by matching the accurate mass and retention time with authentic standards (level 1) or our standard library[51] (level 2), or by searching the accurate mass against the human metabolome database (HMDB)[26] (level 3). Table 4.2 also shows the average $^{12}C/^{13}C$ ratio of these metabolites in each species, with the %RSD included in the parentheses. We can see that the %RSD is generally in the range of 10-30%, indicating a good reproducibility was obtained, despite the combined variations from biological replicates, sample preparation and instrumental analysis. However, we note that for some compounds (e.g., compound **16**), the %RSD is considerably large for all three species, and a detailed examination of the data variations reveals that the large %RSD is mainly attributed to the biological variations (data not shown). Therefore, compounds like this one should not be selected as a discriminator, if one or several individual metabolites are used for differentiating the cell type.

From the above results of the median peak ratio analysis of individual metabolites and the PCA/PLS-DA analysis of the metabolic profiles, we can conclude that the proposed sample amount normalization method can be used for bacterial metabolic comparison to reveal subtle differences. We noticed that some of the metabolite distribution patterns among the three different cells can be rationalized. For example, diaminopimelic acid, a key component of the bacterial cell wall that is incorporated into the peptidoglycan structure of Gram (-) bacteria and Gram (+) bacilli,[220] has been detected in all three species, with the ratios in the two Gram (+) species being markedly larger than in *E. coli* (Figure 4.12). This can be explained by considering that Gram (+) bacteria have thicker cell wall structures and that peptidoglycan is highly abundant in Gram (+) bacteria cell wall.[220,221] We also found that this ratio is larger in *B. megaterium* than in *B. subtilis*, possibly because of the larger size of *B. megaterium*. For many other compounds, for example, glutamine, it is difficult to provide a simple explanation for the different amounts present in the three species because glutamine is involved in several metabolic pathways. However, as long as the amount found in each bacteria species is very consistent and that they are significantly different from the other species (as indicated by the *p*-values), this compound can still serve as a good marker for bacterial identification. Thus, by matching the ratios of many of these marker metabolites obtained in an unknown sample with the data generated from a known species, we would be able to identify the unknown species with high confidence.

Of course, the applicability of this approach of bacterial identification depends on the presence of an informative library containing specific panels of metabolite-markers from a wide variety of bacteria species that are of interest to a practical application. For

example, for diagnosis of bacterial infection, a metabolite-marker library of bacteria found in clinical samples may be constructed. This strategy is similar to the use of protein-markers for bacterial identification that is done mainly using matrix-assisted laser desorption ionization (MALDI) MS[222-224] and has been adapted in some clinical diagnosis laboratories.[23,225-227] Metabolite-based analysis may offer an alternative or complementary tool to the protein-based method for bacterial differentiation or identification.

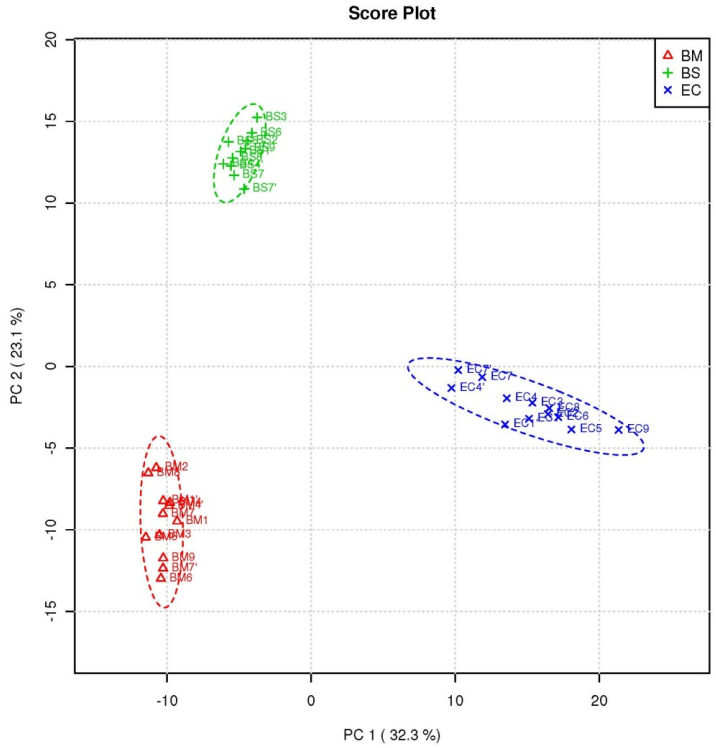


Figure 4.10 PCA score plots for three standard bacterial cultures. BM, *B. megaterium*, in red; BS*, B. subtilis*, in green; EC, *E. coli*, in blue. For each species, 9 biological replicates and 3 experimental replicates were presented.

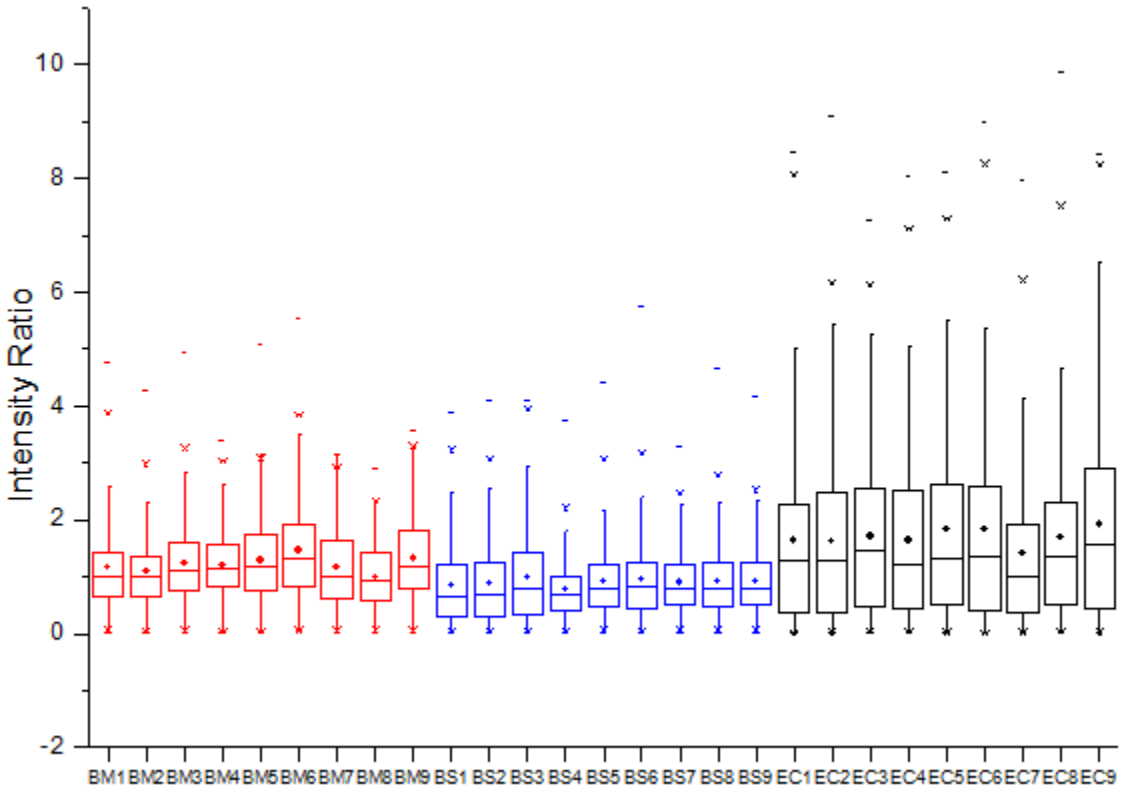


Figure 4.11 Distribution of the intensity ratios of all peak pairs detected in at least 50% of the samples for each biological replicate. BM, *B. megaterium*, in red; BS, *B. subtilis*, in blue; EC, *E. coli*, in black.

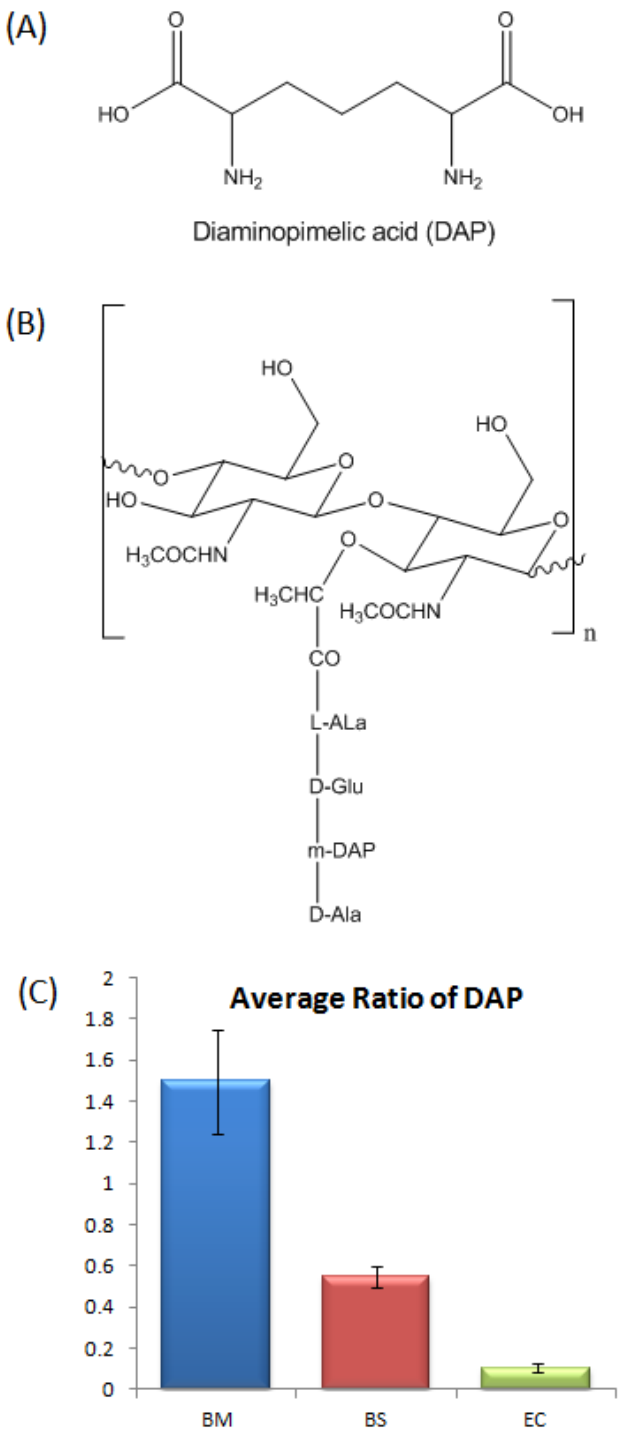


Figure 4.12 (A) Molecular structure of diaminopimelic acid; (B) One structure unit of the cell wall peptidoglycan; (C) Column plot showing the average ratio of diaminopimelic acid detected in the three bacteria species.

Table 4.2. Summary of the top 20 discriminant metabolites (VIP score>1.5) determined by PLS-DA for differentiation between *E. coli*, *B. Subtilis* and *B. Megaterium*.

	Rt (min)	m/z light	$^{12}C/^{13}C$ in BM*	$^{12}C/^{13}C$ in BS	$^{12}C/^{13}C$ in EC	Compound ID	ID level#
1	14.87	329.1059	1.49 (20)	0.55 (9)	0.10 (17)	Diaminopimelic acid	1
2	5.84	380.1274	1.28 (18)	0.82 (24)	0.06 (14)	Glutamine	1
3	16.90	352.0996	0.51 (18)	1.03 (13)	1.50 (11)	Methionyl-Serine, Serinyl-Methionine	3*
4	9.59	478.1645	1.57 (20)	0.45 (31)	0.02 (25)	Hydroxyprolyl-Hydroxyproline, (2S,4S)-Pinnatanine	3
5	9.73	323.1053	1.23 (19)	0.54 (18)	0.12 (17)	Alanine	1
6	20.31	389.1277	0.52 (17)	0.65 (21)	2.62 (22)	Histidine	1
7	13.90	346.0855	0.04 (27)	0.74 (29)	2.19 (27)	Uracil	3*
8	18.68	300.1031	1.11 (26)	0.93 (20)	0.09 (29)	Ornithine	1
9	8.40	394.1432	1.08 (21)	1.02 (22)	0.19 (13)	Alanyl-Alanine, 4-Acetamido-2-aminobutanoic acid	
10	7.95	353.1164	1.11 (20)	0.60 (23)	0.08 (18)	Threonine	1
11	11.41	399.0681	0.29 (23)	1.26 (15)	3.09 (16)	Unknown	\
12	12.37	315.1161	1.76 (23)	0.71 (17)	0.21 (17)	2-Methylpyrrole	3
13	19.41	367.6065	0.45 (28)	0.71 (26)	3.36 (32)	Adenosine, Deoxyguanosine, Vidarabine	3*
14	5.64	440.1488	0.40 (23)	0.12 (12)	5.34 (12)	Serinyl-Threonine, Threoninyl-Serine	3
15	10.44	453.1693	0.69 (14)	0.71 (18)	2.98 (18)	Pantothenic acid	1
16	7.22	438.1333	1.75 (33)	0.21 (51)	0.02 (46)	Aspartyl-Alanine, Alanyl-Aspartate, 5-L-Glutamylglycine	3
17	13.98	393.1352	1.91 (21)	0.11 (25)	0.64 (12)	Unknown	\
18	24.29	378.0830	0.65 (18)	0.64 (24)	3.33 (20)	Unknown	\
19	22.46	384.6401	0.07 (27)	0.48 (46)	1.49 (22)	Unknown	\
20	11.19	392.1275	1.14 (15)	1.29 (13)	0.23 (20)	1-(Hydroxymethyl)-5,5-dimethyl-2,4-imidazolidinedione	3*

*the number in parenthesis refers to %RSD.

#an ID level labeled with an asterisk (*) indicates an unlikely match due to the lack of a correct number of amine or phenol functional groups in the structure. The criteria for accurate mass match is within a mass error of less than 5 ppm.

4.3.7 Analysis of Bacteria in Human Urine.

In addition to the presence of a library of metabolite markers, a reliable identification process also depends on the consistent detection of the bacterial metabolic profiles in various real samples. In order to investigate whether a single bacterial strain isolated from a real biological sample can give a similar metabolic profile as the standard cultures, we spiked 1×10^5 cells/mL of *E. coli* cultured in LB medium into 1 mL of human urine samples and spread the urine sample onto the nutrient agar plates to isolate pure *E. coli* strains, as a model system to mimic a real clinical sample (denoted as ECU). This was done with three healthy individuals' urine samples, each with three replicate experiments. The use of urine samples from different people can provide us an insight into whether differences in urine would have an effect on the metabolic profile of the bacteria. The isolated *E. coli* strains were then extracted, labeled and analyzed in the same way as the standard bacteria cultures. The resulting data were compared with the data obtained from the three standard cultures using PCA, as shown in Figure 4.13. It can be seen that *E. coli* strains isolated from spiked urine samples can still be clustered close together with the standard *E. coli* cultures, and that they can be clearly separated from the two *Bacilli* 'false strains'. In addition, there is no significant distinction between the *E. coli* strains obtained from the three different urine samples, indicating that the composition of urine would not affect the bacterial metabolite profile.

The PCA score plot is useful for visual inspection of the data. However, in order to obtain a more confident conclusion on the similarity of the 'real sample bacteria' to the standard bacteria cultures, a comparison of the ratios of all commonly detected peak pairs was carried out. For a total of 454 commonly detected peak pairs, we assigned a score of

1 to the closest match and a score of 0 to the other two non-matches for each peak pair. This procedure gave a score of 83 for *B. megaterium*, 89 for *B. subtilis* and 282 for *E. coli*. The much higher score for *E. coli* is another indication that this urinary bacteria can be differentiated from the 'false strains' and thus be correctly identified.

As an example, Table 4.3 shows the ratios of 10 definitively identified metabolites (level 1) detected in the four bacteria strains, from which we can see that for all these 10 metabolites, the average ratios found in the urinary bacteria were closest to *E. coli* and were significantly different from that of *B. subtilis* and *B. megaterium*. The %RSDs for different urinary bacteria samples were all below 30%, indicating again that the metabolic profile of bacteria cells was largely independent of the urine composition.

The success of identifying the bacteria strain in the spiked urine samples is an important illustration on the prospect of our method to practical applications, which requires sample-independent, un-biased analyses. Our hypothesis was that regardless of the source of the bacteria, by growing the cells on the same medium under exactly the same conditions, similar metabolic profiles could be obtained. The results of the current experiment support this hypothesis, leading to the possibility of using this method for real sample analysis which we will undertake in the future.

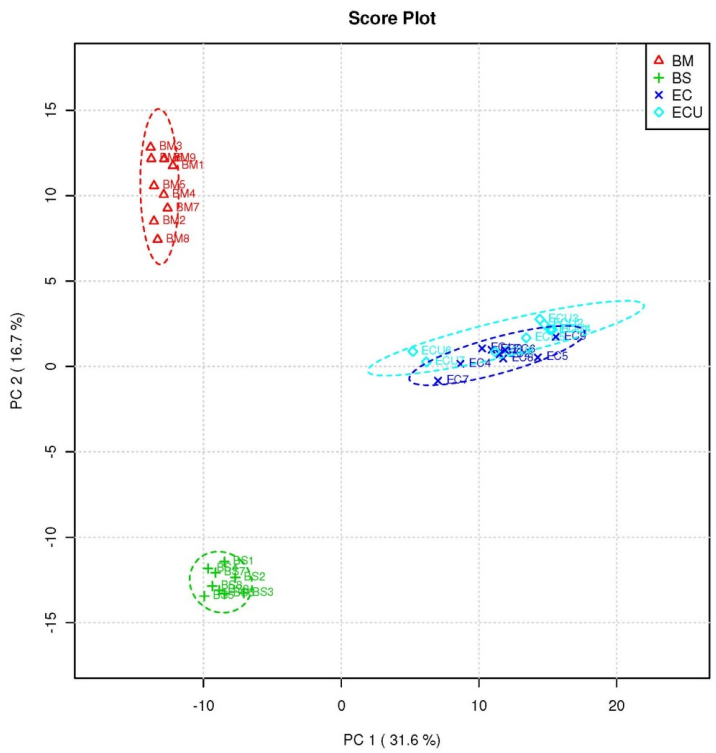


Figure 4.13 PCA score plots for three standard bacterial cultures plus the urinary bacteria. BM, *B. megaterium*, in red; BS*, B. subtilis*, in green; EC, *E. coli*, in blue; and ECU, *E. coli* from urine, in light blue. Only the 9 biological replicates were shown for each species.

Table 4.3. Comparison of the average $^{12}C/^{13}C$ ratio in urinary bacterial samples with standard cultures.

Compound Name	$^{12}C/^{13}C$ in BM*	$^{12}C/^{13}C$ in BS	$^{12}C/^{13}C$ in EC	$^{12}C/^{13}C$ in ECU
Pantothenic acid	0.69 (14)	0.71 (18)	2.98 (18)	2.91 (14)
Diaminopimelic acid	1.49 (20)	0.55 (9)	0.10 (17)	0.13 (23)
Glutamine	1.28 (18)	0.82 (24)	0.06 (14)	0.08 (29)
Serine	0.72 (19)	1.18 (25)	0.19 (32)	0.28 (24)
Ornithine	1.11 (26)	0.93 (20)	0.09 (29)	0.13 (27)
Threonine	1.11 (20)	0.60 (23)	0.08 (18)	0.12 (25)
Alanine	1.23 (19)	0.54 (18)	0.17 (14)	0.21 (21)
Histidine	0.52 (17)	0.65 (21)	2.62 (22)	2.87 (21)
Lysine	0.10 (17)	1.17 (9)	2.51 (20)	2.85 (19)
Glutamic acid	0.33 (24)	2.11 (2)	0.13 (23)	0.12 (14)

*the number in parenthesis refers to %RSD.

4.3.8 Detection Sensitivity.

In this study, we have not carried out the specific experiments to determine the sensitivity of this method, but we note that cells grown under the conditions indicated in the Experimental Section for a period of 24 hours, which corresponds to the order of 1×10^9 cells, could generate sufficiently high signals (i.e., more than 1000 peak pairs detected with S/N > 20). We also noticed that although *E. coli* was used as the model species for method optimization, the peak intensity for the majority of peaks were on the same order of magnitude for all three bacteria species, indicating that this extraction-labeling method works equally well for Gram (+) and Gram (-) species.

4.4 Conclusion

A method for quantitative metabolome profiling of bacterial cells with relatively high metabolome coverage has been developed. The optimized method involves the following major steps: cell sampling and washing, methanol extraction with ultrasonication, isotope labeling of the extracted metabolites, LC-UV quantification of the labeled metabolites, mixing of the isotope-differentially labeled sample and control, LC-MS analysis of the mixture, and data analysis of the peak pairs or putative metabolites found in comparative samples. We have targeted the analysis of amine- and phenol-containing metabolites by applying the dansylation labeling strategy and have optimized the extraction solvent and disruption method for detection of the dansyl-labeled compounds. This method has also incorporated a LC-UV quantification step targeting the labeled metabolites, which is very straight-forward and simple to use. We demonstrated that this method allows the detection of an average of 1052±58 peak pairs of putative metabolites from *E. coli* using a 25-min LC-MS run and over 2484 unique peak pairs from a combination of different solvents and sample disruption methods. Application of this method was successfully demonstrated on the differentiation of three different bacteria species, as well as identification of bacteria cells spiked in urine samples. The possibility of applying the current method in clinical applications has been discussed, which will be our future research focus. We envisage that another important application of this method is in the area of biological metabolomics where specific changes of the metabolome within cells after exposing to different environmental conditions can be probed for functional studies of cellular metabolisms and networks.

Chapter 5

Metabolic Signature of Autism Spectrum Disorders Revealed by High Performance Isotope Labeling LC-MS

5.1 Introduction

Autism spectrum disorder (ASD) is a set of neurodevelopmental disorders developed in the early childhood. It is characterized by impaired social interaction and communications, as well as abnormal behaviors like restricted and repetitive activities. There's been a growing research interest in the study of ASD because of its increasing prevalence over recent years, which is estimated to affect 1 in 50 children in 2011-2012.[228] The exact pathophysiology of autism is still not well understood,[229] and the disorder is believed to be related to a complex interaction of genetic and environmental factors, with the heritability accounting for approximately 40-60%.[230,231] Diagnosis of ASD is largely determined by clinical evaluations of observed behaviors, which is inevitably subject to the expertise of clinicians. In addition to abnormal behaviors, autistic patients were often found to exhibit altered functions and metabolisms in the body's basic systems including the neurologic, gastrointestinal, immunologic and toxicologic systems.[232] Comorbid conditions like epilepsy and depression are also commonly observed in autistic individuals.[233] Because of the complexities in the pathogenesis of this disorder and the high clinical heterogeneity, it is necessary to establish a more objective method for diagnosis of ASD based on biological markers. However, to date, there's still lack of reliable biomarkers that can be used clinically for diagnosis of autism. In recent years, several studies have been reported that aims to identify specific proteins or metabolites as potential biomarkers for clinical diagnosis.[234-

[238] However, most of these studies were based on a relatively small sample size (less than 50 autistic patients and healthy controls) and a validation phase based on an independent sample set is often absent.

Metabolomics is now a rapidly growing field that allows high-throughput profiling of metabolites at the global level. As the end product of all biological processes, the status of metabolites could be used to reflect pathophysiological changes in diseases, and consequently disease-specific metabolic signatures could be readily identified by metabolomics studies.[239] In addition, the comprehensive analysis of the metabolome allows simultaneous detection and quantification of multiple metabolite signatures which when combined together, could provide a better diagnostic performance compared to the use of a single biomarker. Because of the high complexity of biological samples, a robust and sensitive analytical method capable of detecting and quantifying a large number of metabolites is required for metabolomics studies. To this end, our group has developed a $^{12}C/^{13}C$-dansylation labeling technique for the analysis of the amine- and phenol-containing sub-metabolome, which allows separation of polar and ionic metabolites on RPLC while providing a signal enhancement of 10- to 1000-fold.[146] This method enables detection and relative quantification of hundreds to over a thousand of metabolites and has been applied in the metabolic profiling of various biofluids including urine,[146], cerebrospinal fluid[49] and saliva.[156]

In this work, we applied this dansylation labeling technique for metabolomic profiling of 427 plasma samples from children aged 2-10 with ASD and age-matched controls (Table 5.1). Plasma, compared to other biofluids such as urine and cerebrospinal fluid (CSF), has the following features for studying disorders related to the neurological

system: 1) it is less invasive than CSF and can be readily obtained in clinical laboratories; 2) the metabolite levels in plasma were found to be correlated to those in CSF[239] and therefore it may provide more valuable information on the neurological system compared to urine samples; 3) plasma metabolite concentrations are highly regulated so that metabolite levels can be directly quantified and compared across different samples. In addition to the selection of biofluid type, a few more considerations were made to ensure a more homogeneous sample set. Firstly, children were preferred subjects than adults because metabolite levels in adult patients are more likely affected by disease duration.[240] Secondly, healthy control samples were carefully chosen so that both the age and gender match with the autistic group. Finally, since the metabolite profile is sensitive to exogenous treatments such as medication, only drug-naïve subjects were recruited. Our results demonstrate a significant difference in the metabolic profiles of ASD patients and controls and a set of metabolite signatures were identified and validated. To the best of our knowledge, this is the largest sample set that has been reported for metabolomics studies of ASD.

Table 5.1 Clinical characteristics of study participants

	First set		**Second set**		**Blind test set**	
	HC*	ASD	HC	ASD	HC	ASD
Number of samples	100	100	100	100	19	8
Age (years)	6.2 ± 2.5	5.3 ± 2.6	5.9 ± 2.4	4.6 ± 2.4	5.2 ± 2.7	2.9 ± 0.8
Male/Female	50/50	50/50	80/20	80/20	10/9	5/3

*HC: Healthy control; ASD: Autism spectrum disorder

5.2 Experimental

5.2.1 Chemicals and Reagents.

^{12}C-dansyl chloride (DnsCl) and formic acid were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). The isotopic compound used to synthesize ^{13}C-dansyl chloride was purchased from Cambridge Isotope Laboratories (Cambridge, MA, USA). ^{13}C-dansyl chloride was synthesized as described previously[146] and the other chemicals used to synthesize this isotope reagent were purchased from Sigma-Aldrich. Dipeptide standards were purchased from Sigma-Aldrich, MP Biomedicals(Santa Ana, CA, USA) and Chem-Impex (Wood Dale, IL, USA). All the other standards were purchased from Sigma-Aldrich. LC-MS grade water, acetonitrile (ACN) and methanol (MeOH) were purchased from Thermo Fisher Scientific (Edmonton, AB, Canada).

