

University of Alberta

**Isotope Labeling Liquid Chromatography Mass Spectrometry: Metabolite
Identification and Human Salivary Metabolome Profiling**

by

Jiamin Zheng

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Chemistry

©Jiamin Zheng

Fall 2012

Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

Abstract

The objective of this work was to use paired labeling reagents that are chemically identical but isotopically different to provide a simple and robust way of mass spectrometry (MS)-based metabolome profiling. A differential ^{13}C -/ ^{12}C -isotope dansyl chloride (DnsCl) derivatization strategy has been further developed and applied for qualitative and quantitative profiling of amine- and phenol-containing metabolites by liquid chromatography electrospray ionization Fourier Transform ion cyclotron resonance MS (LC-ESI FT-ICR-MS). For labeled metabolites, a new technique using skimmer-region fragmentation, followed by tandem MS analysis of formed fragment ions, was developed for generating structural information. In addition, a new method based on use of this isotope labeling LC-MS platform was developed for human salivary metabolome analysis. It offered superior performance over other reported methods for saliva metabolome profiling. Its potential utility for disease biomarker discovery was demonstrated in the analysis of metabolomic differences between normal individuals and diseased-individuals with mild cognitive impairment (MCI).

Acknowledgement

I would like to thank my supervisor, Dr. Liang Li, for his guidance, support, and encouragement. It has been a pleasure and an honour learning more about mass spectrometry under his direction. I have found his enthusiasm and passion for analytical chemistry to be very inspirational.

I would also like to thank my supervisory committee members, Professors Jed Harrison and Roger Dixon, for reviewing my thesis and for their helpful suggestions.

I would like to thank Dr. Randy Whittle and the entire staff of the University of Alberta's Mass Spectrometry Laboratory for their instruction and advice on the usage of the FT-ICR-MS instrument.

I would like to thank the members of the Li research group that I have worked with during my graduate program: Lu Chen, Jared Curle, Margot Dawe, Dr. Andrea De Souza, Lara Ebert, Feifei Fu, Dr. Kevin Guo, Tao Huan, Chad Iverson, Zhendong Li, Dr. Andy Lo, Jun Peng, Dr. Avalyn Stanislaus, Difei Sun, Yanan Tang, Tran Tran, Chiao-li Tseng, Dr. Nan Wang, Dr. Fang Wu, Yiman Wu, Dan Xu, Dr. Mingguo Xu, Dr. Xiaoxia Ye, Ruokun Zhou, and Dr. Azeret Zuniga. Working with all of these great individuals made my graduate experience enjoyable. I would especially like to thank Dr. Avalyn Stanislaus and Dr. Azeret Zuniga for their mentorship at the beginning of my research and for their patience, insight and support; and also especially thank Dr. Xiaoxia Ye for her help on my research and personal life through the entire research period.

I would like to thank Dr. David Wishart and all the members of Human Metabolome Project for providing the standards used in my research.

The University of Alberta, Department of Chemistry, Genome Alberta, and Genome Canada are acknowledged for financial support.

Thank you to my parents, Yuezhen and Guorong, to my aunt and uncle, Jenny and Mingwei, to my cousin, Raymond, and to my grandparents, Baogen, Jingui and Tingyao for their love, support and helpful discussions and suggestions. Thank you to my good friends for all of their support and encouragement. Finally, thank you to my husband, Lun Zhang, for his patient waiting, extraordinary understanding and great love.

Table of Contents

Chapter 1: Introduction.....	1
1. General Introduction.....	1
1.1. Mass Spectrometry	2
1.1.1. Electrospray Ionization	4
1.1.2. Mass Analyzer	8
1.1.2.1. Quadrupole MS.....	9
1.1.2.2. Fourier Transform Ion Cyclotron Resonance MS	15
1.1.3. Flow Injection.....	20
1.1.4. Liquid Chromatography.....	22
1.2. Metabolomics	27
1.2.1. Targeted Metabolomics	30
1.2.2. Dansylation Labeling.....	33
1.2.3. Quantitative Differential Isotope Labeling	35
1.3. Overview of Thesis.....	38
1.4. Literature Cited.....	40
Chapter 2: Fragmentation of Protonated Dansyl-labeled Amines for Structural Analysis of Amine-containing Metabolites.....	47
2.1. Introduction.....	47
2.2. Experimental.....	49
2.2.1. Chemical and Reagents.....	49
2.2.2. Labeling Reaction	49

2.2.3. Direct Flow Injection-MS/MS	50
2.2.4. Fragmentation Pattern Analysis	50
2.3. Results and Discussion	51
2.3.1. Neutral Loss of Unlabeled Amines.....	52
2.3.2. Fragment Ions of Unlabeled Amines	56
2.3.3. MS ² Fragmentation Patterns of Dansylated Amines	57
2.3.4. MS ³ Fragmentation Patterns of the Fragment Ions from Dansylated Amines	60
2.3.5. <i>pseudo</i> -MS ³ fragmentation patterns of dansylated amines and comparison to those of unlabeled amines	61
2.3.6. Relevance to Metabolome Profiling	66
2.4. Conclusions	68
2.5. Literature Cited.....	69

Chapter 3: Development and Application of Isotope Labeling LC-MS for Human Salivary Metabolomics	71
3.1. Introduction.....	71
3.2. Experimental.....	73
3.2.1. Chemical and reagents	73
3.2. Sample collection and processing.....	73
3.2.1. Sample collection:.....	73
3.2.2. Sample processing	74
3.2.3. Labeling reaction	74
3.2.4. LC-FTICR-MS.....	75
3.2.5. LC-UV	75

3.2.6. Quantitative data processing and statistical analysis	76
3.2.7. Differentially expressed metabolites identification	76
3.3. Results and discussion	77
3.3.1. ¹² C-/ ¹³ C-labeled saliva sample mixing volume normalization	77
3.3.2. Initial sample size optimization	81
3.3.3 Determination of optimal amount injection.....	94
3.3.4 Method reproducibility	95
3.3.5 Pilot test of metabolome profiling	95
3.3.6 Sample storage effect.....	96
3.3.7 Metabolome comparison of healthy and MCI saliva samples	100
3.3.7 Metabolite identification.....	105
3.4. Conclusions	106
3.5. Literature Cited.....	108
Chapter 4: Conclusions and Future Work.....	110
Appendix	113

List of Tables

Table 1.1.	Description of triple quadrupole and QqLIT operation modes.....	14
Table 3.1.	Summary of total number of metabolites detected using four different initial saliva volumes.....	85
Table 3.2.	Comparison of metabolite extraction efficiency with different initial saliva volumes examined by UV absorbance.....	87
Table 3.3.	T-test results (p-values) showing the intra- and inter-individual comparisons, where the intra-comparison was performed on different storage conditions within each individual, and the inter-comparison was performed to compare different individuals.....	99
Table 3.4.	Intra- and inter-individual pearson's correlation coefficient (r) calculation results showing the correlations between different storage conditions within each individual and those between different individuals.....	100
Table 3.5.	Summary of the discriminant metabolites from VIP scores of OPLS-DA model for variations between MCI diseased group and normal healthy group.....	105

List of Figures

- Figure 1.1. Schematic representation of the process of generating gas phase ions in ESI with positive ion mode (adapted from Quantitative Chemical Analysis 7th Edition).....7
- Figure 1.2. The a-q stability diagram ($m_1 > m_2 > m_3$). (adapted from Chemistry 505 class notes, Fall 2009, Prof Liang Li, University of Alberta).....10
- Figure 1.3. Schematic of QqLIT (QTRAP).....12
- Figure 1.4. Ion cyclotron motion. Moving path of the positive ion in the plane is bent into a circle by the Lorentz magnetic force generated by a homogenous magnetic field perpendicular to the plane. (adapted from Marshall et al., 1998²²).....17
- Figure 1.5. Simplified depiction of an ICR cell. The plates positioned on the front and the back are two excitation plates, on the two sides are the trapping plates, and the two on the top and bottom are the detection plates. (adapted from Marshall et al., 1998²²).....18
- Figure 1.6. Ion cyclotron orbital motion of excitation and detection.....20

Figure 1.7.	Workflow for relative quantification of amine- and phenol-containing metabolites using dansylation labeling LC-ESI FT-ICR-MS.....	38
Figure 2.1.	(A) CID MS/MS spectrum of the protonated 3-aminobenzoic acid. (B) <i>pseudo-MS</i> ³ spectrum of the skimmer-fragment ion with <i>m/z</i> 138.1 with enhanced product ion (EPI) scan in the QTRAP. (C) CID MS/MS spectrum of the protonated dansyl-3-aminobenzoic acid. The collision energy used for each spectrum reported in this work is given in the original spectra presented in appendix.....	53
Figure 2.2.	(A) MS/MS spectrum of the protonated methionine sulfoxide. Peak with <i>m/z</i> 166.0 represents methionine sulfoxide molecular ion after protonation, and peak with <i>m/z</i> 149.0 shows the fragment ion after neutral loss of 17 from the protonated methionine sulfoxide ion. (B) <i>MS</i> ³ spectrum of the deaminated ion of methionine sulfoxide. Peak with <i>m/z</i> 149.1 represents the deaminated ion after neutral loss of 17 from the protonated methionine sulfoxide precursor ion, and peak with <i>m/z</i> 121.0 shows the fragment ion after neutral loss of 28 from the deaminated ion.....	55
Figure 2.3.	(A) MS/MS spectrum of the protonated dansyl-threonine. (B) skimmer-fragmentation spectrum of the protonated dansyl-threonine.	58
Figure 2.4.	Fragmentation scheme of dansyl-threonine.....	59

Figure 2.5.	(A) MS/MS spectrum of the protonated N-acetylputrescine. (B) <i>pseudo</i> -MS ³ spectrum of the skimmer-fragment ion with <i>m/z</i> 131.2 with enhanced product ion (EPI) scan in QTRAP. (C) MS/MS spectrum of the protonated dansyl-N-acetylputrescine.....	64
Figure 2.6.	Fragmentation scheme of skimmer-fragment ion generated from the dansyl-N-acetylputrescine ion with the same <i>m/z</i> value as that of the protonated N-acetylputrescine.....	65
Figure 3.1.	Calibration curve built with ¹² C-dansyl chloride labeled amino acid standard mixture solution.....	79
Figure 3.2.	Calibration curve built with ¹² C-dansyl chloride labeled pooled saliva sample.....	80
Figure 3.3.	A representative total ion chromatogram of dansyl-labeled saliva sample solution from LC-ESI FT-ICR-MS analysis.....	83
Figure 3.4.	A representative mass spectrum at a specific retention time from an entire LC-ESI FT-ICR-MS analysis for dansyl-labeled saliva sample.....	84
Figure 3.5.	Effect of the concentration of unlabeled metabolites on the number of metabolites detected (start with 5 μL saliva and apply dansylation labeling in half-scale)	89

Figure 3.6.	Effect of the concentration of unlabeled metabolites on the number of metabolites detected (start with 10 μ L saliva and apply dansylation labeling in normal-scale)	91
Figure 3.7.	Effect of concentration process of the labeled saliva solution on the number of metabolites detected.....	93
Figure 3.8.	PCA analysis result for the pilot test with 5 individual saliva samples. Each colour represents 9 replicates of the same individual sample.....	96
Figure 3.9.	PCA analysis result for the storage condition comparison test with 3 individual saliva samples. Each color indicates one specific storage condition, and each shape represents twenty-seven LC-MS injections of one individual sample: nine replicates of freshly collected analysis (red), nine replicates of analysis after four weeks room temperature storage (green), and nine replicates of analysis after four weeks of -80 $^{\circ}$ C freezer storage (blue)	98
Figure 3.10.	(A) The PCA plot of all the data obtained from the 40 individuals (B) The scores plot of the OPLS-DA model demonstrating the separation between the MCI diseased group and the normal healthy group.....	103

Figure 3.11. Box plots of two discriminant metabolites in differentiating MCI from normal healthy control (A) One representative of down-regulated discriminant metabolite with m/z 226.1687 (B) One representative of up-regulated discriminant metabolite with m/z 145.1104.....104

List of Abbreviations

2D	Two-dimensional
A β	Beta-amyloid peptide
AC	Alternating current
ACN	Acetonitrile
AD	Alzheimer's disease
APCI	Atmospheric pressure chemical ionization
API	Atmospheric pressure ionization
B	Homogeneous magnetic field
CE	Capillary electrophoresis
CE	Collision energy
CI	Chemical ionization
CID	Collision-induced dissociation
CRM	Charge residue model
CSF	Cerebrospinal fluid
CUR	Curtain gas
CV	Coefficient of variation
Da	Dalton
dc	Direct current
DIL	Differential isotope labeling
DmPA	p-dimethylaminophenacyl
DnsCl	Dansyl chloride
DP	Declustering potential

EI	Electron impact ionization
EI-MS	Electron impact ionization mass spectrometry
EMS	Enhanced MS
EPI	Enhanced product ion
ESI	Electrospray ionization
FAB	Fast atom bombardment
FFT	Fast Fourier transformation
FT-ICR-MS	Fourier Transform ion cyclotron resonance mass spectrometry
GC-MS	Gas chromatography mass spectrometry
GS1	Ion source gas 1
GS2	Ion source gas 2
HILIC	Hydrophilic interaction chromatography
HMDB	Human Metabolome Database
HPLC	High performance liquid chromatography
IEM	Ion evaporation model
IS	IonSpray Voltage
iTRAQ	Isobaric tag for relative and absolute quantitation
LC	Liquid chromatography
LC-MS	Liquid chromatography mass spectrometry
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LIT	Linear ion trap
m/z	Mass-to-charge ratio
MALDI	Matrix-assisted laser desorption ionization
MCI	Mild cognitive impairment
MeOH	Methanol

MRM	Multiple reaction monitoring
MS	Mass spectrometry
MS ⁿ	Tandem mass spectrometry
MS/MS	Tandem mass spectrometry
NMR	Nuclear magnetic resonance
OPLS-DA	Orthogonal partial least squares-discriminant analysis
PCA	Principal component analysis
Q-TOF-MS	Quadrupole-time of flight-mass spectrometry
QIT	Quadrupole ion trap
QqQ	Triple quadrupole
r	Pearson's correlation coefficient
rf	Radio frequency
RPLC	Reversed phase liquid chromatography
SIDT	Single ion in droplet theory
SIL	Stable isotope labeled
SPE	Solid phase extraction
t _R	Retention time
TEM	Temperature
TLC	Thin layer chromatography
TOF	Time of flight
UV-Vis	Ultraviolet-visible
VIP	Variable importance of the projection

Chapter 1: Introduction

1. General Introduction

Metabolomics is a research field focusing on the identification and quantification of all metabolites in a biological system. Liquid chromatography combined with mass spectrometry (LC-MS) has become an important analytical tool for metabolome analysis. Liquid chromatography is a chromatographic technique that separates a mixture of compounds with the purpose of purifying, identifying and quantifying each individual component in the mixture. It can reduce ion suppression in MS caused by co-elution of compounds, isobaric interferences in use of low-resolving mass analyzers, and sometimes can separate isomers. Mass spectrometry, when coupled with effective sample preparation and chromatographic separation, can offer highly selective and sensitive quantitative analysis, and provide the potential to identify unknown metabolites. An ideal LC-MS platform for metabolomics would identify and quantify all the metabolites in a biological sample, such as blood, urine, saliva, etc. Recent advances in MS instrumentation have provided extraordinarily great progress in the analysis of metabolome. However, due to the vast diversity of physiochemical properties of metabolites, it is difficult to detect and identify all the metabolites with LC-MS-based metabolome analysis. Methods were developed to analyze a group of metabolites sharing similar structural moieties. With this targeted approach, a larger number of metabolites can be identified, thereby enlarging the overall metabolome coverage.

The analysis of human amine- and phenol-containing metabolites has been of great interest for biological studies and disease biomarker discovery using

metabolomics. Comprehensive analysis of amine- and phenol-containing metabolites by LC-MS is challenging due to the great complexity of their physiochemical properties and chemical structures, and matrix and ion suppression effects. The work presented in this thesis was directed towards the development and application of a qualitative and quantitative method for profiling the endogenous amine- and phenol-containing metabolites in human saliva, based on differential ^{13}C -/ ^{12}C -isotope dansylation labeling, with sensitive analysis by liquid chromatography electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (LC-ESI FT-ICR-MS).

1.1. Mass Spectrometry

Mass spectrometry (MS) is an analytical technique that measures the mass-to-charge ratio (m/z) of charged particles. It is used for determining masses of particles, the elemental composition of a sample or molecule, and for elucidating the chemical structures of molecules. In the early history of MS at the beginning of the 20th century, MS was used to probe fundamental aspects of atomic and molecular structure, driving the determination of atomic weights of elements and the discovery of stable isotopes. The birth of MS is commonly attributed to the physicist J. J. Thomson with his discovery of the electron, using an electric field inside a cathode ray tube. His success led him to develop a crude “mass spectrograph” to measure atomic weights of elements.¹ In the first decade of the 20th century, Thomson went on to construct the first mass spectrometer (then called a parabola spectrograph) for the determination of m/z of ions. In this instrument, ions generated in discharge tubes were passed into electric and magnetic fields, which made the ions move through

parabolic trajectories. The rays were then detected on a fluorescent screen or photographic plates. In 1912, as part of his exploration into the composition of canal rays, Thomson and his research assistant F. W. Aston channeled a stream of ionized neon through a magnetic and an electric field and measured its deflection by placing a photographic plate in its path. They observed two patches of light which suggested two different parabolas of deflection, and concluded that neon was composed of two different atomic masses (^{20}Ne and ^{22}Ne), in other words, two isotopes. This was the first evidence for isotopes of a stable element, and the separation of these two neon isotopes by mass using the parabola spectrograph was the first example of MS.²

F. W. Aston went on to work at the University of Cambridge and designed a mass spectrometer in which ions were dispersed by mass and focused by velocity. This improved the MS resolving power by an order of magnitude over the resolution Thomson had been able to achieve.³ Around 1920, A. J. Dempster at the University of Chicago further developed a magnetic deflection instrument with directional focusing of the ions into an electrical collector, which was the format later adopted and still in use today.⁴ He also developed the first electron impact source, which ionized volatilized molecules with a beam of electrons from a hot wire filament. Since the 1930s, a variety of mass spectrometers have been developed and widely used.

Generally, MS instruments consist of three modules: an ion source, a mass analyzer, and a detector. In a typical MS procedure, a sample is loaded onto the MS instrument and undergoes vaporization. The components of the sample are then ionized, which leads to the formation of charged particles. Next, the generated ions are separated according to their m/z in the mass analyzer by electromagnetic fields.

Finally, the ions are detected, and the ion signal is processed into mass spectra. Along with the development of MS, a variety of ion sources, mass analyzers and detectors are available to be used for different study purposes. The choice of instrumentation is usually dependent on the nature of the sample to be analyzed, the analyte of interest and the expected performance, and sometimes based upon the cost.

1.1.1. Electrospray Ionization

The key to using MS for solutions is the ability to transfer the analytes firstly into gas phase, and then into the vacuum region of the mass spectrometer as ionic species. This process is handled by the ionization source. Several types of ionization sources are available including chemical ionization (CI), electron impact ionization (EI), atmospheric pressure chemical ionization (APCI), fast atom bombardment (FAB), matrix-assisted laser desorption ionization (MALDI), and electrospray ionization (ESI). EI is the oldest and best-characterized of all the ionization methods. For the EI process, a beam of electrons passes through the gas phase sample. An electron that collides with a neutral analyte molecule can knock off another electron, resulting in a positively charged ion. This ionization process can either produce a molecular ion which will have the same molecular weight and elemental composition of the starting analyte, or it can produce a fragment ion that corresponds to a smaller piece of the analyte molecule. This technique is well-understood, and able to generate reproducible mass spectra. Fragmentation through EI can provide rich structural information, and libraries of mass spectra are available to be searched for EI mass spectral “fingerprint”. However, the use of EI is limited to thermally volatile and stable molecules, and the signals of molecular ions may be weak or even absent for

many compounds. The great interest in larger biological macromolecule analysis led to the development of a different type of ionization source, for example, MALDI and ESI. J. B. Fenn and K. Tanaka were jointly awarded the Nobel Prize in Chemistry in 2002 “for the development of methods for identification and structure analyses of biological macromolecules” relating to ESI and MALDI, respectively.^{5,6}

The phenomenon of electrospray has been known about for hundreds of years, and was first reported by J. Zeleny.⁷ It was not until the early parts of the 20th century that its significance was fully understood. The pioneering experiments by Dole et al. demonstrated the use of electrospray to ionize intact chemical species and the technique of ESI was invented.⁸ In 1984, work in the laboratory of J. Fenn for the first time showed the use of ESI for the ionization of high mass biologically important compounds and the subsequent analysis by MS.⁹ In the late 1980’s, J. Fenn and his co-workers succeeded in demonstrating the basic experimental principles and methodologies of the ESI technique, such as soft ionization of non-volatile and thermally labile compounds, multiple charging of proteins and intact ionization of complexes. In this thesis work, ESI was chosen as the ionization technique for the analysis of human metabolites, due to its high sensitivity of detecting small molecules and readiness to interface with LC.

In ESI, the sample is introduced into the source in solution either from a syringe pump or from liquid chromatography. The sample solution passes through the electrospray needle to which a high potential is applied at the tip, typically 2 - 4 kV. The induced electric field E_c at the end of the needle is given by:

$$E_c = \frac{2V_c}{r_c \ln\left(\frac{4d}{r_c}\right)} \quad (Eq.1.1)$$

where: V_c = potential applied to the capillary

r_c = radius of the capillary

d = distance from the counter electrode

This makes the sample solution become charged, with the polarity the same as that on the needle. When the charged solution leaves the needle tip, a Taylor cone is formed. Surface tension of the liquid will then resist the increased surface generated by the formation of the cone. At high enough E_c , the cone becomes unstable, and forms a fine filament. Downstream, the liquid filament becomes unstable and breaks into a spray of fine droplets. As the droplets traverse the space to the entrance of a mass spectrometer, solvent evaporation occurs. During the solvent evaporation process, the droplets shrink until they reach the point that the surface tension can no longer sustain the charge (the Rayleigh limit) at which point a “Coulombic explosion” occurs, leading to the fission of the droplets. Finally, gas phase ions are generated and directed into the mass analyzer for analysis (see Figure 1.1). This whole process has been described by Kebarle and Tang.¹⁰

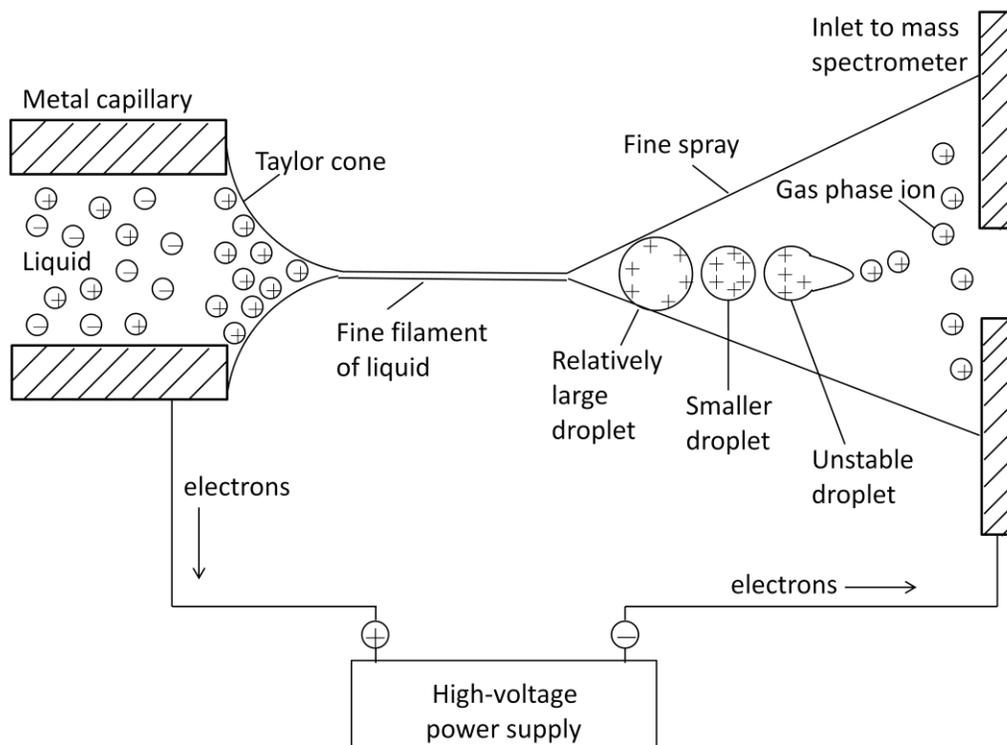


Figure 1.1. Schematic representation of the process of generating gas phase ions in ESI with positive ion mode. (adapted from Quantitative Chemical Analysis 7th Edition)

Two different mechanisms have been proposed to elucidate the generation of gas phase ions from charged droplets. One proposed by Dole et al. in the late 1960s is known as “single ion in droplet theory (SIDT), also referred to as the charge residue model (CRM).⁸ It describes a process of droplet fissure until the microdroplet contains one charge that it emits into the gas phase. To explain in detail, surface-charge density would be increased by the solvent evaporation until the surface tension was comparable with the Coulombic repulsion forces. At this point, a series of “Coulombic explosions” would occur and produce extremely small droplets which contained only one molecule of analyte. By carrying some of the droplet’s

charge, a free gas phase ion was produced. Compared with SIDT, the other theory proposed by Iribarne and Thomson in the late 1970s, ion evaporation model (IEM) is more widely accepted.¹¹ It suggested spontaneous emission of multiple gas phase ions from the droplet surface as long as the droplet reached a certain critical diameter. The IEM was developed from transition state theory. In order for an ion to evaporate out of the solvent, a certain energy barrier due to opposing electrostatic forces must be exceeded. When the droplet radius decreased, the repulsion of escaping ion by other charges was increasing. With the decrease of droplet radius to a certain level, this repulsion would overcome the surface tension at the droplet surface. Elastic deformation of the droplet would then occur where a charged site on a protonated molecule that had approached the droplet surface was moved a distance outside the droplet, facilitating the escape of the ion into the gas phase.

1.1.2. Mass Analyzer

The mass analyzer is the heart of the mass spectrometer. It separates ions according to their m/z and subsequently records their intensities. The performance of mass analyzers is characterized by several parameters, such as resolution, mass accuracy, mass range, possibility of tandem mass spectrometry (MS/MS) measurement, ion transmission, spectral recording speed, and precision isotopic pattern, etc. Four basic types of mass analyzers are widely used in modern MS for small molecule analysis: time-of-flight (TOF), quadrupole and a recent derivative, quadrupole linear ion trap (LIT), quadrupole ion trap (QIT), and FT-ICR. In this work, the first studies were carried out using a hybrid triple quadrupole-linear ion trap MS

instrument and further qualitative and quantitative studies were performed using an FT-ICR-MS.

1.1.2.1. Quadrupole MS

The quadrupole mass analyzer was developed in parallel with the quadrupole ion trap by the third Nobel prize winning mass spectrometry pioneer, Wolfgang Paul,^{12, 13} and is still the most commonly used mass analyzer in analytical laboratories today. It works on the basis of electric field generated between a set of four axial rods through which ions pass on their way to the detector. The voltages applied to these rods consist of direct current (dc) and radio frequency (rf) components. They create a hyperbolic field to allow a certain m/z range to pass through for a given combination of potentials; and ions outside that m/z range will hit the rods, be discharged and not be transmitted. Therefore, the quadrupole is also referred to as a mass filter, into which ions are accelerated from the ion source by a small potential. The applied potential to the rods is of the form: $\Phi_0 = U + V\cos(\omega t)$, where U is the dc voltage and $V\cos(\omega t)$ is the rf potential with frequency $\omega/2\pi$. Two rf waveforms are applied to two pairs of opposing parallel rods, with a 180 degree phase shift. As a result, the trajectory of an ion fluctuates constantly as it travels along the rods until it hits the detector. In a full-scan mode, the dc and rf components are ramped at a constant ratio, and ions entering from the ion source are enabled to pass through the rod assembly successfully. For a specific combination of dc and rf, only a very small m/z range can pass through. Ultimately, ion trajectories and stabilities inside the quadrupole mass filter are directly dependent on the applied dc and rf voltages, as defined by Mathieu equations and as given by the a-q stability diagram (see Figure 1.2). The canonical

form of Mathieu's differential equation has two types of solutions: an unbounded solution and a bounded solution. The former solution corresponds to an unstable trajectory in a quadrupole analyzer, while the latter corresponds to stable ion trajectory in the quadrupole and the motion of the ion in the z-direction can be summarized by a stability diagram given in terms of the Mathieu coordinates a and q , where $a = 4eU/mr^2\omega^2$ and $q = 2eV/mr^2\omega^2$. The m/z range of stable ion motion can be made wider by decreasing the slope of the mass scan line ($a/q \sim 2U/V$), or be made narrower by increasing the slope, so that the mass scan line can only intersect the tip of the a - q diagram. If U/V is kept constant, then a/q or the slope of the mass scan line is kept constant as well. In this situation, if the values of U and V are increased simultaneously, then ions with increasing mass within the certain defined mass window will be brought to the tip of a - q diagram, and take turns to pass the mass filter achieving mass scanning.^{14, 15}

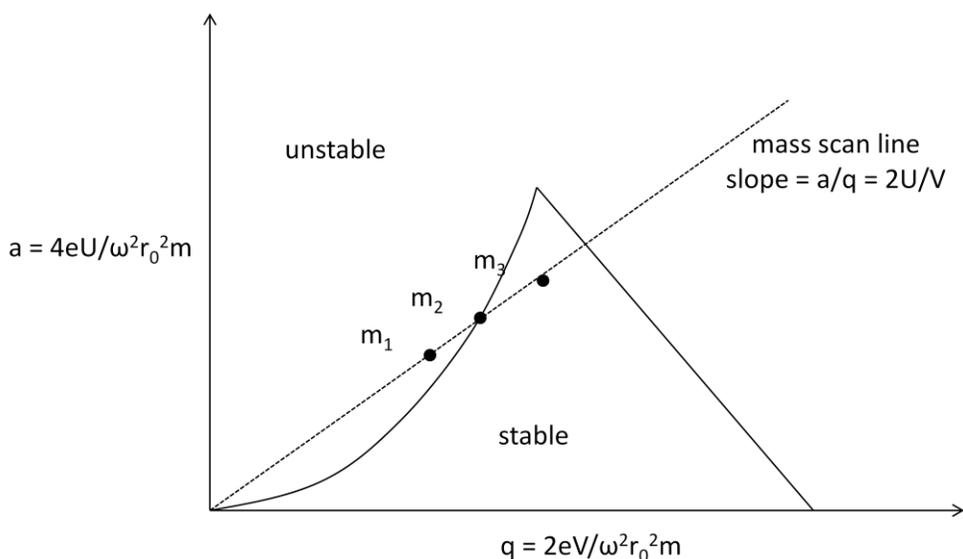


Figure 1.2. The a - q stability diagram ($m_1 > m_2 > m_3$). (adapted from Chemistry 505 class notes, Fall 2009, Prof Liang Li, University of Alberta)

When a single quadrupole instrument is operated with ESI source, selectivity is always not good enough. It can be increased when fragmentation is induced in the ion source region (in-source fragmentation). This procedure refers to the activation of ions in the region between the ion source and the mass analyzer, in which fragmentation can be initiated by collisions with residual gas molecules at intermediate pressures. Therefore, more fragment ions can be produced for structure elucidation and conformation. However, as no precursor ions will be selected and fragment ions cannot be specifically linked to the precursor ions of interest, structure assignments are made more difficult especially when fragment ions could originate from several co-eluted compounds in one LC-MS peak. Currently, the most common mass analyzer for quantitative bioanalytical analyses is the triple quadrupole. This tandem mass spectrometry instrument consists of three sequential quadrupoles (Q1-q2-Q3). This configuration allows additional ion activation in q2 after ions of interest have been selected in Q1. The second quadrupole q2 is operated in rf-only mode, and can be filled with a neutral gas such as N₂ or Ar to act as a collision cell. The fragment ions generated due to collision-induced dissociation (CID) occurred in q2 will then be analyzed with the third mass analyzed Q3. This is the simplest form of an MS/MS experiment on a triple quadrupole MS, called product ion scan. Depending on the different modes Q1 and Q3 are used, other MS/MS scan modes are also available, such as precursor ion scan, neutral loss scan and multiple reaction monitoring (MRM), etc.

LIT is another emerging mass analyzer being widely used. Ions are confined radially by a two-dimensional (2D) rf field, and axially by stopping dc potentials applied to end electrodes. It can either act as a mass analyzer or be used for storing

ions, with higher injection efficiencies and higher ion storage capacities. For the hybrid triple quadrupole-linear ion trap-mass spectrometry (QTRAP®), the triple quadrupole ion path can be reconfigured such that the final quadrupole Q3 will be configured as a LIT (QqLIT, see Figure 1.3). With this configuration, Q1 selects the precursor ions, which are accelerated into a transmission collision cell to induce fragmentation. The ions generated in q2 are then trapped in the Q3 LIT and subsequently mass selectively scanned out of the trap toward the detector. The fragmentation patterns generated from this new configuration are the same as those from a conventional triple quadrupole mass spectrometer. However, the effects of increasing ion density in the Q3 LIT are significantly enhancing the detection sensitivity and having less detrimental effect on mass spectral peak shape as well as mass assignment, thereby helping greatly structure elucidation of unknown compounds

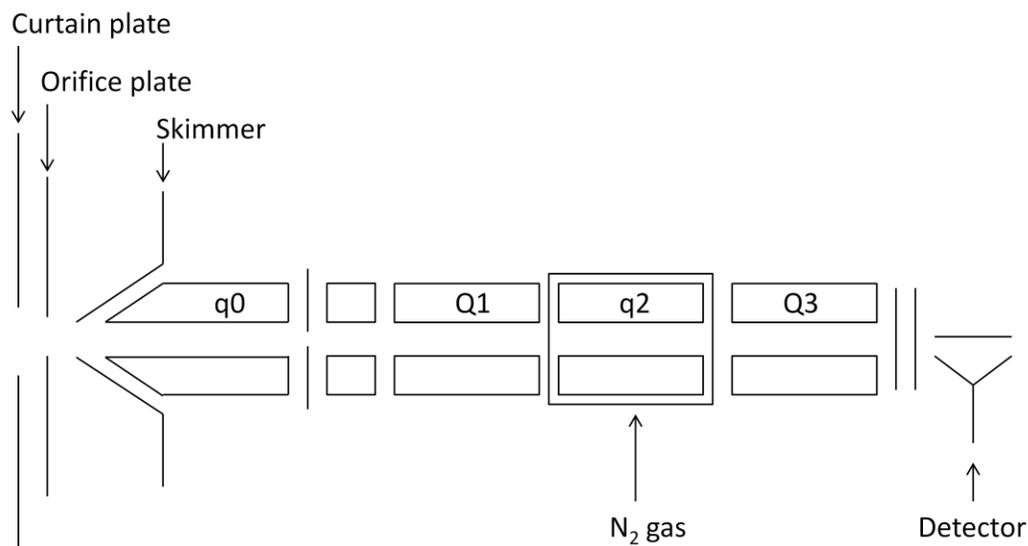


Figure 1.3. Schematic of QqLIT (QTRAP).

In this work, a QTRAP instrument was used to perform tandem MS experiments (MS^n). Various modes of operation could be achieved by using this instrument (listed in Table 1.1). Specific scan functions of traditional triple quadrupole such as product ion scan (PI), constant neutral loss scan (NL), precursor ion scan (PC) and selected reaction monitoring (SRM) mode are still maintained in this instrument. With the third quadrupole Q3 performed in LIT mode, the term enhanced is introduced in the modes of enhanced MS scan (EMS), enhanced product ion scan (EPI), etc. In addition, two particular modes to perform MS/MS experiment can also be performed, which are the time-delayed fragmentation (TDF) and the enhanced multiply charged (EMC) scan mode that selects multiple charged ions in the trap mode. In the qualitative work presented in Chapter 2, an EMS scan was performed first to find the protonated molecular ion, followed by an EPI scan to generate an MS/MS spectrum. EPI scans were also carried out to generate MS^3 , or we called *pseudo-MS³* spectra by increasing the skimmer voltage. Normally, the ionization process is sufficiently violent to leave the resulting ions with sufficient internal energy to fragment within the mass spectrometer. If the product ions persist in their non-equilibrium state for a moderate amount of time before auto-dissociation, this process is called metastable fragmentation. Skimmer fragmentation refers to the induction of in-source fragmentation by increasing the skimmer voltage. Although it allows for fragmentation analysis, it is not technically tandem mass spectrometry unless the metastable ions are mass analyzed or selected before auto-dissociation and a second stage analysis is performed on the resulting fragments. By coupling the in-source fragmentation with tandem mass spectrometry to allow for two steps of fragmentation, a *pseudo-MS³*-type of experiment is performed, and the resulting mass spectrum is

considered as a *pseudo-MS*³ spectrum. Due to the lack of precursor ion selection, skimmer fragmentation was not widely used for structural analysis. However, in some cases, such as the generation of precursor ions of unlabeled metabolites from the dansylation labeled metabolites in the study introduced in Chapter 2, skimmer fragmentation was particularly useful to provide unlabeled metabolites precursor ions with abundance that was high enough to perform further fragmentation.

Table 1.1. Description of triple quadrupole and QqLIT operation modes.

Mode of Operation	Q1	q2	Q3
Q1 Scan	Resolving	RF-only	RF-only
Q3 Scan	RF-only	RF-only	Resolving
Product Ion Scan	Resolving	Fragment	Resolving
Precursor Ion Scan	Resolving	Fragment	Resolving
Neutral Loss Scan	Resolving	Fragment	Resolving
Selected Reaction Monitoring	Resolving	Fragment	Resolving
Enhanced MS Scan	RF-only	No Fragment	Trap/Scan
Enhanced Product Ion Scan	Resolving	Fragment	Trap/Scan
Time Delayed Fragmentation	Resolving	Trap/No Fragment	Fragment/Trap /Scan
Enhanced Multiply Charged	RF-only	No Fragment	Trap/Scan

1.1.2.2. Fourier Transform Ion Cyclotron Resonance MS

Fourier Transform Ion Cyclotron Resonance MS (FT-ICR-MS) is a type of mass analyzer for the determination of m/z of ions based on the cyclotron frequency of the ions in a fixed magnetic field, which was derived from the cyclotron principles first introduced by Lawrence.¹⁶ FT-ICR-MS is an analytical technique with ultrahigh resolution in that masses of ions can be measured with very high accuracy. It is also because a superconducting magnet is much more stable than rf voltage that the FT-ICR-MS being able to achieve higher levels of resolution than other mass analyzers.¹⁷ This high resolution allows the signals of two ions with similar m/z to be detected as distinct ions, which is highly important in analyzing complex mixtures.¹⁸⁻²⁰ Different from other trapping mass analyzers that have to scan one frequency at a time to generate the spectrum, the ICR cell used here is able to collect an entire spectrum at once. In the 38 years since its inception in 1974 by Comisarow and Marshall,²¹ FT-ICR-MS has been widely used and is now still one of the most sensitive methods of ion detection in existence. The FT-ICR-MS used in this study the Bruker 9.4 T Apex-Qe FT-ICR-MS.

As the commonly used ionization technique coupled with FT-ICR, ions are usually introduced to the mass spectrometer *via* ESI. The ions then pass through a series of pumping stages at increasingly high vacuum until they enter the ICR cell in which the pressure is in the range of 10^{-10} to 10^{-11} mBar with temperature close to absolute zero. In the cell where a spatially uniform magnetic field is present, a moving ion will face a force, the Lorentz force, given by *Eq. 1.2*

$$\text{Force} = (\text{mass})(\text{acceleration}) = m \left(\frac{dv}{dt} \right) = qv \times B \quad (\text{Eq. 1.2})$$

where m , q , and v represent ionic mass, charge and velocity, and the vector cross product term “ $\times B$ ” demonstrates that the direction of the magnetic component of Lorentz force is perpendicular to the plane established by v and B . If no collisions occur and the ion keeps a constant speed, the magnetic field will bend the ion path into a circle, of which the radius is r (see Figure 1.4).²² Because the angular acceleration in the plane perpendicular to B is expressed as $dv/dt = v_{xy}^2/r$, and the angular velocity (ω) is defined as $\omega = v_{xy}/r$, then Eq. 1.2 becomes $m\omega^2 r = qB\omega r$, or is simplified to give Eq. 1.3

$$\omega = \frac{qB}{m} \quad (\text{Eq. 1.3})$$

From this equation, in which the “unperturbed” ion cyclotron frequency is denoted, it is illustrated that all ions of a given m/z , or here m/q , have the same ICR frequency that is independent of their velocity. As translational energy is no longer essential for the precise mass measurement, this feature meanwhile makes ICR particularly useful for MS.

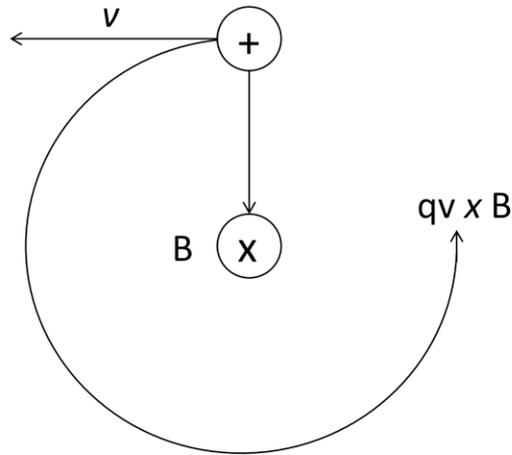


Figure 1.4. Ion cyclotron motion. Moving path of the positive ion in the plane is bent into a circle by the Lorentz magnetic force generated by a homogenous magnetic field perpendicular to the plane. (adapted from Marshall et al., 1998²²)

The commonly used ICR trap introduced by Comisarow^{23, 24} consists of six electrodes (three pairs). One pair of three provides the dc trapping potential, while the other two pairs supply potential for excitation/ejection and detection (see Figure 1.5). When a small symmetric electric field is applied onto the trapping plates which are perpendicular to the magnetic field, a potential well will be generated to trap the ions inside the cell, causes the harmonic oscillation of ions along the z-axis of the magnetic field, and confines the ions axially. Meanwhile, with the spatially uniform magnetic field, the ions are confined in the x,y-plane.

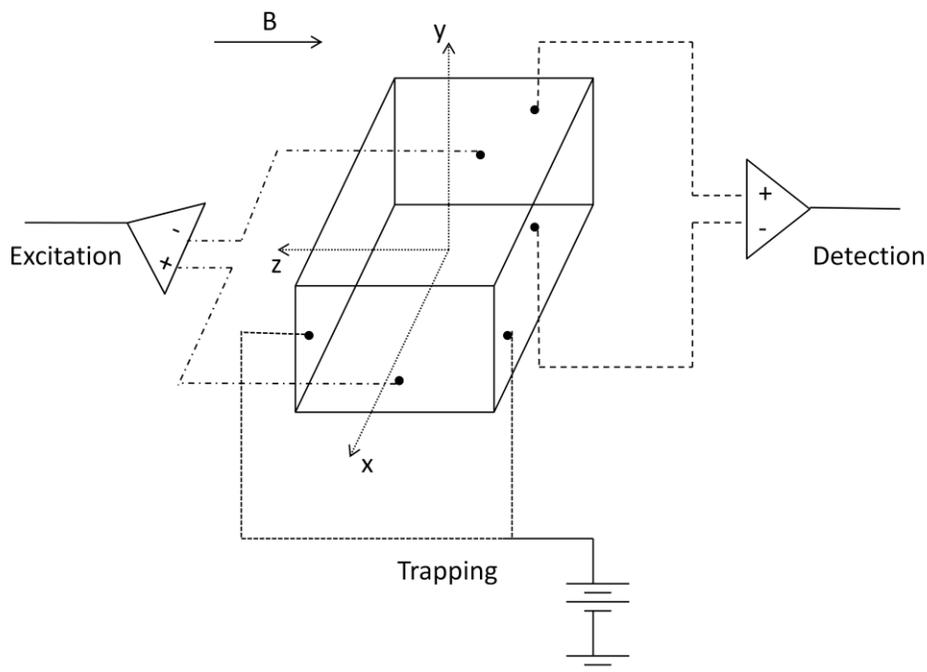


Figure 1.5. Simplified depiction of an ICR cell. The plates positioned on the front and the back are two excitation plates, on the two sides are the trapping plates, and the two on the top and bottom are the detection plates. (adapted from Marshall et al., 1998²²)

The sequence of four events that occur in an FT-ICT-MS experiment is to first quench, in which process a large negative voltage will be applied to the trapping plates to remove all ions leftover in the cell may be from a previous experiment. This is followed by ionization as the second step. An electron beam or a laser beam will be used to ionize the molecules inside the cell or a packet of ions will be introduced into the cell. A fast rf sweep will be applied to excite all ions to larger cyclotron orbits. Finally, decay of the cyclotron motions will induce image current that can be amplified, digitized and stored in memory. Deconvolution of this signal by fast Fourier transformation (FFT) methods will result in the deconvoluted frequency *vs.*

intensity spectrum which is then converted to the mass *vs.* intensity spectrum, known as a mass spectrum.

In this sequence, an important step is to excite the ions in order to be detected. After ions are formed and trapped inside the ICR cell, they often have only a small amount of energy, therefore the radius of an ion's cyclotron orbit is usually small compared with the dimensions of the cell. In order to detect the ions, they are excited by applying an alternating current (AC) electric field to the two excitation plates that are parallel to the axis of the magnetic field. The ions will spiral outwards when its cyclotron frequency is in resonance with the frequency of the applied rf electric field. By continuously applying the rf potential, the ion will spin outward all the way until they strike an excitation or detection plate and be neutralized (see Figure 1.6). All ions of the same m/z are excited coherently, grouped tightly after excitation, and experienced cyclotron motion as a packet. This packet then induces a differential current between the two opposing detection plates that also lie parallel to the magnetic axis. This alternating current is called the image current, and its amplitude is proportional to the population of ions in the packet with same m/z . To accomplish broadband detection that can detect ions of many masses simultaneously, many frequencies are applied during the excitation process and a rapid frequency sweep, in other words, an rf chirp is carried out. This will cause all the ions with cyclotron frequencies in the programmed range to be excited almost simultaneously, and finally lead to a composite image current and the resulting mass spectrum.

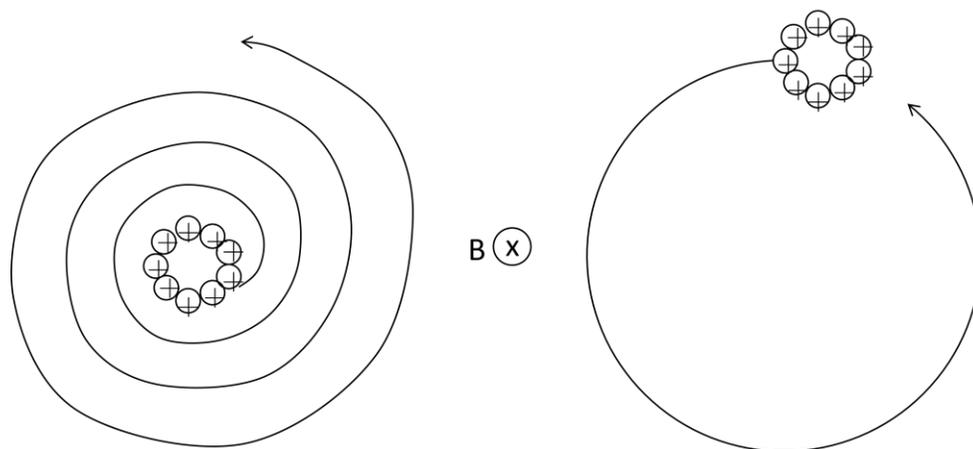


Figure 1.6. Ion cyclotron orbital motion of excitation and detection.

With the ion trap nature of FT-ICR-MS, it is possible to measure ions without destroying them. This feature enables the ions to be performed with further experiments, such as MS/MS or even MS^n for unknown compounds structural elucidation. To use FT-ICR-MS for tandem mass spectrometry experiments, excitation waveforms can be manipulated to excite ions in a mass-selective fashion.

1.1.3. Flow Injection

Flow injection mass spectrometry analysis, also referred to as direct injection mass spectrometry analysis, describes the injection or infusion of a sample solution into the ionization source of a mass spectrometer without chromatographic separation. This technique is normally used with atmospheric pressure ionization (API) techniques, especially ESI. Flow injection is an approach with high-throughput, which is able to process a sample within a few minutes. For the past few decades, flow injection has been performed a lot using ESI and nominal mass resolution mass analyzers. Lewis et al. characterized a humanized monoclonal antibody by flow

injection ESI in positive mode and triple-quadrupole mass spectrometer.²⁵ Castrillo et al. also used direct injection ESI in positive mode with triple-quadrupole mass spectrometer to analyze yeast intracellular metabolites.²⁶ Blok-Tip et al. elucidated the structure of unknown compounds present in herbal products using ESI in positive mode and an ion trap mass spectrometer.²⁷

In order to distinguish between isobars that sharing the same nominal mass, high-resolution mass spectrometers are of choice. They also allow accurate mass measurement and help the calculation of empirical formulas. TOF mass spectrometers and FT-ICR mass spectrometers have been used to perform the flow injection high-resolution mass spectrometry. Allen et al. used TOF-MS with ESI in positive mode for metabolic footprinting analysis of yeast cell media samples.²⁸

Hybrid instruments, such as QTRAP, have also been used with flow injection. Structural elucidation of compounds can be performed by MS/MS experiments without being connected to another instrument. Using the first quadrupole as a scanning device, precursor ions can be selected for fragmentation in the second quadrupole, which is served as a collision cell. Product ions are then scanned and detected in the third quadrupole, used as a linear ion trap, to produce product ion spectra. Another hybrid instrument, quadrupole-time of flight-mass spectrometry (Q-TOF-MS) has been used frequently. Scholz et al. used this instrument for rapid screening analysis of plant extracts.²⁹ With the TOF being used as mass analyzer, accurate mass measurement with high mass resolution can be achieved. This accurate mass measurement of precursor and product ions provides clarification of fragmentation processes, hence dramatically facilitating the spectra interpretation.

Although flow injection-MS is a high-throughput technique for metabolome analysis, it has several disadvantages. Chemical isomers cannot be distinguished as they have exactly the same mass, thereby need chromatographic separation. Formation of in-source fragmentation ions or that of adduct ions is complicated. Besides, in the case of ESI, ion suppression is also a severe concern. Since all sample analytes are injected simultaneously into the ionization source, signal suppression can occur to those analytes that are relatively difficult to be ionized. This may prevent the data analysis to be complete and meaningful.

1.1.4. Liquid Chromatography

Liquid chromatography (LC) can be coupled with mass spectrometry to separate the analytes in a mixture before ionization and detection, so that more individual components in the sample mixture can be identified and quantified. For metabolome analysis that requires to analyze thousands of analytes in a single sample, separation such as LC is especially important to be used in line with the mass analyzer to achieve identification and quantification of individual metabolites. The most commonly used form of LC for current metabolomics studies is reversed phase LC (RPLC), derived from the opposite technique of high performance liquid chromatography (HPLC), as a standard tool for the separation of medium polar and non-polar analytes. RPLC has a non-polar stationary phase inside a column (a cylindrical tube) and an aqueous, moderately polar mobile phase composed of water with an organic solvent, typically acetonitrile (ACN) or methanol (MeOH), being pumped through the column. After injection of sample for analysis, it is carried into the column by the mobile phase and interacts with the stationary phase while

traveling along the column. One common stationary phase is porous silica (typically 10 nm pore size for use in small molecule studies), which is normally 1.5-5 μm in diameter. It has been surface-modified with RMe_2SiCl , where R is a straight chain alkyl group such as $\text{C}_{18}\text{H}_{37}$. Different analytes will be carried by mobile phase to go through the column, enter the pores by diffusion and interact with the C18 groups. The separation of analytes in the column is achieved by differential migration. With such stationary phase, retention time will be longer for molecules that are less polar, as they will have more interaction with C18 stationary phase, hence retained more and will elute later from the column. For relatively polar analytes, they will migrate through the column with a higher velocity compared with less polar components, thereby eluting from the column earlier. The longer time the analytes migrate through the column, the further separated of them along the column. However, as it is not possible for all the analytes to all have different polarities, analytes with same polarity will form a certain zone, often referred to as a band, during the migration. When a band is eluted out from the column to be detected, signal will be displayed as a chromatographic peak. For an ideal separation, each peak represents a specific analyte. The retention time of an analyte, t_R , is taken as the elapsed time between the time of injection of the analyte and that of elution of the peak maximum. It is a unique characteristic of the analyte and can indicate the identity of the analyte. Associated with the analyte peak, peak area is representative of the analyte concentration in the sample. Therefore, using LC in combination with MS, identification and quantification techniques for analytes are improved.

Up-to-date development in LC/MS coupling technique makes the combination of LC with MS user friendly. LC/MS interface provides the connection between the

LC separation system and the mass spectrometer. It is responsible for the reliable and efficient transfer of analytes from the solution phase eluted from the column to the gas phase. It is also responsible for a critical process in mass spectrometry: ionization; and the most frequently used is ESI. In gradient elution LC, the mobile phase is changed to less polar gradually throughout the whole elution period by increasing the proportion of organic solvent in the mobile phase, thereby getting the less polar analytes eluted out from the column. On the other hand, addition of organic solvent into mobile phase decreases the surface tension of ESI spray droplets, and increases the detection sensitivity.³⁰ The overall sensitivity of ESI-MS is also limited by the ionization efficiency. By coupling LC with MS to separate the analytes and reduce the complexity of the sample introduced to ESI, the commonly observed ion suppression effect in ESI is reduced as well, which then increases the analysis sensitivity. However, even with the prior separation done by LC to reduce ion suppression, it is still a major issue.

LC/MS and liquid chromatography tandem mass spectrometry (LC-MS/MS) have been established as the most sensitive and selective analytical techniques for biological studies. Regardless of the sensitivity or selectivity of the mass analyzer of choice, ion suppression is always a form of matrix effect that LC-MS techniques suffer from. Ion suppression effect is the negative effect that matrix complexity can have on analyte ionization in electrospray, and will negatively affect several analytical figures of merit, such as detection capability, precision and accuracy. To further explain this, ion suppression occurs in the early stage of the ionization process in the LC-MS interface, when a component eluted from the HPLC column influences the ionization of a co-eluted analyte. This suppressed ionization will consequently

lead to analytes not being fully ionized nor transmitted to mass analyzer, therefore influence the detection capability, and subsequently affect the precision and accuracy of the analysis.

Many studies have reported difficulties with reproducibility and accuracy when analyzing small quantities of analytes in complex samples such as biofluids. Most of the difficulties were pointed out to be related to ion suppression.³¹⁻³³ Even though the origin and mechanisms of ion suppression are not fully understood or proven, many possible sources for ion suppression are proposed, including endogenous compounds and exogenous substances from the sample matrices, and molecules from contamination during sample preparation, such as polymers extracted from plastic vials, etc.³⁴ Some other factors are also considered to make a compound a primary candidate for inducing ion suppression, for example, high concentration, mass and basicity.³⁵ Many possible mechanisms of ion suppression have been proposed as well. Although ion suppression can have potential deleterious effects on both ESI and APCI, evidence indicated that the electrospray interface was more likely to be impacted.³⁶ Due to a limited amount of charge available on ESI droplets or to the saturation of ESI droplets with analytes on the surface at high concentrations, the ions trapped inside the droplets are prevented from ejecting. Besides, in multi-component samples with high analyte concentrations, competition for either space or charge is most likely occurring, result in the suppression of signals. This is mainly attributed to the characteristic, such as surface activity and basicity, of an analyte. Because biological samples often contain large amounts of basic endogenous compounds, ion suppression is always inevitably observed. In response to this problem, several approaches can be used to reduce ion suppression, for example, by using an

optimized LC separation method to reduce analytes co-elution, by maintaining clean LC system and sample preparation steps to reduce the introduction of impurities, etc. Some sample preparation clean-up procedures such as solid phase extraction (SPE) can be used as well to reduce sample complexity, hence reducing matrix effects and ion suppression. These approaches are often not feasible, as analytes might be incompletely extracted and losses will occur if the extract is too concentrated. The addition of a trace amount of acid to both the sample and mobile phase can also improve analyte ionization efficiency by changing the basicity of the basic species in the biological matrix. Finally, to correct for the ion suppression effect on quantitative analysis, an internal standard can be used, especially where the analyte and the internal standard co-elute to suffer same extent of ion suppression.³⁵

Another mechanism for ion suppression in ESI is explained by the increased viscosity and surface tension of the droplets from the high concentration of interfering compounds. This reduces the solvent evaporation as well as the ability of the analytes to get into the gas phase.^{37, 38} It has been pointed out that existence of non-volatile components such as salts in the sample can cause ion suppression as well. Non-volatile compounds can decrease the efficiency of droplet formation, therefore preventing the droplets from reaching the critical radius necessary for the gas phases to be emitted.³⁵ Some studies have also referred the cause of ion suppression to the higher mass and higher polarity of analytes. This will suppress the ionization efficiency of those analytes with lower mass and lower polarity.^{35, 39}

So far, there is no universal solution to solve ion suppression issue. However, numerous chromatographic and sample preparation techniques have been introduced and further optimized to prevent the occurrence of ion suppression. Some other

calibration or normalization methods are also developed to balance the effects or to minimize the consequences. To eliminate the risk of ion suppression, more attention should be paid to the optimization of sample preparation, analytes chromatographic separation and calibration techniques. Among all of the mentioned improvement approaches, it is obvious that the key is to achieve better chromatographic separation of all analytes in the sample prior to ionization.

1.2. Metabolomics

Metabolomics is the comprehensive analysis of all endogenous metabolites in a biological system. Metabolites are small molecules that participate in general metabolic reactions and that are required for the maintenance, growth and normal function of a cell. They are the end products of cellular processes, and their concentrations can be treated as the definitive response of a biological system to genetic or environmental influences.^{40, 41} To thoroughly understand a living system, the detection, identification and quantification of all metabolites are of importance. The complete set of metabolites in an organism is defined as the metabolome. Although metabolome analysis is still in its infancy, it is obvious that metabolomics is playing a more and more central role in biological studies, as well as serving as a powerful tool in disease biomarker discovery and drug discovery.⁴²⁻⁴⁴

The beginning of metabolomics can be traced back all the way to 2000-1500 B.C. when traditional Chinese doctors started using ants to evaluate the urine of patients to determine if the urine contained high glucose of diabetics. The fundamentals of metabolomics have also been in practice as scientific revolution for a long history. For example, it was recognized that body fluids could be used to

determine a disease state, such as with the identification of a high concentration of glucose in diabetic urine in the 17th century. In the early 1970s, GC-MS based approaches for small molecule analysis were firstly introduced by Dalglish et al. in 1966, describing GC-MS based approach for profiling of urine and tissue extracts for a wide range of metabolites.⁴⁵ These experiments were then expanded by Mamer et al.⁴⁶ and Horning et al.⁴⁷ in 1971 to perform further GC-MS based urine metabolome profiling. In the same year, Pauling and Robinson investigated biological variability being explained by ranges of nutritional requirements, and published the first paper on metabolomics.⁴⁸ These early experiments lead to an entire generation of metabolomics studies being performed by GC-MS. It is still in use today for metabolome profiling studies. However, this field of study currently relies more on nuclear magnetic resonance (NMR) and LC-MS.

In the past decade, the field of metabolomics has been greatly developed, and regarded as an important field of “omics” research. The term metabolome was first coined in 1998 by Oliver.⁴⁹ In 2001 Fiehn first defined metabolomics as the comprehensive and quantitative analysis of all metabolites that could help in the understanding of biosystems and revealing of their metabolome.⁵⁰ Metabolomics represents an interface between genetic pre-disposition and environmental influence, occupies a unique position in the biology system hierarchy, and has evolved rapidly from a low key research area to a mainstream study field. This should also be attributed to the improved analytical techniques for small molecule analysis, such as liquid chromatography, mass spectrometry and nuclear magnetic resonance spectroscopy.

It is desirable to generate a global metabolome profile with one approach. Comprehensive LC-MS based techniques have been of interest in the area of untargeted and unbiased metabolomics studies. With MS to be used for detection, the improved separation and ion production through LC has helped to increase the coverage of metabolome. Hence LC-MS has also become an analytical tool to be chosen for metabolome profiling.

Metabolome profiling focuses on the analysis of a particular group of metabolites either related to a specific metabolic pathway or a class of compounds. An even more directed approach is targeted analysis that aims at measurement of selected analytes such as disease biomarkers. The results of metabolome profiling are normally quantitative and ideally independent of the technology used for data acquisition. Consequently, results obtained from several different analytical platforms can be compiled to give a global overview of a particular biosystem metabolome, or used to build databases that can be integrated with pathway maps or other “omics” data to enhance biological understanding. This former approach was used by Wishart et al. in the profiling of human cerebrospinal fluid (CSF) by compiling the results of NMR, GC-MS, and LC-MS analyses. With this approach, they successfully identified 308 metabolites.⁵¹

The disadvantage of metabolome profiling is that this study system is not a truly global “omics” approach. An ideal metabolomics study should provide a comprehensive, qualitative and quantitative overview of all metabolites in a biological system. Due to the sheer number of metabolites present in the system, and their large dynamic range and great diversity in physiochemical properties, the universal detection of all metabolites is extraordinarily challenging. However,

numerous quantitative metabolic profiling methods focusing on different metabolite classes have already been developed and frequently used. A powerful metabolome profiling approach, targeted metabolomics, will still evolve.