5.2.2 Clinical sample collection.

Children, aged 2 to 10 years, suspected and diagnosed with ASD were recruited from the Department of Paediatrics and Adolescent Medicine, Duchess of Kent Children Hospital in Hong Kong. The diagnosis of autism was made by developmental pediatricians based on the diagnostic criteria in The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR). Samples were collected with informed consent from both non-autistic and autistic children using the same procedure, and handled in accordance with the Declaration of Helsinki's "Ethical Principles for Medical Research Involving Human Subjects". Altogether 400 patient samples (200 each for non-autistic and autistic children) were collected and used in system training and validation.

Venous blood samples were collected, by single-use venipuncture device in 1-mL tubes containing lithium heparin as anticoagulant. After collection, blood samples were inverted 5 times for adequate mixing of blood and anticoagulant. Samples were

centrifuged at 3500 g for 20 min at 4^{o}C. De-identified plasma samples from autistic and non-autistic children were obtained and stored at -20^{o}C until analysis.

For the blind test, additional 27 patient samples were collected following the above samples collection procedures.

5.2.3 Plasma Metabolite Extraction and Labeling.

Metabolites were extracted from plasma samples via protein precipitation with methanol. Three volumes of ice-cold methanol were added into 30 μL of plasma, vortexed and incubated on ice for 15 min. This was followed by centrifugation at 20 817 g for 15 min. The supernatants were dried using a SpeedVac and resuspended in 30 μL of 5:1 v/v water:ACN. For the labeling step, the extracted solution was mixed with sodium carbonate/sodium bicarbonate buffer and ACN. The solutions were vortexed, spun down and mixed with 25 μL freshly prepared ^{12}C-dansyl chloride solution (18 mg/mL) (for light labeling) or ^{13}C-dansyl chloride solution (18 mg/mL) (for heavy labeling). The reaction was allowed to proceed for 1 hr at 40°C. After 1 hr, NaOH was added to the reaction mixture to quench the excess dansyl chloride. The solution was then incubated at 40°C for another 10 min. Finally, formic acid in 50/50 ACN/H_2O was added to consume excess NaOH and to make the solution acidic. Experimental duplicate was carried out for each plasma sample.

5.2.4 Preparation of Labeled Samples for LC-MS Analysis.

Aliquots of individual plasma samples were mixed together to generate a pooled sample which was subsequently labeled by ^{13}C-dansyl chloride to serve as the reference sample. The individual samples were then labeled with ^{12}C-dansyl chloride and combined with an equal amount of ^{13}C-labeled reference sample prior to LC-MS analysis. Quality

control samples were prepared by 1:1 volume mix of a ^{12}C-labeled and a ^{13}C-labeled pooled sample.

5.2.5 LC-MS-based Metabolic Profiling.

The labeled metabolites were analyzed using a Bruker Maxis Impact QTOF mass spectrometer (Bruker, Billerica, MA) linked to an Agilent 1100 series binary HPLC system (Agilent, Palo Alto, CA). The samples were injected onto an Agilent reversed phase Eclipse Plus C18 column (2.1 mm × 10 cm, 1.8 μm particle size, 95 Å pore size) for separation. Solvent A was 0.1% (v/v) formic acid in 5% (v/v) acetonitrile, and solvent B was 0.1% (v/v) formic acid in acetonitrile. The chromatographic conditions were: t = 0 min, 20% B; t = 3.5 min, 35% B; t = 18 min, 65% B; t = 24 min, 99% B; t = 28 min, 99% B. The flow rate was 180 μL/min. All MS spectra were obtained in the positive ion mode with a scan range of 150 to 1000 *m/z*. The capillary voltage was 4500 V and nebulizer pressure was 1.8 bar. The dry gas flow was set to 8 l/min and the dry gas temperature was set to 230 ^{o}C. Quality control samples were analyzed between every 25 sample runs to monitor the instrument performance.

5.2.6 Data Process and Statistical Analysis.

Data obtained from LC-MS acquisition was internal-mass calibrated and processed using our in-house peak-pair picking software, IsoMS, written in R language. This program eliminated the false positive peaks, such as isotopic peaks, common adduct ions, and multiply charged ions. Only the protonated ion pairs with S/N greater than 10 were exported for further analysis. The peak ratios of individual peak pairs were calculated by the program as a measure of the concentration differences of the putative metabolites. Average values of experimental duplicates were used for data analysis in

order to minimize variations during sample preparation, data acquisition and data processing. The extracted peak-pair data from LC-MS were aligned by retention time within 30 seconds and accurate mass within 5 ppm. Only those peak-pair features shared by no less than 50% of both sample groups were retained for statistic analysis. Principle component analysis was performed using SIMCA P+12 (Umetrics, Umeå, Sweden) and volcano plot was performed by Metaboanalyst (www.metaboanalyst.ca)[241] using the criteria of $p < 0.01$ and fold change > 2. The online tool ROC Curve Explorer and Tester (www.roccet.ca)[242] was employed for receiver operating characteristic (ROC) curve analysis and model evaluation. The data were mean-centered and auto-scaled (unit variance) prior to analysis.

5.2.7 Metabolite Identification and Quantification.

The accurate masses of selected metabolites were searched against the human metabolome database (HMDB)[26] to provide a list of amine- or phenol-containing compounds within 5 ppm accuracy. The MS/MS fragmentation patterns were then obtained for these selected metabolites using scheduled MRM with a mass window of ±1 Da and collision energy of 23 eV. The scan range was 80 to 800 *m/z*. The fragmentation data were interpreted to narrow down the list to only one or a few candidates and standards correspond to these candidate compounds were purchased if available. Metabolite identification was carried out by running standards under exactly the same chromatographic conditions and matching the accurate mass, retention time and fragmentation patterns with that of the samples. For matched metabolites, 10 μM standard solutions were first labeled with ^{12}C dansyl chloride and combined with ^{13}C labeled pooled plasma. The concentration of metabolites in the pooled sample was

estimated based on the $^{12}C/^{13}C$ ratio. Then the standard solutions were diluted to 2, 1, 0.5, 0.25 and 0.125 folds of the estimated concentration to ensure the actual concentration falls in this range. The ^{12}C labeled serially diluted standard solutions were mixed with ^{13}C labeled pooled sample and a calibration curve of $^{12}C/^{13}C$ ratio versus the standard concentration was built for each standard. The threshold metabolite concentration was calculated from the corresponding calibration curve and the optimal threshold ratio (Table 5.4).

5.3 Results and Discussion

The major objective of this work is to identify a set of metabolite signatures to distinguish ASDs from the controls. In this work, differential isotope dansylation labeling combined with LC-MS has been used as the analytical platform. Aliquots of individual plasma samples were mixed together to generate a pooled sample which was subsequently labeled by ^{13}C-dansyl chloride to serve as the reference sample. The individual samples were labeled with ^{12}C-dansyl chloride and combined with an equal amount of ^{13}C-labeled reference sample for LC-MS analysis. The peak ratios of an individual peak pair found in the mass spectra were calculated to measure the concentration differences of the putative metabolite in these samples. This differential isotope labeling strategy offers three unique attributes for discovery of metabolite signatures. First of all, a larger number of metabolites can be profiled which allows a more comprehensive search for the differentiating metabolites. Secondly, metabolite quantities can be evaluated and compared with minimum bias from instrumental fluctuations as each metabolite can be referenced to the same ^{13}C internal standard.

Finally, data acquired from different batches or even different LC-MS platforms can be readily compared by adjusting different reference samples to the same reference value.

5.3.1 Analytical and biological variability.

One challenge associated with human metabolomics analysis is the analytical variability caused by the analytical method used to generate the metabolome data and the biological variability due to inherent metabolome differences of different subjects within the same group. In biomarker discovery, it is typical to apply empirical thresholds to select metabolites that show significant changes between two or more phenotypic groups. However, the importance of evaluating the effect of analytical and biological variability on selection of differentiating metabolites is often overlooked. If the variability in the sample set is very large, a more stringent threshold should be applied to avoid false biomarker discovery caused by analytical and biological variations. In this study, because of the use of a relatively large number of subjects in one group, the analytical variations of the analysis platform and the biological variations of the sample set were evaluated prior to performing the statistical analysis.

Table 5.2 summarizes the measured analytical and biological variations. For analytical variations, four experimental replicate runs were performed for one plasma sample in the control group and one sample in the ASD group. The coefficient of variations (CV) was calculated using peak ratios of the peak pairs commonly detected in all four replicate runs. The mean CV was 18% for a control sample and 19% for an ASD sample. The distribution of CVs for analytical variations was plotted in Figure 5.1A and C, which shows a high density at below 20%. This result indicates a good precision of the

current analytical platform, and the reproducibility is comparable to other LC-MS analytical platforms typically employed for metabolomics analysis.[243]

Table 5.2 Measured analytical and biological variations in plasma samples

Sample phenotype*	Replicate type	Mean CV (%)	Median CV (%)
HC	Analytical	18	14
ASD	Analytical	19	14
HC	Biological	48	45
ASD	Biological	49	40

*HC: Healthy control; ASD: Autism spectrum disorder

For biological variations, CV was calculated from all 100 samples that belong to the same group. A mean CV of 48% was obtained for the control group and 49% for the ASD group, and Figure 5.1B and D show the distribution of CVs for the biological variations. Compared to the analytical variations, there is a wider envelope of the distribution with the center shifts far to the right. This suggests that the biological variations are more significant, which would more likely cause false discoveries than analytical variations. Therefore, the magnitude of significant changes caused by biological variability was evaluated. To do this, 100 samples of the same group were randomly divided into two groups of 50 and a volcano plot was generated based on this random grouping (Figure 5.2). It can be seen that for both groups, a *p*-value of less than 0.01 could preclude most of the features from being significantly different, and when a filter of greater than two-fold change is applied, all features were excluded. Based on this result, a filter of $p < 0.01$ and fold change (FC) > 2 was applied as the threshold for discovery of metabolite signatures between two different groups, which should in principle exclude most of the biological variations inherently present in the sample sets.

5.3.2 Metabolomic profiling of plasma samples.

Metabolomic profiling was performed independently on two batches of plasma samples collected at the biomarker discovery and verification phases, each consists of plasma from 100 ASD children and 100 aged-matched controls. Experimental duplicate was carried out for each sample. On average, 760 ± 102 peak pairs or putative metabolites were detected from one sample. Principle component analysis (PCA) was first applied to the two batches to provide an overview of the dataset (Figure 5.3).

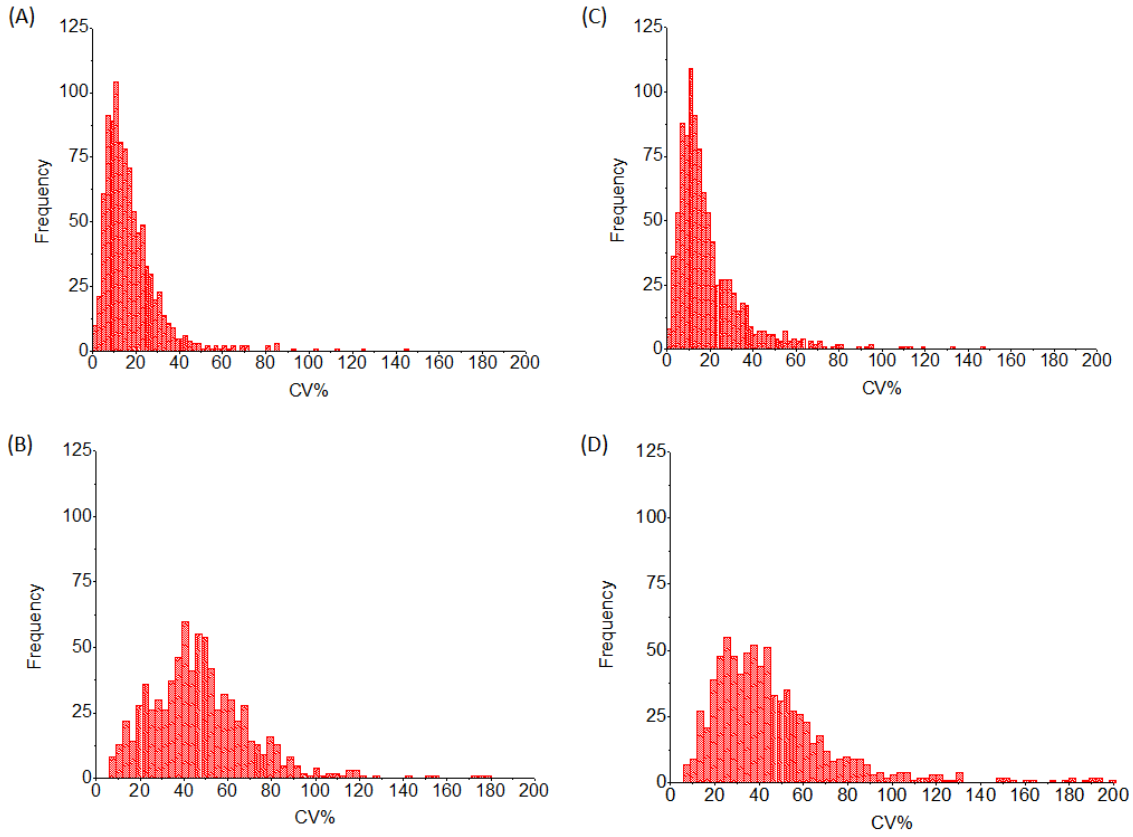


Figure 5.1 Distribution of coefficient of variations (CV) for (A) analytical variations in control group; (B) biological variations in control group; (C) analytical variations in ASD group; (D) biological variations in ASD group.

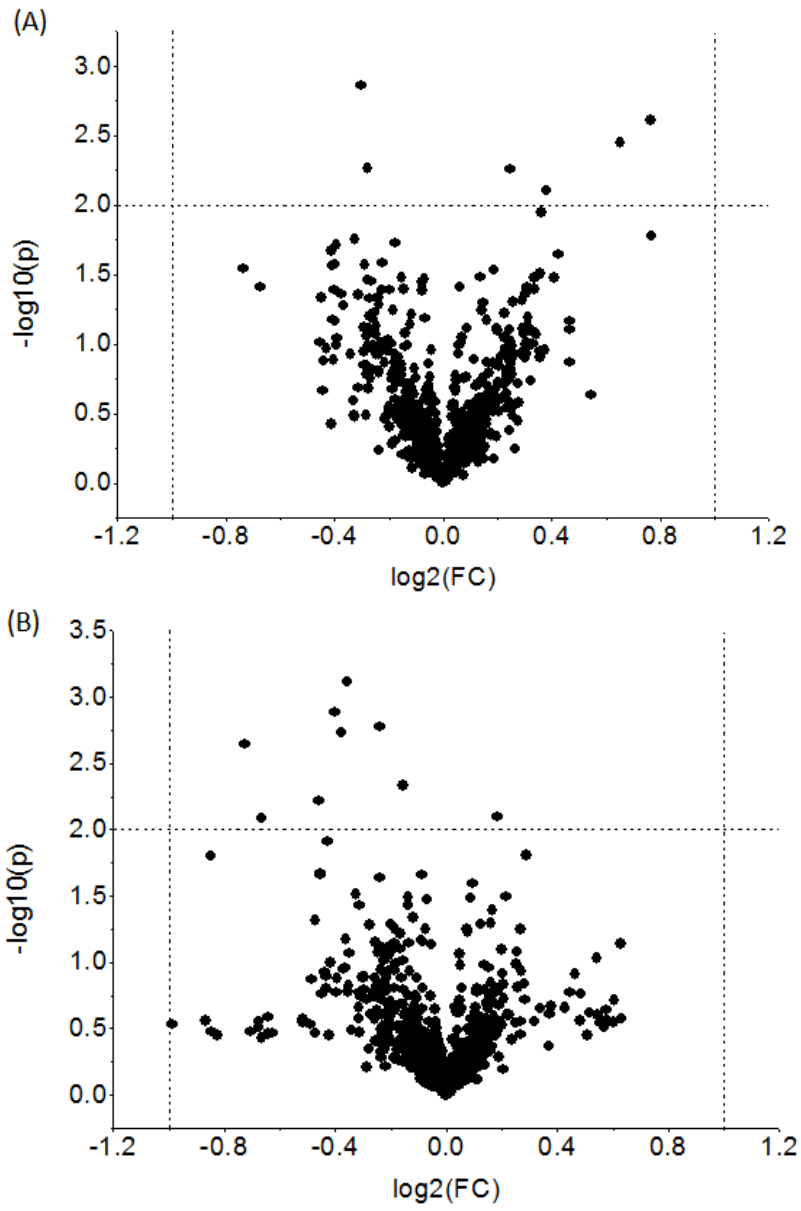


Figure 5.2 Volcano plot for within-group comparisons based on random grouping: (a) comparisons within control group; (b) comparisons within ASD group. The horizontal dotted line represents cutoff at $p = 0.01$ and the vertical dotted line represents cutoff at FC = 2.

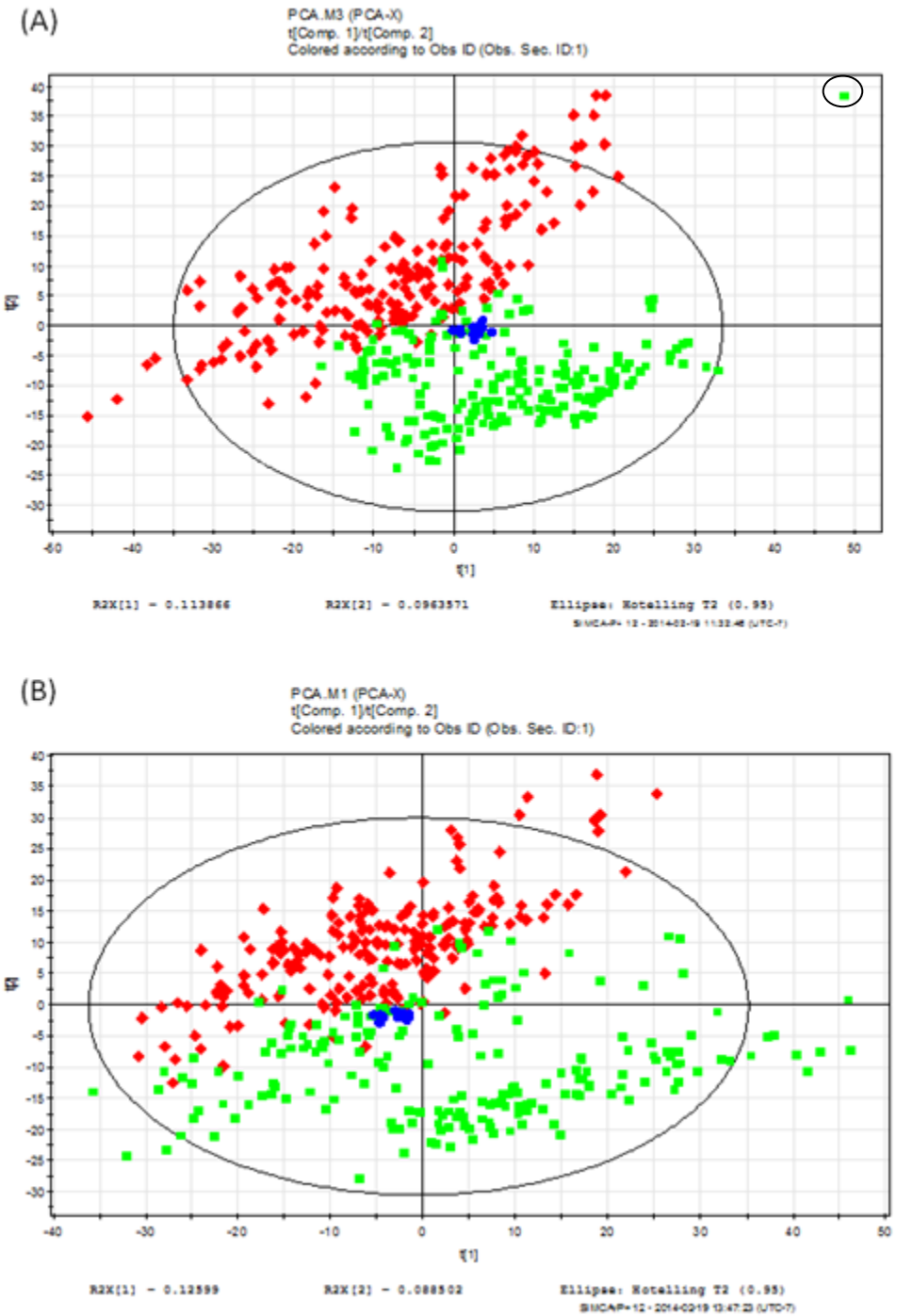


Figure 5.3 PCA score plots for (A) sample batch 1 and (B) sample batch 2. Red diamonds and green boxes represent health control and ASD groups, respectively. Quality control samples were shown as blue dots. One sample with significant deviation (circled) is identified in sample batch 1.

Three observations were made from the PCA score plot. First of all, the QC samples were clustered close together for both batches, indicating that the instrument performance was stable during the entire LC-MS analysis. Secondly, one significant outlier was detected from the first batch based on the Hotelling's T^2 (Figure 5.3A) and distance to model (DModX) measure (Figure 5.4), which was found to be caused by an experimental error. Therefore, this data was rejected from data analysis. No other data points were identified as outliers. For the following data analysis, average values of experimental duplicates were used in order to minimize variations during sample preparation, data acquisition and data processing. Since the outlier was removed from the sample set, only one experimental replicate was used for that sample. Finally, for both data sets, the control group and the ASD group were well separated, and the score plot did not show specific trends for other biological features such as age and gender (Figure 5.5). This suggests that there's a significant difference in their metabolic profiles and that the difference between the two phenotypes overwhelms other biological differences.

The data was then interrogated using volcano plot based on the criteria of FC > 2 and $p < 0.01$. As shown in Figure 5.6, 53 metabolites were selected from the first batch while 44 metabolites were selected from the second batch, with an overlap of 32 metabolites that covers 49.2% of the total number of differentially expressed metabolites. For the 33 metabolites that were selected by only one batch, 30 (90.9%) of them were still significantly different ($p < 0.01$) in the other batch, with a fold change between 1.4 and 2. This result illustrates good consistency between the two data sets. For discovery and identification of metabolite signatures, only the 32 common metabolites were studied

as they are expected to exhibit better discrimination power and therefore better prediction performances. A summary of these 32 metabolites is listed in Table 5.3.

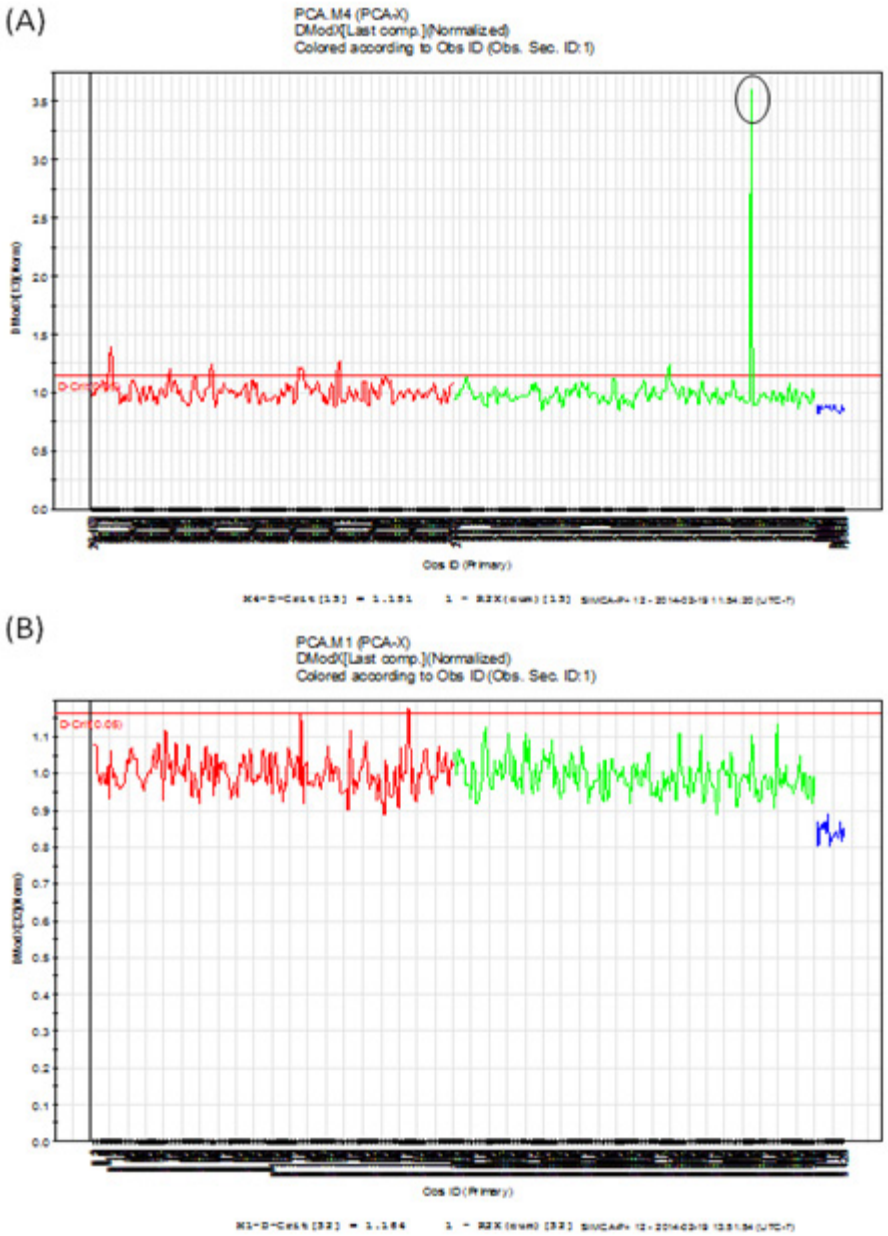
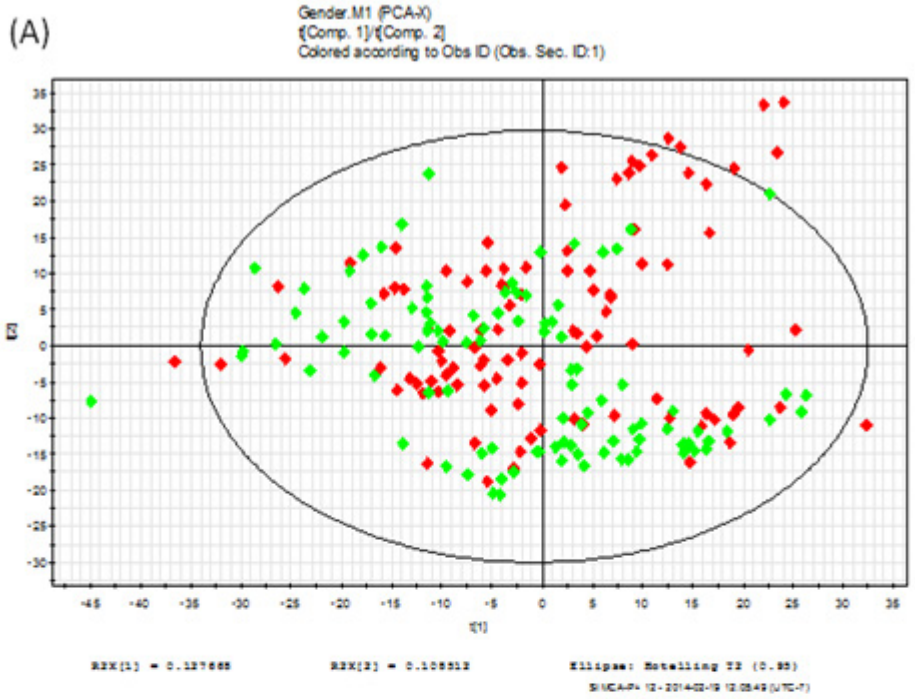
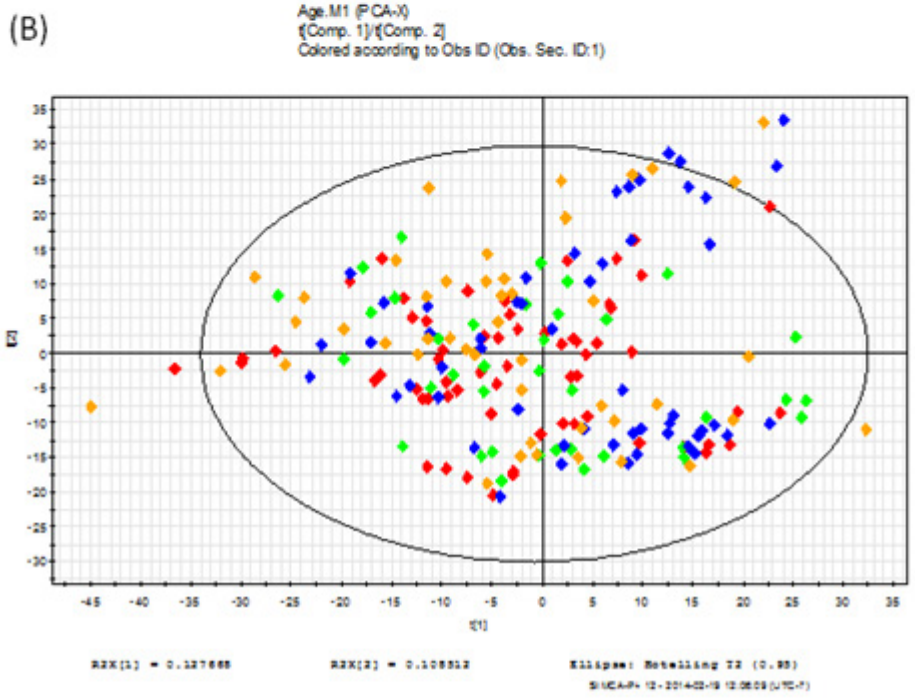


Figure 5.4 Distance to model analysis of (A) sample batch 1 and (B) sample batch 2. The x-axis represents individual observations and y-axis shows the residual standard deviation (DModX) value. Color coding is the same as in Figure 5.3. One sample with significant deviation (circled) is identified in sample batch 1.