1.2.1. Targeted Metabolomics

Targeted metabolomics approaches refer to a method in which a specified list of metabolites is measured, typically focusing on one or more related pathways of interest, or a specific group of metabolites. The metabolome of a biological system can be viewed as the combination of many sub-metabolomes according to the common properties that the metabolites share. To get a more complete and accurate profile of the metabolome, a targeted metabolome approach is applied. The strategy is to fractionate the whole metabolome into several sub-metabolomes according to hydrophobicity, chemical structures, or other properties, analyze each group separately, and finally compile all the results.

Developments in MS and NMR provide great advantages for performing metabolomics studies due to their specificity and quantitative reproducibility. Moreover, there are many other analytical tools in principle that could be considered for metabolites measurement, such as ultraviolet-visible (UV-Vis) spectroscopy. However, no single analytical platform is suited to provide a comprehensive view of the metabolome as a whole. Therefore, targeted metabolome analysis is getting more and more popular. Among all the targeted metabolomics studies, disease biomarker discovery research has been discussed extensively.⁵²⁻⁵⁶ Introduction of MS/MS experiments by using triple quadrupole (QqQ) mass spectrometry and improvement in LC separation techniques further improved the limits of detection and

quantification of diagnostic metabolites in low abundance. Applications of such analytical platforms to targeted metabolomics studies have provided a highly sensitive and robust method for measuring a significant number of biologically important metabolites with relatively high throughput, in other words, disease screening and identification. For example, in diabetes-related research, targeted metabolomics approaches were used to investigate patient response to glucose challenge.⁵⁷ In other studies, targeted screening also revealed citric acid metabolites and a small group of essential amino acids as metabolic signatures of myocardial ischaemia and diabetes, respectively.^{58, 59}

Our group has been developing a metabolome profiling platform whereby metabolites of a sub-metabolome sample are selectively labeled with an isotope reagent that reacts with a specific functional group. The labeling reagents are rationally designed to improve the performance of LC separation (i.e., better retention on a high-efficiency reversed-phase column) and electrospray ionization (i.e., better detectability and higher sensitivity in overall detection). We reported the ¹³C₂- and ¹²C₂-dansylation chemistry for profiling amine- and phenol-containing metabolites⁶⁰ and ¹³C₂- and ¹²C₂-p-dimethylaminophenacyl (DmPA) bromide chemistry for profiling carboxylic acid-containing metabolites⁶¹. Our work on the use of ¹⁴N₂-/¹⁵N₂-dansylhydrazine for profiling ketones, aldehydes, and sugars has been submitted for publication. These three isotope labeling chemistries offer a convenient and quantitative route to profiling a large number of metabolites; more than 80% of the 8,000 known human metabolites in the Human Metabolome Database or HMDB⁶² contain one or more of the targeted functional groups. Studies are still on-going for the isotopic labeling of other sub-metabolomes. The studies of LC-MS based

quantitative isotopic profiling of human salivary amine- and phenol-containing sub-metabolomes with ^{13}C -/ ^{12}C -dansylation labeling are presented in this thesis.

In the analysis of endogenous metabolites, several analytical platforms have been used. NMR is a popular technique. It has potential for high-throughput analysis, minimal requirements for sample preparation, and non-discriminating and non-destructive nature of the technique. However, only medium to high abundance metabolites will be detected. GC-MS has traditionally been an approach of choice. The combination of gas chromatography with electron impact ionization mass spectrometry (EI-MS) provides high chromatographic resolution, analyte-specific detection, and the ability to identify unknown compounds. Unfortunately, a sufficient vapor pressure and thermal stability of the analytes are required, though derivatization at the functional group could be applied to some analytes to reduce the polarity and increase the thermal stability and volatility. Capillary electrophoresis (CE) is a powerful technique for charged metabolites with high resolution. Coupling of CE with MS makes it an ideal tool for metabolome analysis. However, not a lot of metabolome studies have been done by CE-MS so far. Currently, LC-MS is still the technique best suited for the analysis of biological samples that are extremely complicated. Among a variety of different mass analyzers, FT-ICR mass spectrometer is of choice most frequently. It offers very high resolution (100,000-1,000,000) and highest available mass accuracy (0.1-1 mamu), very low limits of detection in range of attomole to femtomole, and MS^n capabilities.⁶³⁻⁶⁵ In targeted metabolome studies of biological systems, both known and unknown metabolites are expected to be detected, identified and quantified. The ultrahigh resolving power of FT-ICR-MS is superior in determining their structures.

Identification and quantification of amine- and phenol-containing metabolites in complex biological samples are critical for disease biomarker discovery. For example, homocysteine was found to be a risk factor for stroke,⁶⁶ and also to have influence on the occurrence of Alzheimer's disease (AD).⁶⁷ As another example, amine accumulation is proposed to be an attribute of Parkinson's disease.⁶⁸ In addition, polyamines are reported to be essential for cellular growth and function;⁶⁹ and can regulate the cell-death process called apoptosis as well.⁷⁰ All these studies have shown that different amine- and phenol-containing metabolites have the potential to be biomarkers for a variety of diseases. Therefore, advanced analytical techniques are essential to be developed for qualitative and quantitative analysis of amine- and phenol-containing metabolites.

1.2.2 Dansylation Labeling

A major challenge for metabolome profiling is the analysis of highly polar metabolites present in majority in a specific metabolome of biological system.⁷¹ Metabolites with high polarity or hydrophilicity, for example, amine-containing metabolites, are poorly retained on a RPLC stationary phase, and will elute almost at the initial void. However, ESI-MS detection sensitivity near the void is normally severely reduced. This is because of the poor ESI desolvation performance due to the high percentage of aqueous mobile phase in the RP gradient at the beginning period. Significant ion suppression may be further created by the co-elution of polar species and salts, which decreases even more the ESI signal of targeted polar analytes. Besides RPLC, other different separation mechanisms have also been reported to be useful, such as hydrophilic interaction chromatography (HILIC).^{72, 73} Compared to

RPLC, however, the separation efficiency of HILIC is comparatively low for separating complicated mixtures. Alternatively, changing the chemical properties of analytes, such as their hydrophobicity, could be a potential way to have the hydrophilic analytes separated with high efficiency by RPLC and compatible to ESI-MS. This alteration can be accomplished through chemical derivatization.

Dansyl chloride is a commonly used derivatizing agent for amine- and phenol-containing compounds. Dansylation is simple, robust, and has been routinely performed for many years as pre-column derivatization for the quantification of amino acids, biogenic amines and phenolic hydroxyls by thin layer chromatography (TLC) and HPLC separation followed by fluorescence or UV detection.⁷⁴⁻⁷⁹ It has also been used to form derivatives of targeted analytes, followed by LC-MS analysis, for the detection of fenfluramine and phentermine,⁸⁰ β -estradiol and estrone,⁸¹ and four phenol-containing metabolites of a drug.⁸²

Through dansylation, a relatively hydrophobic naphthalene moiety is introduced to an analyte, allowing the polar analyte to be eluted during RPLC gradient runs at a much higher percentage of organic mobile phase. With a higher organic solvent, the ionization desolvation efficiency and the electrospray stability are both improved, resulting in a higher ESI response. In addition, the hydrophobic naphthalene moiety tagged to analyte increases the droplet surface affinity of the analyte, thus increasing the surface activity of the analyte. Generally, as long as a molecule is chargeable, the analyte containing more hydrophobic groups will have a higher electrospray response than that with less hydrophobic groups,⁸³ and this in turn will also lead to a higher ESI response. Finally, introduction of a more easily protonated dimethylamino moiety makes the competition more favorable for the limited amount of charges on

the droplet surfaces. Especially with the naphthalene conjugation structure to stabilize the protonated charge at the tertiary amine moiety, the ESI response is also expected to improve. Besides the improvement on ESI response, dansylation has some other benefits as well. For example, the signal-to-background ratio of the analytes can be improved by shifting m/z out of the low mass region that normally exhibits significant background noise.⁸⁴ It also increases the stability of metabolites for LC-MS analysis. In-source fragmentation is rarely observed, and a peak showing up in the spectrum can be assigned to a metabolite ion instead of a fragment ion with confidence. As a result, with the dansylation derivatization, a much greater number of amine- and phenol-containing metabolites can be detected by RPLC-MS, and a more comprehensive sub-metabolome profile can be obtained.

1.2.3. Quantitative Differential Isotope Labeling

To achieve accurate and precise quantification of a relatively small number of analytes, and meanwhile to overcome ion suppression effect in LC-ESI-MS, using an isotope-labeled internal standard for each targeted analyte is the most ideal way. However, it is not practical for metabolome analysis, as the isotope-labeled internal standards are not always available, affordable, or easy to synthesize. In the cases where the targeted analytes are too many, or the identities of the analytes are not even known, the use of isotope-labeled internal standard is unrealistic.

A differential isotope labeling (DIL) methodology has been developed for the analyses of sub-metabolomes. It uses a chemical reaction to introduce an isotope tag to the analytes in one sample and another mass-difference isotope tag to the same analytes in another comparative sample or the corresponding standards, followed by

mixing the two labeled samples for mass spectrometric analysis. The peak intensity ratios of the isotope labeled analyte pairs provide the basis of quantification of the analytes. Relative quantification can be achieved by labeling one sample with a light tag, and the other comparative sample with a heavy tag. Absolute quantification of the analytes in a sample can be obtained if the other sample is a set of standards with known concentrations. While DIL is widely used for quantitative proteome analysis,⁸⁵⁻⁸⁷ however, only a few reports are on the use of DIL in the field of metabolomics studies.

One early report on using DIL for metabolomics study was the use of isobaric tag for relative and absolute quantitation (iTRAQ) reagent, which was originally used to label peptides for quantitative proteomics, to quantitatively analyze amino acids in urine and blood samples.⁸⁸ Yang et al. introduced an LC-MS based method for analyzing amino acids.⁸⁹ It involved derivatization with an N-hydroxysuccinimide ester of N-alkylnicotinic acid where hydrogen in the alkyl chain was replaced by deuterium to give a differential isotope tag. Fukusaki et al. reported the use of ^{13}C -/ ^{12}C -methylation for relative quantification of flavonoids.⁹⁰ Ji et al. described the use of acetaldehyde- d_4 to quantify the monoamine neurotransmitters in rat brain microdialysates.⁹¹ Shortreed et al. and Abello et al. have developed ^{13}C -/ ^{12}C -based tags for differential isotope labeling of amine-containing metabolites.^{92, 93} Guo et al. from our research group has also reported a series of ^{13}C -/ ^{12}C -based differential isotope labeling methods for the analyses of amine-, phenol- and carboxyl-containing metabolites.^{60, 61, 72} Although LC-MS is the mainstream technique for metabolome analysis, GC-MS has been combined with chemical isotope derivatization as well, for example, for the quantification of amino acids, fatty acids and organic acids done by

Huang et al.⁹⁴ In addition, NMR-based differential isotope labeling has also been another choice for targeted metabolomic profiling studies.⁹⁵

Use of stable isotope labeled (SIL) compounds as internal standards for targeted metabolite quantification by LC-MS always requires the SILs and analytes to have identical behavior in sample preparation and ionization process. However, deuterated-SILs often show different retention properties from the corresponding analytes in RPLC,^{96,97} they are not co-eluted or ionized simultaneously.⁹⁸ Therefore, the SILs and the analytes can still suffer matrix and ion suppression effect differently, leading to different signal responses. By using ¹³C-SILs, such as those developed in our group, the isotopic effect can be mitigated. The other beauty of our ¹³C-/¹²C-isotope labeling approach is that the isotope label has been incorporated into the tag, hence an isotope labeled internal standard for each target metabolite could be generated simply in one single labeling reaction with the differential isotope labeling reagent. As an example, the workflow of our differential ¹³C-/¹²C-isotope dansylation labeling strategy for saliva metabolome analysis is presented in Figure 1.7. This strategy will be discussed in detail in Chapter 3.

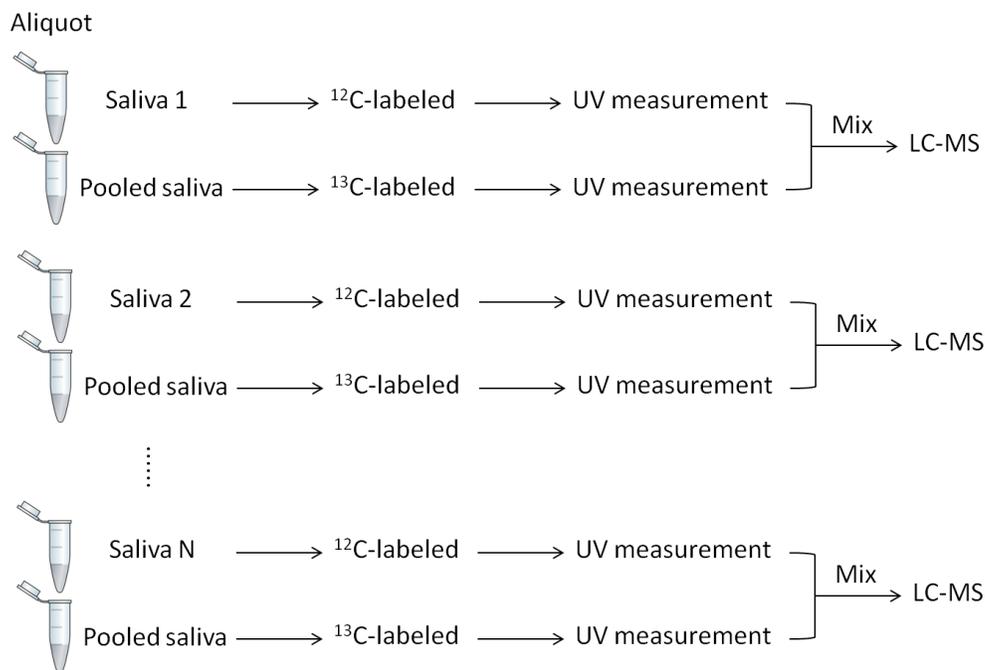


Figure 1.7. Workflow for relative quantification of amine- and phenol-containing metabolites using dansylation labeling LC-ESI FT-ICR-MS.

1.3. Overview of Thesis

This thesis focuses on the development of LC-MS methods for quantitative amine- and phenol-containing metabolome profiling with differential isotope labeling. While rational design of chemical labeling can improve the overall detection of the metabolome, it is also critical to consider another important aspect of the metabolome profiling work, i.e., metabolite identification. One important question that needs to be addressed is: can the labeled metabolites produce useful fragment ions during the tandem MS analysis for structural analysis? ESI-MS/MS fragmentation pathways of 32 amine-containing metabolites and comparison of their fragmentation patterns before and after dansylation labeling were investigated. In most cases, there were a few fragment ions observed from the loss of H_2O , NH_3 , HCOOH from the original

amine molecule, but these ions were usually not informative for deducing or confirming the chemical structures of amines. However, a *pseudo*-MS³ spectrum of the skimmer-fragment ion that had the same *m/z* value as the protonated unlabeled amine could be generated for structural analysis. They normally contained same types of fragment ions as those in the MS/MS spectra of the protonated unlabeled amines, and could be directly compared to the library spectra of unlabeled amines for potential spectral matching and structure elucidation. This work is presented in Chapter 2.

A robust, quantitative metabolome profiling technique was developed using differential ¹³C-/¹²C-isotope dansylation labeling with LC-MS for the analysis of amine- and phenol-containing metabolome. The purpose of differential isotope derivatization was to generate a ¹³C-stable isotope labeled internal standard for each corresponding ¹²C-labeled amine- and phenol-containing analyte. Each pair of two dansylation labeled isoforms were co-eluted and simultaneously detected, hence no isotopic effect was observed, allowing for reliable relative and absolute quantification. An analytical protocol was developed and optimized for the use of minimal saliva sample amount to generate the maximal metabolome information. Application of this protocol for the discovery of disease biomarkers using human saliva metabolome profiling was performed. In this work, our interest was to look for biomarkers for diagnosis or prognosis of mild cognitive impairment (MCI) disease. Relative quantification of amine- and phenol-containing metabolites present in the saliva samples collected from 20 healthy old adults and 20 diseased old adults was performed using the differential isotope dansylation labeling. This quantitative work

is presented in Chapter 3. Finally, conclusions of the thesis work and future directions are presented in Chapter 4.

1.4. Literature Cited

- (1) Thomson, J. J. *Philos. Mag.* **1897**, *44*, 293-316.
- (2) Aston, F.W. *Philos. Mag.* **1920**, *39*, 449-455.
- (3) Aston, F.W. *Philos. Mag.* **1919**, *38*, 707-714.
- (4) Dempster, A. J. *Phys. Rev.* **1918**, *11*, 316-325.
- (5) Fenn, J. B. *Nobel Lecture, Decemeber 8, 2002*.
- (6) Tanaka, K. *Nobel Lecture, Decemeber 8, 2002*.
- (7) Zeleny, J. *Phys. Rev.* **1914**, *3*, 69-91.
- (8) Dole, M.; Mack, L. L.; Hines, R. L. *J. Chem. Phys.* **1968**, *49*, 2240-2249.
- (9) Yamashita, M.; Fenn, J. B. *J. Phys. Chem.* **1984**, *88*, 4451-4459.
- (10) Kebarle, P.; Tang, L. *Anal. Chem.* **1993**, *65*, 972A-986A.
- (11) Iribarne, J. V.; Thomson, B. A. *J. Chem. Phys.* **1976**, *64*, 2287-2294.
- (12) Paul, W.; Steinwedel, H. *Z. Naturfors. Sect. A-J. Phys. Sci.* **1953**, *8A*, 448-450.
- (13) Paul, W. *Angew. Chem.-Int. Edit.* **1990**, *29*, 739-748.
- (14) Paul, W. *Nobel Lecture, Decemeber 8, 1989*, 22.
- (15) Jonscher, K. R.; Yates, J. R. *Anal. Biochem.* **1997**, *244*, 1-15.
- (16) Lawrence, E. Q.; Livingston, M. S. *Phys. Rev.* **1932**, *40*, 19-35.
- (17) Shi, S. D.-H.; Drader, J. J.; Freitas, M. A.; Hendrickson, C. L.; Marshall, A. G. *Int. J. Mass Spectrom.* **2000**, *195-196*, 591-598.

- (18) Sleno, L.; Volmer, D. A.; Marshall, A. G. *J. Am. Soc. Mass. Spectrom.* **2005**, *16*, 183-198.
- (19) Bossio, R. E.; Marshall, A. G. *Anal. Chem.* **2002**, *74*, 1674-1679.
- (20) He, H.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2001**, *73*, 647-650.
- (21) Comisarow, M. B.; Marshall, A. G. *Chem. Phys. Lett.* **1974**, *25*, 282-283.
- (22) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17*, 1-35.
- (23) Comisarow, M. B. *Adv. Mass Spectrom.* **1980**, *8*, 1698-1706.
- (24) Comisarow, M. B. *Int. J. Mass Spectrom. Ion. Phys.* **1981**, *37*, 251-257.
- (25) Castrillo, J. I.; Hayes, A.; Mohammed, S.; Gaskell, S. J., Oliver, S. G. *Phytochemistry.* **2003**, *62*, 929-937.
- (26) Lewis, D. A.; Guzzetta, A. W.; Hancock, W. S. *Anal. Chem.* **1994**, *66*, 585-595.
- (27) Blok-Tip, L.; Zomer, B.; Bakker, F.; Hartog, K. D.; Hamzink, M.; Hove, J.; Vredendregt, M.; De Kaste, D. *Food. Addit. Contam.* **2004**, *21*, 737-748.
- (28) Allen, J.; Davey, H. M.; Broadhurst, D.; Heald, J. K.; Rowland, J. J.; Oliver, S. G.; Kell, D. B. *Nat Biotechnol.* **2003**, *21*, 692-696.
- (29) Scholz, M.; Gatzek, S.; Sterling, A.; Fiehn, O.; Selbig, J. *Bioinformatics.* **2004**, *20*, 2447-2454.
- (30) Kebarle, P.; Verkerk, U. H. *Mass Spectrom. Rev.* **2009**, *28*, 898-917.
- (31) Ikonomou, M. G.; Blades, A. T.; Kebarle, P. *Anal. Chem.* **1990**, *62*, 957-967.
- (32) Kebarle, P.; Tang, L. *Anal. Chem.* **1993**, *65*, 972A-986A.
- (33) Buhrman, D. L.; Price, P. I.; Rudewicz, P. J. *J. Am. Soc. Mass. Spectrom.* **1996**, *7*, 1099-1105.

- (34) Mei, H.; Hsieh, Y.; Nardo, C.; Xu, X.; Wang, S.; Ng, K.; Korfmacher, W. A. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 97-103.
- (35) Annesley, T. M. *Clin Chem* **2003**, *49*, 1041-1044.
- (36) King, R.; Bonfiglio, R.; Fernandez-Metzler, C.; Miller-Stein, C.; Olah, T. J. *Am. Soc. Mass. Spectrom.* **2000**, *11*, 942-950.
- (37) Mallet, C. R.; Lu, Z.; Mazzeo, J. R. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 49-58.
- (38) Antignac, J. P.; de Wasch, K.; Monteau, F.; De Brabander, H.; Andre, F.; Le Bizec, B. *Anal. Chim. Acta.* **2005**, *529*, 129-136.
- (39) Sterner, J. L.; Johnston, M. V.; Nicol, G. R.; Ridge, D. P. *J. Mass Spectrom.* **2000**, *35*, 385-391.
- (40) Fiehn, O. *Plant Mol.Biol.* **2002**, *48*, 155-171.
- (41) Rochfort, S. *J. Nat. Prod.* **2005**, *68*, 1813-1820.
- (42) Li, X.; Yang, S.; Qiu, Y.; Zhao, T.; Chen, T.; Su, M.; Chu, L.; Lv, A.; Liu, P.; Jia, W. *Metabolomics.* **2010**, *6*, 109-118.
- (43) Beger, R. D.; Sun, J.; Schnackenberg, L. K.; *Toxicol Appl Pharm.* **2010**, *243*, 154-166.
- (44) 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*, 2207-2217.
- (45) Dalglish, C. E.; Horning, E. C.; Horning, M. G.; Knox, K. L.; Yarger, K. *Biochem. J.* **1966**, *101*, 792-810.
- (46) Mamer, O. A.; Crawhall, J. C.; Santjoa, S. *Clin. Chim. Acta.* **1971**, *32*, 171-184.
- (47) Horning, E. C.; Horning, M. G. *Clin. Chem.* **1971**, *17*, 802-809.

- (48) Pauling, L.; Robinson, A. B.; Teranishi, R.; Cary, P. *Proc. Nat. Acad. Sci.* **1971**, *68*, 2374-2376.
- (49) Oliver, S. G.; Winson, M. K.; Kell, D. B.; Baganz, F. *Trends Biotechnol.* **1998**, *16*, 373-378.
- (50) Fiehn, O. *Compar. Funct. Genom.* **2001**, *2*, 155-168.
- (51) Wishart, D. S.; Lewis, M. J.; Morrissey, J. A.; Flegel, M. D.; Jeroncic, K.; Xiong, Y. P.; Cheng, D.; Eisner, R.; Gautam, B.; Tzur, D.; Sawhney, S.; Bamforth, F.; Greiner, R.; Li, L. *J. Chromatogr. B* **2008**, *871*, 164-173.
- (52) Astarita, G.; Ahmed, F.; Piomelli, D. *Methods Mol. Biol.* **2009**, *579*, 201–219.
- (53) 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*, 743–752.
- (54) Yanes, O.; Tautenhahn, R.; Patti, G. J.; Siuzdak, G. *Anal. Chem.* **2011**, *83*, 2152–2161.
- (55) Pike, L. S.; Smift, A. L.; Croteau, N. J.; Ferrick, D. A.; Wu, M. *Biochim. Biophys. Acta* **2011**, *1807*, 726–734.
- (56) Buxton, D. B.; Schwaiger, M.; Nguyen, A.; Phelps, M. E.; Schelbert, H. R. *Circ. Res.* **1988**, *63*, 628–634.
- (57) Shaham, O.; Wei, R.; Wang, T. J.; Ricciardi, C.; Lewis, G. D.; Vasan, R. S.; Carr, S. A.; Thadhani, R.; Gerszten, R. E.; Mootha, V. K. *Mol. Syst. Biol.* **2008**, *4*, 214-222.
- (58) Sabatine, M. S.; Liu, E.; Morrow, D. A.; Heller, E.; McCarroll, R.; Wiegand, R.; Berriz, G. F.; Roth, F. P.; Gerzeten, R. E. *Circulation* **2005**, *112*, 3868–3875.

- (59) Wang, T. J.; Larson, M. G.; Vasan, R. S.; Cheng, S.; Rhee, E. P.; McCabe, E.; Lewis, G. D.; Fox, C. S.; Jacques, P. F.; Fernandez, C.; O'Donnell, C. J.; Carr, S. A.; Mootha, V. K.; Florez, J. C.; Souza, A.; Melander, O.; Clish, C. B.; Gerszten, R. E. *Nature Med.* **2011**, *17*, 448–453.
- (60) Guo, K.; Li, L. *Anal. Chem.* **2009**, *81*, 3919-3932.
- (61) Guo, K.; Li, L. *Anal. Chem.* **2010**, *82*, 8789-8793.
- (62) Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J. G.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De Souza, A.; Zuniga, A.; Dawe, M.; Xiong, Y. P.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhtudinov, R.; Li, L.; Vogel, H. J., Forsythe, I. *Nucleic Acids Res* **2009**, *37*, D603-D610.
- (63) Ohta, D.; Kanaya, S.; Suzuki, H. *Curr. Opin. Biotechnol.* **2010**, *21*, 35-44.
- (64) Han, J.; Danell, R. M.; Patel, J. R.; Gumerov, D. R.; Scarlett, C. O.; Speir, J. P.; Parker, C. E.; Rusyn, I.; Zeisel, S.; Borchers, C. H. *Metabolomics.* **2008**, *4*, 128-140.
- (65) Brown, S. C.; Kruppa, G.; Dasseux, J. L. *Mass Spectrom. Rev.* **2005**, *24*, 223-231.
- (66) Casas, J. P.; Bactivista, L. E.; Smeeth, L.; Sharma, P.; Hingorani, A. D. *Lancet.* **2005**, *365*, 224-232.
- (67) Morris, M.S. *Lancet Neurology.* **2003**, *2*, 425-428.
- (68) Willis, G. L. *Neurosci. Biobehav. R.* **1987**, *11*, 97-105.
- (69) Heby, O. *Differentiation.* **1981**, *19*, 1-20.

- (70) Nitta, T.; Igarashi, K.; Yamamoto, N. *Exp. Cell. Res.* **2002**, *276*, 120-128.
- (71) Dettmer, K.; Aronov, P. A.; Hammock, B. D. *Mass Spectrom. Rev.* **2007**, *26*, 51-78.
- (72) Guo, K.; Ji, C.; Li, L. *Anal. Chem.* **2007**, *79*, 8631-8638.
- (73) Kind, T.; Tolstikov, V.; Fiehn, O.; Weiss, R. *Anal. Biochem.* **2007**, *363*, 185-195.
- (74) Seiler, N.; Deckardt, K. *J. Chromatogr.* **1975**, *107*, 227-229.
- (75) Seiler, N.; Knodgen, B.; Eisenbeiss, F. *J. Chromatogr.* **1978**, *145*, 29-39.
- (76) Loukou, Z.; Zotou, A. *J. Chromatogr. A* **2003**, *996*, 103-113.
- (77) Minocha, R.; Long, S. *J. Chromatogr. A* **2004**, *1035*, 63-73.
- (78) Zezza, F.; Kerner, J.; Pascale, M. R.; Giannini, R.; Martelli, E. A. *J. Chromatogr.* **1992**, *593*, 99-101. (79) K52
- (79) Kaddoumi, A.; Nakashima, M.; Wada, M.; Kuroda, N.; Nakahara, Y.; Nakashima, K. *J. Liq. Chromatogr. Relat. Technol.* **2001**, *24*, 57-67.
- (80) Xia, Y.; Chang, S.; Patel, S.; Bakhtiar, R.; Karanam, B.; Evans, D. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 1621-1628.
- (81) Dalvie, D.; O'Donnell, J. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 419-422.
- (82) Cech, N. B.; Enke, C. G. *Mass Spectrom. Rev.* **2001**, *20*, 362-387.
- (83) Keller, B.; Suj, J.; Young, A.; Whittal, R. *Anal. Chim. Acta* **2008**, *627*, 71-81.
- (84) Leitner, A.; Lindner, W. *Proteomics* **2006**, *6*, 5418-5434.
- (85) Ong, S.; Foster, L.; Mann, M. *Methods* **2003**, *29*, 124-130.
- (86) Gygi, S.; Rist, B.; Gerber, S.; Turecek, F.; Gelb, M.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994-999.

- (87) Karlsson, K. E.; Novotny, M. *Anal. Chem.* **1988**, *60*, 1662-1665.
- (88) Yang, W.; Regnier, F.; Sliva, D.; Adamec, J. *J. Chromatogr., B* **2008**, *870*, 233–240.
- (89) Fukusaki, E.; Harada, K.; Bamba, T.; Kobayashi, A. *J. Biosci. Bioeng.* **2005**, *99*, 75–77.
- (90) Ji, C. J.; Li, W. L.; Ren, X. D.; El-Kattan, A. F.; Kozak, R.; Fountain, S.; Lepsy, C. *Anal. Chem.* **2008**, *80*, 9195–9203.
- (91) Shortreed, M. R.; Lamos, S. M.; Frey, B. L.; Phillips, M. F.; Patel, M.; Belshaw, P. J.; Smith, L. M. *Anal. Chem.* **2006**, *78*, 6398–6403.
- (92) Abello, N.; Geurink, P.; van der Toorn, M.; van Oosterhout, A.; Lugtenburg, J.; van der Marel, G.; Kerstjens, H.; Postma, D.; Overkleeft, H.; Bischoff, R. *Anal. Chem.* **2008**, *80*, 9171–9180.
- (93) Huang, X.; Regnier, F. *Anal. Chem.* **2008**, *80*, 107–114.
- (94) Ye, T.; Mo, H.; Shanaiah, N.; Gowda, G. A. N.; Zhang, S.; Raftery, D. *Anal. Chem.* **2009**, *81*, 4882-4888.
- (95) Wade, D. *Chem. Biol. Interact.* **1999**, *117*, 191–217.
- (96) Turowski, M.; Yamakawa, N.; Meller, J.; Kimata, K.; Ikegami, T.; Hosoya, K.; Tanaka, N.; Thornton, E. R. *J. Am. Chem. Soc.* **2003**, *125*, 13836–13849.
- (97) Zhang, R.; Sioma, C. S.; Thompson, R. A.; Xiong, L.; Regnier, F. E. *Anal. Chem.* **2002**, *74*, 3662–3669.

Chapter 2: Fragmentation of Protonated Dansyl-labeled Amines for Structural Analysis of Amine-containing Metabolites

2.1. Introduction

Liquid chromatography (LC) mass spectrometry (MS) has been increasingly used for metabolome profiling as it is highly sensitive and specific.¹⁻³ However, due to the great diversity of physiochemical properties of metabolites, it is difficult to detect and identify all the metabolites with the LC-MS-based metabolome analysis. One approach to meeting this challenge is to classify or fractionate all the metabolites into different groups according to their functional groups, followed by targeted analysis of the individual groups of metabolites using LC-MS.

Our group has been developing a metabolome profiling platform whereby a metabolome sample is selectively labeled with an isotope reagent that reacts with a specific functional group. The labeling reagents are rationally designed to improve the performance of LC separation (i.e., better retention on a high-efficiency reversed-phase column) and electrospray ionization (i.e., better detectability and higher sensitivity in overall detection). We reported the ¹³C₂- and ¹²C₂-dansylation chemistry for profiling amine- and phenol-containing metabolites⁴ and ¹³C₂- and ¹²C₂-p-dimethylaminophenacyl (DmPA) bromide chemistry for profiling carboxylic acid-containing metabolites.⁵ Our work on the use of ¹⁴N₂-/¹⁵N₂-dansylhydrazine for profiling ketones, aldehydes, and sugars has been submitted for publication. These

A form of this chapter was published as: Jiamin Zheng and Liang Li, "Fragmentation of Protonated Dansyl-labeled Amines for Structural Analysis of Amine-containing Metabolites", *Int. J. Mass Spectrom.*, **2012**, 316-318, 292-299.

three isotope labeling chemistries offer a convenient and quantitative route to profiling a large number of metabolites; more than 80% of the 8,000 known human metabolites in Human Metabolome Database or HMDB⁶ contain one or more of the targeted functional groups. As an example, using a two-dimensional LC-MS system, more than 3500 putative amine- and phenol-containing metabolites could be detected from a human urine sample.⁷ The labeled metabolites show significant improvement in detectability over their unlabeled counterparts (i.e., 10 to 1000 fold signal enhancement can be obtained by dansylation).

While rational design of chemical labeling can improve the overall detection of the metabolome, it is also critical to consider another important aspect of the metabolome profiling work, i.e., metabolite identification. With chemical labeling, particularly using a large molecular tag to affect the physiochemical properties of unlabeled metabolites so as to improve their chromatography retention, enhance ionization efficiency, and reduce low-mass background interference in ESI, one important question that needs to be addressed is: can the labeled metabolites produce useful fragment ions during the tandem MS analysis for structural analysis?

In this work, we report a study of ESI-MS/MS fragmentation pathways of 32 amine-containing metabolites and compare their fragmentation patterns before and after dansylation labeling. We illustrate that MS/MS analysis of the fragment ions produced in the skimmer region from the protonated dansyl amine with the m/z value corresponding to the protonated unlabeled amine can generate similar fragmentation patterns to those of the unlabeled metabolites, suggesting that structural information can be obtained from tandem MS analysis of the dansylated compounds.

2.2. Experimental

2.2.1. Chemical and Reagents

All chemicals and reagents were purchased from Sigma-Aldrich Canada (Markham, ON, Canada) except those otherwise noted. LC-MS grade water and acetonitrile (ACN) were purchased from Thermo Fisher Scientific (Edmonton, AB, Canada).

2.2.2. Labeling Reaction

The synthesis of ^{13}C -dansyl chloride as the isotope labeling reagent has been described by Guo *et al.*⁴ The dansylation labeling reaction has also been described,⁴ but with some minor changes. Briefly, amine standard compounds were dissolved with ACN/H₂O (50:50) at a concentration of 100 μM . Fifty μL of standard solutions were mixed with 25 μL sodium carbonate/sodium bicarbonate buffer (500 mM, pH 9.4) and 25 μL ACN in reaction vials. ^{12}C -dansyl chloride solution in ACN (18 mg/mL) or ^{13}C -dansyl chloride solution in ACN (18 mg/mL) was then added, and the reaction stood for 1 hour at 60 $^{\circ}\text{C}$. After 60 min, 10 μL NaOH (250 mM) was added to the reaction mixture to consume the excess dansyl chloride and quench the labeling reaction. After additional 10 min incubation at 60 $^{\circ}\text{C}$, 50 μL of formic acid in ACN/H₂O (425 mM) was added to neutralize the solution. Finally, the dansylation labeled solutions were diluted 5 folds with ACN/H₂O (10:90) containing 0.1% formic acid for MS analysis.

2.2.3. Direct Flow Injection-MS/MS

An AB Sciex 2000 QTRAP LC-MS/MS system (AB Sciex, Toronto) was used. The sample solutions were infused directly by a syringe pump at a flow rate of 5 $\mu\text{L}/\text{min}$. The MS instrument was operated under the following conditions: Curtain Gas (CUR) 15 psi, IonSpray Voltage (IS) 4800 V, Temperature (TEM) 250 $^{\circ}\text{C}$, Ion Source Gas 1 (GS1) 20 psi, Ion Source Gas 2 (GS2) 15 psi. The mass range was set at m/z 50-1000. An enhanced MS (EMS) scan was performed first to find the protonated molecular ion, followed by an enhanced product ion (EPI) scan to generate an MS/MS spectrum. EPI scans were also carried out to generate MS^3 spectra of the high intensity fragment ions observed in MS/MS spectrum. Declustering potential (DP) and collision energy (CE) were adjusted for each different scan. To generate the *pseudo-MS*³ spectra from the labeled amines, the fragment ions were first produced in the skimmer region by raising DP to 45 V, followed by selecting the skimmer-fragment ions using the first quadrupole mass analyzer (Q1) that were then subjected to collision-induced dissociation (CID) in Q2. The second generation product ions were analyzed by EPI scan in the quadrupole linear trap (Q3). All the MS, MS/MS, MS^3 and *pseudo-MS*³ spectra were obtained in the positive ion mode.

2.2.4. Fragmentation Pattern Analysis

The MS/MS and MS^3 data from the unlabeled amines were compared to look for common fragmentation patterns including diagnostic neutral losses and common fragment ions. For the dansyl-labeled compounds, *pseudo-MS*³ spectra of the fragment ions at the same m/z value as the protonated original unlabeled amines were generated. Comparison between the *pseudo-MS*³ spectra of the labeled amines and

the corresponding MS/MS spectra obtained from the unlabeled amines was carried out to determine any similarity in fragmentation patterns before and after dansylation labeling.

2.3. Results and Discussion

The main objective of this work is to determine whether any useful fragment ions can be generated from the dansyl-labeled amines that can be used for structural analysis. Aside from deducing the chemical structure of a completely unknown metabolite, generation of characteristic fragment ions from the dansyl-labeled amines can also facilitate metabolite identification via spectral matching where the fragment ion spectrum of an unknown is searched against those of standards in a library. MS/MS spectral libraries of metabolites and other small molecules have become available in publicly accessible websites.^{6, 8, 9} The number of entries is expected to grow as more compounds are being identified. However, these libraries are generally constructed using unlabeled compounds. To utilize these resources for metabolite identification, it is important that similar types of fragment ions can be obtained from the labeled and unlabeled metabolites. In this work, we first examine the fragmentation pathways or patterns of amine-containing compounds in ESI-MS/MS. We then compare the fragmentation patterns of the unlabeled and labeled amines to determine whether dansylation affects the types and numbers of the fragment ions generated in tandem MS.

For unknown metabolite identification, it is useful to generate as many different fragment ions as possible to produce structural information on different moieties of a molecule. Thus, whenever possible, MSⁿ of a molecular ion is often conducted. As

the QTRAP MS instrument has the ability to carry out the MS/MS and MS³ experiments,¹⁰ these fragment ion spectra were generated for the standard amine-containing compounds both before and after dansylation labeling. A total of 32 amine-containing compounds were chosen from the Human Metabolome Database (HMDB),⁶ representing a wide range of chemical diversity, with no particular reason or purpose other than the availability of these standards. While these standards only represent a small set of amine-containing metabolites, the observed fragmentation behaviours described in this work should be representative of most amines, with a caution that some exceptions may likely be encountered for some amines. The fragment ion spectra of the 32 labeled and unlabeled compounds are provided in the appendix.

2.3.1. Neutral Loss of Unlabeled Amines

Molecules containing an amino group or primary amines are protonated in positive ion ESI at the nitrogen atom¹¹ and get cleaved at the C-N bond in CID. As a result, a neutral loss of a nominal mass of 17 Da in the form of NH₃ is observed. Thus, in the MS/MS spectra of primary amines, the molecular ion, [M+H]⁺, for a singly charged species along with a fragment ion, [M+H-17]⁺, are commonly observed. One example is shown in Figure 2.1 panel A for the CID MS/MS spectrum of 3-aminobenzoic acid (HMDB01891). The peak at *m/z* 138.1 is from the protonated precursor ion, while the fragment ion at *m/z* 121.0 is the deaminated ion. Most of the primary amines studied gave neutral loss of 17. However, in the case of serine (see appendix, under HMDB00187), the loss of 18 was detected, instead of 17. In this case, intramolecular hydrogen-bond involving the primary amino group can be formed to

prevent the loss of amino group in the form of NH_3 .^{12, 13} The loss of H_2O is a preferred route.

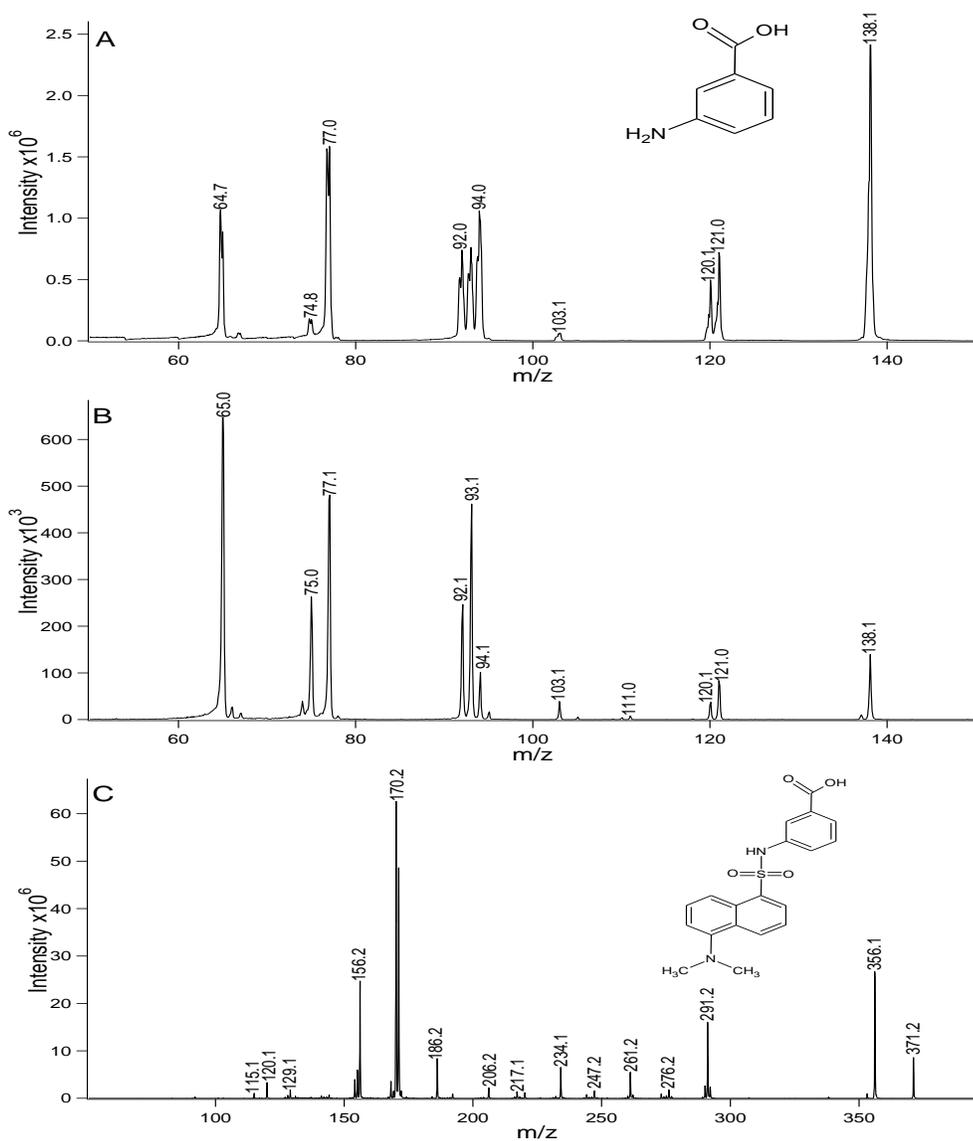


Figure 2.1. (A) CID MS/MS spectrum of the protonated 3-aminobenzoic acid. (B) *pseudo*-MS³ spectrum of the skimmer-fragment ion with m/z 138.1 with enhanced product ion (EPI) scan in the QTRAP. (C) CID MS/MS spectrum of the protonated dansyl-3-aminobenzoic acid. The collision energy used for each spectrum reported in this work is given in the original spectra presented in the appendix.

Neutral loss of 17, followed by neutral loss of 28, is also observed both in MS/MS of the protonated amines or MS³ of the fragment ions from neutral loss of 17 of the protonated amines. This pattern applies mainly to primary amines which also have a carboxy group attached to the same carbon atom, to which the amino group is attached, namely, alpha-amino acid structure. As a primary amine, the charge is retained on the nitrogen atom and cleavage occurs to lose NH₃, resulting in a neutral loss of 17. The resulting deaminated ion, if isolated and fragmented in the MS³ experiment, can further lose the carbonyl group in the original molecule, giving rise a neutral loss of 28 in the MS³ spectrum. Panels A and B in Figure 2.2 show the MS/MS and MS³ spectra of methionine sulfoxide (HMDB02005), respectively. The peak at *m/z* 166.0 is from the protonated parent ion, and the one at *m/z* 149.0 is from the fragment ion after a neutral loss of 17. Furthermore, the deaminated ion with *m/z* 149.0 was isolated for further dissociation. In Figure 2.2 panel B, the peak at *m/z* 121.0 is from the fragment ion generated from a neutral loss of 28 from *m/z* 149.1. In addition, some other neutral losses are also observed in the MS³ spectrum, such as neutral loss of 44 and 46, corresponding to the loss of CO₂ and HCOOH, respectively. Similar results were obtained for selenomethionine (HMDB03966).

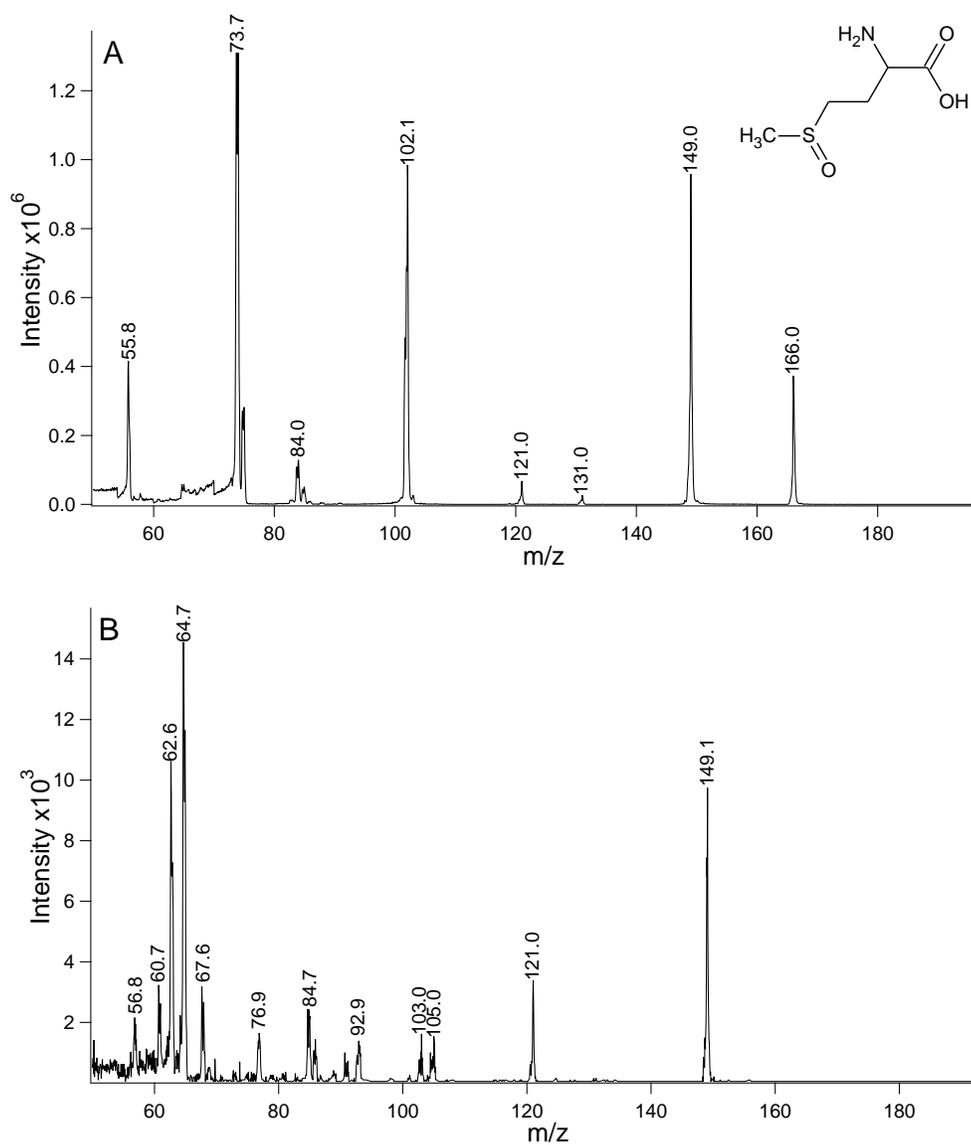


Figure 2.2. (A) MS/MS spectrum of the protonated methionine sulfoxide. Peak with m/z 166.0 represents methionine sulfoxide molecular ion after protonation, and peak with m/z 149.0 shows the fragment ion after neutral loss of 17 from the protonated methionine sulfoxide ion. (B) MS³ spectrum of the deaminated ion of methionine sulfoxide. Peak with m/z 149.1 represents the deaminated ion after neutral loss of 17 from the protonated methionine sulfoxide precursor ion, and peak with m/z 121.0 shows the fragment ion after neutral loss of 28 from the deaminated ion.

2.3.2. Fragment Ions of Unlabeled Amines

The MS/MS spectra shown in panels A and B in Figure 2.1 are typical of the CID spectra that can be obtained from the protonated amines. Notably there are several fragment ion peaks detected from low to high masses. These fragment ions can be readily assigned to the chemical structures of the amines. More importantly, these ions provide the chemical signature that is needed for compound identification based on fragment ion spectral match. There are no characteristic core fragment ions, representative of diverse structures of amines, found from the MS/MS spectra of the 32 compounds. However, a couple of interesting observations are worth noting. One is related to a common fragment ion of m/z 72 observed in the MS/MS spectra of long chain primary amines, such as N-acetylputrescine (HMDB02064) and agmatine (HMDB01432), which has the amino group present at one end of the molecular chain, and connected with four methylene groups ($\text{H}_2\text{N}-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{R}$). In this type of amine-containing compounds, positive charge is initially retained at the nitrogen atom. Upon CID, cleavage occurs at the C-R bond, forming the ion with the structure of $\text{H}_2\text{C}=\text{CH}-\text{CH}_2-\text{CH}_2-\text{NH}_3^+$ at m/z 72.

Another common fragment ion with m/z 102 is observed in the MS/MS spectra of primary amines which have the alpha-amino acid structure present in the original molecule and have at least two methylene groups ($-\text{CH}_2-\text{CH}_2-\text{R}$) connected to the carbon atom, to which both the amino group and the carboxy group are attached. As in the case of molecules that give a fragment ion at m/z 72, in this group of amines, protonation also occurs at the nitrogen atom. Cleavage of the protonated amines occurs at the C-R bond, resulting in the fragment ion which also forms a double-bond between the two methylene groups mentioned above, with its m/z value of 102. This

is exemplified by the MS/MS fragmentation patterns of methionine sulfoxide (HMDB02005), selenomethionine (HMDB03966), and biocytin (HMDB03134).

2.3.3. MS² Fragmentation Patterns of Dansylated Amines

For the dansyl-labeled amines, the fragment ions detected in the MS/MS spectra are mostly from the dansyl part, as the charge after protonation is carried by the nitrogen atom in the dansyl moiety. In most cases, there is no fragment ion containing the part belonging to the original amine molecule without the dansyl moiety. In some cases, fragment ions containing the dansyl group from the neutral loss of a moiety from the original amine molecule are observed, which can provide some structural information on the amine molecule.

As an example, Figure 2.3 panel A shows the MS/MS spectrum of dansyl-threonine (HMDB00167). The major peaks detected can be assigned and the fragmentation scheme of dansyl-threonine is given in Figure 2.4. Most peaks are from the fragment ions containing the dansyl group. Since the charge is located mainly on the dansyl group, no fragment ions corresponding to the original threonine part are observed in the MS/MS spectrum. In this case, the peak at m/z 353.2 is from the protonated dansyl-threonine. At the high m/z region, several fragment ions are detected including the peaks at m/z 338.1, 307.1 and 291.2. The peak at m/z 338.1 is from the loss of the CH₃ group in the dansyl moiety. The peak at m/z 307.1 is generated after losing HCOOH from the original threonine molecule. From the comparison between the MS/MS spectra of the labeled and unlabeled amines, it was also found that the neutral loss fragmentation pathways from the original amine compounds, as described in 2.3.1, can be applied to those of the dansyl-labeled

amines. However, since only neutral loss of simple molecules, such as H₂O and HCOOH, from the amine itself are detected in the MS/MS spectra of dansyl-labeled amines, these spectra alone are not sufficient for compound identification with high confidence.

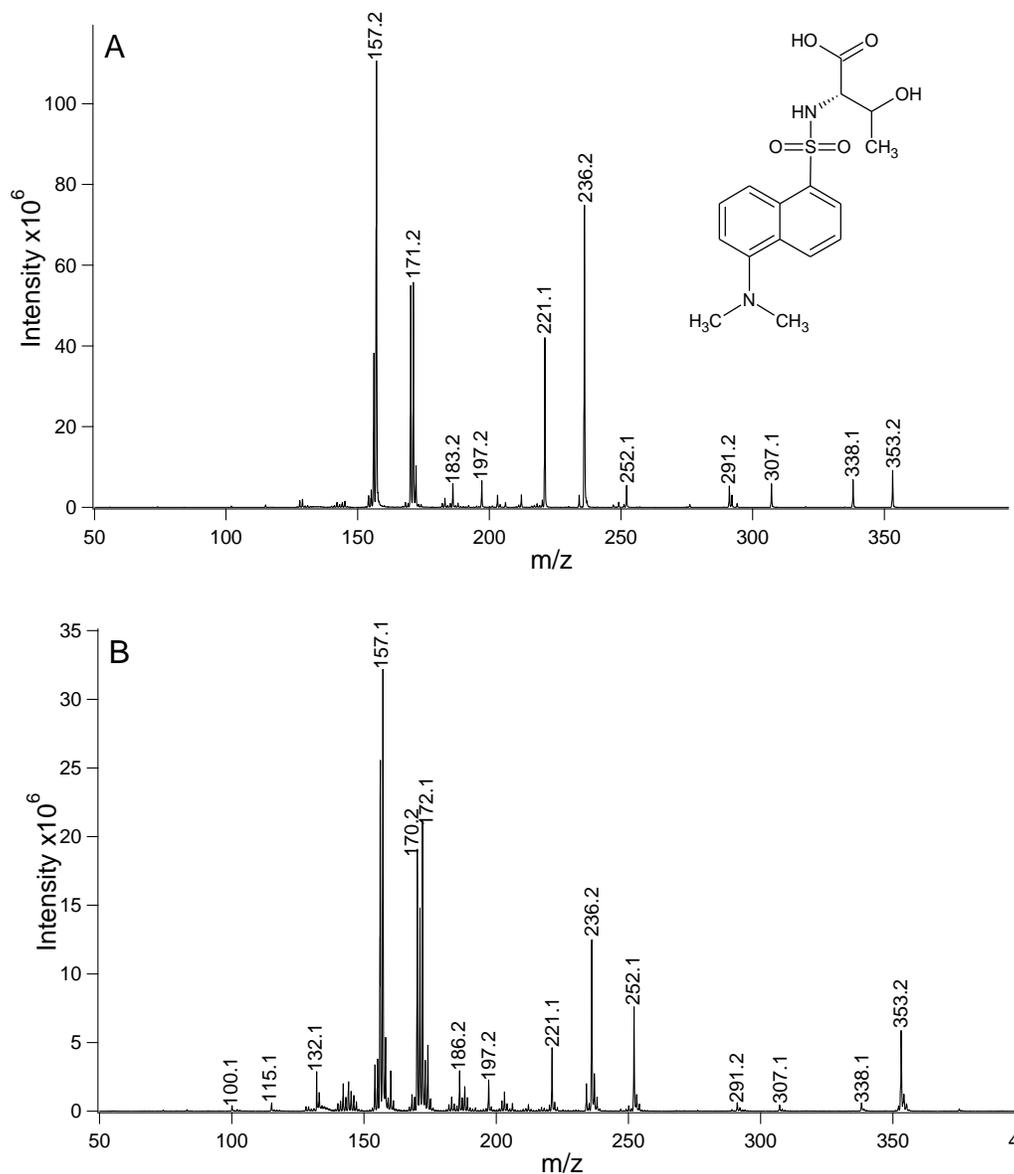


Figure 2.3. (A) MS/MS spectrum of the protonated dansyl-threonine. (B) skimmer-fragmentation spectrum of the protonated dansyl-threonine.

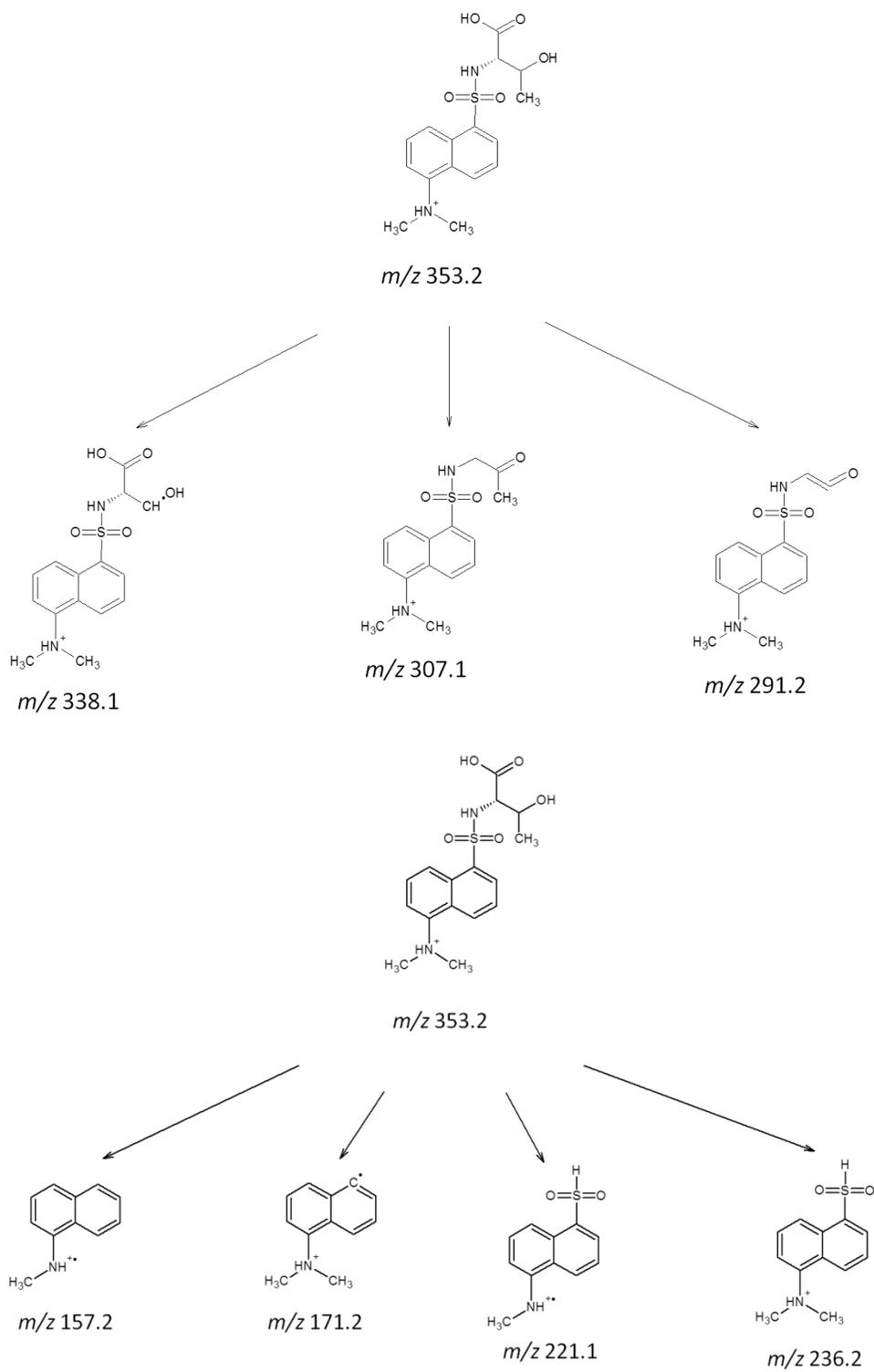


Figure 2.4. Fragmentation scheme of dansyl-threonine.

In the MS/MS spectrum of dansyl-threonine, there is no peak detected at the same m/z value as the protonated unlabeled threonine (m/z 119.06). We attempted to use a more energetic skimmer-fragmentation by applying a high declustering potential in the QTRAP instrument to dissociate the protonated dansyl-threonine, followed by the detection of the fragment ions generated. Figure 3B shows the skimmer-fragmentation spectrum. The fragmentation pattern appears to be similar to the CID spectrum shown in Figure 2.3 panel A. Although more intense low-mass fragment ions are observed, no peaks corresponding to the protonated threonine or its characteristic fragment ions are detected. Similar results are obtained for the other dansyl-labeled amines.

2.3.4. MS³ Fragmentation Patterns of the Fragment Ions from Dansylated Amines

The neutral loss fragment ions from the dansyl-labeled amines can be selected for further fragmentation in the linear trap. For example, in Figure 2.3 panel A, the ions at m/z 338.1, 307.1 and 291.2 generated by the loss of CH₃, HCOOH, and HCOOH+CH₄, respectively, from the protonated dansyl-threonine, were individually selected for fragmentation. However, the MS³ spectra of these species (not shown) do not show any further fragment ions from the original amines. Other high-intensity low-mass ions shown in Figure 2.3 panel A are mainly from the fragmentation of the dansyl group itself and their MS³ spectra do not give useful fragment ion information on the original amine molecules.