(A)


Gender.M1 (PCA-X)
t[Comp. 1]/t[Comp. 2]
Colored according to Obs ID (Obs. Sec. ID:1)
t[2]
t[1]
R2X[1] = 0.127668
R2X[2] = 0.108512
Ellipse: Hotelling T2 (0.95)

(B)


Age.M1 (PCA-X)
t[Comp. 1]/t[Comp. 2]
Colored according to Obs ID (Obs. Sec. ID:1)
t[2]
t[1]
R2X[1] = 0.127668
R2X[2] = 0.108512
Ellipse: Hotelling T2 (0.95)

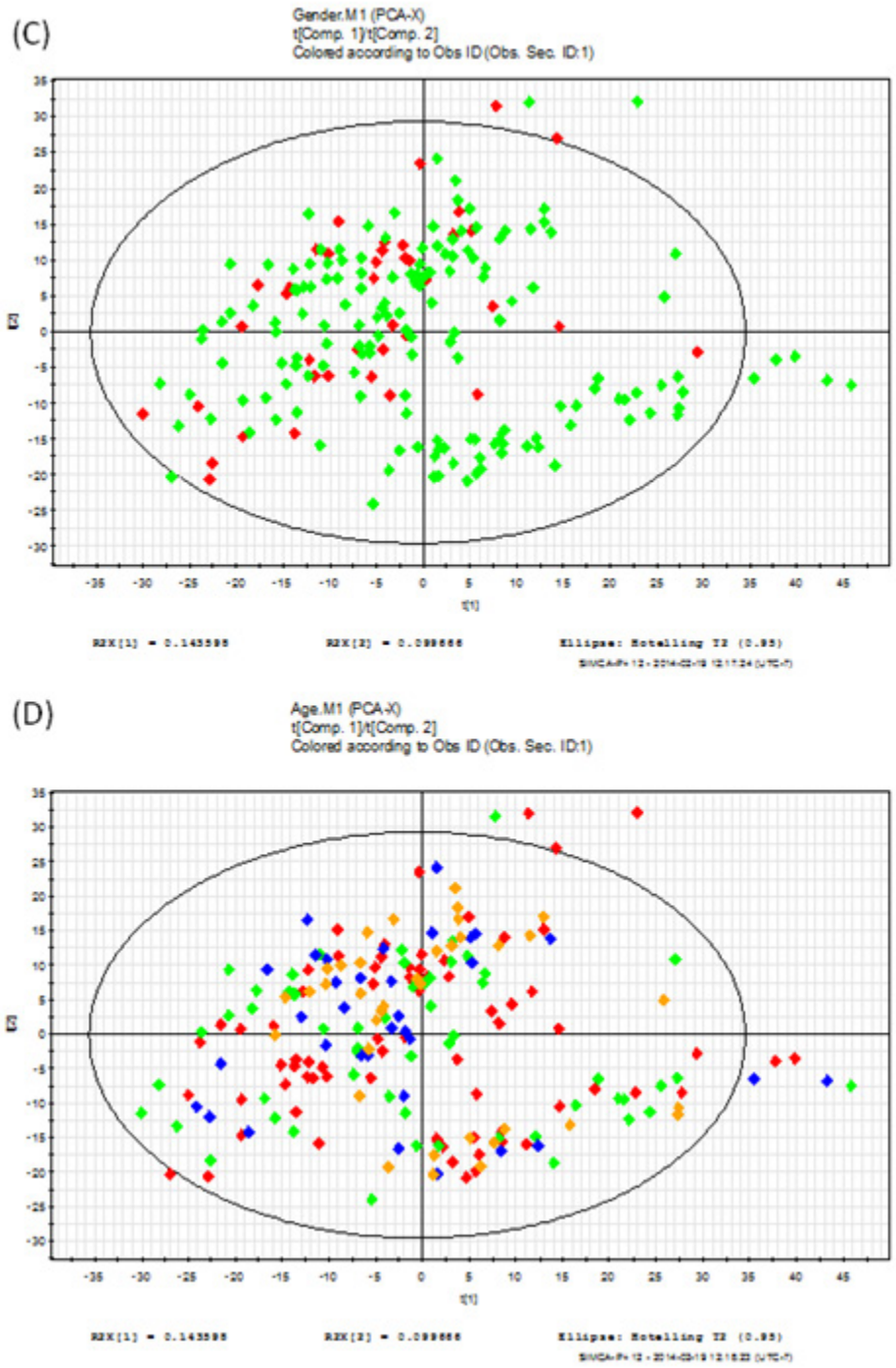


Figure 5.5 PCA score plots color coded by (A) gender in batch 1; (B) age in batch 1; (C) gender in batch 2; (D) age in batch 2. For gender, red and green represent female and male respectively. Age is divided into four intervals: 2-4, in red; 4-6, in green; 6-8, in blue; 8-10, in yellow.

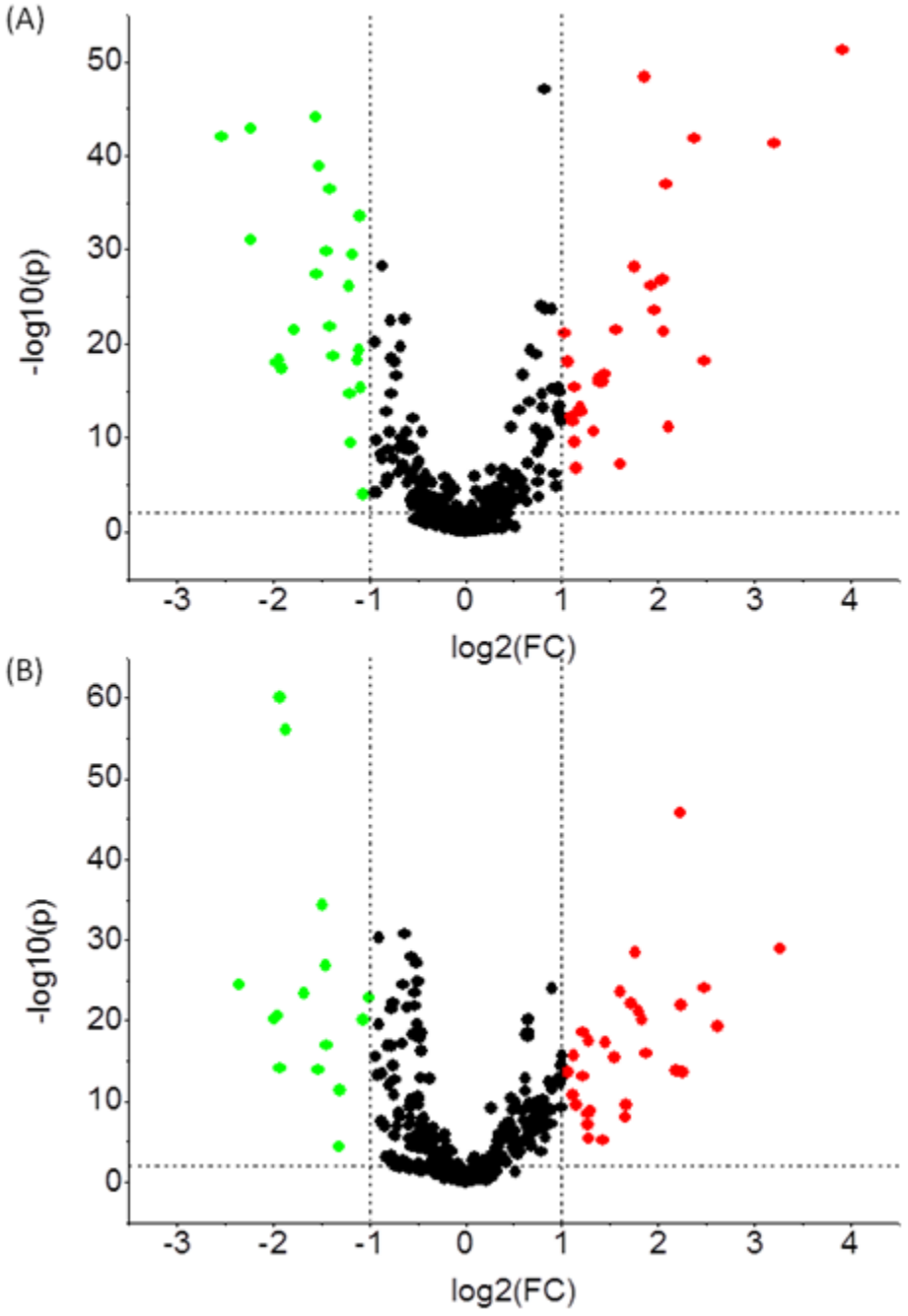


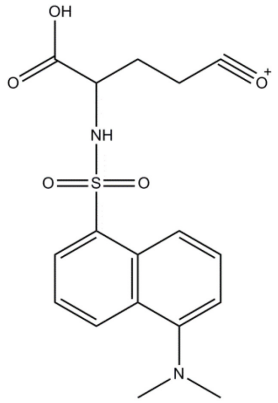
Figure 5.6 Volcano plot for inter-group comparisons: (A) comparisons for the first batch; (B) comparisons for the second batch. Metabolites that were up-regulated or down-regulated with $p < 0.01$ and FC > 2 were shown in red and green, respectively.

Table 5.3 Summary of the 32 discriminant metabolites.

	Rt (min)	Neutral mass	FC in 1st set	*p*-value in 1st set	FC in 2nd set	*p*-value in 2nd set	Compound ID	ID level*
1	2.17	169.00304	0.43	1.92E-32	0.27	8.67E-57	Cysteic acid	2
2	2.65	179.0778	2.23	7.76E-08	2.33	2.55E-19		
3	3.27	234.08364	4.24	4.16E-63	3.40	3.64E-29	γ-Glutamyl-serine	2
4	3.95	276.09451	4.15	2.46E-27	6.16	6.00E-20	γ-Glutamyl-glutamic acid	1
5	3.99	248.09949	3.38	2.40E-36	3.49	7.38E-22	γ-Glutamyl-threonine	2
6	4.01	204.07329	5.18	1.99E-61	4.70	1.88E-46	γ-Glutamyl-glycine	1
7	4.76	165.04492	3.81	1.66E-33	4.72	1.14E-22	Methionine sulfoxide	1
8	5.33	218.0889	2.20	2.64E-14	2.43	3.46E-18	γ-Glutamyl-alanine	1
9	5.63	133.03675	4.10	5.37E-34	2.45	1.56E-09	Aspartic acid	1
10	6.77	100.06326	0.35	5.02E-31	0.50	1.62E-23	Glutamine – HCOOH	1
11	7.25	218.12419	2.30	1.71E-18	2.73	6.05E-18		
12	7.69	246.12038	15.08	1.74E-87	9.62	1.08E-29	γ-Glutamyl-valine	2
13	7.72	203.03522	0.34	1.46E-54	0.48	8.50E-21		
14	8.15	212.11482	4.16	2.94E-31	3.67	1.32E-16		
15	9.35	260.13574	9.23	1.08E-68	5.59	9.38E-25	γ-Glutamyl-leucine	2
16	9.48	129.04156	3.62	1.17E-57	2.33	9.08E-14	Glutamic acid – H_2O	1
17	9.99	188.11474	3.05	4.91E-07	3.17	3.03E-10	Leucyl-glycine	1
18	10.45	426.08553	0.26	9.30E-27	0.26	7.93E-15	Cysteineglutathione disulfide	1
19	11.42	149.05013	0.26	1.86E-26	0.36	1.56E-27	Methionine	1
20	11.59	261.13035	2.19	4.31E-13	2.22	3.19E-10		
21	11.89	275.14576	5.58	2.29E-34	4.56	1.71E-14	γ-Glutamyl-ε-lysine	1
22	11.92	112.02626	0.43	1.55E-20	0.35	1.24E-14		
23	12.31	261.13037	3.89	1.60E-39	3.30	7.58E-23	γ-Glutamyl-ornithine	2
24	13.07	275.14557	4.32	1.51E-37	3.56	7.56E-21	γ-Glutamyl-lysine	2
25	13.09	297.04293	0.21	1.37E-59	0.20	3.83E-25	Cysteinylglycine disulfide	1
26	14.75	240.02169	0.27	4.52E-48	0.26	2.50E-21	Cystine	1
27	16.29	100.01529	0.34	2.86E-43	0.36	4.61E-35		
28	16.37	244.17733	2.73	2.72E-15	2.93	4.03E-16	Leucyl-leucine	1
29	21.40	167.06835	2.51	4.01E-12	4.79	2.65E-14		
30	23.48	195.00143	0.37	2.79E-31	0.37	1.11E-17		
31	24.15	195.00144	0.17	7.12E-62	0.25	7.21E-21		
32	25.09	228.96167	0.21	2.49E-56	0.31	4.85E-24		

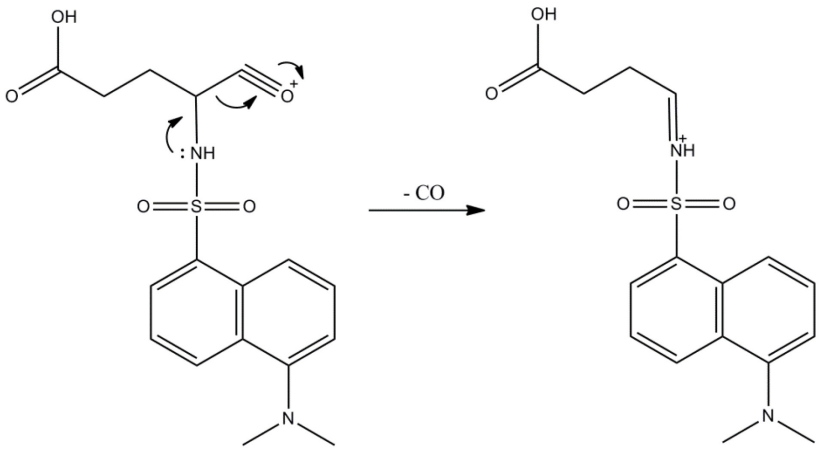
*ID level 1: Identification based on accurate mass, retention time and fragmentation match to standards; ID level 2: Identification based on fragmentation interpretation and/or prediction of retention behavior without standards.

Identification of the 32 common metabolites was based on accurate mass search within 5 ppm against the human metabolome database (HMDB), retention time information as well as MS/MS fragmentation. 21 metabolites were successfully identified, in which 14 of them were matched with authentic standards. For identification without standards, the best hit was given in the table among all isomers based on fragmentation interpretation and prediction of retention behavior. γ-Glutamyl dipeptides were identified based on the characteristic fragment peak at m/z 363.101, which is found to be absent or with very low intensity in their α-glutamyl isomers. Figure 5.7 compares the MS/MS spectra of three γ-glutamyl dipeptides and their corresponding α-glutamyl isomers. It can be seen that the fragment peak at m/z 363.101 is evident in all three γ-glutamyl dipeptides but is almost absent in the MS/MS spectra of α-glutamyl dipeptide. One proposed structure for the peak at m/z 363.101 in the γ-glutamyl dipeptide is

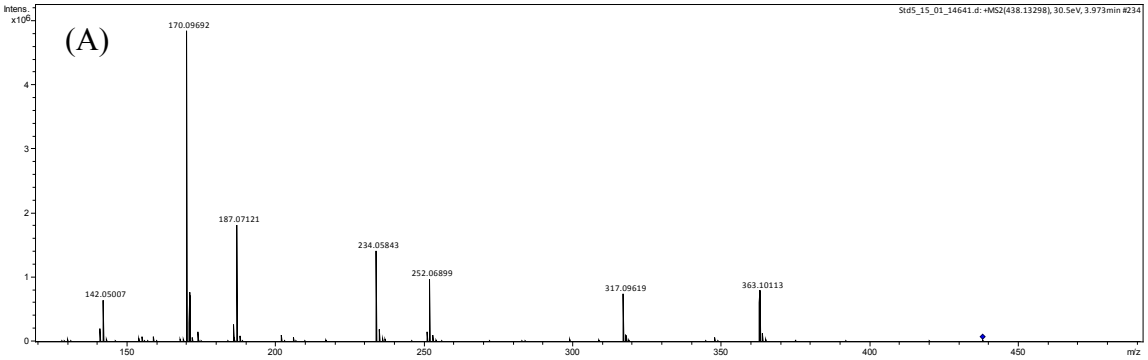


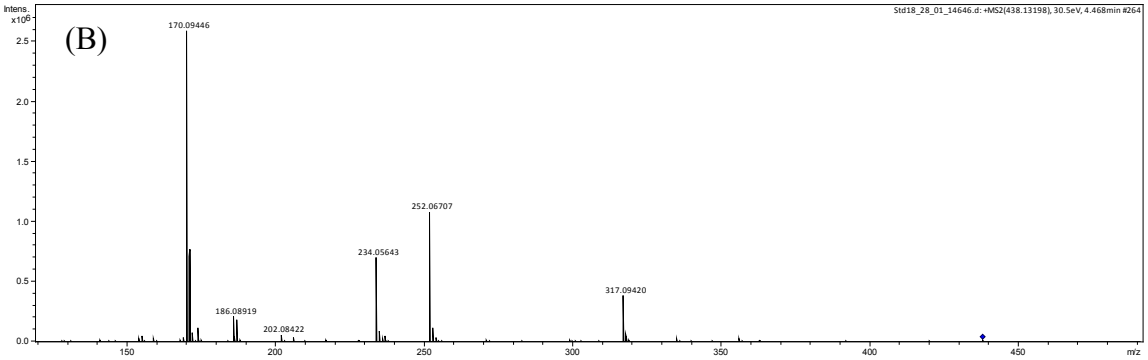
Exact Mass: 363.1009

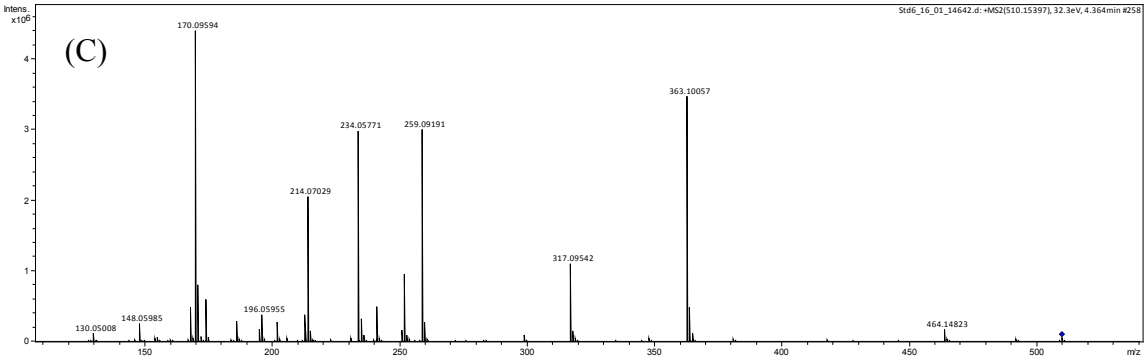
A possible explanation for the absence of this peak in the α-glutamyl isomer is that the corresponding structure is not as stable since the CO group may be readily lost by the following mechanism, and as a result the *m/z* 363.101 cannot be observed.



Exact Mass: 363.1009 Exact Mass: 335.1060







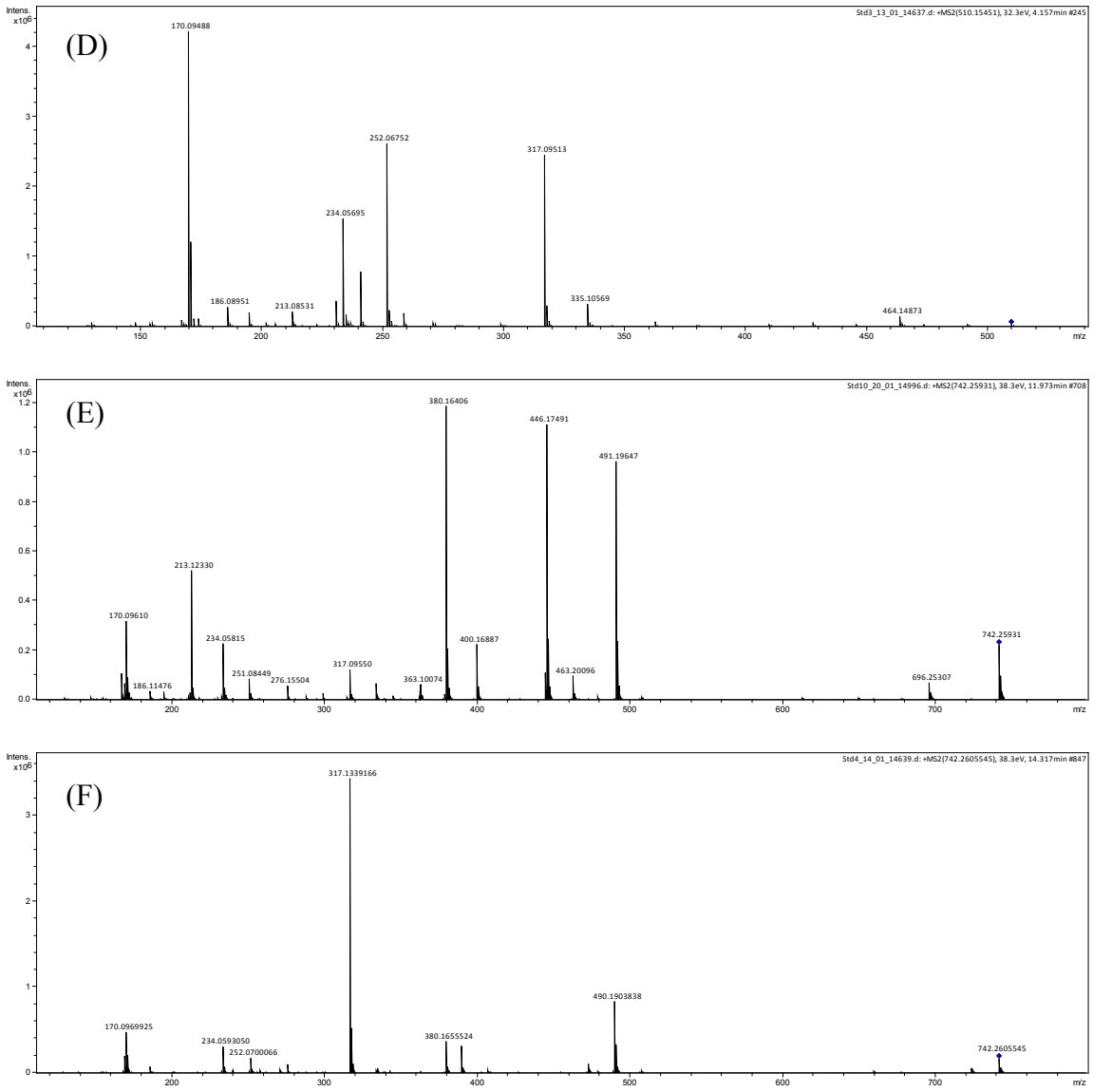


Figure 5.7 MS/MS spectra of (A) γ-Glutamyl-glycine; (B) α-Glutamyl-glycine; (C) γ-Glutamyl-glutamic acid; (D) α-Glutamyl-glutamic acid; (E) γ-Glutamyl-ε-lysine; (F) α-Glutamyl-lysine.

It is noted that some of these compounds were identified as derivatives of a metabolite (e.g., glutamic acid – H_2O), which is most likely formed during the labeling reaction as the same behavior was observed with the standards. Nevertheless, the levels of these derivatives are always positively correlated with the original metabolites and

therefore a significant change in these derivatives suggests that there's also a significant change in the original metabolite. The original metabolites were present at high levels in the sample, and as a result the mass signals were saturated. This possibly explains why these original metabolites were not picked up by the volcano plot. Thiol-containing metabolites are easily oxidized under the labeling conditions and were detected almost exclusively in the disulfide form.[244] Therefore, the amount of cystine should represent the total amount of cysteine and cystine in the sample. Cysteineglutathione disulfide and cysteinylglycine disulfide are most likely formed from their corresponding free thiols in plasma. As glutathione and cysteinylglycine are present at much lower levels in the plasma than cysteine, the plasma cysteineglutathione disulfide and cysteinylglycine disulfide levels should be positively correlated with glutathione and cysteinylglycine levels.

5.3.3 Evaluation of metabolite signatures.

In principle, all these differentially expressed metabolites could serve as metabolite signatures of ASD. However, in practice, it is always desirable to use a small set of metabolites for building a prediction model. In this experiment, data from two batches of samples were acquired and analyzed independently. The first data set was used as the discovery set to build the model, while the second data set served as the validation set to evaluate the prediction performance of the model. For each batch, the ^{13}C-labeled reference was produced by pooling equal amount of all individual samples within that batch. The peak ratios used for comparison between the control and ASD group were calculated by taking ratios between the ^{12}C and ^{13}C peaks. However, for different batches, the ^{13}C reference samples may be different and therefore the peak ratios are not directly

comparable. In order to compare the two sample sets, we prepared two adjustment samples: the first sample combines ^{12}C-labeled pool from second batch (^{12}C P2) and ^{13}C-labeled pool from first batch (^{13}C P1) in 1:1 ratio, while the second sample is a 1:1 mix of ^{12}C and ^{13}C-labeled second batch pool (^{12}C P2 and ^{13}C P2). The ratios of individual samples in the second batch (^{12}C I2) to the ^{13}C reference in the first batch (^{13}C P1) can be calculated using the following equation:

$$\frac{^{12}C\ I2}{^{13}C\ P1} = \frac{^{12}C\ I2}{^{13}C\ P2} \times \frac{^{12}C\ P2}{^{13}C\ P1} \div \frac{^{12}C\ P2}{^{13}C\ P2} \tag{5.1}$$

where the three fractions on the right of the equation are obtained from peak ratios of batch 2 and the two adjustment samples. The result is then directly comparable to the peak ratios of the first batch.

Model building and performance evaluation were accomplished using the online tool ROC Curve Explorer & Tester.[242] in two steps. In the first step, only the discovery set was used. ROC curve analysis was performed on individual metabolites. For each metabolite, the AUC value, optimal threshold and threshold concentration, as well as specificity and sensitivity at the optimal threshold were listed in Table 5.4. It should be noted that this differential isotope labeling strategy enables absolute quantification by using the ^{13}C-labeled sample as the internal standard. In this work, standard solutions were prepared at five different concentrations, and a calibration curve of ^{12}C/^{13}C ratio versus the standard concentration has been established for each standard (see Figure 5.8 for examples of leucyl-glycine and aspartic acid). The metabolite concentration in each sample can then be calculated using its corresponding ^{12}C/^{13}C ratio. Similarly, absolute concentration at the threshold has been calculated from the determined optimal threshold ratio.

Table 5.4 Summary of ROC curve analysis of 32 discriminant metabolites

	Rt (min)	Neutral mass	AUC (95% CIs)	Optimal threshold	Sensitivity	Specificity	Threshold Conc. (μM)
1	2.17	169.00304	0.955 (0.926-0.977)	0.745	0.90	0.87	
2	2.65	179.0778	0.787 (0.722-0.849)	0.832	0.64	0.91	
3	3.27	234.08364	0.989 (0.969-1)	0.606	1.00	0.96	
4	3.95	276.09451	0.999 (0.997-1)	0.188	0.98	0.99	0.42
5	3.99	248.09949	0.951 (0.92-0.973)	0.665	0.92	0.82	
6	4.01	204.07329	0.974 (0.944-0.995)	0.720	0.97	0.95	2.6
7	4.76	165.04492	0.917 (0.87-0.951)	0.618	0.80	0.91	10
8	5.33	218.0889	0.814 (0.748-0.872)	0.878	0.70	0.82	0.83
9	5.63	133.03675	0.924 (0.884-0.954)	0.740	0.79	0.93	18
10	6.77	100.06326	0.939 (0.902-0.967)	0.905	0.86	0.88	520
11	7.25	218.12419	0.913 (0.858-0.961)	0.798	0.94	0.90	
12	7.69	246.12038	1(1-1)	0.355	1.00	1.00	
13	7.72	203.03522	0.995 (0.988-0.999)	0.613	0.95	0.98	
14	8.15	212.11482	0.935 (0.896-0.964)	0.505	0.81	0.93	
15	9.35	260.13574	0.999 (0.996-1)	0.439	0.99	0.99	
16	9.48	129.04156	0.983 (0.966-0.993)	0.712	0.90	0.97	150
17	9.99	188.11474	0.868 (0.813-0.915)	0.167	0.76	0.87	0.017
18	10.45	426.08553	0.972 (0.954-0.987)	0.522	0.88	0.92	0.23
19	11.42	149.05013	0.891 (0.841-0.934)	0.615	0.75	0.95	8.0
20	11.59	261.13035	0.883 (0.832-0.932)	0.571	0.78	0.84	
21	11.89	275.14576	0.937 (0.895-0.967)	0.411	0.88	0.83	0.32
22	11.92	112.02626	0.844 (0.779-0.894)	0.728	0.84	0.72	
23	12.31	261.13037	0.971 (0.944-0.991)	0.583	0.94	0.93	
24	13.07	275.14557	0.949 (0.912-0.975)	0.655	0.94	0.88	
25	13.09	297.04293	0.997 (0.99-1)	0.476	0.97	1.00	3.8
26	14.75	240.02169	0.985 (0.967-0.997)	0.610	0.96	0.98	23
27	16.29	100.01529	0.964 (0.937-0.987)	0.712	0.94	0.92	
28	16.37	244.17733	0.983 (0.966-0.993)	0.492	0.92	0.94	0.069
29	21.40	167.06835	0.929 (0.883-0.96)	0.188	0.95	0.84	
30	23.48	195.00143	0.986 (0.964-0.998)	0.678	0.98	0.96	
31	24.15	195.00144	0.99 (0.974-1)	0.587	1.00	0.97	
32	25.09	228.96167	0.988 (0.968-1)	0.727	0.99	0.97	

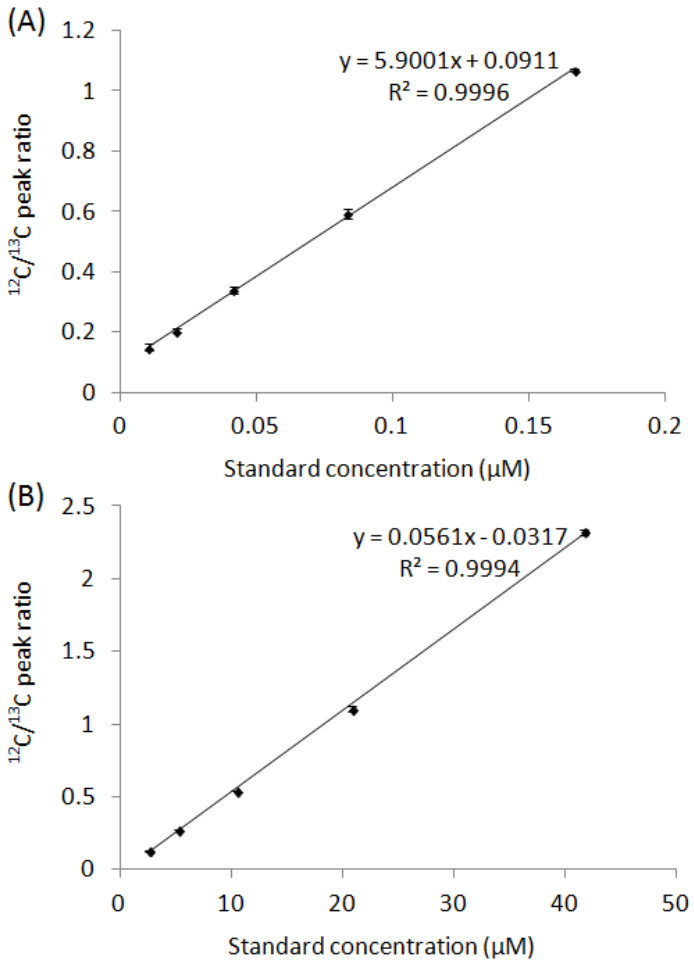


Figure 5.8 Calibration curves of $^{12}C/^{13}C$ ratio versus standard concentration for (A) leucyl-glycine and (B) aspartic acid.

Metabolites signatures were then selected using forward selection: metabolites were added to the classifiers based on decreasing AUC values one at a time until there's no significant improvement in the prediction performance as assessed by Monte-Carlo cross-validation. Linear supporter vector machine (SVM) was used as algorithm for creating and evaluating the classifier. As a result, the final model was built with 3 metabolites: γ-Glutamyl-valine, γ-Glutamyl-leucine and Cysteinylglycine disulfide (Figure 5.9).

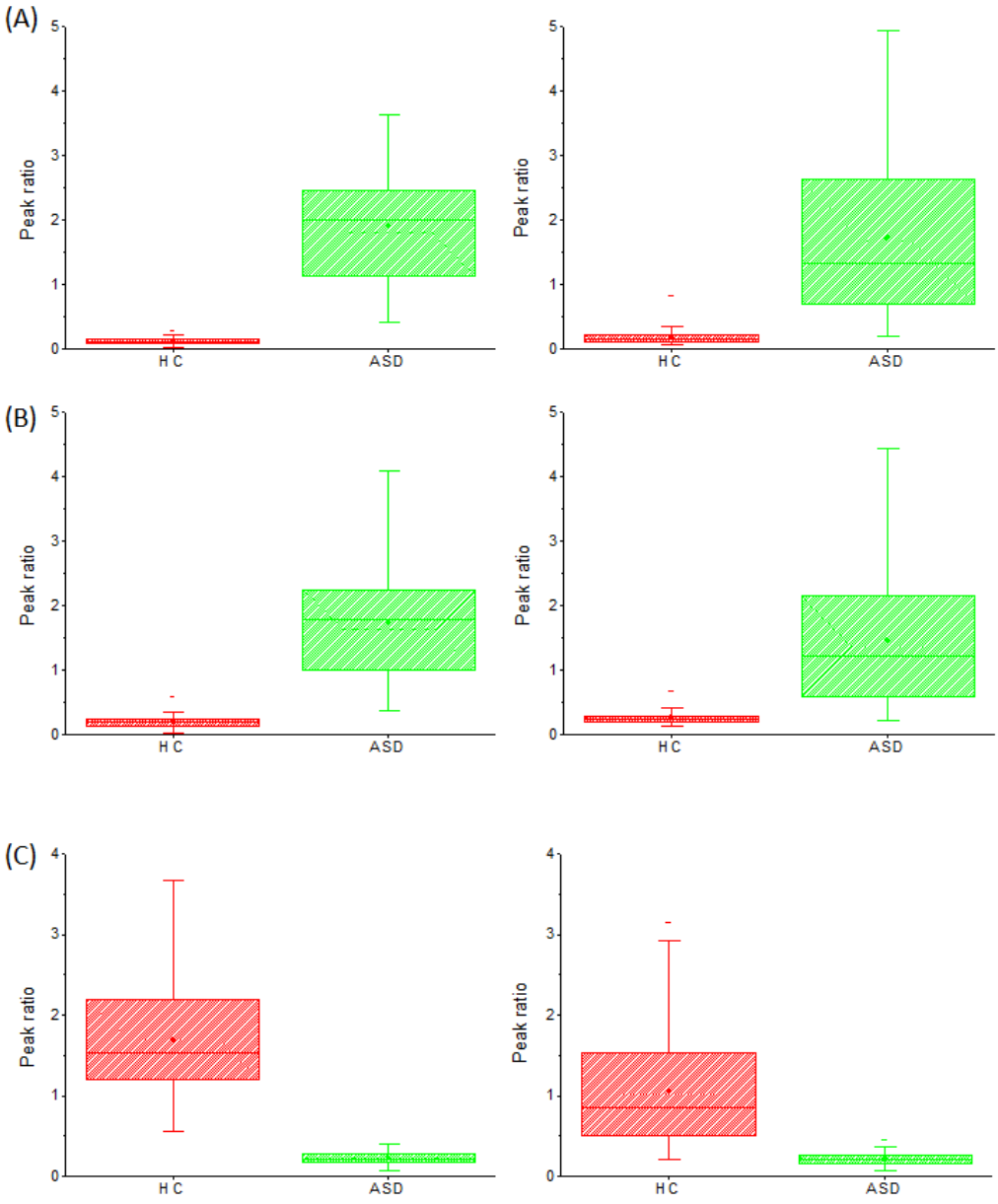


Figure 5.9 Box plots showing peak ratios of (A) γ-Glutamyl-valine; (B) γ-Glutamyl-leucine and (C) Cysteinylglycine disulfide in health control (HC) and autism spectrum disorder (ASD) groups. The range of the box is 25 to 75 percentile. The line in the box represents the median value, and the mean value is shown as a dot in the box.

In the second step, the two data sets were combined and uploaded into ROCCET. The first set served as model builder and the second set was specified as the "hold-out" sample set which did not participate in model creation but acted solely as model evaluator. The prediction performance was evaluated using the three selected metabolite signatures. The model has yielded a highly robust area under the curve (AUC) of 1 for the discovery phase and 0.992 for the validation phase (Figure 5.10A).The result showed that 190 out of 200 samples (accuracy = 0.95) were correctly classified, with a sensitivity of 0.99, a specificity of 0.91 and AUC value of 0.992. Permutation test also indicated that the model is significant with $p < 0.001$ based on 1000 permutations.

Since the selected metabolic signatures were differentially expressed with FC > 2 and $p < 0.01$ in both the discovery phase and the verification phase, it may not seem surprising to have such high prediction accuracy on the verification dataset. In order to have a better assessment on the clinical utility of these potential biomarkers, we performed a blind test on an additional set of 27 samples that is completely independent of discovery and verification phases. The 27 plasma samples from healthy and ASD children were distributed in random orders and the experiments were performed with no information on the sample type. By using the model created based on the three selected metabolic signatures in the discovery set, we were able to correctly identify all 27 samples (accuracy = 1) and obtain an AUC of 1 (Figure 5.10B). This analysis suggests the promise of using these metabolic signatures as biomarkers for clinical diagnosis of ASD with high prediction accuracy.

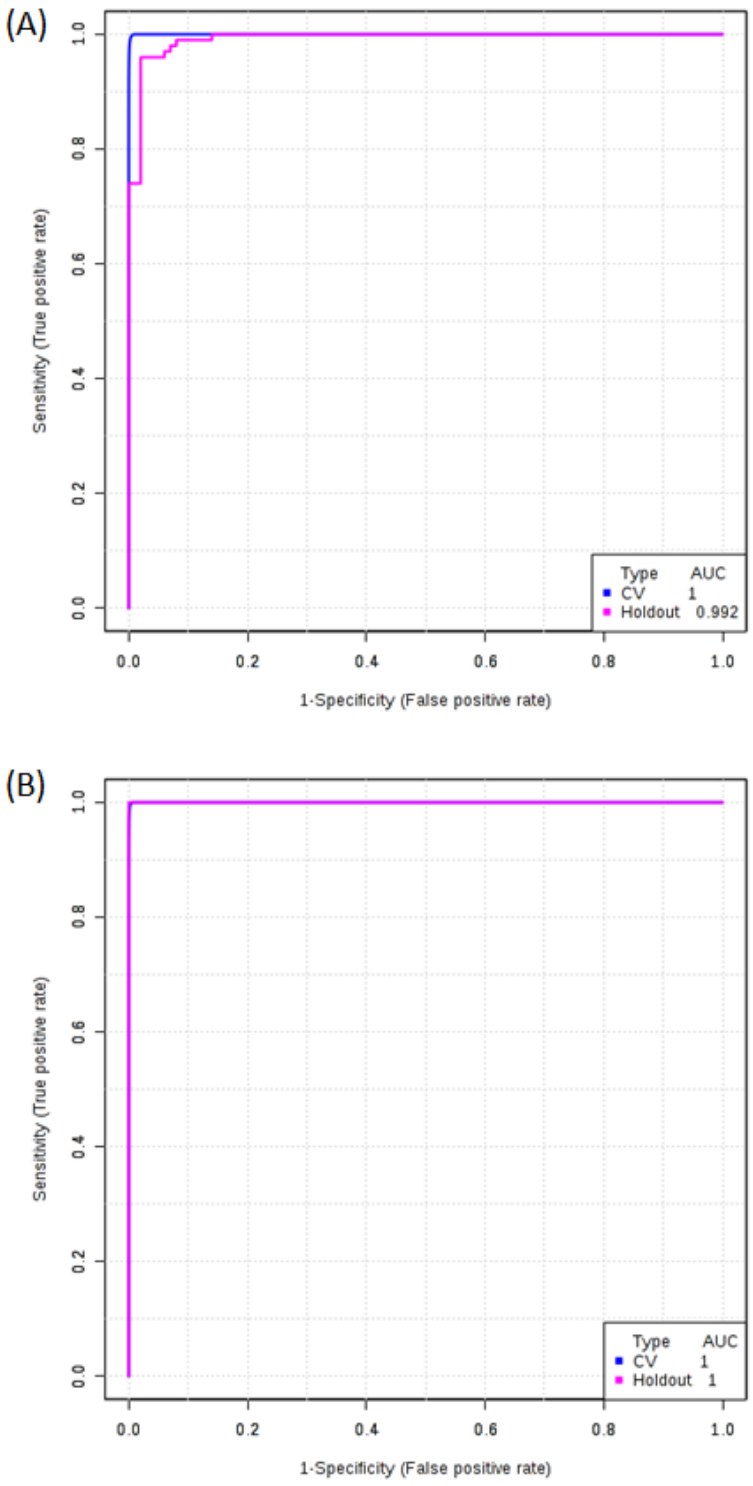


Figure 5.10 Receiver operating characteristic (ROC) analysis for classification assessment. The discovery phase was used to create the classifier model, which was

applied to (A) the validation phase (in pink), and (B) the blind test set (in pink). For the discovery phase, an ROC curve was also created, based on the averaged results of 100 cross-validations (in blue).

5.3.4 Biological implications of differentiating metabolites.

Autism spectrum disorder is a complex neurodevelopmental disease that is thought to be associated with a number of metabolism dysregulations. In this study, significant alterations of peripheral plasma metabolite levels have been observed, suggesting perturbations of several metabolic pathways, including those related to neurotransmission and oxidative stress.

We observed a significant increase of glutamate and aspartate and a decrease of glutamine in plasma of children with ASD. This finding is consistent with other studies and supports the hyperglutamatergic hypothesis of autism.[240,245,246] Glutamate is the major excitatory neurotransmitter in the brain which plays a key role in modulating synaptic plasticity and maintaining memory and learning functions. However, excess glutamate can cause excitotoxin damage and lead to abnormal development of the central nervous system.[247] The blood glutamate level has been shown to be positively correlated with brain glutamate levels.[248] Therefore, the high glutamate level in plasma suggests an excess amount of brain glutamate that would activate the glutamate receptors and result in neurotoxicity and neuron death. Aspartate is also an important excitatory amino acid which can bind to glutamate receptors and therefore may play a similar role in causing excitotoxic cell death in the brain. The increased glutamate and decreased glutamine are indicative of increased gliosis in the brain of subjects with autism.[240] Increase in gliosis is characterized by activation of astrocytes and microglia, which dysregulates the

glutamine/glutamate cycle by down-regulating expression of glutamine synthetase and up-regulating expression of glutaminase,[249,250] causing excessive conversion of glutamine to glutamate.

Increased oxidative stress has been reported to play a role in the pathogenesis of autism.[251] Many of the metabolite changes detected in this study are associated with increased oxidative stress in ASD children. One commonly observed reaction under oxidative stress conditions is the oxidation of methionine, wherein methionine can react readily with reactive oxygen species to produce methionine sulfoxide.[138] In accordance with this mechanism, our data show a remarkable reduction in methionine levels in combination with a significant increase in methionine sulfoxide concentration in autistic children. The low methionine concentration was also shown to be related to oxidative inactivation of methionine synthase[252] as well as reduced methylation capacity,[253] but the concurrent increase of methionine sulfoxide may suggest that oxidative stress is a key factor in modulating methionine concentrations.

One remarkable observation in our study is that 10 out of the 21 identified metabolites are γ-glutamyl dipeptides and all of them were up-regulated in children with ASD. In addition, we found reduced concentrations of total cysteine, glutathione and cysteinylglycine in plasma as reflected by down-regulation of their disulfides. This is an indication of dysregulation of the γ-glutamyl cycle. Under oxidative stress conditions, the activity of the membrane-bound enzyme γ-glutamyl transferase (GGT) is increased to speed up degradation of extracellular glutathione.[254] This process transfers the γ-glutamyl moiety of glutathione to an amino acid to produce a γ-glutamyl dipeptide and cysteinylglycine. The cysteinylglycine produced breaks down further into cysteine and

glycine. For autistic patients, flux from the transsulfuration pathway is reduced, leading to a decrease in cellular cysteine levels.[255] Therefore, production and uptake of extracellular cysteine into the cells is important for maintaining the intracellular redox homeostasis. The increased γ-glutamyl dipeptide levels and decreased total cysteine, glutathione and cysteinylglycine levels observed in this study are consistent with this mechanism. The fact that more than half of the identified metabolites are involved in the γ-glutamyl cycle indicates that this cycle has been strongly affected under oxidative stress conditions for patients with ASD and therefore may provide useful mechanistic information on the metabolite biomarkers discovered in this work.

5.4 Conclusion

In the present study, a robust and sensitive differential isotope labeling LC-MS method has been applied for the metabolic profiling of 400 plasma samples from children with ASD and age-matched controls. The observed metabolite changes were consistent with the proposed mechanisms of autism and some novel metabolite signatures were identified. Validation performance of these metabolite signatures has been evaluated, which shows a high accuracy. We envisage the potential of these metabolite signatures as candidate biomarkers for clinical diagnosis of ASD. Future work will focus on translation of the current technique into a high-throughput targeted monitoring method, as well as extensive validations on samples from different clinical laboratories across the world.

Chapter 6

Development of High-Performance Chemical Isotope Labeling LC-MS for Parallel Metabolomic Profiling of Cerebrospinal Fluid and Serum in Acute Spinal Cord Injury

6.1 Introduction

Acute spinal cord injury (SCI) has emerged as a major public health issue in modern society. In Canada, the incidence of acute traumatic SCI is over 1,500 per year,[256] and the estimated annual expense on SCI is over $2 billion.[257] Acute SCI often causes severe paralysis and disability, for which there are no effective treatments. Clinical assessment of SCI relies on functional measures of neurologic impairment, which are inherently imprecise, slow, and often impossible to assess in the acute setting.[258] Because of the difficulties associated with current clinical testing in acute SCI, there is clearly an urge to identify biomarkers that can accurately classify injury severity and precisely predict neurologic outcome, in order to facilitate clinical evaluation and treatment. The pathophysiologic mechanism triggered after SCI is complex and involves multiple disturbances in the human metabolic network, including oxidative stress, neuroinflammation, glycolysis, amino acid and lipid metabolism.[259-262] This emphasizes the importance of performing global profiling of the metabolic network in the biomarker discovery process.

Metabolomics is an emerging field for high-throughput global profiling of the collection of all metabolites in a biological system (i.e., the metabolome). Recently, it has been shown that metabolomic screening of rat plasma samples can be used to establish an injury severity evaluation model based on the identified metabolomic fingerprints.[263]

However, metabolomic analysis of human SCI samples has not been widely reported. While cerebrospinal fluid (CSF) is considered as a more appropriate biofluid for study SCI because of its proximity to the central nervous system and its metabolic simplicity,[264,265] it is not easily accessible in practice. In contrast, blood (plasma or serum) samples are used more widely in clinical applications because of their relative ease of collection. Although the metabolic profile in blood has been reported to resemble that in CSF in other neurological diseases,[239,266] the relationship between metabolic changes in CSF and blood following SCI, and whether these changes are significant, remain to be investigated.

Here we demonstrate, for the first time, parallel metabolomic profiling of CSF and serum from human patients for discovery of SCI biomarkers using differential dansylation isotope labeling LC-MS. This method targets at the amine- and phenol-containing sub-metabolome, and has been reported to generate high metabolome coverage for human CSF using 1:1 mixing of the same ^{12}C and ^{13}C-labeled CSF samples (i.e., $^{12}C/^{13}C$ peak ratio of 1).[49] However, the quantitative performance of this method in the case of non-uniform $^{12}C/^{13}C$ ratios, and the optimal injection amount for serum and CSF samples have not be studied. In this work, the analytical aspects of this method were first examined, including repeatability and peak detectability, to evaluate capability of this approach for high-performance metabolomic profiling of CSF and serum. The CSF and serum metabolome were then compared in terms of metabolite concentration and coverage. Finally, statistically significant metabolic changes associated SCI were identified and correlated in these two biofluids.

6.2 Experimental

6.2.1 Chemicals and Reagents.

All chemicals and solvents were purchased from Sigma-Aldrich Canada (Markham, ON, Canada) unless otherwise stated. The isotopic compound used to synthesize ^{13}C-dansyl chloride was purchased from Cambridge Isotope Laboratories (Cambridge, MA, USA). ^{13}C-dansyl chloride was synthesized in our lab as described previously.[146]

6.2.2 Clinical Sample Collection.

Intrathecal catheters were inserted into acute SCI patients with AIS A (n = 10), B (n = 11) and C (n = 7) injury severities and left in place for 5 days, during which CSF/serum samples were drawn approximately every 8 hrs. The CSF and serum samples were centrifuged at the bedside and the supernatant was immediately frozen on dry ice. The collected samples were shipped on dry ice and stored at -80°C until further use.

6.2.3 Metabolite Extraction and Labeling.

Metabolites were extracted from serum and CSF samples via methanol protein precipitation. Three volumes of ice-cold methanol were added into 25 µL of serum/CSF, vortexed and incubated on ice for 15 min. This was followed by centrifugation at 20 817 g for 15 min. The supernatants (75 µL for serum and 90 µL for CSF) were dried using a SpeedVac and resuspended in 25 µL of water. The labeling step was performed as described in Chapter 5.

6.2.4 Preparation of Labeled Samples for LC-MS Analysis.

In the analysis of analytical variability, a 90 µL aliquot was taken from one CSF sample of each patient group (AIS A, B, C and healthy control). Three of the 25 µL portions from each sample were labeled with ^{12}C-dansyl chloride as three experimental

replicates, while the remaining liquids were mixed together to generate a pooled sample which was subsequently labeled by ^{13}C-dansyl chloride. The same procedure was performed for serum samples. In all other studies, aliquots of all individual CSF/serum samples were mixed together to generate a pooled CSF/serum sample. For comparison between the serum and CSF metabolome, equal aliquots of pooled CSF and serum were combined and labeled with ^{13}C-dansyl chloride, while the pooled CSF and pooled serum samples were labeled with ^{12}C-dansyl chloride. For metabolomic profiling of the entire sample set, all individual samples were labeled with ^{12}C-dansyl chloride, and the pooled CSF and pooled serum samples were labeled with ^{13}C-dansyl chloride. Prior to LC-MS analysis, the ^{12}C-labeled samples were combined with an equal amount of the correponding ^{13}C-labeled reference. Quality control samples were prepared by 1:1 volume mix of a ^{12}C-labeled and a ^{13}C-labeled pooled sample.