2.3.5. *pseudo*-MS³ fragmentation patterns of dansylated amines and comparison to those of unlabeled amines

As was pointed out above, the MS/MS spectra and the skimmer-fragmentation spectra of dansylated amines do not show fragment ions from the original amine molecules. However, for structural analysis or comparison with the fragment ion spectral library of unlabeled amines, it is critical to produce fragment ions directly from the amines. We resorted to the use of *pseudo*-MS³ to produce these fragment ions. Specifically, EPI scans of the second generation of CID product ions of the fragment ions produced from the skimmer-fragmentation of the protonated dansyl amine were conducted to generate the *pseudo*-MS³ spectrum for each skimmer-fragment ion. We found that all the neutral loss skimmer-fragment ions containing the dansyl group did not show any characteristic product ions belonging to the original amine molecule, which is consistent with the MS³ experiment results discussed in 2.3.3. However, by selecting the skimmer-fragment ions with the same m/z value as the unlabeled protonated amine for CID, it was possible to generate a *pseudo*-MS³ spectrum containing peaks corresponding to the characteristic product ions that were found in the CID spectrum of the protonated unlabeled amine. While the skimmer-fragmentation spectrum did not show much signal at m/z of the protonated unlabeled amine, it was observed that, by raising the declustering potential to induce skimmer-fragmentation while selecting this particular ion by the first quadrupole mass analyzer, this skimmer-fragment ion could be detected in EPI. This indicates that skimmer-fragmentation did produce this ion, but its intensity was too low to be detected in the skimmer-fragmentation spectrum. However, by selecting this ion for detection or CID, the linear trap allowed sensitive detection of this ion

and its product ions. We speculate that the enhanced formation of the protonated unlabeled amine in the skimmer region was likely due to the collision of the dansylated metabolite ions with many types of gaseous species in the skimmer interface, such as solvents, salts and other neutral or ionic molecules.

The above observations are not surprising as, in QTRAP, product ion spectra of low-intensity precursor ions can often be generated using selected reaction monitoring (SRM) to trigger EPI scans, even if the precursor ions are not detectable or buried with the background ions using the normal MS scan.^{10, 14} We note that, in a few cases, such as dansyl-adenine, dansyl-5-hydroxy-L-tryptophan and dansyl-cytosine, the CID spectra as well as the skimmer-fragmentation spectra of the dansyl amines do contain a low-intensity peak with m/z corresponding to the protonated unlabeled amine. For the 32 compounds tested in this work, *pseudo-MS*³ spectra of this type of ion can all be generated (see the middle spectrum in each file in appendix). One example of such a spectrum is shown in Figure 1B. Note that, as in Figure 2.1, each file in the appendix for a particular amine contains three spectra. The spectrum at the top (e.g., Figure 2.1 panel A) is the MS/MS spectrum of the protonated unlabeled amine, the one in the middle (e.g., Figure 2.1 panel B) is the *pseudo-MS*³ spectrum of the fragment ion of the labeled amine which has the same m/z value as the protonated unlabeled amine, and the one at the bottom (e.g., Figure 2.1 panel C) is the MS/MS spectrum of the protonated dansyl-labeled amine. By comparing the first and second fragmentation spectra corresponding to the same standard amine, it was found that the fragmentation patterns, in terms of the number and types of fragment ions, before and after dansylation labeling are similar, while for

some of the compounds, additional fragment ions can be detected to provide further structural information on the amines.

One example is shown in Figure 2.1 for 3-aminobenzoic acid. The MS/MS spectrum (Figure 2.1 panel C) of dansyl-3-aminobenzoic acid does not show any characteristic peaks from the fragmentation of the 3-aminobenzoic acid molecule. Based on this MS/MS spectrum alone, it is difficult to confirm the chemical structure of 3-aminobenzoic acid. However, the *pseudo*-MS³ spectrum (Figure 2.1 panel B) of the ion at m/z 138.1 displays the same types of fragment ions as those observed in the MS/MS spectrum of the unlabeled 3-aminobenzoic acid at m/z 138.1 (Figure 2.1 panel A). Thus, in this case, the *pseudo*-MS³ spectrum would provide the needed fragment ion information to identify this compound, if this were an unknown metabolite.

Another example is shown in Figure 2.5 for N-acetylputrescine (HMDB02064). Figure 2.5 panel A shows the MS/MS spectrum of the protonated N-acetylputrescine, in which the peak at m/z 131.2 is from the protonated N-acetylputrescine precursor ion, while the peaks at m/z 114.0 and 71.8 represent the two fragment ions generated. Figure 2.5 panel B shows the *pseudo*-MS³ spectrum of the skimmer-fragment ion generated from the dansyl-N-acetylputrescine ion with the same m/z value as that of the protonated N-acetylputrescine. In this spectrum, the fragment ions with m/z 114.1 and 72.1 are the same ions generated as above, while the peak at m/z 103.1 is from an extra fragment ion generated from the labeled compound (the fragmentation scheme for peak assignment is shown in Figure 2.6). Note that, in this particular case, the same fragment ions at m/z 114.1 and 72.1 are detected in the skimmer-fragmentation spectrum as shown in Figure 4C. However, the *pseudo*-MS³ spectrum generated from

the precursor ion at m/z 131.2 is more reliable in determining or confirming the chemical structure of N-acetylputrescine, particularly when the MS/MS spectrum of the protonated N-acetylputrescine is known.

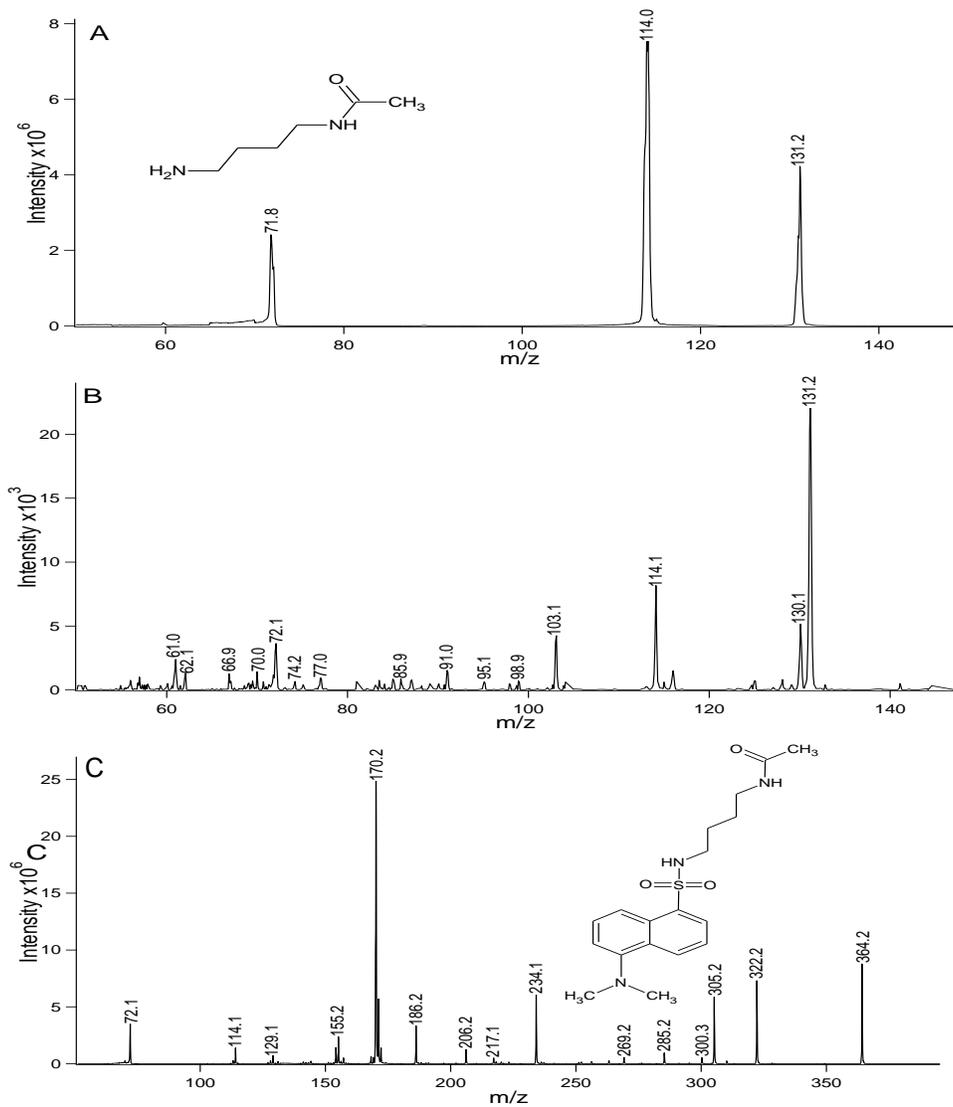


Figure 2.5. (A) MS/MS spectrum of the protonated N-acetylputrescine. (B) pseudo-MS³ spectrum of the skimmer-fragment ion with m/z 131.2 with enhanced product ion (EPI) scan in QTRAP. (C) MS/MS spectrum of the protonated dansyl-N-acetylputrescine.

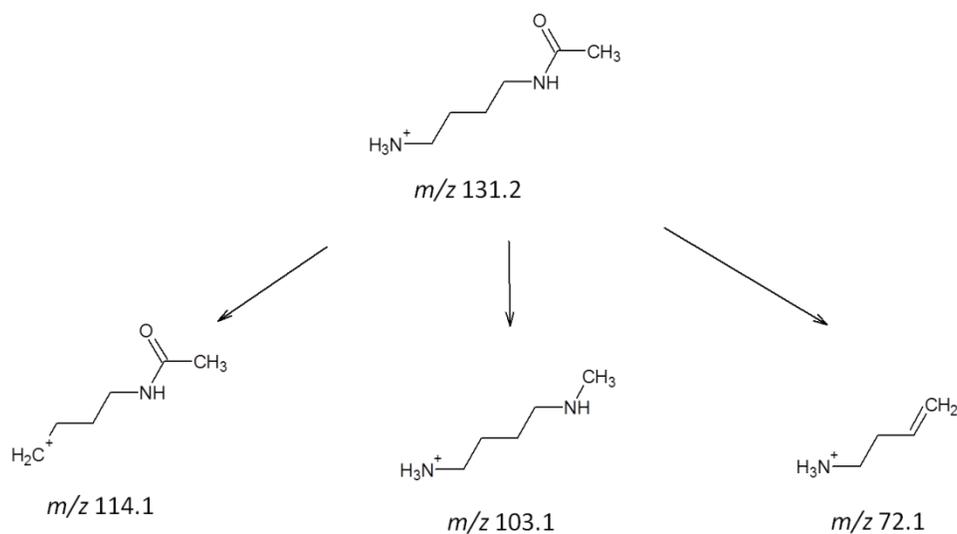


Figure 2.6. Fragmentation scheme of skimmer-fragment ion generated from the dansyl-N-acetylputrescine ion with the same m/z value as that of the protonated N-acetylputrescine.

Among the 32 compounds investigated, all, except one, give similar fragmentation patterns in their *pseudo*-MS³ spectra to those from the MS/MS spectra of unlabeled amines. The exception is for dansyl-*agmatine* (HMDB01432). In this case, the characteristic skimmer-fragment ions with m/z corresponding to the protonated *agmatine* was detected (see the middle spectrum in appendix, under HMDB01432). However, little fragment ions were detected. Fortunately, for this compound, the MS/MS spectrum of dansyl-*agmatine* itself contains the two major fragment ion (m/z 72.1 and 114.1) (see the bottom spectrum in appendix) that can be used for identification of this compound. The reason of this compound did not give a good *pseudo*-MS³ spectrum is unknown. It may be related to the structure of this unique molecule where the amine group may form a stable structure through a hydrogen bridge with another nitrogen in the guanidine moiety. During the

dansyl-arginine dissociation process, this stable structure might be formed while in the protonated form as in MS/MS of the unlabeled amine a more open structure was formed.

2.3.6. Relevance to Metabolome Profiling

Skimmer fragmentation is not widely used for structural analysis due to the lack of precursor ion selection. However, for this work, we use the skimmer fragmentation to generate the precursor ion of the unlabeled metabolite from the dansylated metabolite, followed by selecting the precursor ion using MS1 for further fragmentation to generate the MS/MS spectrum. Thus, the precursor selection of the more useful, unlabeled metabolite ion is still performed. As a result, we generate the MS/MS spectrum of the precursor ion of the unlabeled metabolite, which is more important than the precursor ion of the labeled metabolite.

Our results indicate that a *pseudo*-MS³ spectrum of the skimmer-fragment ions that has the same *m/z* value as the protonated unlabeled amine can be used for structural analysis. The intensity of the protonated unlabeled amine generated is generally low. As a result, the sensitivity of the *pseudo*-MS³ method is not as good as the MS detection of the intact dansylated amine or the MS/MS spectral acquisition from the dansylated amine molecular ion. However, this low sensitivity in producing the *pseudo*-MS³ spectrum should not be a major barrier in the overall workflow of the dansylation LC-MS metabolome profiling method. In the dansylation LC-MS method, quantitative metabolome profiling of many comparative samples (e.g., diseased vs. healthy group) is carried out first based on the MS data only, not the MS/MS data. Statistics analysis of the relative metabolome abundance differences among these

samples is then conducted to determine which metabolite features (each feature with specific retention time and accurate mass) give the most separation of the two groups. Usually only a dozen or so of these features are found. At last, research efforts will be devoted to the identification of these features, i.e., determining the structures of these metabolites. Thus, the proposed *pseudo-MS*³ approach is only used in the last step. To compensate for the low sensitivity of the method, samples can be pooled to increase the analyte concentration for LC-MS injection. Alternatively, the LC fraction containing the important metabolite features is collected from a larger column separation or from multiple injections of a small column separation, followed by *pseudo-MS*³ analysis of the collected fraction.

It should also be noted that there are some significant advantages of using the *pseudo-MS*³ spectra for structural analysis over the use of MS/MS spectra of dansylated compounds. As it was shown earlier, the MS/MS spectra of many dansylated metabolites were not informative about the original structures of the amines (e.g., only loss of H₂O or CO₂ was found). Thus, the usefulness of these spectra for spectral match with the MS/MS spectra of dansylated amine standards for unknown metabolite identification is questionable. In contrast, the *pseudo-MS*³ spectra contain the peaks of fragment ions with the structural signatures of the original amines and, thus, spectral match between the *pseudo-MS*³ spectrum and the MS/MS spectrum of an amine standard is more reliable for compound identification. In addition, the *pseudo-MS*³ spectra can be, in principal, used to facilitate unknown metabolite identification, even if the MS/MS spectrum of the unknown is not present in the spectral library of unlabeled standards. In this case, comparison of fragmentation patterns of different compounds with similar structures (e.g., a core structure with

CH₃O- attached in the standard library vs. the same core structure with CH₃CH₂O- attached in the unknown sample) can be used to reduce the number of metabolite candidates of an unknown metabolite.

2.4. Conclusions

We have investigated the fragmentation behaviors of 32 dansyl-labeled amines with diverse chemical structures and their unlabeled counterparts using an ESI QTRAP mass spectrometer. It was found that the MS/MS spectra of dansyl-labeled amines mainly consist of peaks from the fragment ions containing the dansyl group. In most cases, there were a few fragment ions observed from the loss of H₂O, NH₃, HCOOH from the original amine molecule. These ions are usually not informative for deducing or confirming the chemical structures of amines. In a few cases, fragment ions from the original amine molecule were detected. However, they are not sufficient to identify a compound, if spectral match to a spectral library composed of fragment ion spectra of unlabeled amines is used. However, a *pseudo-MS*³ spectrum of the skimmer-fragment ion that has the same *m/z* value as the protonated unlabeled amine could be generated for structural analysis. In most cases, these spectra contain the same types of fragment ions as those in the MS/MS spectra of the protonated unlabeled amines, while in a few cases additional types of fragment ions are observed. We suggest that the *pseudo-MS*³ spectra generated from the dansyl-labeled amines could be used for metabolite identification in a workflow where the metabolome is labeled with dansylation chemistry for comprehensive profiling. The *pseudo-MS*³ spectra can be directly compared to the library spectra of unlabeled amines for

potential spectral matching. The application of this approach for metabolome analysis will be reported in the future.

2.5. Literature Cited

- (1) Theodoridis, G.; Gika, H. G.; Wilson, I. D. *Mass Spectrom. Rev.* **2011**, *30*, 884-906.
- (2) Dettmer, K.; Aronov, P. A.; Hammock, B. D. *Mass Spectrom. Rev.* **2007**, *26*, 51-78.
- (3) Zehethofer, N.; Pinto, D. M. *Anal. Chim. Acta.* **2008**, *627*, 62-70.
- (4) Guo, K.; Li, L. *Anal. Chem.* **2009**, *81*, 3919-3932.
- (5) Guo, K.; Li, L. *Anal. Chem.* **2010**, *82*, 8789-8793.
- (6) Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J. G.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De Souza, A.; Zuniga, A.; Dawe, M.; Xiong, Y. P.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhutdinov, R.; Li, L.; Vogel, H. J., Forsythe, I. *Nucleic Acids Res* **2009**, *37*, D603-D610.
- (7) Guo, K.; Peng, J.; Zhou, R. K.; Li, L. *J. Chromatogr. A.* **2011**, *1218*, 3689-3694.
- (8) 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*, 747-751.

- (9) 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*, 703-714.
- (10) Hager, J. W.; Le Blanc, J. C. Y. *J. Chromatogr. A.* **2003**, *1020*, 3-9.
- (11) Weissberg, A.; Dagan, S. *Int. J. Mass Spectrom.* **2011**, *299*, 158-168.
- (12) Dookeran, N. N.; Yalcin, T.; Harrison, A. G. *J. Mass Spectrom.* **1996**, *31*, 500-508.
- (13) Zhao, J.; Shoeib, T.; Siu, K. W. M.; Hopkinson, A. C. *Int. J. Mass Spectrom.* **2006**, *255-256*, 265-278.
- (14) Lewis-Stanislaus, A. E.; Li, L. *J. Am. Soc. Mass Spectrom.* **2011**, *21*, 2105-2116.

Chapter 3: Development and Application of Isotope Labeling

LC-MS for Human Salivary Metabolomics

3.1. Introduction

Metabolomics uses a nontargeted analytical approach to study the identities, relative and absolute quantities of a large number of small molecules in a specific biological sample, e.g., blood, urine, saliva, tissue, cell, etc. These small molecules, termed as metabolites, include endogenous metabolites, small peptides, dietary components and others. The composition of metabolites can be altered by environmental factors, such as changes in diet, consumption of drugs, and occurrence of diseases. Thus, metabolomics study currently has been increasingly used in the fields of disease diagnosis,¹ pharmaceutical development,² and biomarker discovery.³ With the advances in separation and detection techniques, along with the sophisticated statistical analysis tools, it is now possible to detect a large number of metabolites in a biological sample. Liquid chromatography (LC) mass spectrometry (MS) has been used more often recently for metabolome profiling work⁴⁻⁶ due to its high sensitivity and specificity. However, it is still difficult to detect and identify all the metabolites with the LC-MS-based metabolome analysis because of the diversity of physiochemical properties of metabolites. One approach for improving the technique is to classify or fractionate all the metabolites into different groups according to their functional groups, followed by targeted analysis of different

A form of this chapter was prepared for publication as: Jiamin Zheng, Roger Dixon, and Liang Li, "Development and Application of Isotope Labeling LC-MS for Human Salivary Metabolomics". Saliva samples used in this study were obtained from Dr. Dixon's lab.

individual groups of metabolites, with the help of chemical derivatization. The work previously done by our group can be viewed as two examples: $^{13}\text{C}_2$ - and $^{12}\text{C}_2$ -dansylation chemistry for profiling amine- and phenol-containing metabolites⁷ and $^{13}\text{C}_2$ - and $^{12}\text{C}_2$ -p-dimethylaminophenacyl (DmPA) bromide chemistry for profiling carboxylic acid-containing metabolites.⁸

Human saliva, as an oral fluid, is secreted mainly from three salivary glands, namely parotid, submandibular and sublingual glands, and also secreted from hundreds of other minor salivary glands. The chemical constitution of saliva is affected by a variety of pathways,⁹ making it promising to be used for health condition monitoring. It is being increasingly treated as a way to assess disease states, and has been regarded as “mirror of the body” in that most of the compounds found in blood also exist in saliva.¹⁰ As a potential diagnostic biofluid that may contain biomarkers for the detection of some diseases, such as oral cancer,¹¹ pancreatic cancer and breast cancer,¹² etc., saliva metabolomics study is gaining more attention recently due to its simple and non-invasive collection, easy processing, as well as its low costs.

One of the potential applications of saliva metabolomics is in the area of diagnosing dementia, such as Alzheimer’s disease (AD). AD is the most common form of dementia, and most often diagnosed in people of over 65 years old.¹³ There is no cure for this disease currently. However, it will greatly benefit the treatment decisions if we can identify individuals who have high probability to progress to dementia using simple diagnostic and prognostic tools. Among the populations, individuals with mild cognitive impairment (MCI) have the tendency to develop AD with a conversion rate of approximately 10% to 15% per year.¹⁴

In this work, we report a high performance isotope-labeling LC-MS approach for quantitative and comprehensive profiling of salivary metabolome. An analytical protocol was developed to handle saliva samples for metabolome profiling with high sensitivity and metabolite coverage. This method was then applied to study the effect of MCI on metabolome changes in saliva, compared to age- and gender-matched controls.

3.2. Experimental

3.2.1. Chemical and reagents

All chemicals and reagents were purchased from Sigma-Aldrich Canada (Markham, ON, Canada) except those otherwise noted. LC-MS grade water and acetonitrile (ACN) were purchased from Thermo Fisher Scientific (Edmonton, AB, Canada).

3.2. Sample collection and processing

3.2.1. Sample collection:

Saliva samples were collected by Oragene[®]•DNA self-collection kit, which was carried out in Professor Roger Dixon's and his clinical collaborators' laboratories. Ethics approval of this work was obtained from the University of Alberta according to the university's health research policy. In collecting the saliva samples, the subjects were asked not to eat, drink, smoke or chew gum for 30 minutes before giving the saliva samples. For the collection, the subjects were required to spit enough amount of liquid saliva into the tube until reaching the fill line on the collection tube. The lid of the tube was then closed tightly while the tube was held

upright, by firmly pushing the lid until a loud click was heard. Again holding the tube upright, unscrewing the tube from the funnel, followed by closing the tube with a small cap coming within the kit. Finally, the tube was shaken for 5 seconds. The samples were stored at room temperature.

3.2.2. Sample processing

Acetone was cooled to -20 °C in advance. An aliquot of saliva sample was placed in 600- μ L Eppendorf tube, and four times the sample volume of cold acetone was added to the tube. The solution was then vortexed, and incubated at -20 °C overnight. After incubation, the solution was centrifuged at 14000 rpm for 30 min. The resulting supernatant was aliquoted out for dansylation labeling.

3.2.3. Labeling reaction

The synthesis of ^{13}C -dansyl chloride as the isotope labeling reagent has been described by K. Guo *et al.*⁷ The dansylation labeling reaction has also been described, but with some minor changes. Briefly, 50 μ L of processed saliva sample solutions were mixed with 25 μ L sodium carbonate/sodium bicarbonate buffer (500 mM, pH 9.4) and 25 μ L ACN in reaction vials. ^{12}C -dansyl chloride solution in ACN (18 mg/mL) or ^{13}C -dansyl chloride solution in ACN (18 mg/mL) was then added, and the reaction stood for 60 min at 60 °C after which, 10 μ L NaOH (250 mM) was added to the reaction mixture to consume the excess dansyl chloride and quench the labeling reaction. After additional 10 min incubation at 60 °C, 50 μ L of formic acid in ACN/H₂O (425 mM) was added to neutralize the solution.

3.2.4. LC-FTICR-MS

An Agilent 1100 series capillary HPLC system (Agilent, Palo Alto, CA) and an Agilent reversed-phase Eclipse C₁₈ column (2.1 mm × 100 mm, 1.8 μm particle size, 95 Å pore size) were used for online LC-MS. LC solvent A was 0.1% (v/v) formic acid in 5% (v/v) ACN, and solvent B was 0.1% (v/v) formic acid in ACN. The gradient elution profile was as follows: t = 0 min, 20% B; t = 3.50 min, 35% B; t = 18.00 min, 65% B; t = 21.00 min, 95% B; t = 21.50 min, 95% B; t = 23.00 min, 98% B; t = 24.00 min, 98% B; t = 26.50 min, 99% B. The flow rate was 180 μL/min, and the sample injection volume was 2 μL. The flow from HPLC was split 1:3 and a 60 μL/min flow was loaded to the ESI source of a Bruker 9.4 Tesla Apex-Qe Fourier-transform ion-cyclotron resonance (FT-ICR) mass spectrometer (Bruker, Billerica, MA, USA), while the rest of the flow was delivered to waste. All MS spectra were obtained in the positive ion mode.

3.2.5. LC-UV

An ACQUITY UPLC[®] system (Waters Corporation, Milford, MA) including binary solvent manager, sampler manager, and photo diode array (PDA) detector, and a Waters ACQUITY UPLC[™] BEH (Ethylene Bridged Hybrid) C₁₈ column (2.1 mm × 50 mm, 1.7 μm particle size) were used for online LC-UV. LC solvent A was 0.1% (v/v) in 5% (v/v) ACN, and solvent B was 0.1% (v/v) formic acid in ACN. The gradient elution profile was as follows: t = 0 min, 0% B; t = 1.00 min, 0% B; t = 1.01 min, 95% B; t = 2.50 min, 95% B; t = 3.00 min, 0% B; t = 6.00 min, 0% B. The flow rate was 450 μL/min, and the sample injection volume was 2 μL. The detection wavelength was set at 338 nm.

3.2.6. Quantitative data processing and statistical analysis

The data files obtained from LC-FTICR-MS analysis were first converted to NetCDF format by Bruker Compass DataAnalysis software (Bruker Daltonics). XCMS, a public-domain software, was used to analyze the NetCDF files to pick up the ^{12}C -/ ^{13}C -ion pairs of the same metabolites in two comparative samples. An abundance ratio between the ^{12}C - and ^{13}C -labeled ion pair was also calculated and reported.

All the processed files by XCMS were aligned together, by an in-house written Perl program, to find common metabolites across all the samples that were listed in one Excel file. By manually checking the raw mass spectra and calculating the abundance ratio, some missing ratios were filled into the generated Excel file, in the case that the ion pairs were indeed existed in the spectra, but not picked up by XCMS. This modified file was then exported to SIMCA-P+ 12.0 software (Umetrics, Umeå Sweden) for multivariate statistical analysis. An orthogonal partial least squares-discriminant analysis (OPLS-DA) was used to identify the discriminant metabolites. Model performance was indicated as cumulative correlation coefficients for the model ($R^2\text{X}[\text{cum}]$ and $R^2\text{Y}[\text{cum}]$) and predictive performance on the basis of seven-fold cross validation calculations ($Q^2[\text{cum}]$). Finally, a list of top ranked important metabolites contributing the most to build the model was generated from the variable importance of the projection (VIP) plot.

3.2.7. Differentially expressed metabolites identification

Several metabolites were selected according to their ranks of VIP value which may serve as potential biomarkers for differentiating two different groups of

individuals. Accurate mass of each underivatized metabolite was calculated by subtracting the mass of the dansyl group from that of dansylation labeled metabolite. An in-house developed web-based software MyCompoundID was used to search the accurate mass within the human metabolome database (HMDB),¹⁵ using a mass accuracy tolerance of 5 ppm. The potential biomarker was definitively identified if both the retention time and the accurate mass could be matched with those of the authentic standard.

3.3. Results and discussion

3.3.1. ¹²C-/¹³C-labeled saliva sample mixing volume normalization

Many biofluids, such as human plasma, cerebrospinal fluid (CSF), urine and saliva, contain a large amount of metabolites that are highly polar and poorly ESI ionizable, making detection of these metabolites by LC separation and ESI-MS difficult. The use of dansylation derivatization overcomes this problem by changing the hydrophobicity of the original amine- or phenol-containing metabolites, thus altering the chromatographic retention behavior and improving the ESI response of these metabolites.

With the advantage of differential isotope labeling, peaks belonging to the labeled amine- or phenol-containing metabolites can be picked up according to the accurate mass difference of the ion pair. For example, for one dansyl group labeled metabolite pair, the mass difference should be 2.0067 Da, with measurement accuracy of < 2 ppm in mass difference. Using the in-house developed program, redundant peaks such as adduct ions of the same metabolite and the noise and

background peaks are eliminated to retain only one ion pair from a putative metabolite at a given retention time.

In applying this isotope labeling technique for profiling the metabolome of saliva samples, our goal was to use a small volume of starting material to generate a maximum number of peak pairs. Thus, we examined several variables that could influence the analytical performance and developed the methods to optimize or control these variables. To compare the effect of individual variables on the total number of peak pairs detected, we needed to ensure that the same amount of labeled metabolites from differently processed samples was injected for comparison. Thus, the first step in our process of developing an optimal saliva metabolome analysis method was to develop a quantification method to measure the total amount or concentration of labeled amine- or phenol-containing metabolites in a sample. Because the quantity or concentration of salivary metabolites was unknown for human individuals and there is no saliva standard where the total metabolite concentration is known, it was not possible to directly determine the total concentration of salivary metabolites using a standard calibration curve.

Because human saliva contains amino acids and amino acids are expected to be present in relatively larger amounts, compared to other metabolites, an amino acid standard mixture solution was labeled with ^{12}C -dansyl chloride, and used to establish a calibration curve to see whether this curve could be utilized to quantify the labeled salivary metabolites. The amino acid mixture calibration curve is shown in Figure 3.1. Meanwhile, another curve was generated from dilution of a labeled pooled saliva sample and is shown in Figure 3.2.

It clearly shows that the two calibration curves have different slopes, indicating that the labeled amino acid mixture and the labeled saliva sample have different absorption coefficients. Thus, the calibration curve of the labeled amino acid mixture cannot be used to determine the absolute concentration of the labeled salivary metabolites. However, for controlling the amount of labeled metabolites being injected for LC-MS analysis, only the relative concentrations among different samples are needed. Therefore, we resorted to the use of the dilution curve of a labeled saliva sample as the calibration curve for relative quantification of the labeled salivary metabolites among different samples, as described below.