6.2.5 LC-MS Analysis and Data Processing.

The labeled metabolites were analyzed using a Bruker Maxis Impact QTOF mass spectrometer (Bruker, Billerica, MA) linked to an Agilent 1100 series binary HPLC system (Agilent, Palo Alto, CA). The LC-MS conditions were the same as described in Chapter 5. Quality control samples were analyzed between every 20 sample runs to monitor instrument performance. The resulting data was internal-mass calibrated and processed as described in Chapter 3.

6.2.6 Statistical Analysis.

The extracted peak-pair data from LC-MS were aligned by retention time within 30 seconds and accurate mass within 5 ppm. Only those peak-pair features shared by no less than 50% of the samples were retained for statistic analysis. Principle component

analysis (PCA) and partial least squares discriminant analysis (PLS-DA) were performed using SIMCA P+12 (Umetrics, Umeå, Sweden). The threshold for selection of significant features in PLS-DA is VIP > 1.5. Volcano plot was performed using Excel with the criteria of $p < 0.01$ and fold change > 1.5, and analysis of variance (ANOVA) was performed by Metaboanalyst (www.metaboanalyst.ca)[241] with $p < 0.05$. The data were mean-centered and auto-scaled (unit variance) prior to analysis. Metabolite identification was based on accurate mass and retention time search against the dansyl standard library with mass difference of less than 5 ppm and retention time shift of less than 20 seconds.

6.3 Results and Discussion

Figure 6.1 illustrates the overall differential isotope labeling metabolomic profiling workflow. In this workflow, CSF and serum samples were analyzed in parallel. Prior to LC-MS analysis, each ^{12}C-labeled individual CSF or serum sample was combined with an equal amount of the corresponding ^{13}C-labeled pooled reference sample. The LC-MS data was processed by the in-house software IsoMS[154] to extract peak ratio information for each individual peak pair found in the mass spectra, and the missing values were retrieved using the zero-fill program.[267] Finally the processed data was subjected to statistical analysis for discovery of differentiating metabolites. Since each sample was only analyzed once (i.e., no experimental replicates), it is important to ensure a good peak detectability and repeatability of this analytical platform. Therefore, these parameters were first examined prior to the metabolomic profiling analysis.

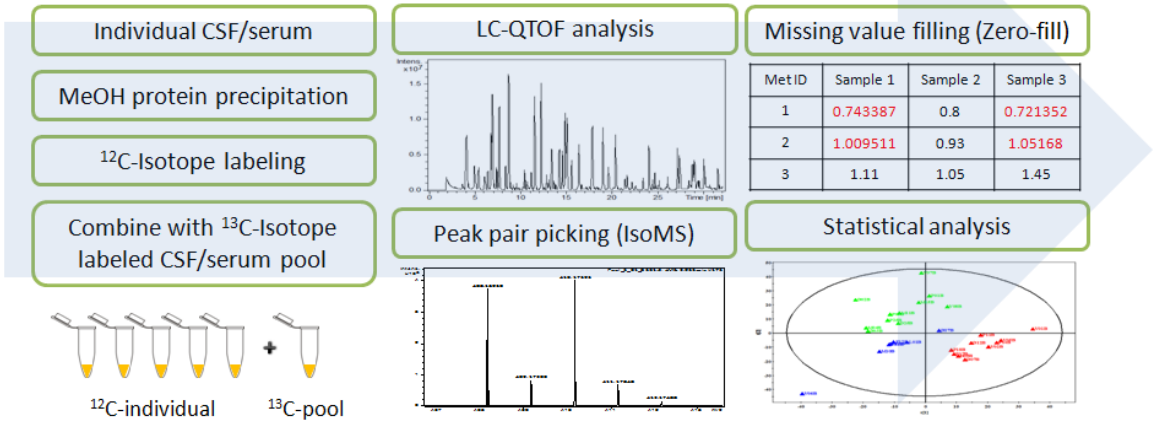


Figure 6.1 Overall workflow for differential isotope labeling metabolomic profiling of CSF and serum.

6.3.1 Evaluation of the Analytical Platform.

To evaluate the analytical variability, the coefficient of variation (CV) was determined from three experimental replicates that is a combination of variations introduced during protein precipitation, labeling, solution mixing, LC-MS detection and data processing. The CV values were calculated by using peak ratios of all peak pairs commonly detected in three replicate runs. Table 6.1 lists the median and average CV values for each sample. For all eight samples, good experimental repeatability was observed, with a median CV of less than 15% and a mean CV of 17% or less. In addition to determination of the CV values, we also examined the percentage of commonly detected peak pairs from experimental triplicates. The non-common peak pairs, or missing values, in replicate runs are usually caused by the presence of borderline metabolites, false positive peak pair identification, or other situations in which the peak picking criteria are not met (e.g., large retention time or mass shift). Therefore, the percentage of common peak pairs reflects both the reproducibility of measurements and

the robustness of the data analysis procedure. Figure 6.2 shows the number distributions of the peak pairs found in three experimental replicates for all eight samples. The percentage overlap of detected metabolites between three experimental replicates was over 92% for all samples examined, suggesting a good reproducibility of this analytical platform. Overall, analysis of the analytical variability indicates the capability of this differential isotope labeling approach for high performance metabolomic profiling of CSF and serum samples.

Table 6.1 Measured analytical variations in CSF and serum samples

Biofluid	Sample group	Mean CV (%)	Median CV (%)
CSF	AIS A	17	12
	AIS B	13	11
	AIS C	17	14
	Control	16	14
Serum	AIS A	16	12
	AIS B	17	14
	AIS C	16	12
	Control	17	14

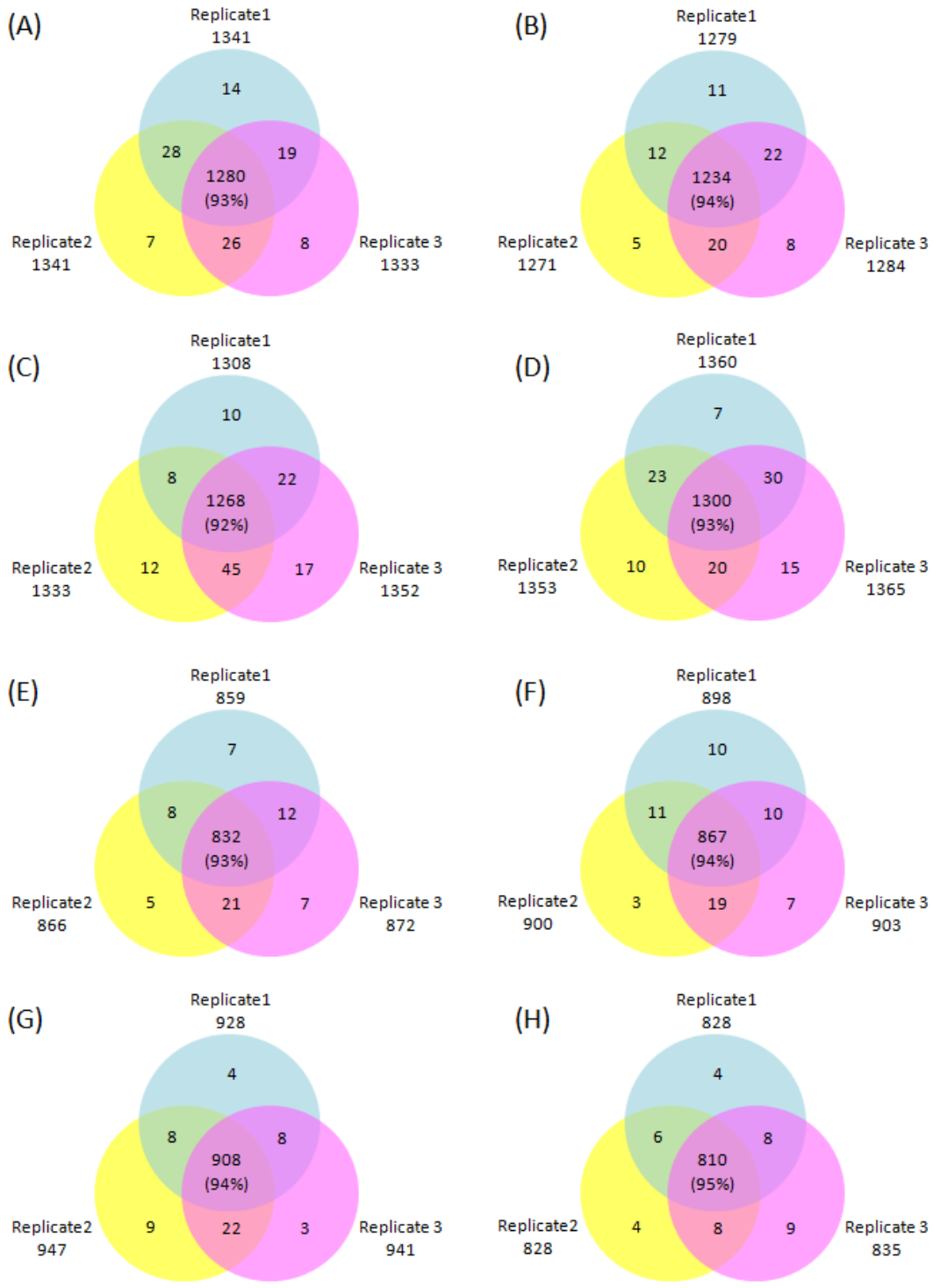


Figure 6.2 Distributions of the number of ion pairs detected in three experimental replicates for (A) serum AIS A; (B) serum AIS B; (C) serum AIS C; (D) serum control; (E) CSF AIS A; (F) CSF AIS B; (G) CSF AIS C; (H) CSF control.

In LC-MS analysis, the sample injection amount plays an important role on the number of metabolites detected. For small injection amounts, metabolites with concentrations close to the instrument detection threshold (known as borderline metabolites)[92] may not be detected. On the other hand, with large injection amounts, signals from low abundance metabolites co-eluting with a high intensity metabolite may be suppressed. Moreover, column saturation and sample carryover problems may occur. In this work, we examined the relationship between the number of peak pairs detected and the injection volume using the pooled CSF and serum samples (Figure 6.3). The injection amount can be calculated by multiplying the injection volume with the nominal total metabolite concentrations of CSF and serum determined from the calibration curve established with a mixture of 17 amino acid standards. For CSF, the number of peak pairs increases by 20% when the injection amount increases from 6 μL to 12 μL, and then levels off. This significant increase in the number of peak pairs is likely attributed to the low metabolite concentration levels in CSF samples, since many of the borderline metabolites cannot be detected when the injection amount is not sufficiently high. In contrast, for serum samples, by increasing the injection amount from 6 μL to 10 μL, the number of peak pairs increases only by 5%, and further increase in the injection amount leads to a decrease in the peak pair number. Compared to CSF, the metabolite levels in serum are considerably higher, and consequently the percentage of borderline metabolites is smaller. On the other hand, at higher injection amounts, ion suppression from high abundance ions becomes noticeable in serum samples.[268] The effect of ion suppression at higher injection amounts is illustrated in Figure 6.4. In this example, the signals of the low abundance peak pair 318.560 and 320.567 increases when the injection amount

increases from 6 μL to 8 μL, and decreases as the injection amount increases further. At an injection amount of 14 μL, this peak pair falls below the detection threshold and becomes almost non-detectable. Considering both effects (borderline metabolites and ion suppression), the optimal injection amount was chosen as the point at which the number of peak pairs levels off or starts to decrease. Based on Figure 6.2, the optimal injection volume was determined to be 12 μL for CSF and 10 μL for serum (the corresponding injection amount was 3.3 nmole and 5.7 nmole), which gave 1213 and 2316 peak pairs for CSF and serum, respectively. It is not surprising to see a smaller number of metabolites in CSF compared to serum. As the primary carrier of small molecules in the body, the metabolite composition in serum is much more complex than CSF, and it also contains a greater number of exogenous compounds. On the other hand, the type of metabolites present in CSF is restricted by the blood brain barrier, and it includes mainly neurotransmitters or related metabolites. The current CSF and serum metabolome database contains 468 and 4651 identified metabolites, respectively.[269,270] The number of peak pairs (or putative metabolites) we detected is higher than the number of identified metabolites in the CSF metabolome database. The high coverage of CSF metabolome is not surprising since many of the neurotransmitters or related metabolites are amines, and can be readily labeled with dansyl chloride. In contrast, more than 3000 serum metabolites are phospholipids and glycerolipids, which are not the targets of our isotope labeling approach. When lipids are excluded, the number of putative metabolites detected in serum using our method is also higher compared to the number of identified compounds in the database, indicating that our analytical platform provides a good coverage of both the CSF and serum metabolome.

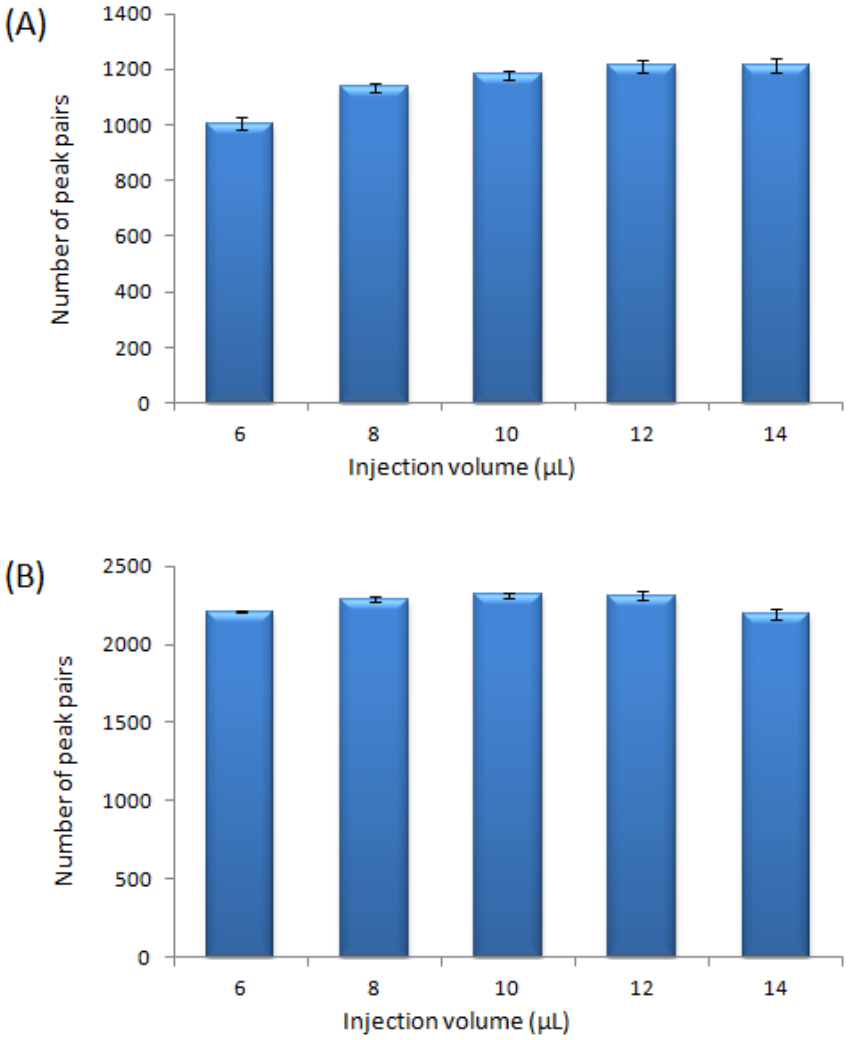


Figure 6.3 Plot of the number of peak pairs detected against the injection volume for (A) CSF and (B) serum.

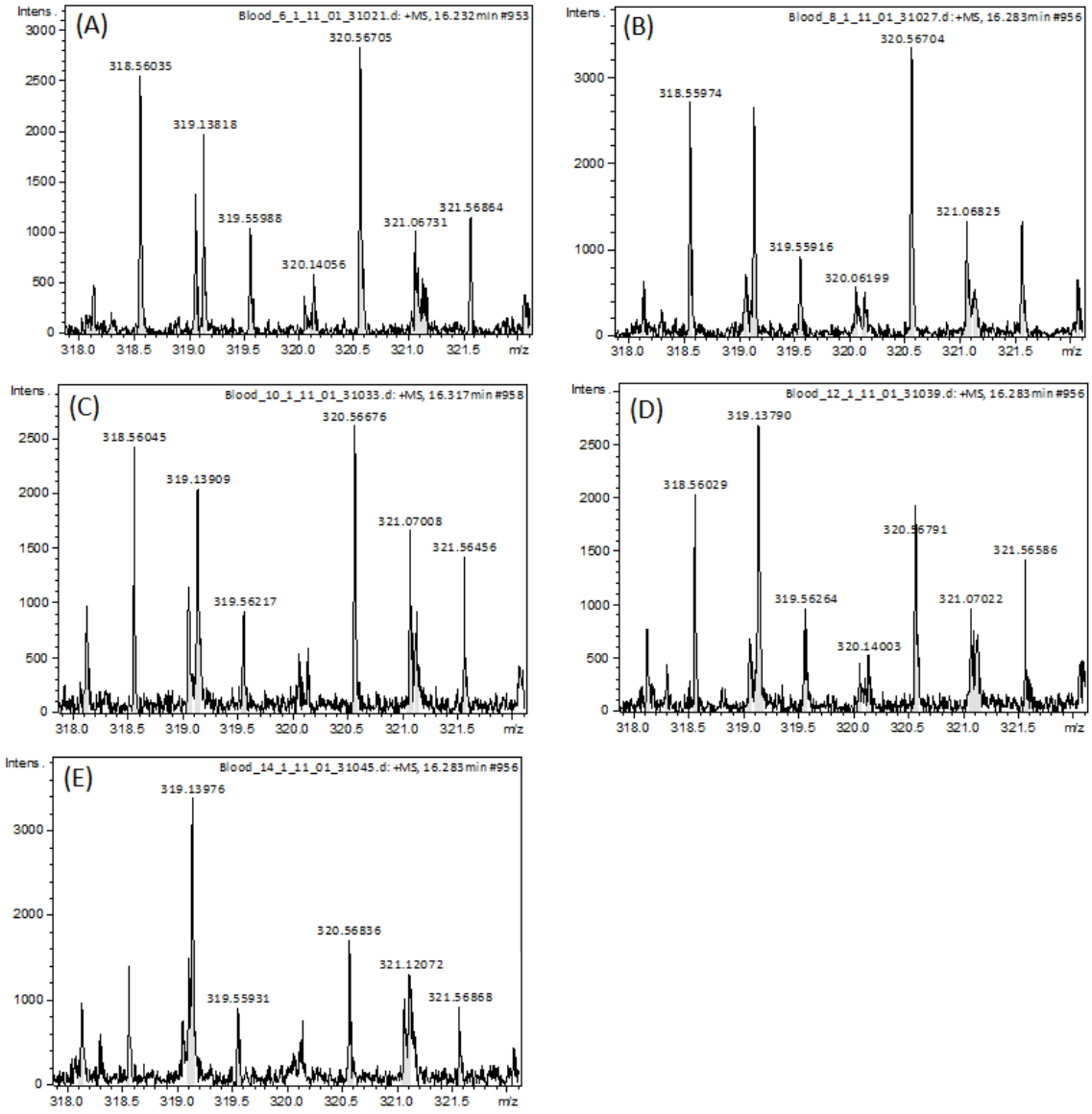


Figure 6.4 An example showing ion suppression effect with larger injection amounts for the peak pair 318.560 and 320.567. The injection amounts are (A) 6 µL; (B) 8 µL; (C) 10 µL; (D) 12 µL and (E) 14 µL.

6.3.2 Comparison of the CSF and Serum Metabolome.

Figure 6.5 shows the overlapped base peak chromatograms (BPCs) for CSF and serum. Most of the high abundance peaks correspond to labeled amino acids (indicated by stars). It is noted that while the majority of the peaks were common to CSF and serum,

the metabolite intensity in serum is much higher than that in CSF. For example, among the fifteen identified amino acids in Figure 6.5, only glutamine gives higher signal in CSF compared to serum, while all the other fourteen amino acids have much higher abundance in serum. In addition, since serum is a lipid rich biofluid, we also observed several phospholipid peaks in the BPC of serum (labeled with dots), which were barely detectable in CSF. These unlabeled compounds appear as singlet peaks in the mass spectra and therefore will not be considered in the current metabolomic profiling workflow.

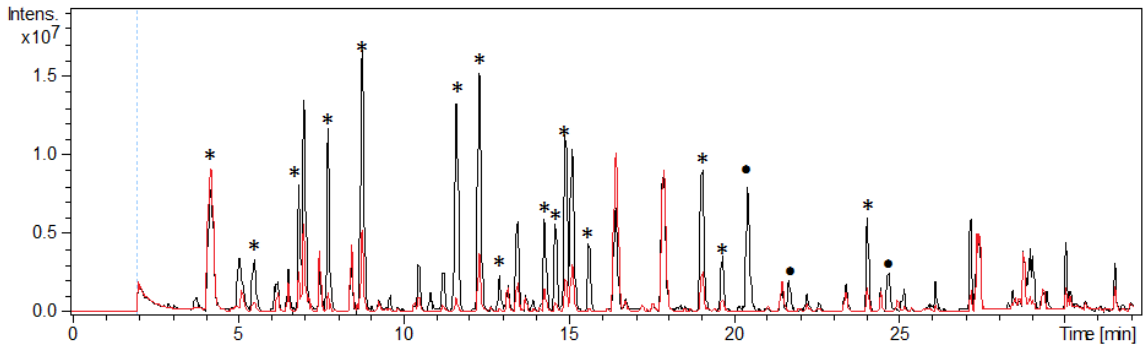


Figure 6.5 Base peak chromatograms of CSF (red) and serum (black). Peaks labeled with a star correspond to amino acids, and peaks labeled with a dot correspond to phospholipids. The first amino acid peak (eluted out at 4 min) is glutamine.

The BPC only shows metabolites with the greatest signal intensities. To compare the relative quantities of all metabolites common to CSF and serum, a "pooled pool" reference sample was generated by combing the pooled CSF and serum samples and then labeled with ^{13}C dansyl chloride. The ^{12}C-labeled CSF and serum pool samples were combined with the ^{13}C-labeled "pooled pool" reference in 1:1 volume ratio, so that the relative intensities of individual metabolites can be assessed based on the ^{12}C to ^{13}C peak ratios of each peak pair. Figure 6.6 shows a scatter plot that compares the relative

metabolite intensities in CSF and serum. For more than half of the peak pairs, the $^{12}C/^{13}C$ peak ratio is larger in serum than in CSF, indicating higher metabolite concentration in serum. In contrast, only less than 30% of the metabolites have higher concentrations in CSF. Moreover, there are a number of data points located along the y = x curve, which correspond to metabolites that have similar intensities in CSF and serum. In addition to these common metabolites, there were 190 and 27 peak pairs which could only be detected in serum and CSF, respectively. Altogether, we have identified 84 metabolites based on accurate mass and retention time match with our current standard library, which are listed in Table 6.2 according to their distributions in CSF and serum.

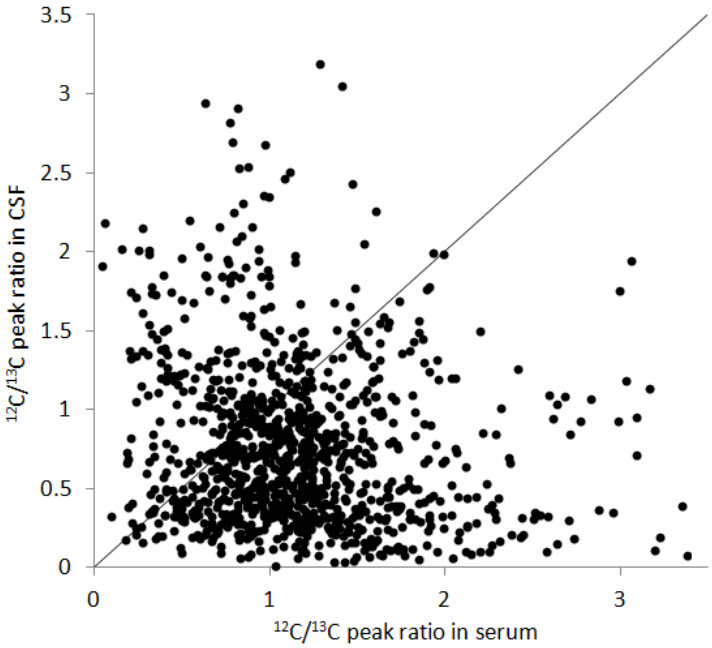


Figure 6.6 Comparison between the relative metabolite intensities in serum and CSF, expressed as $^{12}C/^{13}C$ peak ratios. Each dot represents a peak pair (or putative metabolite).

While CSF is often considered as an ideal biofluid for studying neurological disorders, the low metabolite concentration in CSF can pose a challenge to the biomarker discovery process, as the concentrations of some biologically meaningful metabolites

may fall below the instrument detection limit. For example, we found that serotonin, an important neurotransmitter, was detected as a borderline metabolite with our method. On the other hand, the higher overall metabolite concentration in serum allows for more sensitive detection and identification of biomarkers. In addition, serum also has the advantage of easy accessibility, which makes it an attractive alternative for studying central nervous system disorders.[239] Nevertheless, some of the metabolites can only be detected in CSF (e.g., the brain-specific dipeptide homocarnosine). For investigating metabolic changes associated with these compounds, analysis of CSF is necessary. Therefore, in order to obtain the most comprehensive metabolic coverage, parallel metabolomic profiling of CSF and serum of the same individuals should be the best approach.

6.3.3 Metabolomic Profiling of CSF and Serum in Spinal Cord Injury.

In this work, we performed parallel metabolomic profiling of CSF and serum to study the metabolic changes occur after acute spinal cord injury and to find discriminating metabolites in different injury scales. Principle component analysis (PCA) was first applied to provide an overview of the dataset. Figure 6.7 shows the PCA score plots for all serum and CSF samples analyzed. Quality control (QC) samples were run at the beginning, in the middle (after 18 samples) and at the end. As shown in Figure 6.7, the three QC samples were clustered close together, indicating good instrument stability throughout the LC-MS analysis. In addition, a good separation between uninjured controls and injured patients was observed for both serum and CSF samples.

Table 6.2 List of metabolites identified based on accurate mass and retention time match with dansyl standard library

Group 1[a]	Phe-Phe, Homogentisic acid, Desaminotyrosine
Group 2	3-Methylhistidine or 1-Methylhistidine, Taurine, Arginine, Homoarginine, Asparagine, Citrulline, Methionine sulfoxide, Serine, Threonine, 4-Hydroxyproline, Glutamic acid, Aspartic acid, Aminoadipic acid, Beta-Alanine, 3-Aminoisobutyric acid, 2-Aminobutyric acid, Glycine, Alanine, 5-Aminopentanoic acid, Sarcosine, Methylcysteine, Proline, Methionine, Valine, Tryptophan, Phenylalanine, Pipecolic acid, Cystathionine, 5-Hydroxylysine, Isoleucine, Leucine, Cystine, Theophylline, Salicylic acid, Ornithine, 3-Hydroxyphenylacetic acid, p-Hydroxyphenylacetic acid, Vanillylmandelic acid, Lyine, Histidine, 2-Aminooctanoic acid, Indole-3-carboxylic acid, Tyrosine, 3,4-Dihydroxybenzeneacetic acid, Serotonin, Phenylephrine or 3-Methoxytyramine, Cysteine-glutathione disulfide, Deoxyepinephrine
Group 3	Methylguanidine, Homoserine, N6-Acetyllysine, N-Alpha-acetyllysine, Glycylproline, Uridine, Pantothenic acid, Ribothymidine, Hypoxanthine, Uracil, Hydroxyphenyllactic acid, 2-Aminobenzoic acid, Acetaminophen, Homovanillic acid, 4-Hydroxybenzoic acid, Putrescine, Cadaverine, Guaiacol, Phenol, o-, m- or p-Cresol, 4-Ethylphenol, Ephedrine, 5-Hydroxytryptophan, Thymol
Group 4	Phosphoethanolamine, Glutamine, Ethanolamine, Gamma-Aminobutyric acid, Xanthine, 5-Hydroxyindoleacetic acid, Cysteamine
Group 5	Homocarnosine, gamma-Aminobutyryl-lysine

[a]Group 1: Metabolites detected only in serum; Group 2: Metabolites with higher serum concentration; Group 3: Metabolites that have similar concentration in serum and CSF ($p > 0.05$); Group 4: Metabolites with higher CSF concentration; Group 5: Metabolites detected only in CSF.

However, within the injured samples, there was no clear distinction between the three impairment scales. This indicates that significant metabolic perturbance occurs when spinal cord injury is triggered, but the metabolic changes have a relatively small dependence on the injury severity. The PCA score plots also revealed outliers in the model. For serum, a severe outlier was identified, which corresponds to an AIS C sample. We examined the most significant metabolites that differentiate this sample from others, and putatively identified (based on accurate mass search in HMDB) a drug metabolite (likely from the drug Venlafaxine) that has been highly elevated in this sample. This suggests that the metabolic differences may be attributed to drug consumption. Since this sample showed strong deviation from all other samples, it was excluded from the statistical analysis. We also observed an outlier in CSF, which belongs to the AIS B group. Nevertheless, this deviation is not as significant and it still falls in the Hotelling T^2 ellipse with 99% confidence level (data not shown). Therefore, this sample is retained in the following analysis because of the relatively small sample cohort involved in this study.

We first interrogated the data to identify metabolic differences between injured and uninjured samples using PLS-DA and volcano plot. For both serum and CSF, the two sample groups were well separated on the first principle component in the PLS-DA score plots with high R^2 and Q^2 values (Figure 6.8). 425 and 283 statistically significant putative metabolites were chosen from serum and CSF respectively, based on VIP score of greater than 1.5. In addition to PLS-DA, univariate analysis was also performed using t-test with $p < 0.01$and fold change (FC) > 1.5, which resulted in the selection of 288 and 301 putative metabolites from serum and CSF. In order to improve the confidence of

picking the true differentiating metabolites, only the metabolites commonly selected by PLS-DA and volcano plot were selected for further analysis. The overlap includes 269 and 246 putative metabolites in serum and CSF, respectively, among which 8 metabolites have been identified in each biofluid based on dansyl library search (Table 6.3). Most of the observed metabolic changes have not been reported before in spinal cord injury, but some of the changes can be rationalized. For example, O-phosphoethanolamine is a precursor in phospholipid synthesis, and decreased level of this metabolite suggests a decline in the phospholipid metabolism. This observation is consistent with the reported decrease in phospholipid levels following traumatic spinal cord injury in rat models.[271] Another example is the up-regulation of homocarnosine, which is an important brain-specific dipeptide with antioxidant activity.[272] Oxidative stress is known to contribute to the damage in spinal cord after injury and increased levels of other antioxidants have been observed in related studies.[273] In addition, because of the antioxidant activity, homocarnosine is also known to be neuroprotective,[274] and its elevation may be associated with prevention of neuronal cell death following the injury. While these metabolic changes are location specific (i.e., only observed in one of the biofluids), we have also found 33 (7%) differentiating metabolites common to both biofluids. This overlap is considerably small, suggesting that the majority of metabolic changes are not the same in these two biofuilds. Nevertheless, among the 33 commonly detected differentiating metabolites, 27 of them have similar fold changes in serum and CSF. This indicates that for most of the common differentiating metabolites, changes in their serum metabolic profile correlate positively with those in CSF. As an example, uridine was down-regulated to almost the same extent in serum (FC = 0.39) and CSF (FC = 0.29) in

SCI patients. These metabolites can therefore serve as potential biomarkers in SCI diagnosis regardless of the biofluid to use, which provides some flexibility in clinical applications.

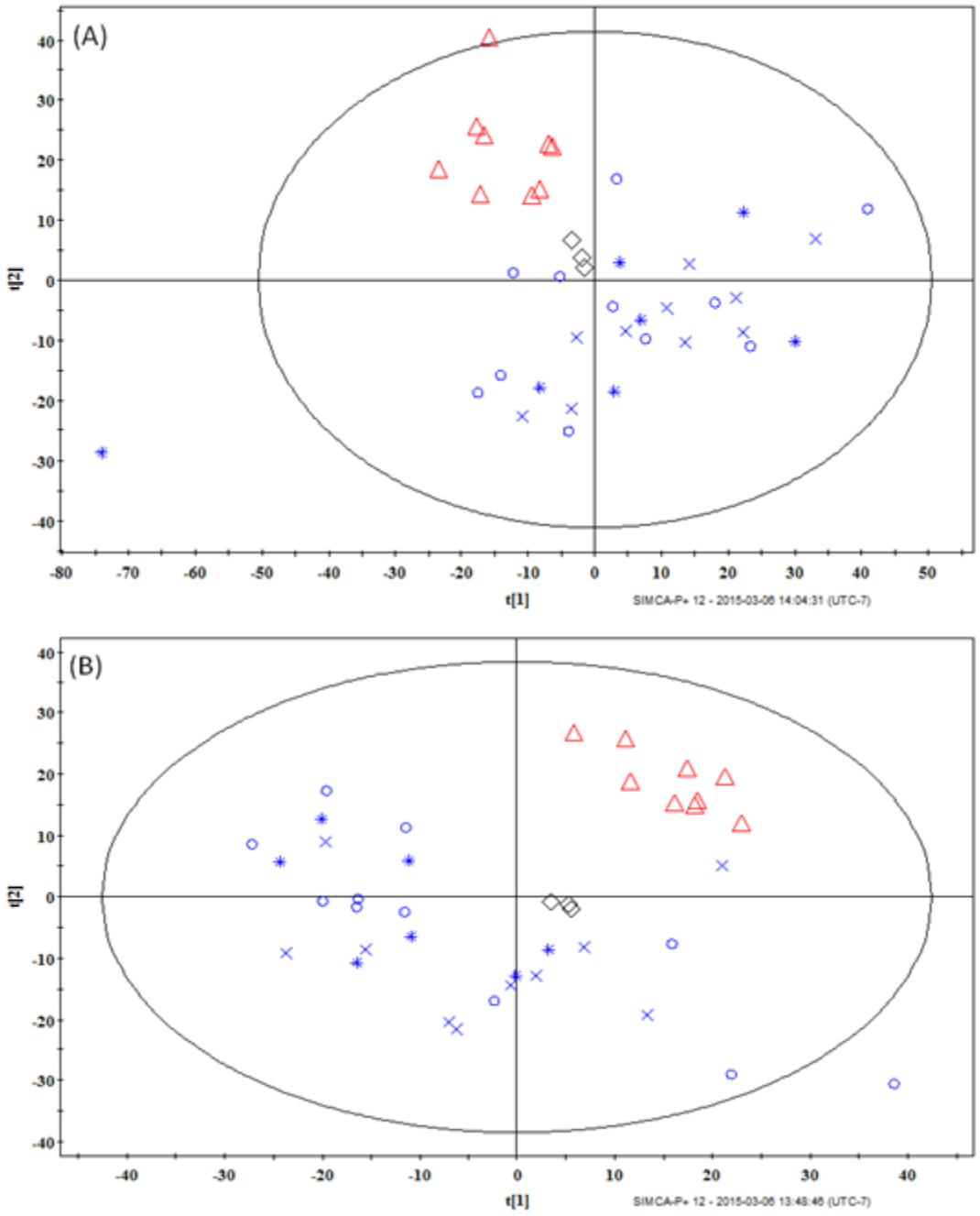


Figure 6.7 PCA score plots for (A) serum and (B) CSF. Red triangle: uninjured control; blue diamond cross: AIS A; blue circle: AIS B; blue star: AIS C; black diamond: quality control.

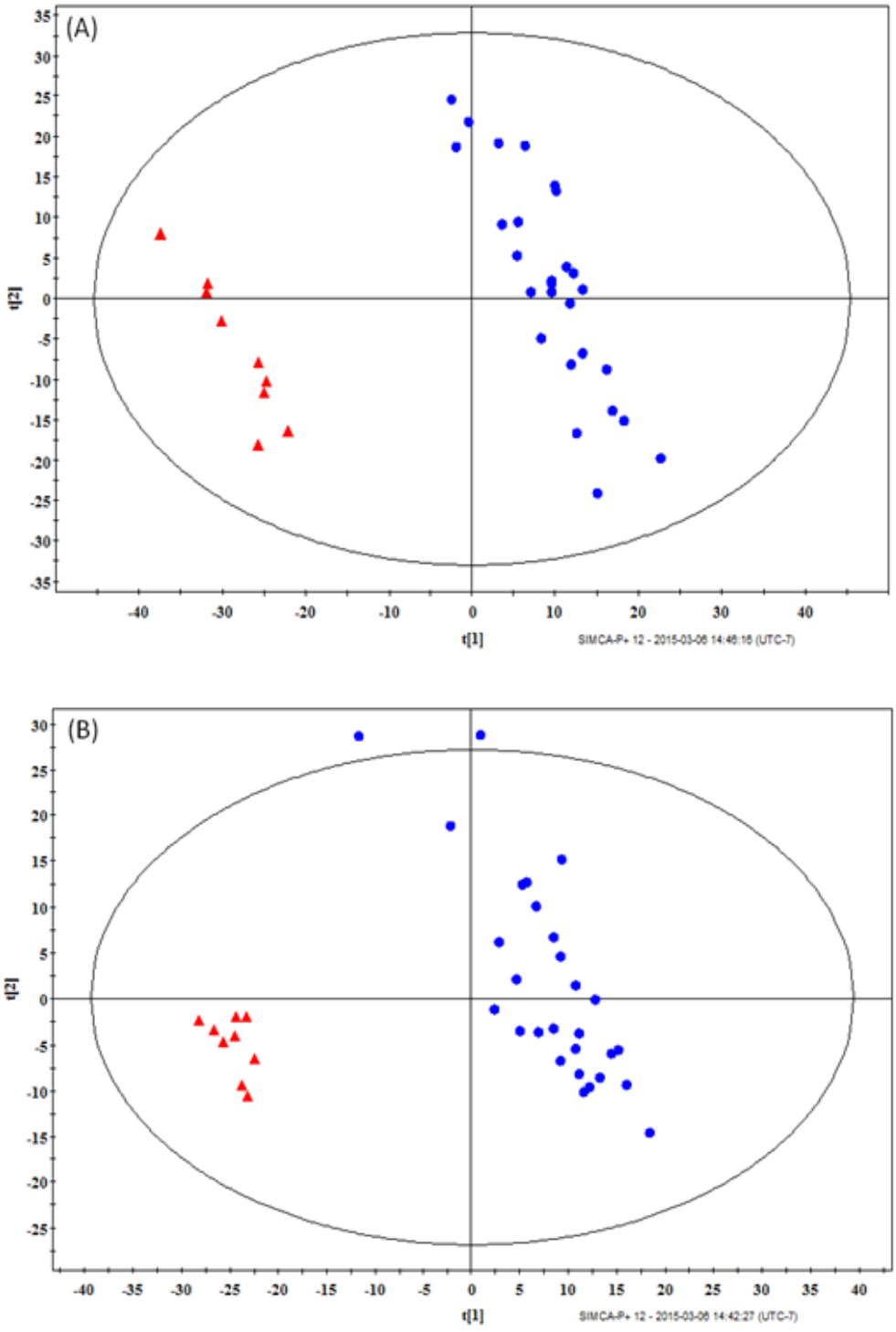


Figure 6.8 PLS-DA score plots for (A) serum ($R^2 = 1$, $Q^2 = 0.93$) and (B) CSF ($R^2 = 0.999$, $Q^2 = 0.931$). Red triangle: uninjured control; blue dot: injured patients.

Table 6.3 Summary of differentiating metabolites between injured and uninjured samples in serum (top) and CSF (bottom)

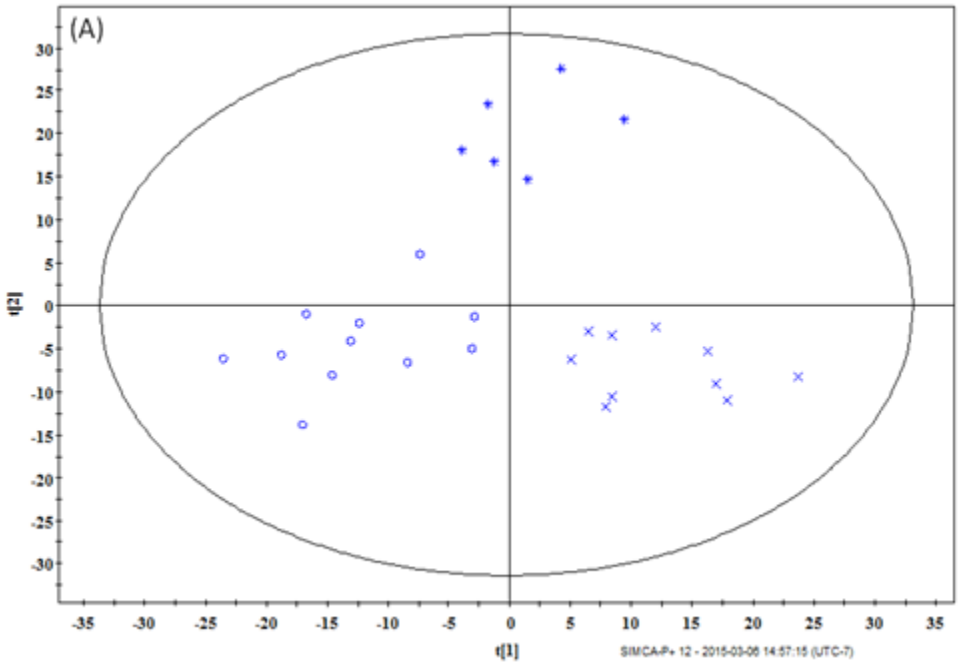
Metabolite ID	VIP score	Fold change[a]	*p* value
Uridine	2.73	0.39	1.85×10^{-10}
Acetominophen	2.73	0.11	1.33×10^{-10}
Phenylalanine	2.66	1.57	5.80×10^{-10}
4-Hydroxyproline	1.94	0.64	3.94×10^{-4}
2-Aminooctanoic acid	1.86	0.51	1.70×10^{-4}
O-Phosphoethanolamine	1.73	0.65	1.27×10^{-3}
Hydroxyphenyllactic acid	1.59	1.51	5.17×10^{-3}
Glycylproline	1.52	3.00	4.30×10^{-3}
Metabolite ID	**VIP score**	**Fold change**	***p* value**
Uridine	2.65	0.29	4.74×10^{-13}
Acetominophen	2.52	0.12	5.96×10^{-10}
Arginine	2.22	0.59	1.87×10^{-7}
Lysine	1.78	0.63	1.17×10^{-4}
N-methylaspartic acid	1.73	0.56	1.84×10^{-4}
Alpha-aminobutyric acid	1.64	1.85	1.83×10^{-3}
2-Aminooctanoic acid	1.59	0.58	4.07×10^{-3}
Homocarnosine	1.50	3.40	1.97×10^{-3}

[a]Fold change was calculated as the ratio of the average peak pair ratio in injured samples to that in uninjured controls.

We then investigated whether there are metabolic signatures for evaluation of the injury scale. We compared the metabolic profiles of AIS A, B and C samples using PLS-

DA and ANOVA. As discussed before, the difference between the three injury levels is much smaller compared to the difference between injured and uninjured samples, and no clear separation was observed in the unsupervised PCA score plot (Figure 6.7). Nevertheless, as shown in Figure 6.9, when the supervised PLS-DA model was applied, the three different patient groups can be separated with good prediction of the testing samples using 7-fold cross-validation ($Q^2 > 0.5$). We extracted metabolites with VIP score > 1.5 and determined their *p* values using ANOVA to evaluate whether there's statistical significance between group means. As a result, 57 and 29 putative metabolites were selected using the criteria of $p < 0.05$, which include metabolites that are significantly elevated in group A, B and C, respectively, or changed progressively from A to C. These metabolites may therefore be used as potential biomarkers for evaluation of injury severity. As an illustration, in each case one metabolite was selected and the average ratio distributions are plotted in Figure 6.10. However, none of these metabolites could be matched with the dansyl standard library, and putative identification based on accurate mass search against HMDB usually gives several possible structures or no matched entries. For example, the accurate mass search results for the four peak pairs illustrated in Figure 6.10 are summarized in Table 6.4, which shows that three of them are matched with more than one compound and the other one could not be matched with any compound within the specified error (5 ppm). For more confident identification, MS/MS experiments and authentic standards will be required. Nevertheless, since major objective of this work is to demonstrate the capability of this isotope labeling technique for SCI severity differentiation and to compare the CSF and serum metabolic profile, further metabolite identification was not performed at this stage. Finally, we have

detected 5 (6%) putative metabolites that were common to serum and CSF, with similar inter-group ratio distributions in the two biofluids. This overlap percentage is similar to what we observed previously for uninjured vs. injured samples, which again indicates a relatively large difference in the metabolic changes at different locations of the body in response to SCI.



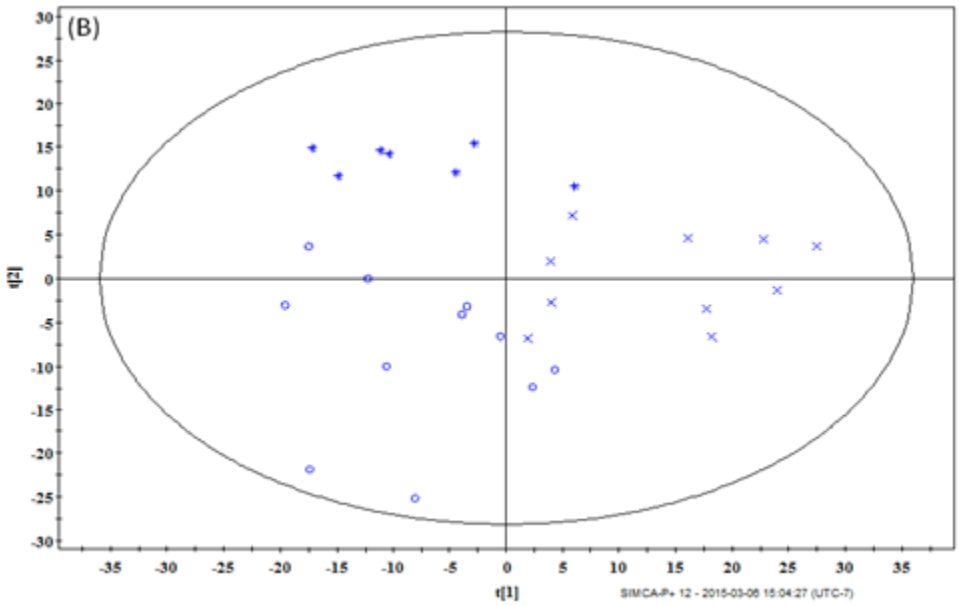


Figure 6.9 PLS-DA score plots for (A) serum (R^2 = 0.999, Q^2 = 0.633) and (B) CSF (R^2 = 1, Q^2 = 0.857). Blue diamond cross: AIS A; blue circle: AIS B; blue star: AIS C.

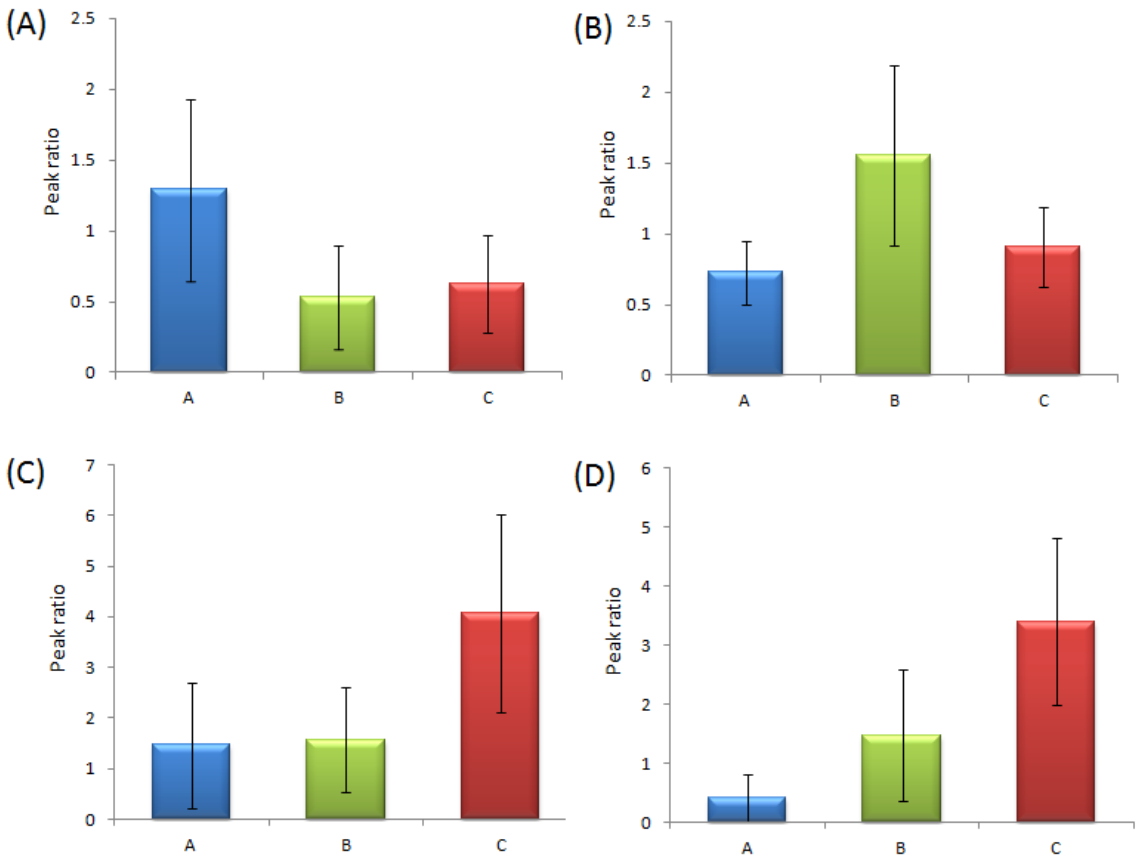


Figure 6.10 Column plot showing the average ratio of four peak pairs that show significant differences among three injury groups. The accurate mass of the four metabolites are (A) *m/z* 101.0918; (B) *m/z* 204.0729; (C) *m/z* 178.1095; (D) *m/z* 103.0634.

Table 6.4 Summary of accurate mass search results for the four peak pairs illustrated in Figure 6.10.

Peak pair	*m/z*	HMDB search result[a]	Mass error (ppm)
A	101.0918	/	/
B	204.0729	Alanyl-Aspartate	3.9
		5-L-Glutamylglycine	3.9
		L-beta-aspartyl-L-alanine	3.9
		Aspartyl-Alanine	3.9
C	178.1095	Glycinexylidide	2.6
		Pseudooxynicotine	2.6
D	103.0634	N-Ethylglycine	0.3
		O-Acetylethanolamine	0.3

[a]Only metabolites containing a primary/secondary amine or phenol group are listed. The list is further filtered by eliminating metabolites that have different retention times as in the dansyl standard library.

6.4 Conclusion

In this study, differential isotope labeling LC-MS has been applied to the metabolomic profiling of serum and CSF samples, which was shown to have good analytical repeatability and peak detectability. While the metabolic compositions in serum and CSF are similar, metabolite concentrations can vary largely in these two biofluids, and therefore parallel metabolomic profiling of CSF and serum is recommended to obtain the most comprehensive metabolite coverage. As an example of the parallel metabolomic analysis, metabolic changes associated with spinal cord injury have been examined in both serum and CSF, and a set of differentiating metabolites common to both biofluids has been discovered. Future work will focus on metabolite

identification, validation on additional sample sets, as well as time-course analysis of the differentiating metabolites, with the aim of getting a deeper insight into the pathophysiology of the injury and discovering more specific biomarkers of SCI.

Chapter 7

Conclusion and Future Work

LC-MS-based metabolomics provides a sensitive and robust methodology for comparative analysis of small molecules in different groups of biological samples. It allows simultaneous detection and quantification of a wide range of metabolites, and has been increasingly applied in biomarker discovery and system biology research. However, because of the great chemical diversity of metabolites, it is difficult to identify and quantify the entire metabolome in one shot. To address this problem, our group has undertaken the divide-and-conquer strategy to increase the metabolome coverage. The metabolome is divided into several sub-metabolomes depending on the functional groups present in the metabolites, and chemical derivatization of specific functional groups is employed to improve chromatographic separation and detection sensitivity of the LC-MS platform. In addition, chemical derivatization enables incorporation of an isotope tag into each metabolite, which can serve as an internal standard to compensate for matrix effect and instrument drift. Currently, our group has successfully developed a differential $^{12}C/^{13}C$-isotope dansylation labeling approach for the amine- and phenol-containing sub-metabolome. My thesis research focuses on applying this labeling approach to biological and clinical samples. Based on the research objectives, my thesis work is composed of two parts. The first part aims at developing a simple and fast dansyl labeling UV-based normalization method for correction of concentration variations in biological samples. The second part emphasizes on real metabolomics applications of this differential isotope labeling technique, including bacterial differentiation and disease biomarker discovery. The major achievements of each research project are summarized below.