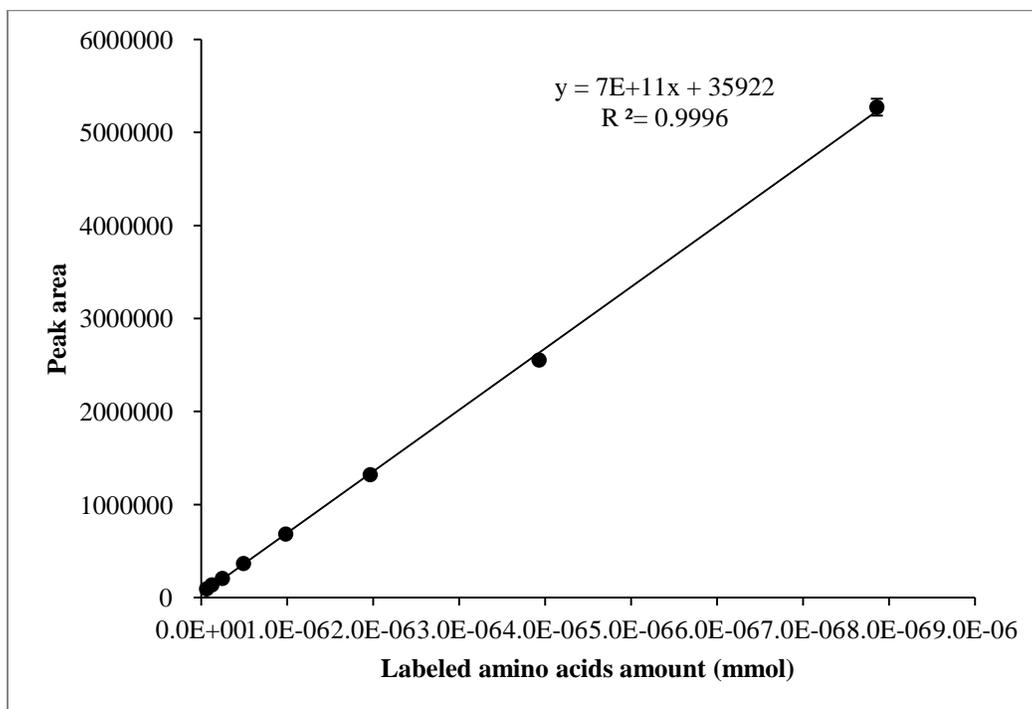


Figure 3.1. Calibration curve built with ^{12}C -dansyl labeled amino acid standard mixture solution.

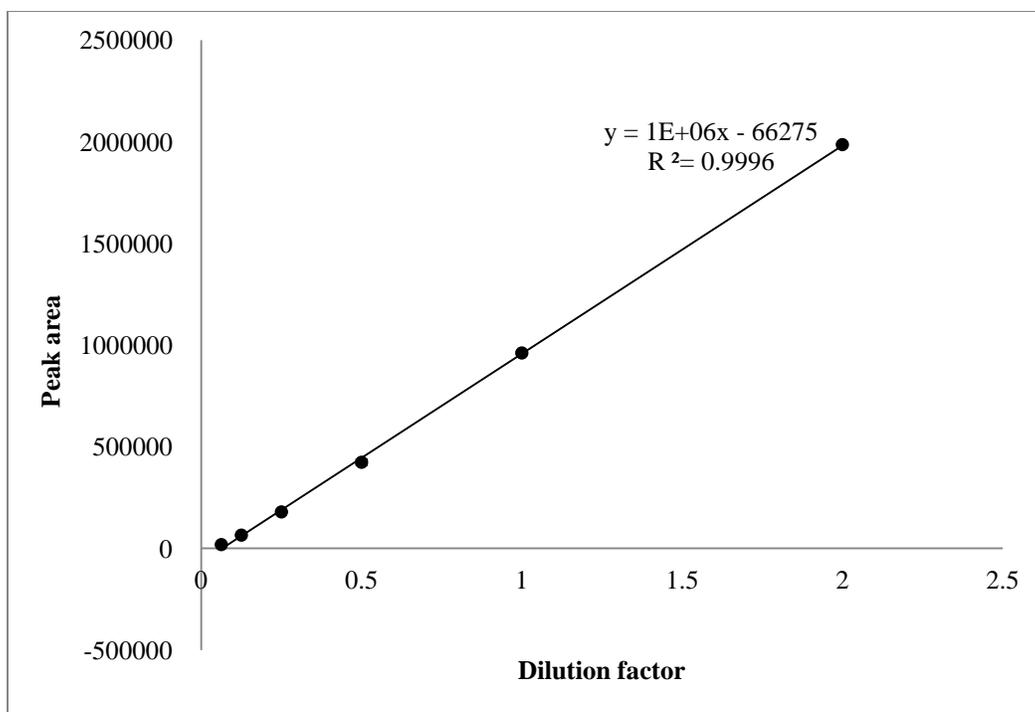


Figure 3.2. Calibration curve built with ^{12}C -dansyl chloride labeled pooled saliva sample.

First of all, five individual saliva samples were labeled with ^{12}C -dansyl chloride and their dilution curves were determined individually. Their curves are similar to that shown in Figure 3.2 with similar slopes, indicating that the absorption coefficient or absorptivity of the labeled individual saliva samples is similar. Statistically, these similarities were evaluated by comparing the slope of each calibration curve generated using individual saliva sample against the one in Figure 3.2, using a modified student t-test at 95% confidence level¹⁷ and the results indicated that there were no statistical difference between each two slopes. Thus, a dilution calibration curve such as the one shown in Figure 3.2 can be used for relative quantification of the total concentration of labeled saliva metabolites in individual samples.

To normalize the total concentration of two comparative samples (e.g., one is ^{12}C -labeled individual sample and another one is ^{13}C -labeled pooled sample), the UV peak area of the individual labeled sample was first measured. After determining the UV peak area of each sample, a “Dilution factor” was calculated according to the saliva calibration curve. This factor gauges the relative amount of labeled metabolites in saliva: the larger the factor the higher concentration of labeled metabolites in a sample. By comparing the dilution factor of each ^{12}C -labeled individual saliva solution with that of another sample (e.g., another ^{12}C -labeled sample or ^{13}C -labeled pooled sample), mixing ratio between the volumes of the two solutions was then determined. The higher the dilution factor, the smaller the volume inversely proportionally to be mixed.

3.3.2. Initial sample size optimization

Now that we knew that the dilution curve could be used to gauge the relative amount of labeled amine- or phenol-containing metabolites in the saliva samples, we could control the amount of labeled-metabolites to be injected for LC-MS analysis. Several different variables that can affect the outcome were studied. The effect of the sample volume used on metabolite detectability was examined first.

To enrich metabolome information that can be obtained from saliva sample, and meanwhile to use a minimal size of the initial saliva sample, different volumes of unprocessed saliva were used to determine the sample volume effects. They were 5 μL , 10 μL , 20 μL , and 50 μL . To meet with the sample volume requirement of dansylation labeling reaction, after removing the proteins from the original saliva with acetone precipitation, the resulting saliva-acetone mixture solutions were dried

down, and each was re-dissolved in 50 μL of water, followed by ^{12}C -dansylation labeling. Another 200 μL of saliva sample was also processed with acetone precipitation, dissolved to final volume of 200 μL using water, and labeled with ^{13}C -dansyl chloride, which served as a standard for comparison. For each ^{12}C -labeled saliva solution and ^{13}C -labeled standard solution, a 30 μL was aliquoted individually, and injected 2 μL for UV measurement. With the measured UV response in terms of peak area, relative quantity of total labeled amine- or phenol-containing metabolites was calculated. As a result, mixing volume ratios were determined to give the amounts of ^{12}C - and ^{13}C -labeled metabolites in each mixture with 1:1 ratio.

Because the optimal injection amount for LC-MS analysis was unknown at this stage, absolute volumes of ^{12}C -labeled solutions to be used were initially determined by the maximal available volume of the most diluted labeled-saliva solution, which was the one prepared with 5 μL saliva as starting material. While the mixing volume of ^{13}C -labeled saliva solution was fixed to be the same for all the samples, volumes of other ^{12}C -labeled solutions were calculated accordingly. In order to inject the same amount of dansylation labeled metabolites for LC-FTICR-MS analysis, all the 1:1 ratio mixtures were dried down, and re-dissolved to 30 μL . Two μL of the mixture was injected for LC-MS analysis. One representative total ion chromatogram and a resulting mass spectrum at one retention time point, from the LC-FTICR-MS analysis of the labeled-saliva sample, were shown in Figure 3.3 and Figure 3.4.

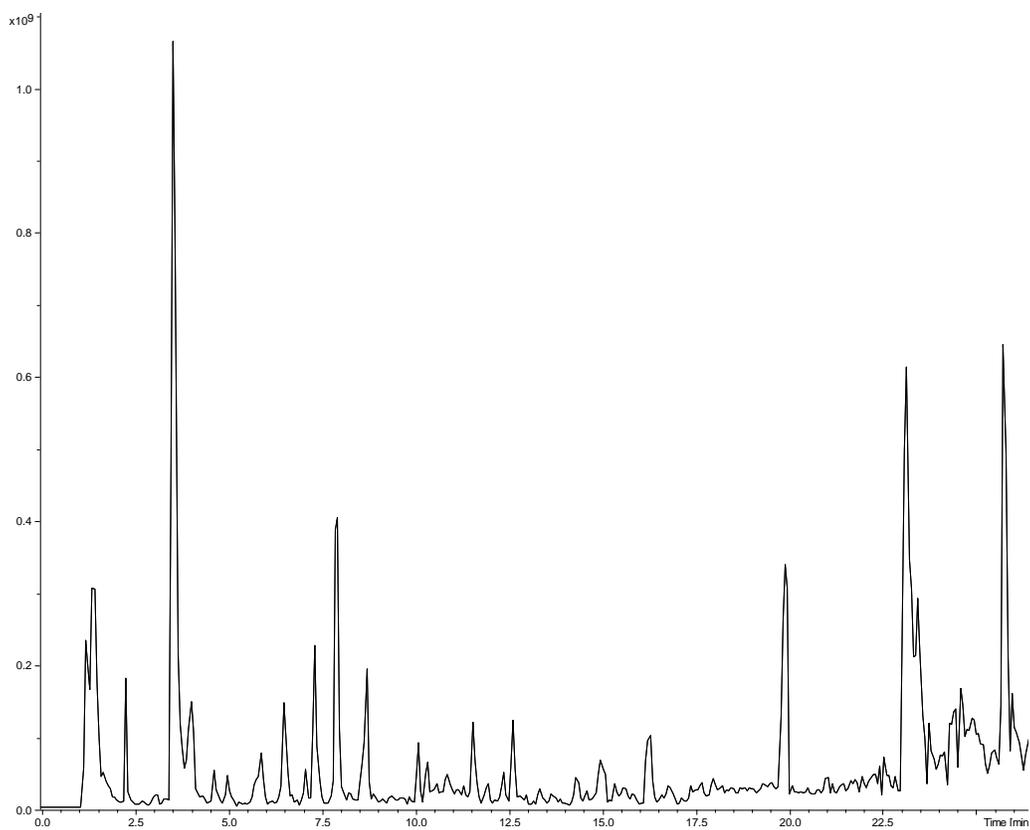


Figure 3.3. A representative total ion chromatogram of dansyl-labeled saliva sample solution from LC-ESI FT-ICR-MS analysis.

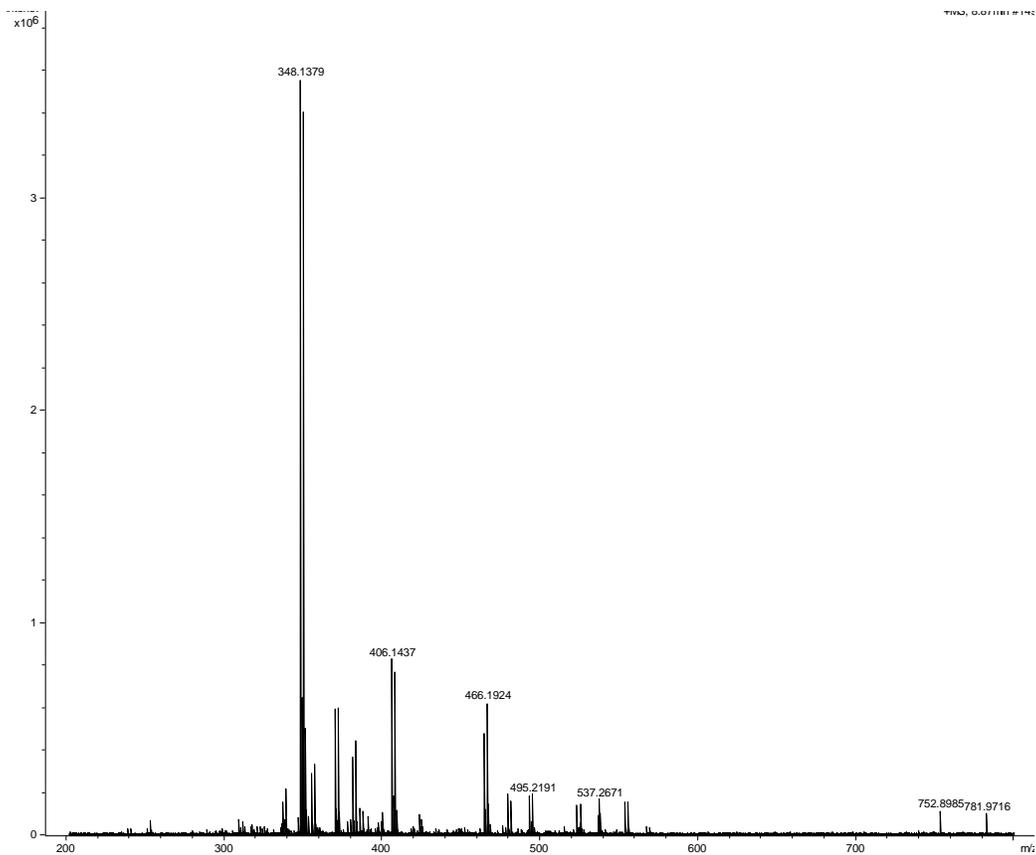


Figure 3.4. A representative mass spectrum at a specific retention time from an entire LC-ESI FT-ICR-MS analysis for dansyl-labeled saliva sample.

Table 3.1. Summary of total number of metabolites detected using four different initial saliva volumes.

Initial saliva volume (μL)	Number of peak pairs					
	5-1	5-2	5-3	Average	SD	RSD (%)
5	213	205	228	223	16	6.97
	231	209	240			
	228	206	248			
Average	224	207	239			
RSD (%)	4.31	1.01	4.22			
10	10-1	10-2	10-3	Average	SD	RSD (%)
	293	321	321	310	20	6.45
	283	296	343			
	303	299	332			
Average	293	305	332			
RSD (%)	3.41	4.47	3.31			
20	20-1	20-2	20-3	Average	SD	RSD (%)
	458	445	440	447	13	2.88
	463	449	420			
	453	456	440			
Average	458	450	433			
RSD (%)	1.09	1.24	2.66			
50	50-1	50-2	50-3	Average	SD	RSD (%)
	483	502	484	505	30	6.03
	488	547	454			
	529	535	524			
Average	500	528	487			
RSD (%)	5.05	4.41	7.21			

Table 3.1 demonstrates the number of peak pairs detected in each sample starting with different initial volumes of saliva. With the same total injection amount of labeled amine- or phenol-containing metabolites, ideally, all of the above four different conditions should give similar results or comparable numbers of peak pairs detected. However, this was not what we observed. To address this issue, four possible causes of the apparent sample volume dependence were investigated:

extraction efficiency during acetone precipitation, sample loss in drying process before dansylation labeling, sample loss in drying process after dansylation labeling while concentrating the solution, and variation of absolute concentration of unlabeled metabolites in the supernatant obtained after acetone precipitation.

First of all, extraction efficiency for different initial saliva volumes was examined. In brief, two aliquots of 20 μL , two aliquots of 10 μL and four aliquots of 5 μL saliva underwent acetone precipitation firstly, and the resulting supernatants were taken out individually after centrifugation. Supernatants from the 20 μL ones were then fractionated into 8 vials, each 25 μL ; those from the 10 μL ones were also fractionated with each 25 μL , into 4 vials; and those from the 5 μL ones were directly transferred to 4 vials, with each 25 μL as well. Another 25 μL of water was added into each of the above aliquots to a total volume of 50 μL for dansylation labeling. UV measurement was carried out after labeling. Table 3.2 displays the concentration of labeled amine- or phenol-containing metabolites in each aliquot in terms of dilution factor. It was clear that all the aliquots gave similar concentration, indicating that the extraction efficiency of metabolite molecules from the original saliva through the acetone precipitation process was almost the same even when different initial volumes of saliva were used.

Table 3.2. Comparison of metabolite extraction efficiency with different initial saliva volumes examined by UV absorbance.

Aliquot	UV peak area	Dilution factor
20-1	462632	0.529
20-2	434492	0.501
20-3	442564	0.509
20-4	418833	0.485
20-5	444456	0.511
20-6	461126	0.527
20-7	433296	0.500
20-8	431449	0.498
10-1	445360	0.511
10-2	459850	0.526
10-3	449206	0.515
10-4	473283	0.539
5-1	416713	0.483
5-2	431406	0.498
5-3	439281	0.506
5-4	416750	0.483
		0.508±0.017

Secondly, the possibility of sample loss during the process of drying down the supernatant after acetone precipitation was studied. According to the volume requirement of an optimized dansylation labeling reaction condition, 50 μL of unlabeled solution was needed for each sample. While dealing with different volumes of initial saliva, especially 20 μL and 50 μL ones, final volumes of the supernatants were always more than 50 μL , leading to the necessity to first dry them down, and re-dissolve to 50 μL for the reaction. The dilution factors of labeled amine- or phenol-containing metabolites in the above mentioned solutions both starting with 5 μL of original saliva was 0.202 on average for the ones undergoing the drying down process before dansylation labeling, and 0.492 for the ones without drying down. This illustrates that there was a possibility of sample loss during the drying down of

supernatants. To further confirm this, LC-FTICR-MS analysis was used to determine the total number of peak pairs that could be detected in aliquot 5-1, 5-2, and 5-3, each with three injections. As a result, 320 ± 8 peak pairs were found as an average of the nine injections in total, which was significantly more than the number of peak pairs detected for those with the drying down, i.e., 223 ± 16 as listed in Table 3.1. Both comparisons of concentrations of labeled amine- or phenol-containing metabolites (i.e., dilution factors) the numbers of detected peak pairs indicate that there was sample loss during the whole process if the supernatants were needed to be dried down before dansylation labeling. To avoid this problem, in all subsequent experiments, supernatants were used directly for labeling after acetone precipitation.

Next, the effect of absolute concentration of unlabeled metabolites on the final number of detected peak pairs was investigated. In our work, it was found that although the number of metabolite peak pairs increased when the 5 μL starting volume of saliva sample was processed without drying down after acetone precipitation, compared to drying down, the number was still not as high as that found in other larger volume samples. Thus, we reasoned that the concentration of unlabeled metabolites in the supernatant from the 5 μL of saliva sample was too low for efficient labeling. To maintain a sufficiently high concentration of metabolites for efficient labeling, we scaled down the labeling reaction, i.e., to reduce the total reaction volume. In this experiment, 25 μL of the supernatant was used directly for dansylation labeling without dilution with water to 50 μL . In other words, the scale of the labeling reaction was halved. UV measurements of the three 5- μL saliva samples labeled using this reduced scale demonstrated that the average dilution factor was 1.010, compared to 0.508 in Table 3.2. One of these three solutions was then mixed

with ^{13}C -dansyl chloride labeled standard prepared from 25- μL saliva, followed by LC-FTICR-MS analysis. The mixture was injected with different volumes to LC-MS: 2 μL , 4 μL , 6 μL , 8 μL and 10 μL . The results are shown in Figure 3.5. The maximum number of peak pairs (966 ± 5) was detected at an injection volume of 4 μL . When the volume of sample injection increased, a decrease in peak pair number was observed, likely due to peak broadening in the chromatographic separation that decreased the detectability of low abundance ion peaks. Nevertheless, the results shown in Figure 3.5 clearly indicate that by keeping the absolute concentration of unlabeled metabolites relatively high a large number of metabolites could still be determined, even when the starting material was about 5 μL of saliva, which is comparable to those obtained from other larger volumes of samples.

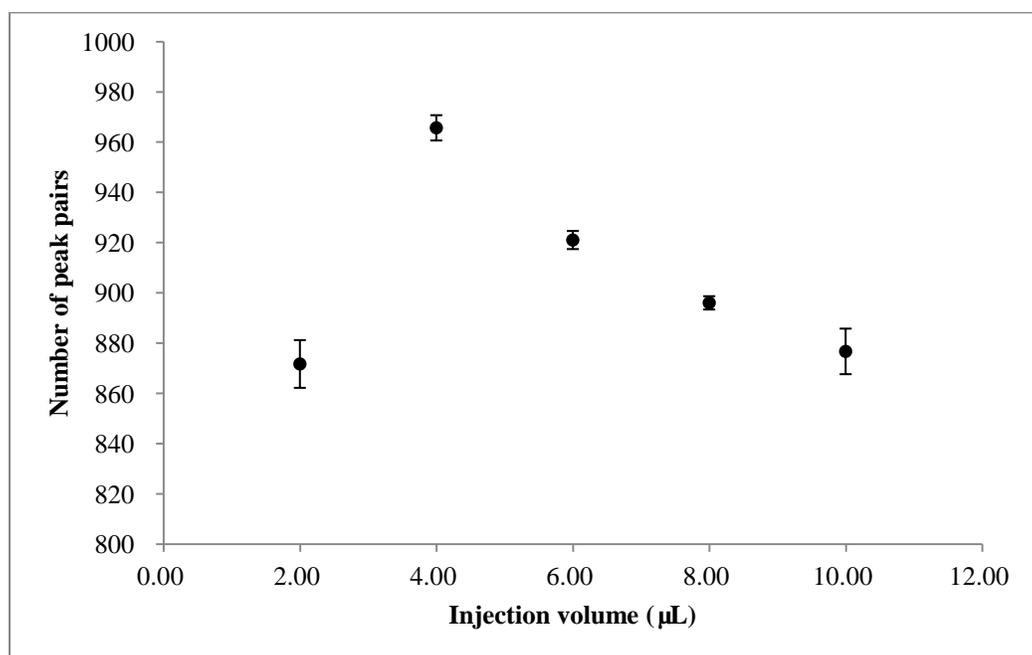


Figure 3.5. Effect of the concentration of unlabeled metabolites on the number of metabolites detected (start with 5 μL saliva and apply dansylation labeling in half-scale).

While the absolute concentration of unlabeled metabolites plays an important role in metabolite detectability, the influence of the concentration of dansyl chloride was also investigated. Instead of 18 mg/mL of dansyl chloride for labeling the saliva sample, a concentration at 36 mg/mL was tested. UV measurement results showed that the average peak area from three replicates was 849639 ± 42407 and 825491 ± 20683 , respectively, for the low and high concentration dansyl labeling. From the comparison, it was clear that the concentration of dansyl chloride used, i.e., 18 mg/mL, was sufficiently high for labeling saliva samples and any increase in dansyl chloride concentration did not improve the product yield.

Finally, the possibility of sample loss in the course of concentrating solutions after dansylation labeling was examined. In this case, three aliquots of 10 μ L saliva were obtained for acetone precipitation, and 50 μ L of supernatant from each aliquot was individually taken out for ^{12}C -dansyl chloride labeling. The average dilution factor of these three solutions was found to be 0.954 ± 0.017 , which was close to those prepared with 5 μ L of saliva and done using half-scale labeling reaction. After spiking in the ^{13}C -labeled standard solution, one of the three replicates was directly injected for LC-FTICR-MS analysis, with an injection volume of 2 μ L, 4 μ L, 6 μ L, 8 μ L or 10 μ L. The results are shown in Figure 3.6, which can be compared with those in Figure 3.5. The two datasets are very similar, further proving that similar results could be obtained with 5 μ L of saliva starting materials, compared to the use of larger volumes of samples.

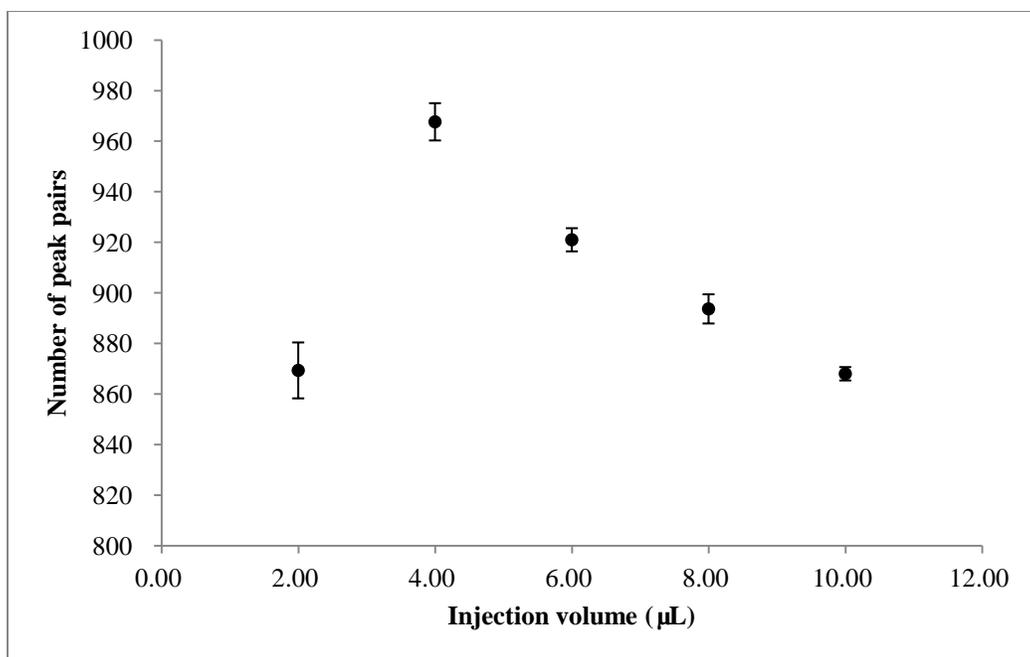


Figure 3.6. Effect of the concentration of unlabeled metabolites on the number of metabolites detected (start with 10 µL saliva and apply dansylation labeling in normal-scale).

The 2nd replicate sample was concentrated to double the concentration by evaporating solvent to reduce the volume by half. The resultant sample was injected in 1 µL, 2 µL, 3 µL, 4 µL, or 5 µL separately for LC-FTICR-MS analysis. The results obtained from the injections of this concentrated solution are shown in Figure 3.7. Interestingly, more ion pairs were detected from 1, 2, 3, 4, or 5 µL of injection of the concentrated sample, compared to 2, 4, 6, 8, or 10 µL of injection of the non-concentrated sample, respectively. This suggests that there was no significant loss of samples during the concentration step. The increase in peak pair number in each volume as well as a different dependence of the number of peak pairs as a function of sample volume indicates that injecting the same amount, but more

concentrated solution, has a benefit of increasing the overall number of peak pairs detected. This effect is likely related to the chromatographic peak shapes in LC-MS. Injecting a more concentrated solution in a lower volume, the chromatographic peaks appeared during LC separation were not as broad as those obtained while injecting a higher volume of a lower concentration solution. Especially, the non-concentrated solution directly after dansylation labeling, which contained 50% of aqueous solvent and 50% of organic solvent, had stronger eluting strength than the initial condition of mobile phase in the LC separation containing 20% of organic solvent. On the other hand, the concentrated solution was mostly aqueous, with little organic solvent remaining. From these results, it can be concluded that concentrating the sample solutions after labeling before LC-FTICR-MS analysis is a preferred approach to increase the number of peak pairs detected.

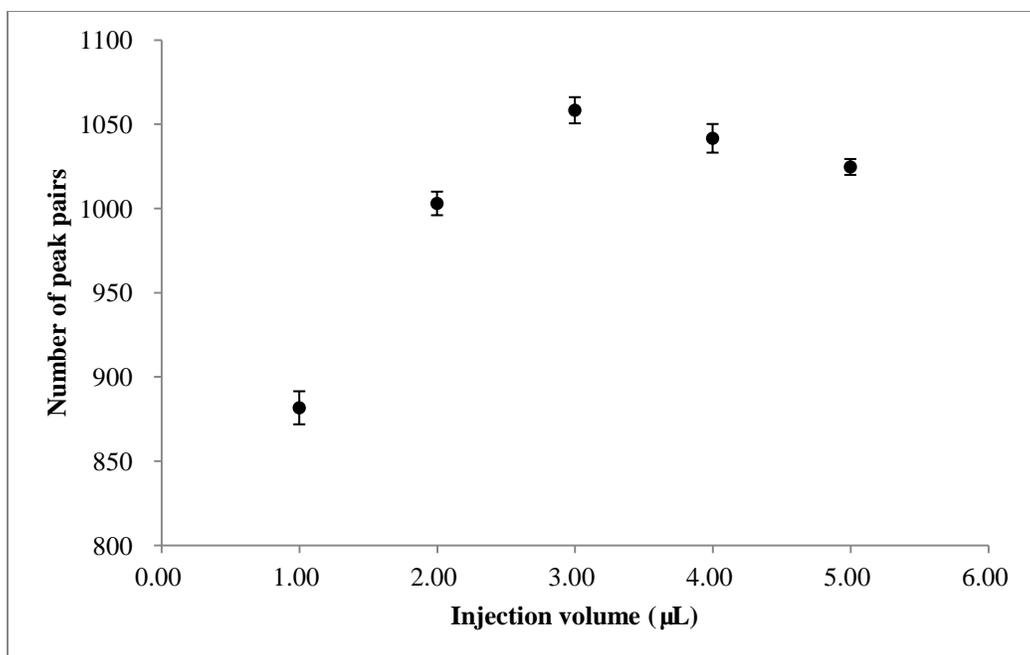


Figure 3.7. Effect of concentration process of the labeled saliva solution on the number of metabolites detected.

In summary, 5 µL of saliva sample could be processed using the protocol described to generate a similar number of peak pairs as those from larger volumes of starting materials. Further reduction of the volume of the starting material is possible. However, overall, 5 µL saliva was found to be convenient to obtain and to work with. There are several advantages working with this volume, including (a) reducing materials (isotope reagents) while working with everything in a smaller scale and (b) avoiding sample loss before dansylation labeling, e.g., if 10 µL or more was used, the drying process before dansylation labeling was required after acetone precipitation to reduce the total volume to 50 µL which was the optimized volume for dansylation labeling. The general protocol of processing 5 µL of saliva involves acetone precipitation to remove proteins. The supernatant after centrifugation is taken out for

direct dansylation labeling. For each individual sample that is ^{12}C -labeled, UV measurement of the labeled sample is carried out to determine the relative concentrations among different samples and the pooled control sample. The volume of a sample needed for mixing with the ^{13}C -labeled standard or pooled sample is normalized so that the same total concentration of the ^{12}C -labeled metabolites in each sample is used for metabolome comparison. The mixture is then concentrated by solvent evaporation in a SpeedVac to a lower volume for LC-MS analysis.