Chapter 1 provides an overview of the basic concepts of metabolomics, LC-MS techniques, sample preparation and normalization, as well as isotope labeling techniques. In particular, a comprehensive review of major normalization methods used in MS-based metabolomics is included in this chapter.

In Chapter 2, a general approach of determining the total concentration of metabolites is developed. It is based on the use of dansyl labeling to attach a UV absorbent to the metabolites to be analyzed, followed by rapid step-gradient LC-UV detection of the labeled metabolites. It has been shown that quantification of the total labeled analytes in a biological sample facilitates the preparation of an appropriate amount of starting materials for MS analysis as well as the optimization of the sample loading amount to a mass spectrometer for achieving optimal detectability. LC-UV quantification of the labeled metabolites could be optimally performed at the detection wavelength of 338 nm. A calibration curve of a mixture of 17 dansyl-labeled amino acid standards was used to determine the total concentration of the labeled metabolites in a sample. A workflow incorporating this LC-UV metabolite quantification strategy was then developed in which all individual urine samples were first labeled with ^{12}C-dansylation and the concentration of each sample was determined by LC-UV. The volumes of urine samples taken for producing the pooled urine standard were adjusted to ensure an equal amount of labeled urine metabolites from each sample was used for the pooling. The pooled urine standard was then labeled with ^{13}C-dansylation. Equal amounts of the ^{12}C-labeled individual sample and the ^{13}C-labeled pooled urine standard were mixed for LC-MS analysis. This way of concentration normalization among different

samples with varying concentrations of total metabolites was found to be critical for generating reliable metabolome profiles for comparison.

Chapter 3 describes a simpler and more rapid method for sample amount normalization. It is also based on dansylation labeling of the amine and phenol sub-metabolome of an individual sample, but the procedure is followed by solvent extraction of the labeled metabolites and ultraviolet (UV) absorbance measurement using a microplate reader. The volume of an aliquot taken from each sample is adjusted according to the measured concentrations so that the same sample amount is taken for subsequent metabolome comparison. As an example of applications, this dansylation metabolite assay method was shown to be useful in comparative metabolome analysis of two different *E. coli* strains using the differential isotope labeling LC-MS platform. Because of the low cost of equipment and reagents and the simple procedure used in the assay, this method can be readily implemented and is applicable to many types of samples for quantitative metabolomics.

In Chapter 4, a simple and robust method for profiling the amine- and phenol-containing sub-metabolome of bacterial cells is described. The overall workflow consists of methanol-based cell lysis and metabolite extraction with ultrasonication, differential isotope dansylation labeling of cellular metabolites, and analysis of the labeled metabolites by LC–MS. Over a thousand peak pairs or putative metabolites can be detected from bacterial cells in a single LC–MS run and near 2500 putative metabolites can be found in one bacterium from combined results of multiple analyses. After careful examination and optimization of the sample preparation process, this method was shown to be effective for both Gram-positive and Gram-negative bacteria. The idea of applying

LC–UV detection to quantify the total amount of labeled metabolites was shown to be effective for normalizing the amounts of metabolites present in different samples for metabolome comparison. The use of differential isotope labeling allowed relative quantification of each individual metabolite, which facilitates comparative metabolomics studies and the generation of a metabolic fingerprint of a bacterium. Finally, this method was demonstrated to be useful for the differentiation of three bacterial species in cultured media and spiked human urine samples.

In Chapter 5, the differential isotope labeling LC-MS technique is applied for the metabolite profiling of plasma samples from autistic children (n = 200) and their age- and gender-matched controls (n = 200). A detailed biomarker discovery workflow, including evaluation of analytical and biological variations, statistical analysis, metabolite identification and absolute quantification, has been discussed. As a result, the levels of 32 metabolites were found to be significantly altered in autistic children compared to the neurotypical controls ($p < 0.01$). A total of 21 differentiating metabolites were identified, including several amino acids such as glutamic acid, glutamine and methionine, as well as a number of γ-glutamyl dipeptides. Alterations of these metabolites were thought to be related to perturbations in neurotransmission and increased oxidative stress in autistic patients. The validation performance of these metabolite signatures was evaluated and high prediction accuracy (0.95) can be achieved by using a combination of three of the metabolites. Overall, this work has lead to the promise of using these metabolites as potential biomarkers for clinical diagnosis of ASD.

Chapter 6 demonstrates application of the differential isotope labeling LC-MS approach for parallel metabolomic profiling of two important biofluids, CSF and serum,

from patients with acute spinal cord injury. Good analytical repeatability was achieved for both biofluids, based on assessment of coefficient of variation and percentage overlap between three experimental replicates. The effect of sample loading amount to the mass spectrometer was evaluated, and it was found that at high injection amounts, ion suppression became noticeable in serum samples. The optimal injection amount was determined to be 3.3 nmole for CSF and 5.7 nmole for serum. A comparison between the CSF and serum metabolome suggests that the chemical compositions of these two biofluids are similar, but there are large differences in their metabolite levels. Finally, metabolomic profiling of CSF and serum has revealed significant metabolic changes in both biofluids following spinal cord injury and moderate differences in the metabolic profiles of different injury severities.

Today metabolomics is still a relatively young research field that requires unceasing advances. While the research work described in this thesis has made positive progresses in the development and applications of the differential dansylation isotope labeling LC-MS approach for metabolomics studies, there are still ways in which the methodology can further improve. Here three important future directions will be addressed. Firstly, the metabolome coverage can be increased further in two ways. The first approach is to develop and apply new labeling chemistries that target at functional groups other than amine and phenol. For example, our group has developed dimethylaminophenacyl labeling for carboxylic acids and dansylhydrazine labeling for aldehydes and ketones. These different labeling strategies can then be applied in combination in one metabolomics study to increase the overall metabolome coverage. This is particularly useful for discovering specific biomarkers (i.e., casting a wider net to

increase the chance of detecting one or a set of specific metabolite biomarkers) as well as pathway analysis, as each metabolic pathway can involve several different classes of metabolites. However, one potential limitation of chemical derivatization methods is the matrix effect in biological samples, which may cause variations in reaction efficiency. One way to reduce the matrix effect is to perform a metabolite extraction step before labeling, so that only the metabolites of interest are subjected to derivatization. Therefore, the future work can also focus on development of metabolite extraction methods prior to chemical derivazation in order to improve labeling specificity and efficiency. Another way is to use two-dimensional LC prior to MS analysis. Two-dimensional LC allows better metabolite separation of complex biological samples, and is expected to improve mass spectrometric detection of low intensity metabolites through reduced ion suppression and matrix effects.

The second future direction is to improve metabolite identification. In our dansyl labeling applications, metabolite identification still presents a bottleneck that limits biological interpretation of observed metabolic changes. Currently we have already built a dansyl standard library containing over 280 amine and phenol compounds. However, considering the huge size of the metabolome, expanding the library by purchasing or synthesizing more standards is needed. In addition, MS/MS spectra interpretation is also helpful in identification of some dansyl-labeled metabolites, such as dipeptides. However, this approach may fail when the dansyl fragment ions dominate the spectrum or when the metabolite intensity is too low. Another possible way to improve identification is to concentrate unknown metabolites by fraction collection, followed by structure analysis with NMR.

The third future direction is to combine the untargeted metabolomic profiling with the targeted workflow. Targeted metabolomic analysis provides improved sensitivity and quantification. Once one or more metabolites associated with a certain pathway are identified to be altered by untargeted metabolomic profiling, changes in other metabolites in the same pathway can be monitored using the targeted approach. Such analysis could provide a more comprehensive picture of the biological processes of interest. Moreover, when one or more potential biomarkers have been identified, development of a high-throughput targeted metabolite monitoring approach would be required for large-scale validation and clinical applications.

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) Fiehn, O. *Plant Mol Biol* **2002**, *48*, 155.

(4) Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica* **1999**, *29*, 1181.

(5) Robertson, D. G. *Toxicol Sci* **2005**, *85*, 809.

(6) Holmes, E.; Wilson, I. D.; Nicholson, J. K. *Cell* **2008**, *134*, 714.

(7) Kaddurah-Daouk, R.; Krishnan, K. R. *Neuropsychopharmacology* **2009**, *34*, 173.

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

(9) Gomez-Casati, D. F.; Zanor, M. I.; Busi, M. V. *Biomed Res Int* **2013**, *2013*, 792527.

(10) Mishur, R. J.; Rea, S. L. *Mass Spectrom Rev* **2012**, *31*, 70.

(11) Armitage, E. G.; Barbas, C. *J Pharm Biomed Anal* **2014**, *87*, 1.

(12) Xu, X. H.; Huang, Y.; Wang, G.; Chen, S. D. *Neurosci Bull* **2012**, *28*, 641.

(13) Carrola, J.; Rocha, C. M.; Barros, A. S.; Gil, A. M.; Goodfellow, B. J.; Carreira, I. M.; Bernardo, J.; Gomes, A.; Sousa, V.; Carvalho, L.; Duarte, I. F. *J Proteome Res* **2011**, *10*, 221.

(14) Wang-Sattler, R.; Yu, Z.; Herder, C.; Messias, A. C.; Floegel, A.; He, Y.; Heim, K.; Campillos, M.; Holzapfel, C.; Thorand, B.; Grallert, H.; Xu, T.; Bader, E.; Huth, C.; Mittelstrass, K.; Doring, A.; Meisinger, C.; Gieger, C.; Prehn, C.; Roemisch-Margl, W.; Carstensen, M.; Xie, L.; Yamanaka-Okumura, H.; Xing, G.; Ceglarek, U.; Thiery, J.; Giani, G.; Lickert, H.; Lin, X.; Li, Y.; Boeing, H.; Joost, H. G.; de Angelis, M. H.; Rathmann, W.; Suhre, K.; Prokisch, H.; Peters, A.; Meitinger, T.; Roden, M.; Wichmann, H. E.; Pischon, T.; Adamski, J.; Illig, T. *Mol Syst Biol* **2012**, *8*, 615.

(15) Kim, K.; Taylor, S. L.; Ganti, S.; Guo, L.; Osier, M. V.; Weiss, R. H. *OMICS* **2011**, *15*, 293.

(16) Ren, Y.; Wang, T.; Peng, Y.; Xia, B.; Qu, L. J. *J Genet Genomics* **2009**, *36*, 621.

(17) Pan, Q.; Wang, Q.; Yuan, F.; Xing, S.; Zhao, J.; Choi, Y. H.; Verpoorte, R.; Tian, Y.; Wang, G.; Tang, K. *PLoS One* **2012**, *7*, e43038.

(18) Ramirez, T.; Daneshian, M.; Kamp, H.; Bois, F. Y.; Clench, M. R.; Coen, M.; Donley, B.; Fischer, S. M.; Ekman, D. R.; Fabian, E.; Guillou, C.; Heuer, J.; Hogberg, H. T.; Jungnickel, H.; Keun, H. C.; Krennrich, G.; Krupp, E.; Luch, A.; Noor, F.; Peter, E.; Riefke, B.; Seymour, M.; Skinner, N.; Smirnova, L.; Verheij, E.; Wagner, S.; Hartung, T.; van Ravenzwaay, B.; Leist, M. *ALTEX* **2013**, *30*, 209.

(19) Robertson, D. G.; Watkins, P. B.; Reily, M. D. *Toxicol Sci* **2011**, *120 Suppl 1*, S146.

(20) Dudley, E.; Yousef, M.; Wang, Y.; Griffiths, W. J. *Adv Protein Chem Struct Biol* **2010**, *80*, 45.

(21) Zhou, B.; Xiao, J. F.; Tuli, L.; Ressom, H. W. *Mol Biosyst* **2012**, *8*, 470.

(22) Patti, G. J.; Yanes, O.; Siuzdak, G. *Nat Rev Mol Cell Biol* **2012**, *13*, 263.

(23) Becker, S.; Kortz, L.; Helmschrodt, C.; Thiery, J.; Ceglarek, U. *J Chromatogr B Analyt Technol Biomed Life Sci* **2012**, *883-884*, 68.

(24) Liu, T.; Qian, W. J.; Gritsenko, M. A.; Xiao, W.; Moldawer, L. L.; Kaushal, A.; Monroe, M. E.; Varnum, S. M.; Moore, R. J.; Purvine, S. O.; Maier, R. V.; Davis, R. W.; Tompkins, R. G.; Camp, D. G., 2nd; Smith, R. D. *Mol Cell Proteomics* **2006**, *5*, 1899.

(25) Hegeman, A. D. *Brief Funct Genomics* **2010**, *9*, 139.

(26) 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.

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

(28) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T. *J Mass Spectrom* **2010**, *45*, 703.

(29) Werner, E.; Heilier, J. F.; Ducruix, C.; Ezan, E.; Junot, C.; Tabet, J. C. *J Chromatogr B Analyt Technol Biomed Life Sci* **2008**, *871*, 143.

(30) 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.

(31) 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.

(32) Tautenhahn, R.; Cho, K.; Uritboonthai, W.; Zhu, Z.; Patti, G. J.; Siuzdak, G. *Nat Biotechnol* **2012**, *30*, 826.

(33) 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.

(34) Jordan, K. W.; Cheng, L. L. *Expert Rev Proteomics* **2007**, *4*, 389.

(35) Barba, I.; Fernandez-Montesinos, R.; Garcia-Dorado, D.; Pozo, D. *J Cell Mol Med* **2008**, *12*, 1477.

(36) Malet-Martino, M.; Holzgrabe, U. *J Pharm Biomed Anal* **2011**, *55*, 1.

(37) Zhang, A.; Sun, H.; Wang, P.; Han, Y.; Wang, X. *Analyst* **2012**, *137*, 293.

(38) Koek, M. M.; Jellema, R. H.; van der Greef, J.; Tas, A. C.; Hankemeier, T. *Metabolomics* **2011**, *7*, 307.

(39) Halket, J. M.; Waterman, D.; Przyborowska, A. M.; Patel, R. K.; Fraser, P. D.; Bramley, P. M. *J Exp Bot* **2005**, *56*, 219.

(40) Lei, Z.; Huhman, D. V.; Sumner, L. W. *J Biol Chem* **2011**, *286*, 25435.

(41) Doerfler, H.; Lyon, D.; Nagele, T.; Sun, X.; Fragner, L.; Hadacek, F.; Egelhofer, V.; Weckwerth, W. *Metabolomics* **2013**, *9*, 564.

(42) Ciborowski, M.; Lipska, A.; Godzien, J.; Ferrarini, A.; Korsak, J.; Radziwon, P.; Tomasiak, M.; Barbas, C. *J Proteome Res* **2012**, *11*, 6231.

(43) Snyder, L. R. K., J.J.; Dolan, J.W. *Introduction to modern liquid chromatography* 3rd ed.; John Wiley & Sons, 2010.

(44) MacNair, J. E.; Lewis, K. C.; Jorgenson, J. W. *Anal Chem* **1997**, *69*, 983.

(45) Wren, S. A.; Tchelitcheff, P. *J Chromatogr A* **2006**, *1119*, 140.

(46) Carr, P. W.; Wang, X.; Stoll, D. R. *Anal Chem* **2009**, *81*, 5342.

(47) Skoog, D. A. H., F.J.; Crouch, S.R. *Principles of Instrumental Analysis*; 6th ed.; Thomson Brooks/Cole, 2007.

(48) Lan, K.; Zhang, Y.; Yang, J.; Xu, L. *J Chromatogr A* **2010**, *1217*, 1414.

(49) Guo, K.; Bamforth, F.; Li, L. *J Am Soc Mass Spectrom* **2011**, *22*, 339.

(50) Wang, N.; Xie, C.; Young, J. B.; Li, L. *Anal Chem* **2009**, *81*, 1049.

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

(52) Yamashita, M.; Fenn, J. B. *J Phys Chem* **1984**, *88*, 4451.

(53) Kebarle, P.; Tang, L. *Anal Chem* **1993**, *65*, A972.

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

(55) Taylor, G. *Disintegration of Water Drops in an Electric Field*, 1964; Vol. 280.

(56) Dole, M.; Mack, L. L.; Hines, R. L. *J Chem Phys* **1968**, *49*, 2240.

(57) Iribarne, J. V.; Thomson, B. A. *J Chem Phys* **1976**, *64*, 2287.

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

(59) Barrow, M. P.; Burkitt, W. I.; Derrick, P. J. *Analyst* **2005**, *130*, 18.

(60) Lawrence, E. O., and Edlefsen, N.E. *Science* **1930**, *72*, 1930.

(61) Hipple, J. A.; Sommer, H.; Thomas, H. A. *Phys Rev* **1949**, *76*, 1877.

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

(63) Marshall, A. G.; Comisarow, M. B.; Parisod, G. *J Chem Phys* **1979**, *71*, 4434.

(64) Amster, I. J. *J Mass Spectrom* **1996**, *31*, 1325.

(65) Comisarow, M. B. *J Chem Phys* **1978**, *69*, 4097.

(66) Scigelova, M.; Hornshaw, M.; Giannakopulos, A.; Makarov, A. *Mol Cell Proteomics* **2011**, *10*, M111 009431.

(67) Ledford, E. B., Jr.; Rempel, D. L.; Gross, M. L. *Anal Chem* **1984**, *56*, 2744.

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

(69) March, R. E. *J Mass Spectrom* **1997**, *32*, 351.

(70) Zivkovic, A. M.; Wiest, M. M.; Nguyen, U. T.; Davis, R.; Watkins, S. M.; German, J. B. *Metabolomics* **2009**, *5*, 507.

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

(72) Sansone, S.-A.; Fan, T.; Goodacre, R.; Griffin, J. L.; Hardy, N. W.; Kaddurah-Daouk, R.; Kristal, B. S.; Lindon, J.; Mendes, P.; Morrison, N.; Nikolau, B.; Robertson, D.; Sumner, L. W.; Taylor, C.; van der Werf, M.; van Ommen, B.; Fiehn, O.; Members, M. S. I. B. *Nat Biotechnol* **2007**, *25*, 846.

(73) Fiehn, O.; Robertson, D.; Griffin, J.; van der Werf, M.; Nikolau, B.; Morrison, N.; Sumner, L. W.; Goodacre, R.; Hardy, N. W.; Taylor, C.; Fostel, J.; Kristal, B.; Kaddurah-Daouk, R.; Mendes, P.; van Ommen, B.; Lindon, J. C.; Sansone, S. A. *Metabolomics* **2007**, *3*, 175.

(74) Alvarez-Sanchez, B.; Priego-Capote, F.; Luque de Castro, M. D. *Trac-Trend Anal Chem* **2010**, *29*, 111.

(75) Boudah, S.; Paris, A.; Junot, C. In *Metabolomics Coming of Age with Its Technological Diversity*; Rolin, D., Ed. 2013; Vol. 67, p 159.

(76) Yu, Z. H.; Kastenmuller, G.; He, Y.; Belcredi, P.; Moller, G.; Prehn, C.; Mendes, J.; Wahl, S.; Roemisch-Margl, W.; Ceglarek, U.; Polonikov, A.; Dahmen, N.; Prokisch, H.; Xie, L.; Li, Y. X.; Wichmann, H. E.; Peters, A.; Kronenberg, F.; Suhre, K.; Adamski, J.; Illig, T.; Wang-Sattler, R. *PLoS One* **2011**, *6*.

(77) Denery, J. R.; Nunes, A. A. K.; Dickerson, T. J. *Anal Chem* **2011**, *83*, 1040.

(78) Nakamizo, S.; Sasayama, T.; Shinohara, M.; Irino, Y.; Nishiumi, S.; Nishihara, M.; Tanaka, H.; Tanaka, K.; Mizukawa, K.; Itoh, T.; Taniguchi, M.; Hosoda, K.; Yoshida, M.; Kohmura, E. *J Neurooncol* **2013**, *113*, 65.

(79) Wikoff, W. R.; Pendyala, G.; Siuzdak, G.; Fox, H. S. *J Clin Invest* **2008**, *118*, 2661.

(80) Wei, J.; Xie, G.; Zhou, Z.; Shi, P.; Qiu, Y.; Zheng, X.; Chen, T.; Su, M.; Zhao, A.; Jia, W. *Int J Cancer* **2011**, *129*, 2207.

(81) Bruce, S. J.; Tavazzi, I.; Parisod, V.; Rezzi, S.; Kochhar, S.; Guy, P. A. *Anal Chem* **2009**, *81*, 3285.

(82) Want, E. J.; O'Maille, G.; Smith, C. A.; Brandon, T. R.; Uritboonthai, W.; Qin, C.; Trauger, S. A.; Siuzdak, G. *Anal Chem* **2006**, *78*, 743.

(83) Bligh, E. G.; Dyer, W. J. *Can J Biochem Physiol* **1959**, *37*, 911.

(84) Rezzi, S.; Vera, F. A.; Martin, F. P. J.; Wang, S.; Lawler, D.; Kochhar, S. *J Chromatogr B Analyt Technol Biomed Life Sci* **2008**, *871*, 271.

(85) Alvarez-Sanchez, B.; Priego-Capote, F.; Luque de Castro, M. D. *Trac-Trend Anal Chem* **2010**, *29*, 120.

(86) Veselkov, K. A.; Vingara, L. K.; Masson, P.; Robinette, S. L.; Want, E.; Li, J. V.; Barton, R. H.; Boursier-Neyret, C.; Walther, B.; Ebbels, T. M.; Pelczer, I.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Anal Chem* **2011**, *83*, 5864.

(87) Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. *Anal Chem* **2006**, *78*, 4281.

(88) Warrack, B. M.; Hnatyshyn, S.; Ott, K. H.; Reily, M. D.; Sanders, M.; Zhang, H. Y.; Drexler, D. M. *J Chromatogr B Analyt Technol Biomed Life Sci* **2009**, *877*, 547.

(89) Edmands, W. M.; Ferrari, P.; Scalbert, A. *Anal Chem* **2014**, *86*, 10925.

(90) Mattarucchi, E.; Guillou, C. *Biomed Chromatogr* **2012**, *26*, 512.

(91) Mattarucchi, E.; Guillou, C. *Anal Chem* **2011**, *83*, 9719.

(92) Chen, Y.; Shen, G.; Zhang, R.; He, J.; Zhang, Y.; Xu, J.; Yang, W.; Chen, X.; Song, Y.; Abliz, Z. *Anal Chem* **2013**, *85*, 7659.

(93) Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.;

Wilson, I. D.; Kell, D. B.; Goodacre, R.; Human Serum Metabolome, H. C. *Nat Protoc* **2011**, *6*, 1060.

(94) Ryan, D.; Robards, K.; Prenzler, P. D.; Kendall, M. *Anal Chim Acta* **2011**, *684*, 8.

(95) Miller, R. C.; Brindle, E.; Holman, D. J.; Shofer, J.; Klein, N. A.; Soules, M. R.; O'Connor, K. A. *Clin Chem* **2004**, *50*, 924.

(96) Walsh, M. C.; Brennan, L.; Malthouse, J. P.; Roche, H. M.; Gibney, M. J. *Am J Clin Nutr* **2006**, *84*, 531.

(97) Saude, E. J.; Adamko, D.; Rowe, B. H.; Marrie, T.; Sykes, B. D. *Metabolomics* **2007**, *3*, 439.

(98) Zuppi, C.; Messana, I.; Forni, F.; Rossi, C.; Pennacchietti, L.; Ferrari, F.; Giardina, B. *Clin Chim Acta* **1997**, *265*, 85.

(99) Hou, H.; Xiong, W.; Zhang, X.; Song, D.; Tang, G.; Hu, Q. *J Anal Methods Chem* **2012**, *2012*, 245415.

(100) Waikar, S. S.; Sabbisetti, V. S.; Bonventre, J. V. *Kidney Int* **2010**, *78*, 486.

(101) James, G. D.; Sealey, J. E.; Alderman, M.; Ljungman, S.; Mueller, F. B.; Pecker, M. S.; Laragh, J. H. *Am J Hypertens* **1988**, *1*, 124.

(102) Heymsfield, S. B.; Arteaga, C.; McManus, C.; Smith, J.; Moffitt, S. *Am J Clin Nutr* **1983**, *37*, 478.

(103) Whiteley, P.; Waring, R.; Williams, L.; Klovrza, L.; Nolan, F.; Smith, S.; Farrow, M.; Dodou, K.; Lough, W. J.; Shattock, P. *Pediatr Int* **2006**, *48*, 292.

(104) Launay-Vacher, V.; Gligorov, J.; Le Tourneau, C.; Janus, N.; Spano, J. P.; Ray-Coquard, I.; Oudard, S.; Pourrat, X.; Morere, J. F.; Deray, G.; Beuzeboc, P. *Breast Cancer Res Treat* **2010**, *124*, 745.

(105) Rubin, R. T. *Psychosom Med* **1971**, *33*, 539.

(106) Crim, M. C.; Calloway, D. H.; Margen, S. *J Nutr* **1975**, *105*, 428.

(107) Davison, J. M.; Noble, M. C. *Br J Obstet Gynaecol* **1981**, *88*, 10.

(108) Alessio, L.; Berlin, A.; Dellorto, A.; Toffoletto, F.; Ghezzi, I. *Int Arch Occup Environ Health* **1985**, *55*, 99.

(109) Burton, C.; Shi, H.; Ma, Y. *Anal Chem* **2013**, *85*, 11137.

(110) Kemperman, R. F.; Horvatovich, P. L.; Hoekman, B.; Reijmers, T. H.; Muskiet, F. A.; Bischoff, R. *J Proteome Res* **2007**, *6*, 194.

(111) Chadha, V.; Garg, U.; Alon, U. S. *Pediatr Nephrol* **2001**, *16*, 374.

(112) Gyamlani, G. G.; Bergstralh, E. J.; Slezak, J. M.; Larson, T. S. *Am J Kidney Dis* **2003**, *42*, 685.

(113) Richmond, W.; Colgan, G.; Simon, S.; Stuart-Hilgenfeld, M.; Wilson, N.; Alon, U. S. *Clin Nephrol* **2005**, *64*, 264.

(114) Haddow, J. E.; Knight, G. J.; Palomaki, G. E.; Neveux, L. M.; Chilmonczyk, B. A. *Clin Chem* **1994**, *40*, 562.

(115) Burton, C.; Shi, H.; Ma, Y. *Clin Chim Acta* **2014**, *435*, 42.

(116) Jacob, C. C.; Dervilly-Pinel, G.; Biancotto, G.; Le Bizec, B. *Metabolomics* **2014**, *10*, 627.

(117) Dervilly-Pinel, G.; Weigel, S.; Lommen, A.; Chereau, S.; Rambaud, L.; Essers, M.; Antignac, J. P.; Nielen, M. W.; Le Bizec, B. *Anal Chim Acta* **2011**, *700*, 144.

(118) Rijk, J. C.; Lommen, A.; Essers, M. L.; Groot, M. J.; Van Hende, J. M.; Doeswijk, T. G.; Nielen, M. W. *Anal Chem* **2009**, *81*, 6879.