3.3.3 Determination of optimal amount injection

In order to detect as many metabolites as possible, it is important to inject an optimal amount of the labeled analytes for LC-FTICR-MS analysis. On one hand, when the injection amount is not sufficient, the analytes with relatively low intensity will not be detected, thus leading to less peak pairs detected. On the other hand, if the injection amount is too high, detection signal can be saturated leading to mass spectrometric peak broadening and ion suppression, resulting in a reduced number of peak pairs detected. As discussed in the last section, a maximum number of peak pairs was detected at 3 μL injection of 2-fold concentrated solution prepared using the optimized protocol, which was equivalent to 6 μL of non-concentrated solution.

In our work, UV measurement of the labeled samples provides a basis for sample injection optimization, i.e., to ensure an optimal amount of sample is injected for LC-MS analysis. The UV measurement results of the labeled saliva solutions used above for the determination of optimal injection amount were 877099 and 806007 for the ^{12}C -labeled solution and the ^{13}C -labeled standard, respectively. According to the saliva calibration curve, their dilution factors were 0.943 and 0.872, indicating that

the volume mixing ratio between the ^{12}C - and ^{13}C -labeled saliva solutions should be 1:1.081. Thus, to prepare the 6 μL mixture of non-concentrated solution, 2.88 μL was taken from the ^{12}C -labeled solution, and the remaining 3.12 μL was taken from the ^{13}C -labeled one. By multiplying either dilution factor with its corresponding sample volume taken for mixing, an index of 2.72 was obtained. Knowing this index, the required volume of any other labeled saliva sample could be calculated based on the value of the specific dilution factor of the sample derived from the saliva calibration curve.

3.3.4 Method reproducibility

Experiments discussed in Section 3.3.2 were all performed in triplicate, and reproducibility was tested for both the labeling reaction and the LC-FTICR-MS analysis method. For all the experiments, good reproducibility was obtained, with coefficients of variation (CVs) of less than 6% for the sample preparation and labeling reaction by judging the UV measurement of the labeled amine- or phenol-containing metabolites in each individual sample prepared under the same condition, and CVs of less than 7% for the LC-MS results in terms of the total number of detected peak pairs.

3.3.5 Pilot test of metabolome profiling

To examine the performance of the developed method for saliva metabolome profiling, 5 different saliva samples obtained from normal individuals were analyzed in a pilot test. A pooled sample was generated by an aliquot from each individual sample and labeled with ^{13}C -dansyl chloride. Each individual saliva sample was

prepared in triplicate and labeled with ^{12}C -dansyl chloride, and injected three times for LC-FTICR-MS analysis. Quantitative data processing and statistical analysis as discussed in Section 3.2.6 were applied to the obtained LC-MS datasets. The PCA result ($R^2X[\text{cum}]$: 0.840, $Q^2[\text{cum}]$: 0.811) is shown in Figure 3.8 which demonstrates that, the data points from 9 injections representing the same individual saliva sample cluster closely, while the 5 different individual saliva samples tested can be distinguished and separated into 5 clusters. This example illustrates that the developed protocol can be used to handle 5 μL of saliva sample and generate reproducible results from individual saliva samples.

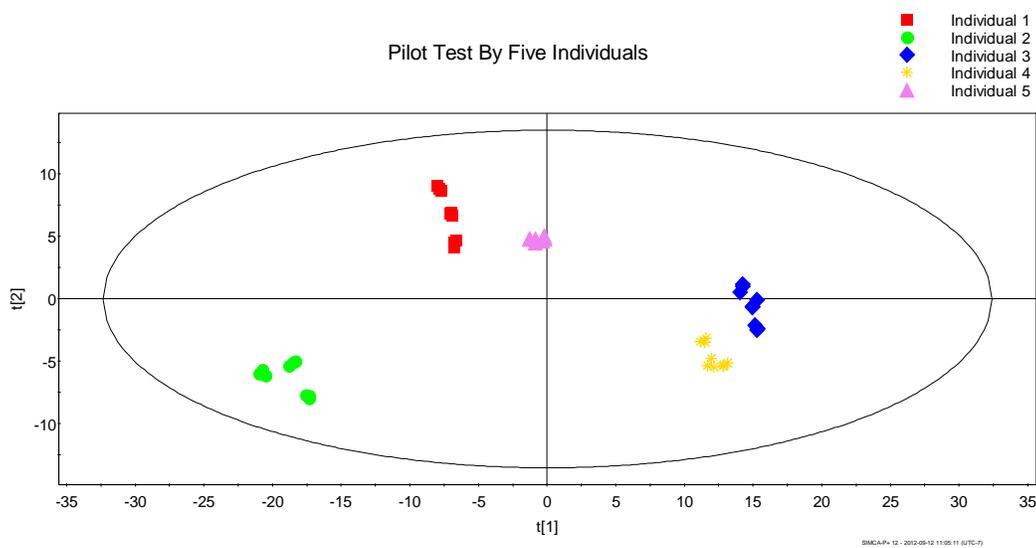


Figure 3.8. PCA analysis result for the pilot test with 5 individual saliva samples. Each colour represents 9 replicates of the same individual sample.

3.3.6 Sample storage effect

As the first application of the developed method for saliva metabolome analysis, we examined the effect of saliva sample storage on the metabolome

profile. Specifically, saliva samples used for genetic testing or other applications are often stored at room temperature prior to analysis. For metabolomics work, because of lower cost for storage and shipping, saliva samples stored at room temperature is preferred over that in a -80 °C freezer. However, during the storage, properties of metabolites may change if the storage conditions are not controlled properly. To address this issue, three freshly collected saliva samples were individually divided into three fractions. One fraction was analyzed immediately after being collected, and the other two fractions were stored at room temperature and in a -80 °C freezer, separately, for 4 weeks.

Figure 3.9 shows the metabolome comparison among all three individual samples, each with three storage conditions described above. The PCA plot (R2X[cum]: 0.837, Q2[cum]: 0.808) shows clear separation between two individual samples (the threshold scores commonly used to define a good separation are 0.500 for both R2X[cum] and Q2[cum]), while the data obtained from all three different storage conditions from the same individual are clustered together. This result indicates that the variations among different individuals are larger than those caused by the sample storage.

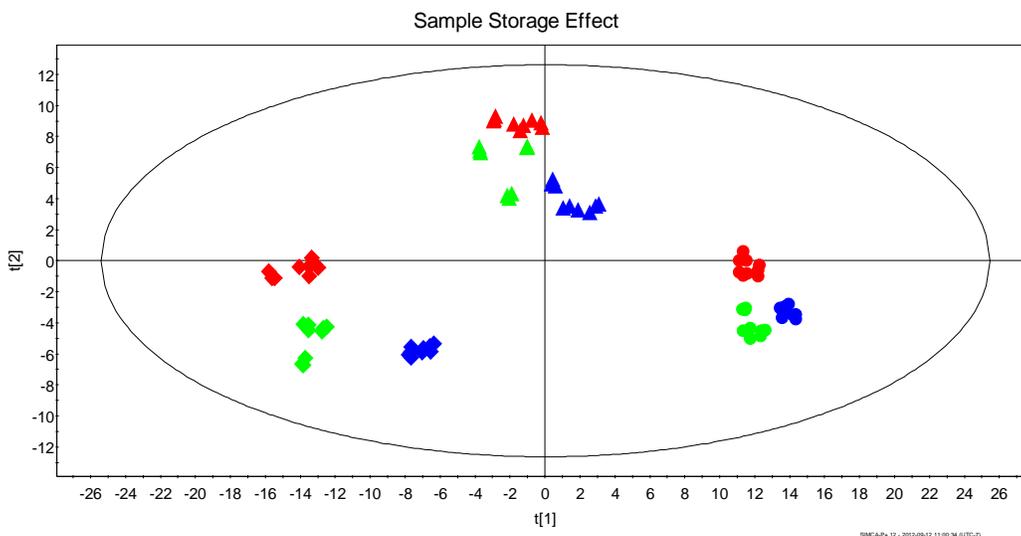


Figure 3.9. PCA analysis result for the storage condition comparison test with 3 individual saliva samples. Each color indicates one specific storage condition, and each shape represents twenty-seven LC-MS injections of one individual sample: nine replicates of freshly collected analysis (red), nine replicates of analysis after four weeks room temperature storage (green), and nine replicates of analysis after four weeks of -80 °C freezer storage (blue).

Tables 3.3 and 3.4 show the results of T-test and calculations of Pearson's correlation coefficient (r), respectively. These statistics analyses further indicate that larger variations were observed among different individuals. T-test results in Table 3.3 show that the p-values obtained from different-individual comparisons are much smaller than those obtained from comparisons of different storage conditions within the same individual. In terms of Pearson's correlation coefficient, the closer the r -value to 1, the more similar the two samples to each other. Calculation results shown in Table 3.4 demonstrate that the r -values of individual-to-individual range from 0.408 to 0.594, while those of different storage condition comparisons of the

same individual are from 0.704 to 0.933. Both tests reach the same conclusion as the visual appearance shown in the PCA plot in Figure 3.9 that room temperature saliva storage does not change the metabolome profile of an individual to a significant extent that affects the comparison of the metabolomes among different individuals. This finding is very significant, as this work suggests that many saliva samples currently collected for other purposes and stored at room temperature could potentially be used for metabolomics studies.

Table 3.3. T-test results (p-values) showing the intra- and inter-individual comparisons, where the intra-comparison was performed on different storage conditions within each individual, and the inter-comparison was performed to compare different individuals.

Individual 1	Fresh	RT	-80 °C
Fresh	\	1.95E-03	3.30E-02
RT	1.95E-03	\	2.59E-08
-80 °C	3.30E-02	2.59E-08	\
Individual 2	Fresh	RT	-80 °C
Fresh	\	2.45E-01	1.27E-17
RT	2.45E-01	\	1.82E-14
-80 °C	1.27E-17	1.82E-14	\
Individual 3	Fresh	RT	-80 °C
Fresh	\	9.99E-09	4.31E-01
RT	9.99E-09	\	1.30E-06
-80 °C	4.31E-01	1.30E-06	\
3 individuals	Individual 1	Individual 2	Individual 3
Individual 1	\	1.40E-99	2.57E-154
Individual 2	1.40E-99	\	0.00E+00
Individual 3	2.57E-154	0.00E+00	\

Table 3.4. Intra- and inter-individual pearson's correlation coefficient (r) calculation results showing the correlations between different storage conditions within each individual and those between different individuals.

Individual 1	Fresh	RT	-80 °C
Fresh	\	0.845	0.814
RT	0.845	\	0.933
-80 °C	0.814	0.933	\
Individual 2	Fresh	RT	-80 °C
Fresh	\	0.874	0.822
RT	0.874	\	0.912
-80 °C	0.822	0.912	\
Individual 3	Fresh	RT	-80 °C
Fresh	\	0.704	0.771
RT	0.704	\	0.849
-80 °C	0.771	0.849	\
3 individuals	Individual 1	Individual 2	Individual 3
Individual 1	\	0.576	0.594
Individual 2	0.576	\	0.408
Individual 3	0.594	0.408	\

3.3.7 Metabolome comparison of healthy and MCI saliva samples

As an example of the potential applications of the developed method for saliva metabolomics, we investigated the metabolome difference between two groups of human subjects, namely 20 individuals with mild cognitive impairment (MCI) and 20 age and gender matched controls. The healthy adults were aged 64-75 and the diseased ones were aged 65-75. To label the individual sample, 5 μL was aliquoted out, went through acetone precipitation to remove the proteins, and then labeled with ^{12}C -dansyl chloride. This was done in duplicate. 5 μL was aliquoted out from each saliva, and pooled to form a control solution which was then subjected to acetone precipitation and labeled with ^{13}C -dansyl chloride. UV measurements of ^{12}C -labeled

individual samples and the ^{13}C -labeled pooled standard were carried out for sample concentration normalization. Each mixture of the ^{12}C -labeled individual sample and ^{13}C -labeled pooled control was analyzed in duplicate by LC-FTICR-MS using the optimized sample injection amount as described earlier. The PCA plot of all the data obtained from the 40 individuals (each individual has 4 replicate points) is shown in Figure 3.10 panel A. The separation of the two groups is not clear in the PCA plot. To determine the separation between the MCI diseased group and the normal healthy group, a supervised statistical analysis method, OPLS-DA, was applied to the metabolome data and the scores plot is shown in Figure 3.10 panel B. This plot displays two clusters separating clearly from each other. Four outliers of data points originated from the same saliva sample were observed and it was found that these data points had a large portion of peak intensity ratio values missed. The reasons of missing ratios are unknown.

To examine the validity of the OPLS-DA model, some statistics criteria were studied. In the model, R^2X and R^2Y represent the fraction of the variance of X matrix and Y matrix, respectively, while Q^2Y indicates the predictive accuracy of the built model upon a seven-fold cross validation conducted by leaving 1/7th samples out in each round. When the cumulative values of R^2X , R^2Y and R^2Q ($R^2X[\text{cum}]$, $R^2Y[\text{cum}]$, and $Q^2Y[\text{cum}]$) are close to 1, it implies an excellent model, while the values above 0.500 were considered to be a validated model. The above three values in our OPLS-DA model were found to be 0.851, 0.958, and 0.920, respectively, and thus the model was valid. An S-plot model was then built to select the significant metabolites that were expressed differently in the diseased group compared to the healthy group.

Table 3.5 lists 18 top important discriminant metabolites with their VIP scores, a measure of their relative influence calculated by T-test, along with their fold change between the diseased group and the healthy group. Both the VIP scores and the p-values obtained from T-test showed a significant difference for each picked discriminant metabolite between the two groups. By judging the fold change for each discriminant metabolite between the two groups, it was obvious that 17 out of the 18 metabolites picked were down-regulated, while one of them was up-regulated. The relative concentration differences of an individual metabolite present in the two groups of samples can be examined by using a box-plot. As examples, the box-plots of one down-regulated and one up-regulated discriminant metabolites out of the eighteen metabolites are shown in Figure 3.11.

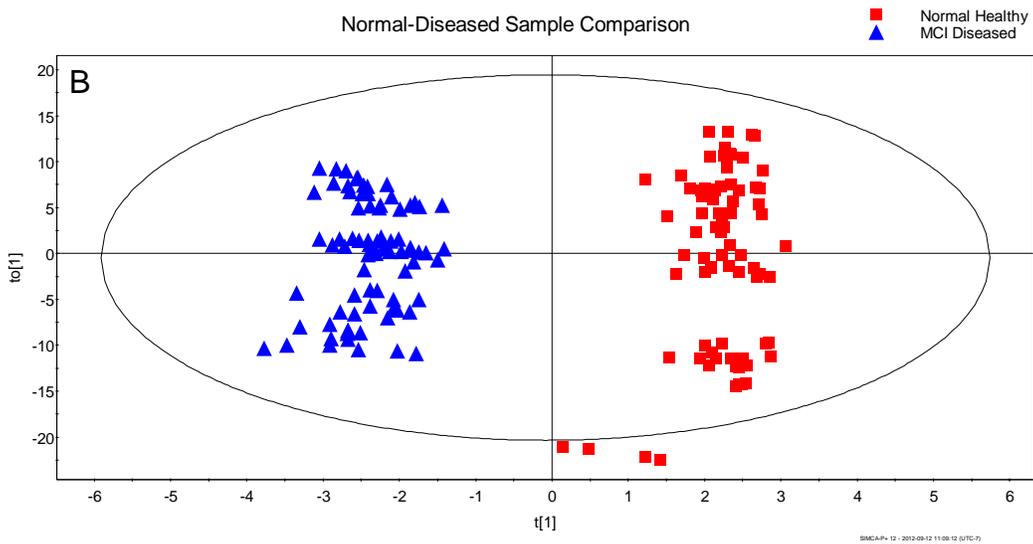
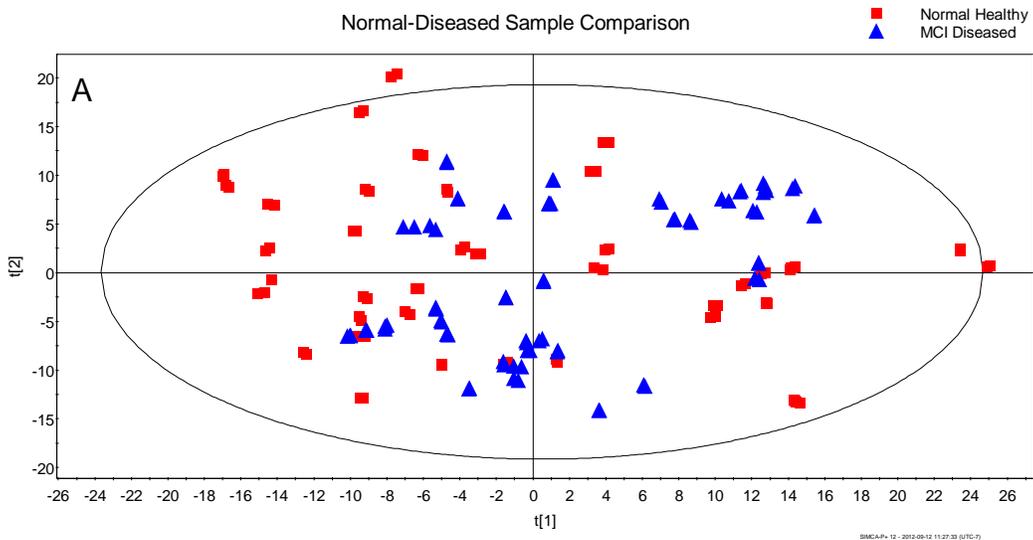


Figure 3.10. (A) The PCA plot of all the data obtained from the 40 individuals. (B) The scores plot of the OPLS-DA model demonstrating the separation between the MCI diseased group and the normal healthy group.

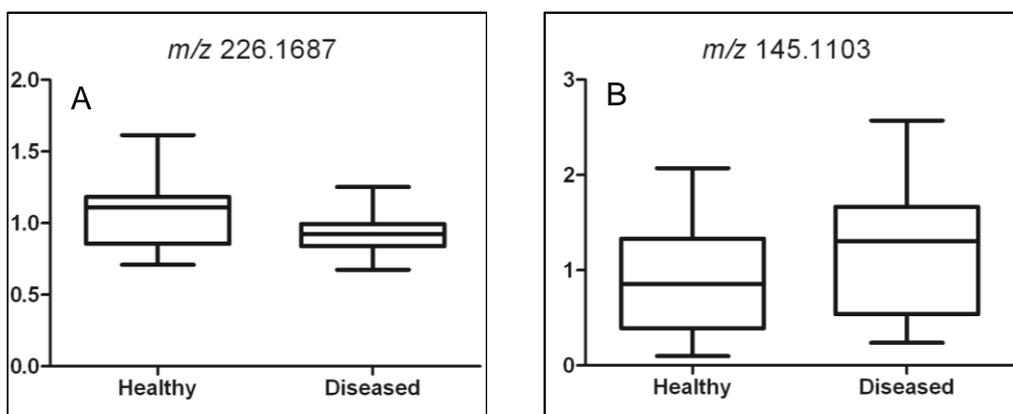


Figure 3.11. Box plots of two discriminant metabolites in differentiating MCI from normal healthy control: (A) a representative of down-regulated discriminant metabolite with *m/z* 226.1687 and (B) an up-regulated discriminant metabolite with *m/z* 145.1104.

Table 3.5. Summary of the discriminant metabolites determined from VIP scores of the OPLS-DA model for variations between MCI diseased group and normal healthy group.

Accurate mass	VIP	p-value	Fold change *	Putative Metabolite Match
226.1687	2.52	1.91E-06	-1.15	1,8-Diazacyclotetradecane-2,9-dione
215.1637	2.33	4.45E-07	-1.30	
250.0954	2.33	2.80E-04	-1.20	
231.1225	2.26	2.22E-04	-1.15	Ala-Ala-Ala, Gly-Gly-Val, Val-Asn, Val-Gly-Gly, Gly-Val-Gly, Asn-Val
262.1323	2.25	2.39E-05	-1.24	Phe-Pro, Pro-Phe
156.0901	2.2	1.70E-03	-1.10	
169.0982	2.19	1.46E-04	-1.13	
213.1118	2.17	3.76E-06	-1.18	
161.1052	2.08	3.24E-04	-1.12	
125.0148	2.06	2.37E-04	-1.31	Taurine
214.0962	2.05	4.98E-05	-1.18	
158.1056	2.04	5.82E-04	-1.16	
192.0750	1.99	4.90E-03	-1.10	
234.1377	1.96	8.04E-04	-1.20	
158.0691	1.94	1.15E-03	-1.13	Ser-Ser
287.1961	1.93	1.26E-03	-1.20	Arg-Leu, Ile-Arg, Leu-Arg, Arg-Ile
145.1104	1.87	2.27E-02	1.26	2-amino-heptanoic acid, L-Alanine-n-butyl ester, N-methyl-isoleucine
152.0950	1.86	1.70E-03	-1.18	4-(Hydroxylamino)-N,N-dimethylaniline

*A positive fold change indicates an up-regulation that has higher metabolite concentration in the MCI group, and a negative fold change represents a down-regulation that has higher metabolite concentration in the healthy group.

3.3.7 Metabolite identification

Among the eighteen selected discriminant metabolites that contributed most to the separation of the MCI diseased group and the normal healthy group based on their

VIP values, we were able to match 7 metabolites (see Table 3.5), based on their accurate masses, to the Human Metabolome Database (HMDB).¹⁵ Moreover, 1 metabolite, taurine, was definitively identified by matching the retention time and accurate mass with its authentic standard under the same experimental condition. Taurine is known to play important roles as an osmoregulator, antioxidant, and neuromodulator.¹⁸ It has also been reported that taurine may prevent the neurotoxicity of beta-amyloid peptide (A β), which is centrally related to the pathogenesis of AD.¹⁹ In accordance with other studies, our results also show a lower taurine expression level in MCI patients. More identification work will be performed in the future upon the availability of standards.

3.4. Conclusions

An isotope labeling LC-MS method has been developed for human salivary metabolome analysis. In this method, 5 μ L of saliva sample could be processed with acetone protein precipitation, dansyl chloride labeling, and then UV measurement of the total concentration of the labeled metabolites. While the absolute concentration of the total metabolites in a saliva sample could not be determined due to the lack of a proper standard for calibration, relative quantification could be performed using a dilution curve of a labeled saliva sample, such as a labeled pooled sample. This method of relative quantification provided a means of normalizing the individual sample concentration by taking varying volumes of samples for labeling and mixing to ensure that the same amount of sample from each individual was used for metabolome comparison. In addition, the UV measurement values could be used to optimize the sample injection amount for LC-FTICR-MS analysis to maximize the

number of metabolites detected. In a differential isotope labeling LC-MS approach where the concentrations of individual metabolites present in ^{12}C -labeled individual samples were compared to those in a ^{13}C -labeled pooled sample, very good reproducibility of both sample processing and LC-MS measurement could be obtained with CVs of less than 7% in terms of total concentration of metabolites and the number of peak pairs detected. In a mixture of ^{12}C -labeled individual sample and ^{13}C -labeled control, the number of peak pairs detected ranged from 1052 to 1067, with an average of 1058. Using this method, the effect of saliva sample storage on metabolome profile changes was investigated and it was found that room temperature sample storage did not cause a significant alteration to the metabolome profile, compared to the use of a freezer for sample storage. Finally, this method was applied for metabolome comparison of two different groups of individuals: normal healthy old adults vs. old adults with MCI disease. Using OPLS-DA, separation between the two groups was clearly observed, leading to the discovery of several discriminant metabolites that contributed most to the separation. Of particular interest, taurine was positively identified as one of the metabolites with lower concentrations in individuals with MCI, compared to the normal old adults. .

Because of the ease of obtaining saliva samples in a non-invasive manner and possibility of sample storage at room temperature, we envisage a wide use of this important biofluid for metabolomics studies, particularly in the field of disease biomarker discovery. The salivary metabolome profiling method described in this work opens the possibility of performing relative quantification of a large number of putative metabolites (up to 1067) using a small volume of starting materials. While this work focused on the use of dansylation chemistry to analyze the amine- and

phenol-containing metabolites in saliva, other labeling chemistries targeted at carboxylic acids,⁸ aldehydes and ketones have been recently developed and application of these labeling chemistries to saliva samples should significantly increase the metabolome coverage.

3.5. Literature Cited

- (1) Li, X.; Yang, S.; Qiu, Y.; Zhao, T.; Chen, T.; Su, M.; Chu, L.; Lv, A.; Liu, P.; Jia, W. *Metabolomics*. **2010**, *6*, 109-118.
- (2) Wishart, D. S. *Drugs R & D*. **2008**, *9*, 307-322.
- (3) Beger, R. D.; Sun, J.; Schnackenberg, L. K. *Toxicol Appl Pharm* **2010**, *243*, 154-166.
- (4) Theodoridis, G.; Gika, H. G.; Wilson, I. D. *Mass Spectrom. Rev.* **2011**, *30*, 884-906.
- (5) Dettmer, K.; Aronov, P. A.; Hammock, B. D. *Mass Spectrom. Rev.* **2007**, *26*, 51-78.
- (6) Zehethofer, N.; Pinto, D. M. *Anal Chim Acta* **2008**, *627*, 62-70.
- (7) Guo, K.; Li, L. *Anal. Chem.* **2009**, *81*, 3919-3932.
- (8) Guo, K.; Li, L. *Anal. Chem.* **2010**, *82*, 8789-8793.
- (9) Kudielka, B. M.; Hellhammer, D. H.; Wust, S. *Psychoneuroendocrino*, **2009**, *34*, 2-18.
- (10) Wong, D. T. *Am Sci.* **2008**, *96*, 37-43.
- (11) 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*, 2207-2217.

- (12) Sugimoto, M.; Wong, D. T.; Hirayama, A.; Soga, T.; Tomita, M. *Metabolomics* **2010**, *6*, 78-95.
- (13) Brookmeyer, R.; Gray, S.; Kawas, C. *Am J Public health* **1998**, *88*, 1337-1342.
- (14) Grundman, M.; Petersen, R. C.; Ferris, S. H.; Thomas, R. G.; Aisen, P. S.; Bennett, D. A.; Foster, N. L.; Jack, C. R.; Galasko, D. R.; Doody, R.; Kaye, J.; Sano, M.; Mohs, R.; Gauthier, S.; Kim, H. T.; Jin, S.; Schultz, A. N.; Schafer, K.; Mulnard, R.; van Dyck, C. H.; Mintzer, J.; Zamrini, E. Y.; Cahn-Weiner, D.; Thal, L. J. *Arch. Neurol.* **2004**, *61*, 59-66.
- (15) Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J. G.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De Souza, A.; Zuniga, A.; Dawe, M.; Xiong, Y. P.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhutdinov, R.; Li, L.; Vogel, H. J., Forsythe, I. *Nucleic Acids Res* **2009**, *37*, D603-D610.
- (16) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779-787.
- (17) Zar, J. H. *Biostatistical Analysis*; 5th ed.; Pearson Prentice Hall Inc., **2010**.
- (18) Huxtable, R. J. *Prog. Neurobiol* **1989**, *32*, 471-533.
- (19) Louzada, P. R.; Lima, A. C.; Mendonca-Silva, D. L.; Noel, F.; De Mello, F. G.; Ferreira, S. T. *Faseb J.* **2004**, *18*, 511-518.

Chapter 4: Conclusions and Future Work

This thesis work focused on the development and application of a differential $^{13}\text{C}/^{12}\text{C}$ -isotope dansylation labeling technique for the quantification and identification of amine- and phenol-containing metabolites in human saliva using LC-ESI FT-ICR-MS. This isotope labeling approach significantly increases the detection sensitivity, and allows for high accuracy and high precision relative quantification of targeted amine- and phenol-containing metabolites in saliva.

In Chapter 2, fragmentation behaviors of 32 dansyl-labeled amines with diverse chemical structures and their unlabeled counterparts were investigated. Some common fragmentation patterns were observed for the unlabeled amine-containing compounds that can be described as follows: (1) a neutral loss of a nominal mass of 17 Da in the form of NH_3 was observed for most of the primary amines; (2) in the cases that formed intramolecular hydrogen-bond involving the primary amino group, which prevented the loss of amino group in the form of NH_3 , a loss of 18 in the form of H_2O was detected instead of 17; (3) in primary amines that also had a carboxyl group attached to the same carbon atom, to which the amino group was attached, neutral loss of 17, followed by neutral loss of 28, was also observed both in MS/MS of the protonated amines or MS^3 of the fragment ions from neutral loss of 17 of the protonated amines. For the dansyl-labeled amines, it was found that their MS/MS spectra mainly consisted of peaks from the fragment ions containing the dansyl group. In most cases, a few fragment ions were observed from the loss of H_2O , NH_3 , HCOOH from the original amine molecule, which were not informative for structural elucidation. In a few cases, fragment ions from the original amine molecule were

detected, which were not enough for the identification of the compounds. However, with skimmer fragmentation, a *pseudo*-MS³ spectrum of the fragment ion that had the same *m/z* value as the protonated unlabeled amine could be generated for structural analysis. In most cases, the same types of fragment ions as those in the MS/MS spectra of the protonated unlabeled amines were generated, while additional types of fragment ions were observed in a few cases.

In Chapter 3, a differential ¹³C-/¹²C-isotope dansylation labeling method, combined with LC-ESI FT-ICR-MS, was developed for relative quantitative profiling of amine- and phenol-containing metabolites in human saliva samples. Optimization of the sample preparation process including dansylation labeling and the LC-MS detection process enabled the use of only 5 µL of saliva sample as a starting material for salivary metabolome analysis. A comparison study was carried out with the developed method on two different groups, the normal healthy old adults group, and the MCI diseased old adults group. Comparative metabolomics using chemometrics tools, such as OPLS-DA, demonstrated obvious separation between the two groups, leading to the discovery of a number of discriminant metabolites influenced most on the separation. By identifying the selected metabolites, potential biomarkers were tentatively found to distinguish the normal old adult and the MCI diseased old adult, among which seven metabolites were putatively identified while one metabolite, taurine, was definitively identified.