(119) Zamora-Ros, R.; Rabassa, M.; Cherubini, A.; Urpi-Sarda, M.; Llorach, R.; Bandinelli, S.; Ferrucci, L.; Andres-Lacueva, C. *Anal Chim Acta* **2011**, *704*, 110.

(120) Gonzalez-Dominguez, R.; Castilla-Quintero, R.; Garcia-Barrera, T.; Gomez-Ariza, J. L. *Anal Biochem* **2014**, *465C*, 20.

(121) Zhou, H.; Yuen, P. S.; Pisitkun, T.; Gonzales, P. A.; Yasuda, H.; Dear, J. W.; Gross, P.; Knepper, M. A.; Star, R. A. *Kidney Int* **2006**, *69*, 1471.

(122) Craig, A.; Cloarec, O.; Holmes, E.; Nicholson, J. K.; Lindon, J. C. *Anal Chem* **2006**, *78*, 2262.

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

(124) Ejigu, B. A.; Valkenborg, D.; Baggerman, G.; Vanaerschot, M.; Witters, E.; Dujardin, J. C.; Burzykowski, T.; Berg, M. *OMICS* **2013**, *17*, 473.

(125) Wu, Y.; Li, L. *Anal Chem* **2012**, *84*, 10723.

(126) Leon, Z.; Garcia-Canaveras, J. C.; Donato, M. T.; Lahoz, A. *Electrophoresis* **2013**, *34*, 2762.

(127) Halama, A. *Arch Biochem Biophys* **2014**, *564*, 100.

(128) Cao, B.; Aa, J.; Wang, G.; Wu, X.; Liu, L.; Li, M.; Shi, J.; Wang, X.; Zhao, C.; Zheng, T.; Guo, S.; Duan, J. *Anal Bioanal Chem* **2011**, *400*, 2983.

(129) Ruiz-Aracama, A.; Peijnenburg, A.; Kleinjans, J.; Jennen, D.; van Delft, J.; Hellfrisch, C.; Lommen, A. *BMC Genomics* **2011**, *12*, 251.

(130) Hutschenreuther, A.; Kiontke, A.; Birkenmeier, G.; Birkemeyer, C. *Anal Methods* **2012**, *4*, 1953.

(131) Marcinowska, R.; Trygg, J.; Wolf-Watz, H.; Mortiz, T.; Surowiec, I. *J Microbiol Methods* **2011**, *87*, 24.

(132) Smart, K. F.; Aggio, R. B.; Van Houtte, J. R.; Villas-Boas, S. G. *Nat Protoc* **2010**, *5*, 1709.

(133) Huege, J.; Krall, L.; Steinhauser, M. C.; Giavalisco, P.; Rippka, R.; Tandeau de Marsac, N.; Steinhauser, D. *Anal Bioanal Chem* **2011**, *399*, 3503.

(134) Teahan, O.; Bevan, C. L.; Waxman, J.; Keun, H. C. *Int J Biochem Cell Biol* **2011**, *43*, 1002.

(135) Silva, A. M.; Cordeiro-da-Silva, A.; Coombs, G. H. *PLoS Negl Trop Dis* **2011**, *5*, e1451.

(136) Silva, L. P.; Lorenzi, P. L.; Purwaha, P.; Yong, V.; Hawke, D. H.; Weinstein, J. N. *Anal Chem* **2013**, *85*, 9536.

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

(138) Fu, F.; Cheng, V. W.; Wu, Y.; Tang, Y.; Weiner, J. H.; Li, L. *J Proteome Res* **2013**, *12*, 4478.

(139) Wu, Y.; Li, L. *Anal Chem* **2014**, *86*, 9428.

(140) Gallardo, E.; Queiroz, J. A. *Biomed Chromatogr* **2008**, *22*, 795.

(141) Mena-Bravo, A.; Luque de Castro, M. D. *J Pharm Biomed Anal* **2014**, *90*, 139.

(142) Appenzeller, B. M.; Schummer, C.; Rodrigues, S. B.; Wennig, R. *J Chromatogr B Analyt Technol Biomed Life Sci* **2007**, *852*, 333.

(143) Xu, W.; Chen, D.; Wang, N.; Zhang, T.; Zhou, R.; Huan, T.; Lu, Y.; Su, X.; Xie, Q.; Li, L. *Anal Chem* **2015**, *87*, 829.

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

(145) Cech, N. B.; Enke, C. G. *Mass Spectrom Rev* **2001**, *20*, 362.
(146) Guo, K.; Li, L. *Anal Chem* **2009**, *81*, 3919.
(147) Keller, B. O.; Sui, J.; Young, A. B.; Whittal, R. M. *Anal Chim Acta* **2008**, *627*, 71.
(148) Guo, K.; Ji, C.; Li, L. *Anal Chem* **2007**, *79*, 8631.
(149) 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.
(150) Guo, K.; Li, L. *Anal Chem* **2010**, *82*, 8789.
(151) Lamos, S. M.; Shortreed, M. R.; Frey, B. L.; Belshaw, P. J.; Smith, L. M. *Anal Chem* **2007**, *79*, 5143.
(152) Yang, W. C.; Adamec, J.; Regnier, F. E. *Anal Chem* **2007**, *79*, 5150.
(153) Manini, P.; Andreoli, R.; Sforza, S.; Dall'Asta, C.; Galaverna, G.; Mutti, A.; Niessen, W. M. *J Chromatogr B Analyt Technol Biomed Life Sci* **2010**, *878*, 2616.
(154) Zhou, R.; Tseng, C. L.; Huan, T.; Li, L. *Anal Chem* **2014**, *86*, 4675.
(155) Tseng, C.-L.; Li, L. *Sci China Chem* **2014**, *57*, 678.
(156) Zheng, J.; Dixon, R. A.; Li, L. *Anal Chem* **2012**, *84*, 10802.
(157) Lu, X.; Zhao, X. J.; Bai, C. M.; Zhao, C. X.; Lu, G.; Xu, G. M. *J Chromatogr B Analyt Technol Biomed Life Sci* **2008**, *866*, 64.
(158) Theodoridis, G. A.; Gika, H. G.; Want, E. J.; Wilson, I. D. *Anal. Chim. Acta* **2012**, *711*, 7.
(159) Theodoridis, G.; Gika, H. G.; Wilson, I. D. *Trac- Trend Anal Chem* **2008**, *27*, 251.
(160) Roux, A.; Lison, D.; Junot, C.; Heilier, J. F. *Clin Biochem* **2011**, *44*, 119.
(161) Tsuchiya, Y.; Takahashi, Y.; Jindo, T.; Furuhama, K.; Suzuki, K. T. *Eur J Pharmacol* **2003**, *475*, 119.
(162) Wagner, B. D.; Accurso, F. J.; Laguna, T. A. *J Cyst Fibros* **2010**, *9*, 212.
(163) Heavner, D. L.; Morgan, W. T.; Sears, S. B.; Richardson, J. D.; Byrd, G. D.; Ogden, M. W. *J Pharm Biomed Anal* **2006**, *40*, 928.
(164) Jatlow, P.; McKee, S.; O'Malley, S. S. *Clin Chem* **2003**, *49*, 1932.
(165) Torgrip, R. J. O.; Aberg, K. M.; Alm, E.; Schuppe-Koistinen, I.; Lindberg, J. *Metabolomics* **2008**, *4*, 114.
(166) Elsebaei, F.; Zhu, Y. *Talanta* **2011**, *85*, 123.
(167) Mowafy, H. A.; Alanazi, F. K.; El Maghraby, G. M. *Saudi Pharm J* **2012**, *20*, 29.
(168) Shibasaki, H.; Okamoto, S.; Inoue, R.; Okita, M.; Yokokawa, A.; Furuta, T. *Anal Bioanal Chem* **2012**, *402*, 2945.
(169) Xia, J. G.; Psychogios, N.; Young, N.; Wishart, D. S. *Nucleic Acids Res* **2009**, *37*, W652.
(170) Seiler, N.; Knodgen, B.; Eisenbeiss, F. *J Chromatogr* **1978**, *145*, 29.
(171) Zar, J. H. *Biostatistical Analysis*; 5th ed.; Pearson Prentice Hall Inc., 2010.
(172) Wang, T. C. L.; Ricca, T. L.; Marshall, A. G. *Anal Chem* **1986**, *58*, 2935.
(173) Westerhuis, J. A.; Hoefsloot, H. C. J.; Smit, S.; Vis, D. J.; Smilde, A. K.; van Velzen, E. J. J.; van Duijnhoven, J. P. M.; van Dorsten, F. A. *Metabolomics* **2008**, *4*, 81.
(174) Zhu, Z. J.; Schultz, A. W.; Wang, J. H.; Johnson, C. H.; Yannone, S. M.; Patti, G. J.; Siuzdak, G. *Nat Protoc* **2013**, *8*, 451.

(175) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W. M.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A. N.; Lindon, J. C.; Marriott, P.; Nicholls, A. W.; Reily, M. D.; Thaden, J. J.; Viant, M. R. *Metabolomics* **2007**, *3*, 211.

(176) Tapuhi, Y.; Schmidt, D. E.; Lindner, W.; Karger, B. L. *Anal Biochem* **1981**, *115*, 123.

(177) Kabra, P. M.; Lee, H. K.; Lubich, W. P.; Marton, L. J. *J Chromatogr* **1986**, *380*, 19.

(178) Loukou, Z.; Zotou, A. *J Chromatogr A* **2003**, *996*, 103.

(179) Quirke, J. M. E.; Adams, C. L.; Vanberkel, G. J. *Anal Chem* **1994**, *66*, 1302.

(180) Corradini, R.; Paganuzzi, C.; Marchelli, R.; Pagliari, S.; Sforza, S.; Dossena, A.; Galaverna, G.; Duchateau, A. *J Mater Chem* **2005**, *15*, 2741.

(181) Peng, J.; Chen, Y.-T.; Chen, C.-L.; Li, L. *Anal Chem* **2014**, *86*, 6540.

(182) Dietmair, S.; Timmins, N. E.; Gray, P. P.; Nielsen, L. K.; Kromer, J. O. *Anal Biochem* **2010**, *404*, 155.

(183) Goodacre, R.; Vaidyanathan, S.; Dunn, W. B.; Harrigan, G. G.; Kell, D. B. *Trends Biotechnol* **2004**, *22*, 245.

(184) Nielsen, J.; Oliver, S. *Trends Biotechnol* **2005**, *23*, 544.

(185) Khoo, S. H. G.; Al-Rubeai, M. *Biotechnol Appl Biochem* **2007**, *47*, 71.

(186) Liebeke, M.; Dorries, K.; Meyer, H.; Lalk, M. In *Functional Genomics: Methods and Protocols, Second Edition*; Kaufmann, M., Klinger, C., Eds.; Humana Press Inc: Totowa, 2012; Vol. 815, p 377.

(187) Mashego, M. R.; Rumbold, K.; De Mey, M.; Vandamme, E.; Soetaert, W.; Heijnen, J. J. *Biotechnol Lett* **2007**, *29*, 1.

(188) Wachsrnuth, C. J.; Almstetter, M. F.; Waldhier, M. C.; Gruber, M. A.; Nurnberger, N.; Oefner, P. J.; Dettmer, K. *Anal Chem* **2011**, *83*, 7514.

(189) Fan, T. W. M.; Bird, J. A.; Brodie, E. L.; Lane, A. N. *Metabolomics* **2009**, *5*, 108.

(190) Olivier, I.; Loots, D. T. *J Microbiol Methods* **2012**, *88*, 419.

(191) Cevallos-Cevallos, J. M.; Danyluk, M. D.; Reyes-De-Corcuera, J. I. *J Food Sci* **2011**, *76*, M238.

(192) Duarte, I. F.; Marques, J.; Ladeirinha, A. F.; Rocha, C.; Lamego, I.; Calheiros, R.; Silva, T. M.; Marques, M. P. M.; Melo, J. B.; Carreira, I. M.; Gil, A. M. *Anal Chem* **2009**, *81*, 5023.

(193) Griffin, J. L. *Curr Opin Chem Biol* **2003**, *7*, 648.

(194) Griffin, J. L.; Bollard, M.; Nicholson, J. K.; Bhakoo, K. *NMR Biomed* **2002**, *15*, 375.

(195) Fiehn, O. *Trac- Trend Anal Chem* **2008**, *27*, 261.

(196) Koek, M. M.; Muilwijk, B.; van der Werf, M. J.; Hankemeier, T. *Anal Chem* **2006**, *78*, 1272.

(197) Buchholz, A.; Takors, R.; Wandrey, C. *Anal Biochem* **2001**, *295*, 129.

(198) Juo, C. G.; Chiu, D. T. Y.; Shiao, M. S. *Biofactors* **2008**, *34*, 159.

(199) Lu, W. Y.; Kimball, E.; Rabinowitz, J. D. *J Am Soc Mass Spectrom* **2006**, *17*, 37.

(200) Yanes, O.; Tautenhahn, R.; Patti, G. J.; Siuzdak, G. *Anal Chem* **2011**, *83*, 2152.

(201) Ramautar, R.; Demirci, A.; de Jong, G. J. *Trac- Trend Anal Chem* **2006**, *25*, 455.

(202) Soga, T.; Ueno, Y.; Naraoka, H.; Ohashi, Y.; Tomita, M.; Nishioka, T. *Anal Chem* **2002**, *74*, 2233.

(203) Kuehnbaum, N. L.; Britz-McKibbin, P. *Anal Chem* **2011**, *83*, 8063.

(204) Winder, C. L.; Dunn, W. B.; Schuler, S.; Broadhurst, D.; Jarvis, R.; Stephens, G. M.; Goodacre, R. *Anal Chem* **2008**, *80*, 2939.

(205) Villas-Boas, S. G.; Hojer-Pedersen, J.; Akesson, M.; Smedsgaard, J.; Nielsen, J. *Yeast* **2005**, *22*, 1155.

(206) Shin, M. H.; Lee, D. Y.; Liu, K. H.; Fiehn, O.; Kim, K. H. *Anal Chem* **2010**, *82*, 6660.

(207) Rabinowitz, J. D.; Kimball, E. *Anal Chem* **2007**, *79*, 6167.

(208) Bolten, C. J.; Kiefer, P.; Letisse, F.; Portais, J. C.; Wittmann, C. *Anal Chem* **2007**, *79*, 3843.

(209) Sellick, C. A.; Hansen, R.; Maqsood, A. R.; Dunn, W. B.; Stephens, G. M.; Goodacre, R.; Dickson, A. J. *Anal Chem* **2009**, *81*, 174.

(210) Siddiquee, K. A.; Arauzo-Bravo, M. J.; Shimizu, K. *Appl Microbiol Biotechnol* **2004**, *63*, 407.

(211) Hajjaj, H.; Blanc, P. J.; Goma, G.; Francois, J. *Fems Microbiology Letters* **1998**, *164*, 195.

(212) Sellick, C. A.; Knight, D.; Croxford, A. S.; Maqsood, A. R.; Stephens, G. M.; Goodacre, R.; Dickson, A. J. *Metabolomics* **2010**, *6*, 427.

(213) Maharjan, R. P.; Ferenci, T. *Anal Biochem* **2003**, *313*, 145.

(214) Villas-Boas, S. G.; Mas, S.; Akesson, M.; Smedsgaard, J.; Nielsen, J. *Mass Spectrom Rev* **2005**, *24*, 613.

(215) Havlicek, V.; Lemr, K.; Schug, K. A. *Anal Chem* **2013**, *85*, 790.

(216) Weibel, K. E.; Mor, J. R.; Fiechter, A. *Anal Biochem* **1974**, *58*, 208.

(217) Mathur, R.; O'Connor, P. B. *Rapid Commun Mass Spectrom* **2009**, *23*, 523.

(218) Stout, S. J.; daCunha, A. R.; Picard, G. L.; Safarpour, M. M. *J Agric Food Chem* **1996**, *44*, 3548.

(219) Young, J. C. *J Agric Food Chem* **1995**, *43*, 2904.

(220) Royet, J.; Dziarski, R. *Nat Rev Microbiol* **2007**, *5*, 264.

(221) Vanheije.J; Elbaz, L.; Dezelee, P.; Petit, J. F.; Bricas, E.; Ghuysen, J. M. *Biochemistry* **1969**, *8*, 207.

(222) Fenselau, C.; Demirev, P. A. *Mass Spectrom Rev* **2001**, *20*, 157.

(223) Toh-Boyo, G. M.; Wulff, S. S.; Basile, F. *Anal Chem* **2012**, *84*, 9971.

(224) Chen, W. J.; Tsai, P. J.; Chen, Y. C. *Anal Chem* **2008**, *80*, 9612.

(225) Kok, J.; Chen, S. C. A.; Dwyer, D. E.; Iredell, J. R. *Pathology* **2013**, *45*, 4.

(226) Khot, P. D.; Couturier, M. R.; Wilson, A.; Croft, A.; Fisher, M. A. *J Clin Microbiol* **2012**, *50*, 3845.

(227) Tan, K. E.; Ellis, B. C.; Lee, R.; Stamper, P. D.; Zhang, S. X.; Carroll, K. C. *J Clin Microbiol* **2012**, *50*, 3301.

(228) Blumberg, S. J.; Bramlett, M. D.; Kogan, M. D.; Schieve, L. A.; Jones, J. R.; Lu, M. C. *Natl Health Stat Report* **2013**, 12 pp.

(229) Blaylock, R. L. *Altern Ther Health Med* **2008**, *14*, 46.

(230) Ruggeri, B.; Sarkans, U.; Schumann, G.; Persico, A. M. *Psychopharmacology* **2014**, *231*, 1201.

(231) Klei, L.; Sanders, S. J.; Murtha, M. T.; Hus, V.; Lowe, J. K.; Willsey, A. J.; Moreno-De-Luca, D.; Yu, T. W.; Fombonne, E.; Geschwind, D.; Grice, D. E.; Ledbetter, D. H.; Lord, C.; Mane, S. M.; Martin, C. L.; Martin, D. M.; Morrow, E. M.; Walsh, C. A.; Melhem, N. M.; Chaste, P.; Sutcliffe, J. S.; State, M. W.; Cook, E. H.; Roeder, K.; Devlin, B. *Mol Autism* **2012**, *3*.

(232) Ratajczak, H. V. *J Immunotoxicol* **2011**, *8*, 80.

(233) Voineagu, I.; Yoo, H. J. *Dis Markers* **2013**, 55.

(234) Kuwabara, H.; Yamasue, H.; Koike, S.; Inoue, H.; Kawakubo, Y.; Kuroda, M.; Takano, Y.; Iwashiro, N.; Natsubori, T.; Aoki, Y.; Kano, Y.; Kasai, K. *Plos One* **2013**, *8*.

(235) Nakamura, K.; Anitha, A.; Yamada, K.; Tsujii, M.; Iwayama, Y.; Hattori, E.; Toyota, T.; Suda, S.; Takei, N.; Iwata, Y.; Suzuki, K.; Matsuzaki, H.; Kawai, M.; Sekine, Y.; Tsuchiya, K. J.; Sugihara, G. I.; Ouchi, Y.; Sugiyama, T.; Yoshikawa, T.; Mori, N. *Int J Neuropsychopharmacol* **2008**, *11*, 1073.

(236) Iwata, K.; Matsuzaki, H.; Miyachi, T.; Shimmura, C.; Suda, S.; Tsuchiya, K. J.; Matsumoto, K.; Suzuki, K.; Iwata, Y.; Nakamura, K.; Tsujii, M.; Sugiyama, T.; Sato, K.; Mori, N. *Mol Autism* **2011**, *2*.

(237) Momeni, N.; Bergquist, J.; Brudin, L.; Behnia, F.; Sivberg, B.; Joghataei, M. T.; Persson, B. L. *Transl Psychiatry* **2012**, *2*.

(238) Emond, P.; Mavel, S.; Aidoud, N.; Nadal-Desbarats, L.; Montigny, F.; Bonnet-Brilhault, F.; Barthelemy, C.; Merten, M.; Sarda, P.; Laumonnier, F.; Vourc'h, P.; Blasco, H.; Andres, C. R. *Anal Bioanal Chem* **2013**, *405*, 5291.

(239) Quinones, M. P.; Kaddurah-Daouk, R. *Neurobiol Dis* **2009**, *35*, 165.

(240) Shimmura, C.; Suda, S.; Tsuchiya, K. J.; Hashimoto, K.; Ohno, K.; Matsuzaki, H.; Iwata, K.; Matsumoto, K.; Wakuda, T.; Kameno, Y.; Suzuki, K.; Tsujii, M.; Nakamura, K.; Takei, N.; Mori, N. *Plos One* **2011**, *6*.

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

(242) Xia, J. G.; Broadhurst, D. I.; Wilson, M.; Wishart, D. S. *Metabolomics* **2013**, *9*, 280.

(243) Crews, B.; Wikoff, W. R.; Patti, G. J.; Woo, H. K.; Kalisiak, E.; Heideker, J.; Siuzdak, G. *Anal Chem* **2009**, *81*, 8538.

(244) Schulze, E.; Neuhoff, V. *Hoppe Seylers Z Physiol Chem* **1976**, *357*, 225.

(245) Moreno-Fuenmayor, H.; Borjas, L.; Arrieta, A.; Valera, V.; Socorro-Candanoza, L. *Invest Clin* **1996**, *37*, 113.

(246) Fatemi, S. H. *Prog Neuropsychopharmacol Biol Psychiatry* **2008**, *32*, 911.

(247) Blaylock, R. L.; Strunecka, A. *Curr Med Chem* **2009**, *16*, 157.

(248) Hassan, T. H.; Abdelrahman, H. M.; Fattah, N. R. A.; El-Masry, N. M.; Hashim, H. M.; El-Gerby, K. M. *Res Autism Spect Dis* **2013**, *7*, 541.

(249) Ortinski, P. I.; Dong, J. H.; Mungenast, A.; Yue, C. Y.; Takano, H.; Watson, D. J.; Haydon, P. G.; Coulter, D. A. *Nat Neurosci* **2010**, *13*, 584.

(250) Pais, T. F.; Figueiredo, C.; Peixoto, R.; Braz, M. H.; Chatterjee, S. *J Neuroinflammation* **2008**, *5*.

(251) Chauhan, A.; Chauhan, V. *Pathophysiology* **2006**, *13*, 171.

(252) James, S. J.; Cutler, P.; Melnyk, S.; Jernigan, S.; Janak, L.; Gaylor, D. W.; Neubrander, J. A. *Am J Clin Nutr* **2004**, *80*, 1611.

(253) Yi, P.; Melnyk, S.; Pogribna, M.; Pogribny, I. P.; Hines, R. J.; James, S. J. *J Biol Chem* **2000**, *275*, 29318.

(254) Pandur, S.; Pankiv, S.; Johannessen, M.; Moens, U.; Huseby, N. E. *Free Radic Res* **2007**, *41*, 1376.

(255) Geier, D. A.; Kern, J. K.; Garver, C. R.; Adams, J. B.; Audhya, T.; Geier, M. R. *Neurochem Res* **2009**, *34*, 386.

(256) Noonan, V. K.; Fingas, M.; Farry, A.; Baxter, D.; Singh, A.; Fehlings, M. G.; Dvorak, M. F. *Neuroepidemiology* **2012**, *38*, 219.

(257) Krueger, H.; Noonan, V. K.; Trenaman, L. M.; Joshi, P.; Rivers, C. S. *Chronic Dis Inj Can* **2013**, *33*, 113.

(258) Tator, C. H. *Neurosurgery* **2006**, *59*, 957.

(259) Long, Y. C.; Kostovski, E.; Boon, H.; Hjeltnes, N.; Krook, A.; Widegren, U. *J Appl Physiol (1985)* **2011**, *110*, 1204.

(260) Dulin, J. N.; Karoly, E. D.; Wang, Y.; Strobel, H. W.; Grill, R. J. *J Neurosci* **2013**, *33*, 652.

(261) Toborek, M.; Malecki, A.; Garrido, R.; Mattson, M. P.; Hennig, B.; Young, B. *J Neurochem* **1999**, *73*, 684.

(262) Panter, S. S.; Yum, S. W.; Faden, A. I. *Ann Neurol* **1990**, *27*, 96.

(263) Peng, J.; Zeng, J.; Cai, B.; Yang, H.; Cohen, M. J.; Chen, W.; Sun, M. W.; Lu, C. D.; Jiang, H. *PLoS One* **2014**, *9*, e93736.

(264) Lubieniecka, J. M.; Streijger, F.; Lee, J. H.; Stoynov, N.; Liu, J.; Mottus, R.; Pfeifer, T.; Kwon, B. K.; Coorssen, J. R.; Foster, L. J.; Grigliatti, T. A.; Tetzlaff, W. *PLoS One* **2011**, *6*, e19247.

(265) Mandal, R.; Guo, A. C.; Chaudhary, K. K.; Liu, P.; Yallou, F. S.; Dong, E.; Aziat, F.; Wishart, D. S. *Genome Med* **2012**, *4*, 38.

(266) Trushina, E.; Dutta, T.; Persson, X. M.; Mielke, M. M.; Petersen, R. C. *PLoS One* **2013**, *8*, e63644.

(267) Huan, T.; Li, L. *Anal Chem* **2015**, *87*, 1306.

(268) Zhou, R.; Li, L. *J Proteomics* **2014**.

(269) Psychogios, N.; Hau, D. D.; Peng, J.; Guo, A. C.; Mandal, R.; Bouatra, S.; Sinelnikov, I.; Krishnamurthy, R.; Eisner, R.; Gautam, B.; Young, N.; Xia, J.; Knox, C.; Dong, E.; Huang, P.; Hollander, Z.; Pedersen, T. L.; Smith, S. R.; Bamforth, F.; Greiner, R.; McManus, B.; Newman, J. W.; Goodfriend, T.; Wishart, D. S. *PLoS One* **2011**, *6*, e16957.

(270) Wishart, D. S.; Lewis, M. J.; Morrissey, J. A.; Flegel, M. D.; Jeroncic, K.; Xiong, Y.; Cheng, D.; Eisner, R.; Gautam, B.; Tzur, D.; Sawhney, S.; Bamforth, F.; Greiner, R.; Li, L. *J Chromatogr B Analyt Technol Biomed Life Sci* **2008**, *871*, 164.

(271) Demediuk, P.; Daly, M. P.; Faden, A. I. *J Neurosci Res* **1989**, *23*, 95.

(272) Kohen, R.; Yamamoto, Y.; Cundy, K. C.; Ames, B. N. *Proc Natl Acad Sci U S A* **1988**, *85*, 3175.

(273) Fujieda, Y.; Ueno, S.; Ogino, R.; Kuroda, M.; Jonsson, T. J.; Guo, L.; Bamba, T.; Fukusaki, E. *PLoS One* **2012**, *7*, e43152.

(274) Tabakman, R.; Jiang, H.; Levine, R. A.; Kohen, R.; Lazarovici, P. *J Neurosci Res* **2004**, *75*, 499.