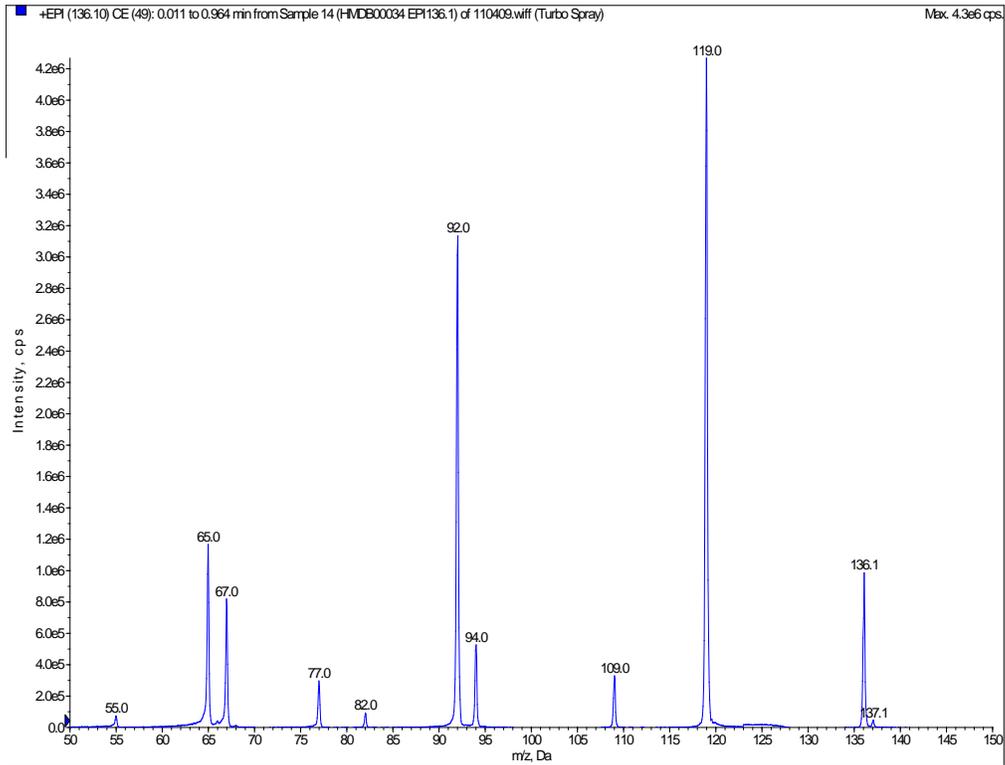
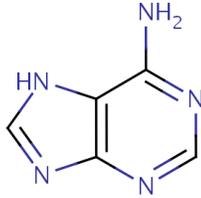
The developed differential ¹³C-/¹²C-isotope dansylation labeling strategy with analysis by LC-ESI FT-ICR-MS was found to be robust for both qualitative and quantitative amine- and phenol-containing metabolome profiling. In the future, combination of different labeling techniques to target different groups of metabolites

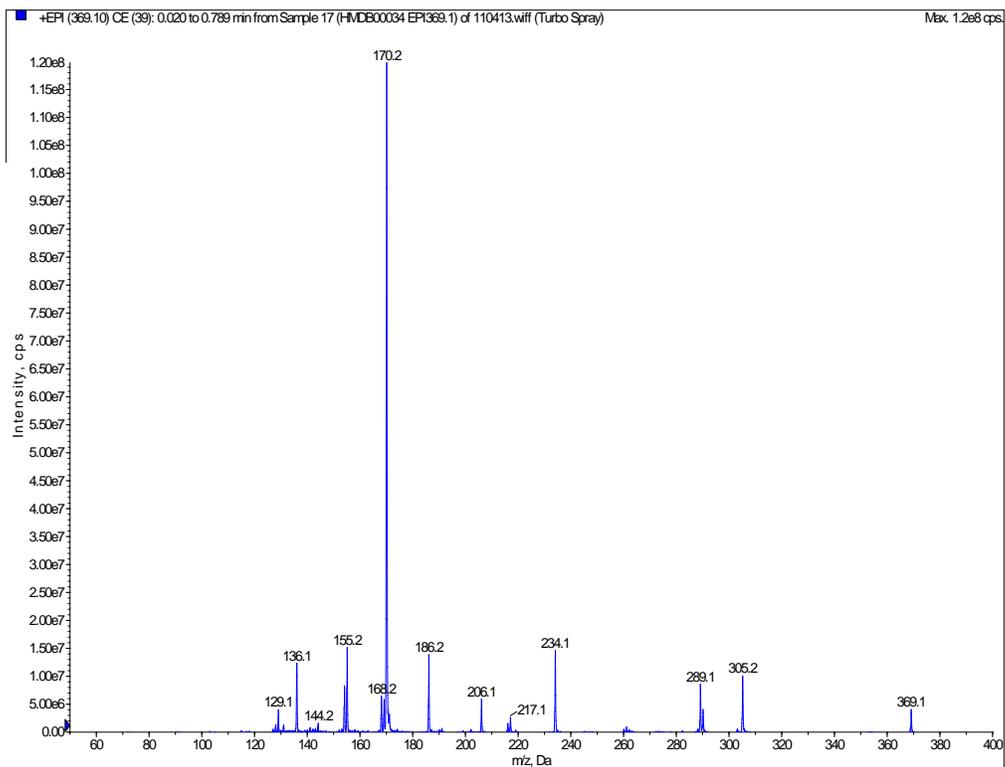
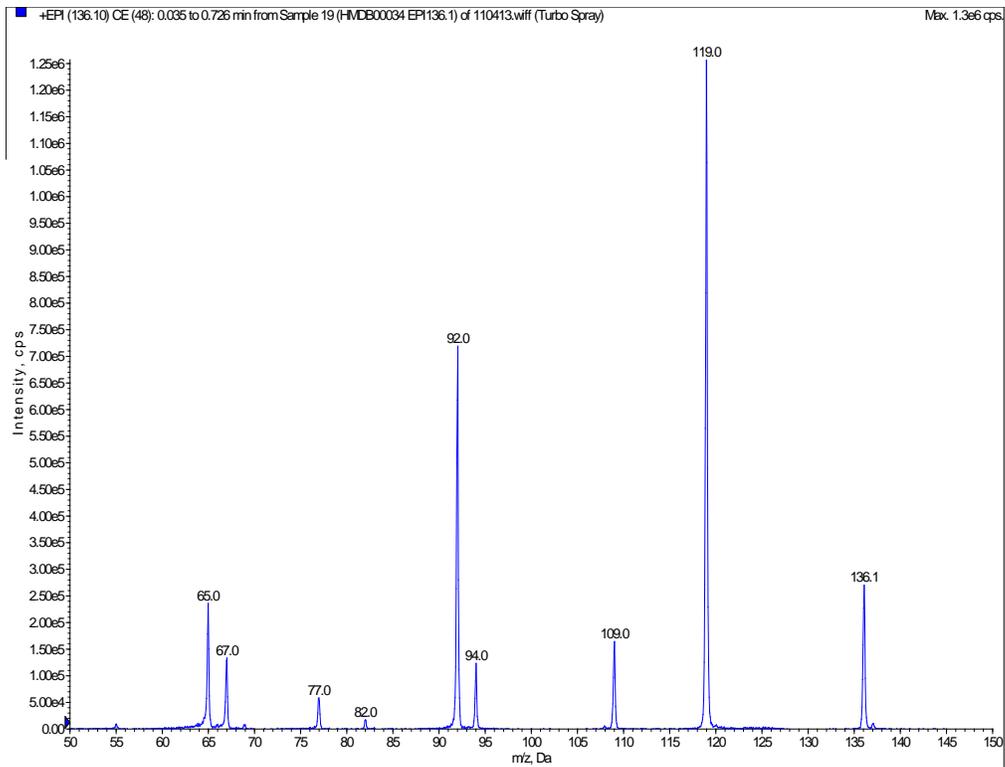
will allow us to profile a greater number of metabolites for metabolomics studies, which will surely expand the understanding of the human salivary metabolome. Another future direction will be on metabolite identification. The method of skimmer-fragmentation followed by MS/MS, as described in Chapter 2, will be applied for identifying unknown metabolites along with other techniques such as NMR if a sufficient amount of a metabolite of interest can be purified. Synthesis of standards to confirm metabolite identification will be the last step in unknown metabolite identification. Because of the great difficulty in identifying an unknown metabolite, compound identification will be carried out only after one or a few metabolites have been confidently determined to a biomarker(s) of a disease. To validate a biomarker of a disease, a large number of samples (thousands) will be required. In this regard, the study of differentiating the MCI individuals from the normal controls needs to be expanded to include a much larger number of populations for metabolome profiling. Nevertheless, the salivary metabolome profiling method described in this thesis forms the foundation from which future work can be carried out by expanding the metabolome coverage and by analyzing a great number of samples in many disease biomarker discovery projects.

Appendix

HMDB00034

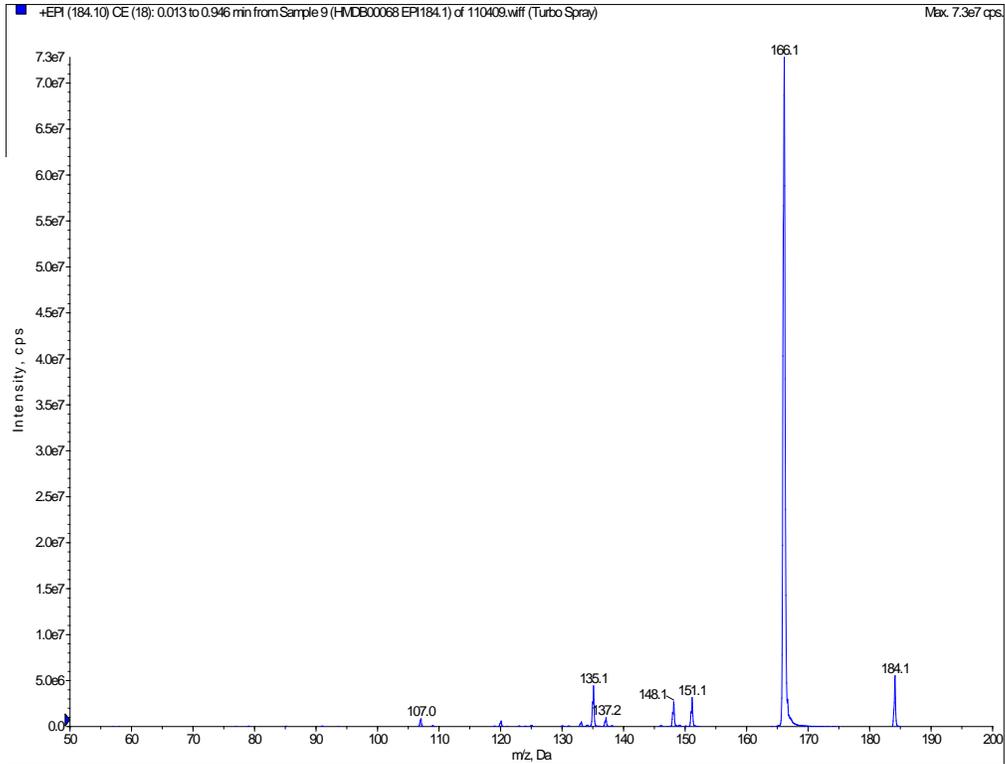
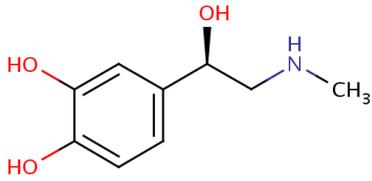
Adenine (7H-purin-6-amine)

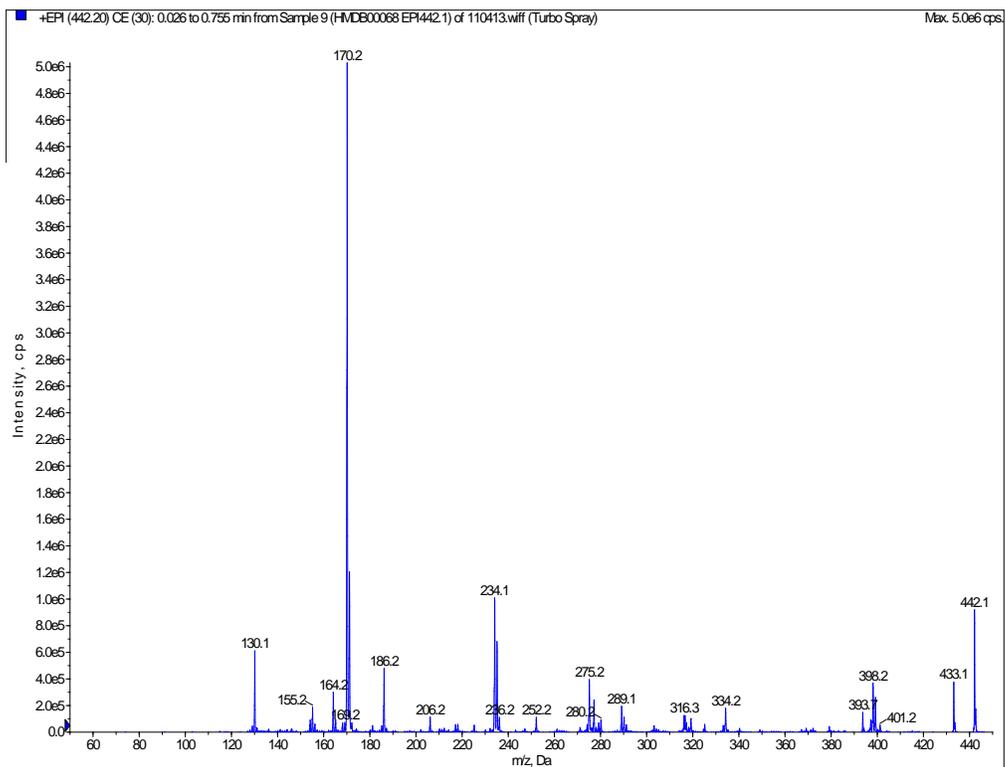
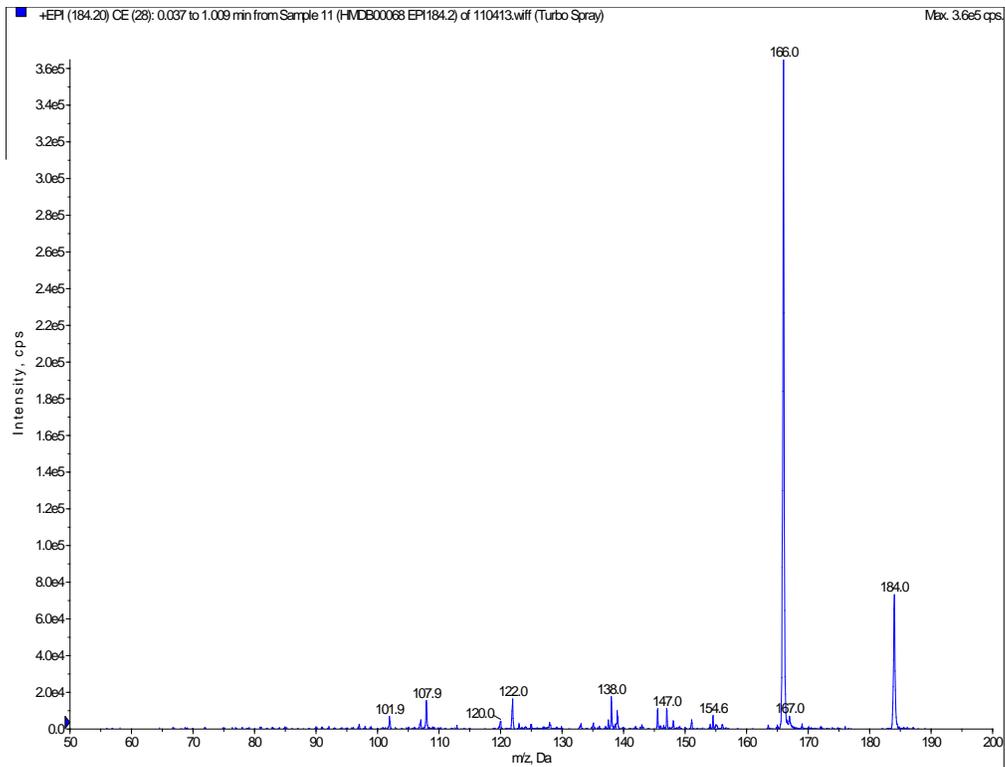




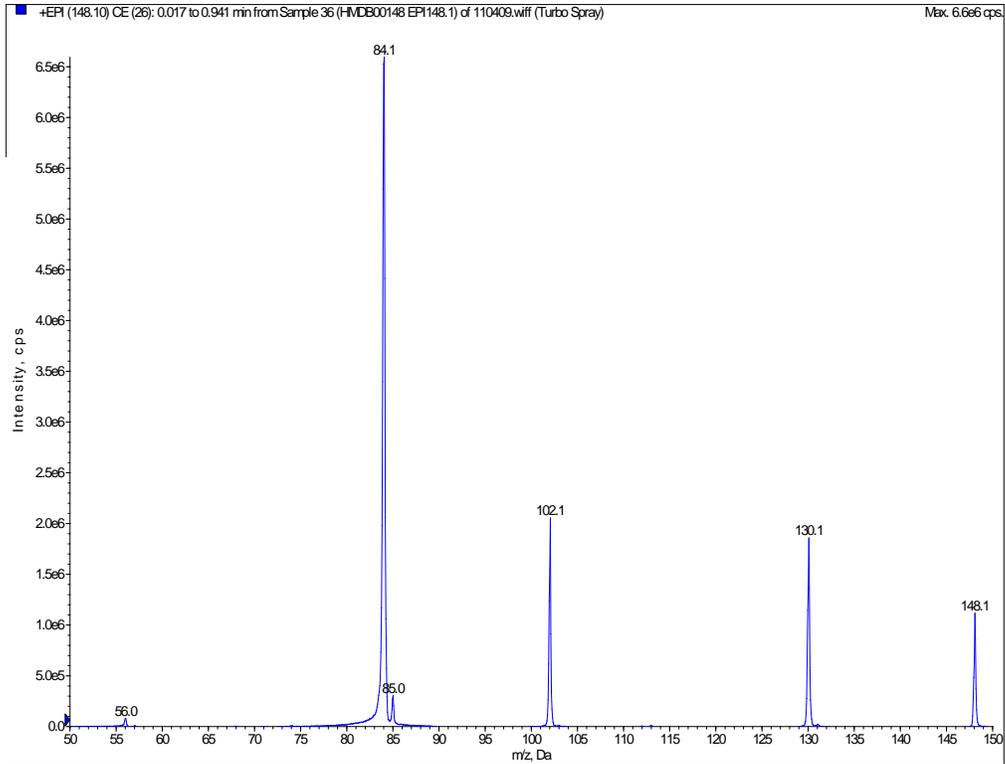
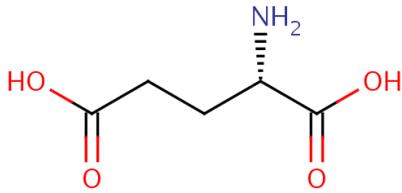
HMDB00068

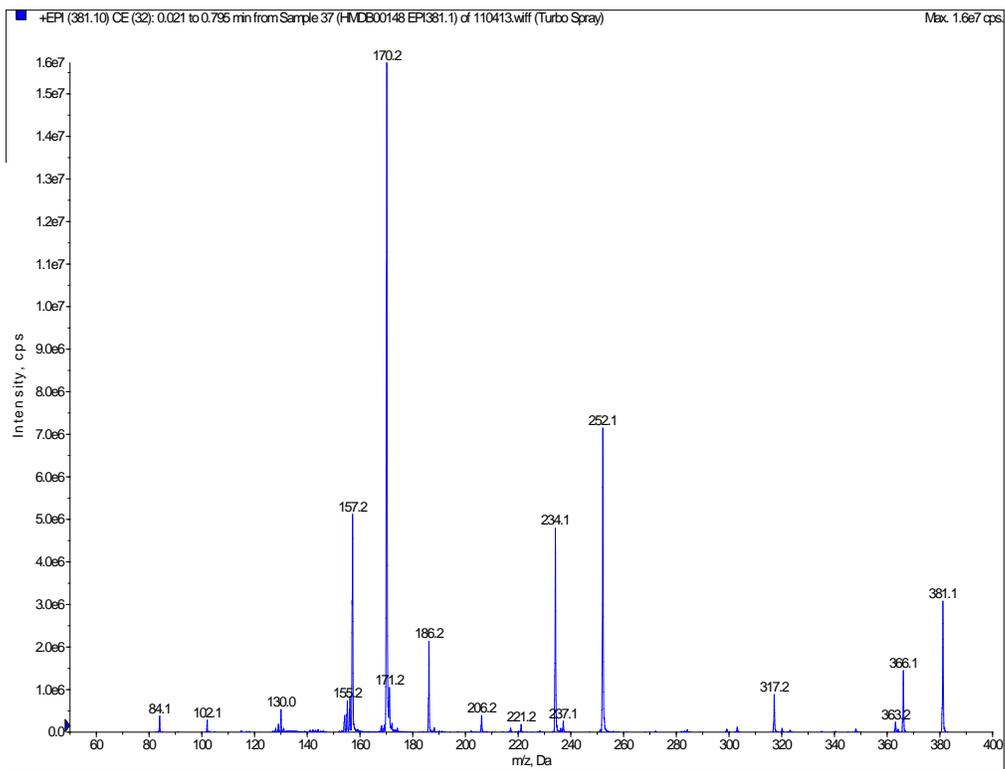
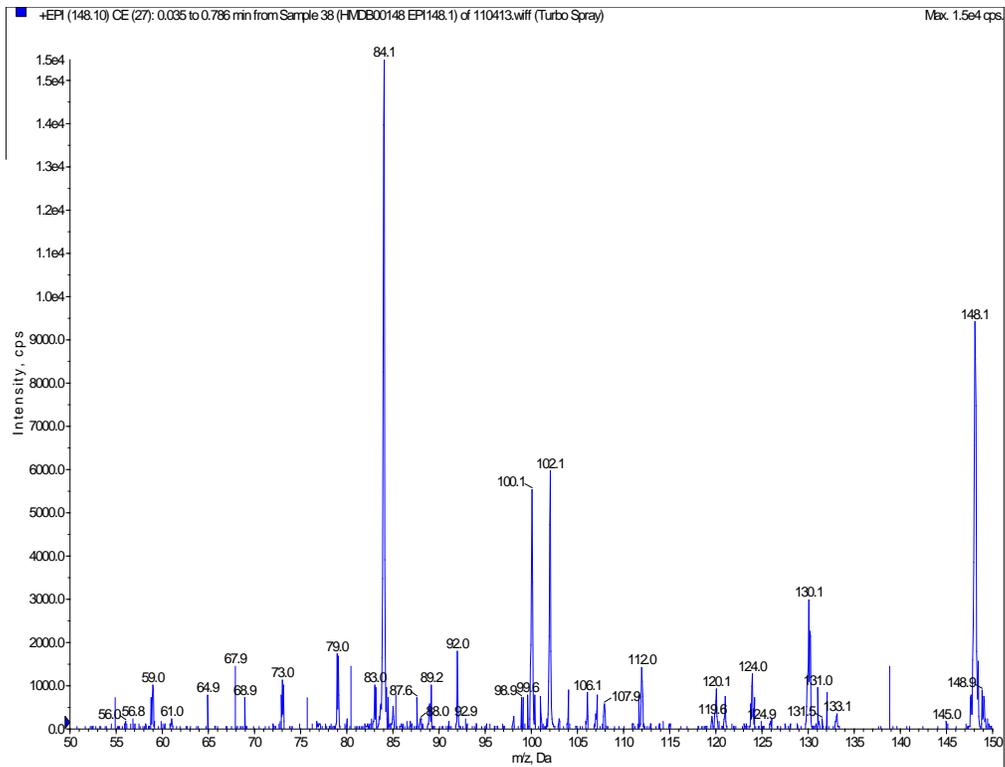
Epinephrine (4-[(1R)-1-hydroxy-2-methylamino-ethyl]benzene-1,2-diol)



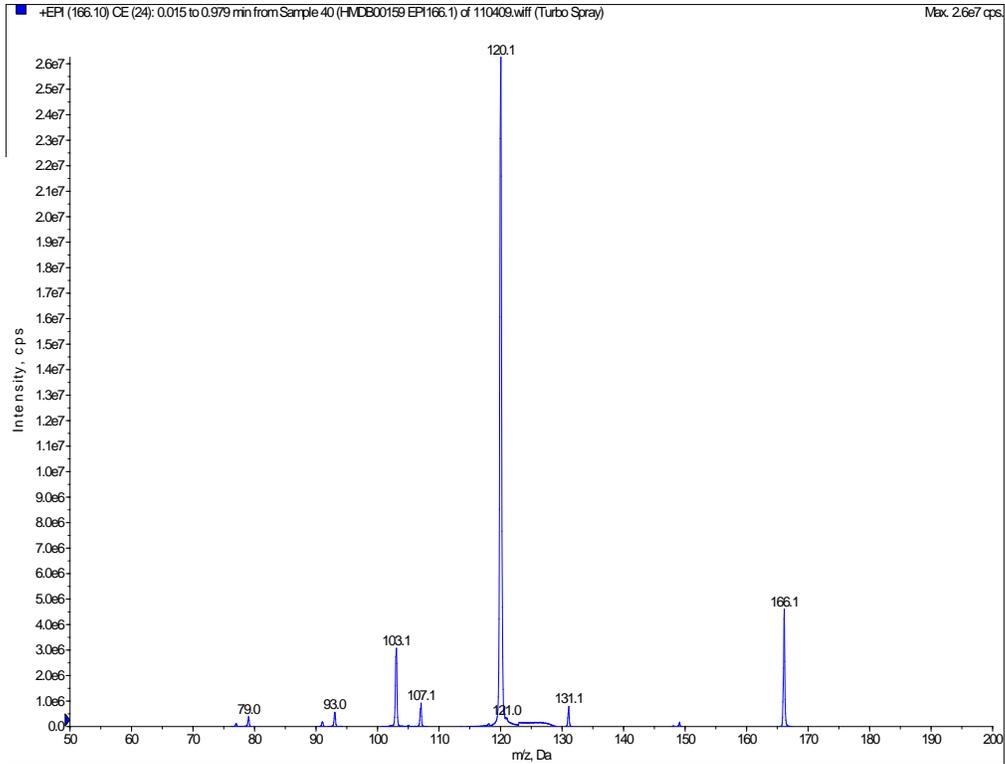
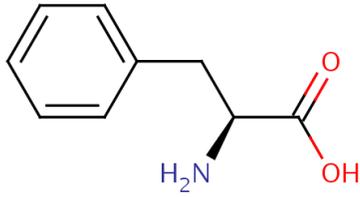


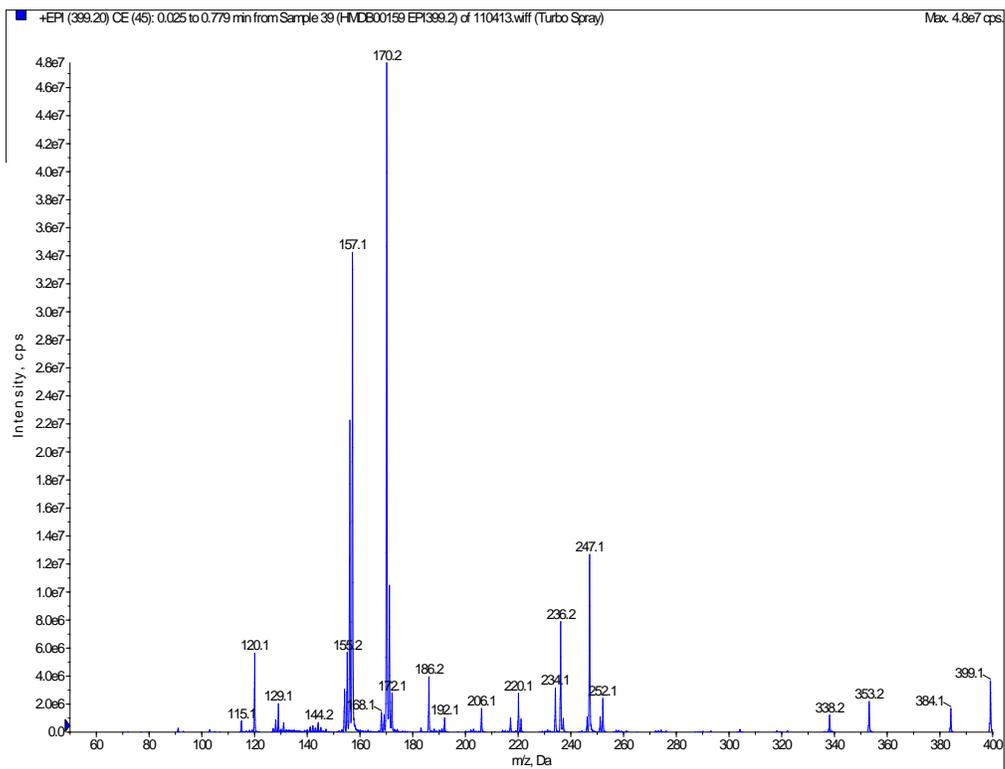
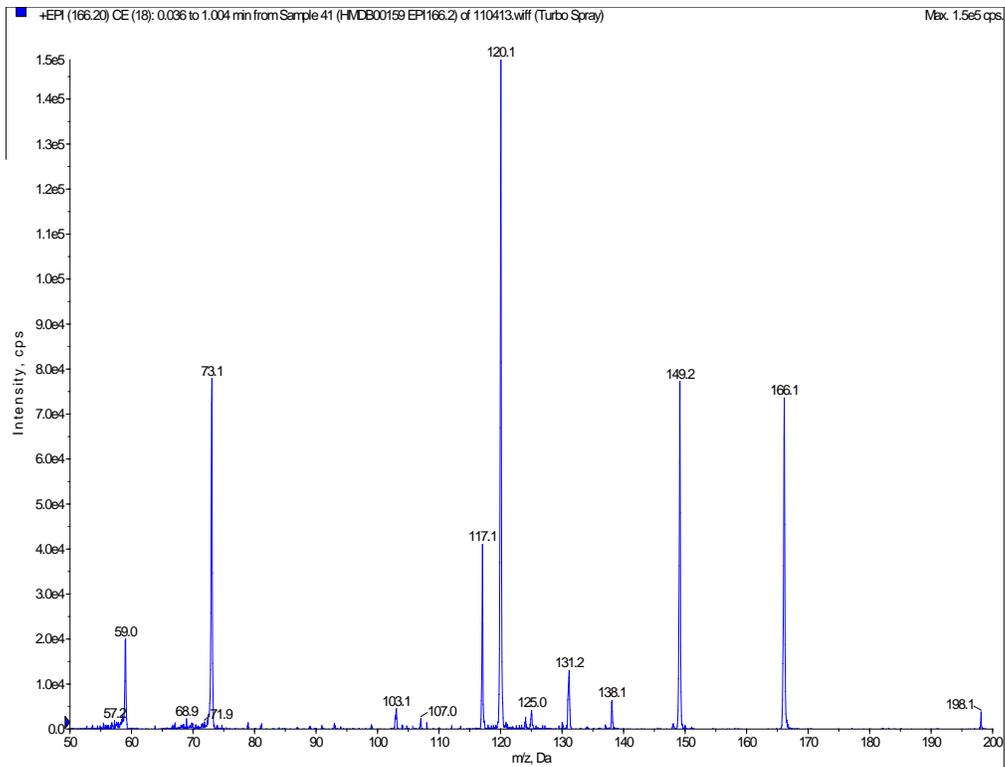
HMDB00148
L-Glutamic acid (2-aminopentanedioic acid)





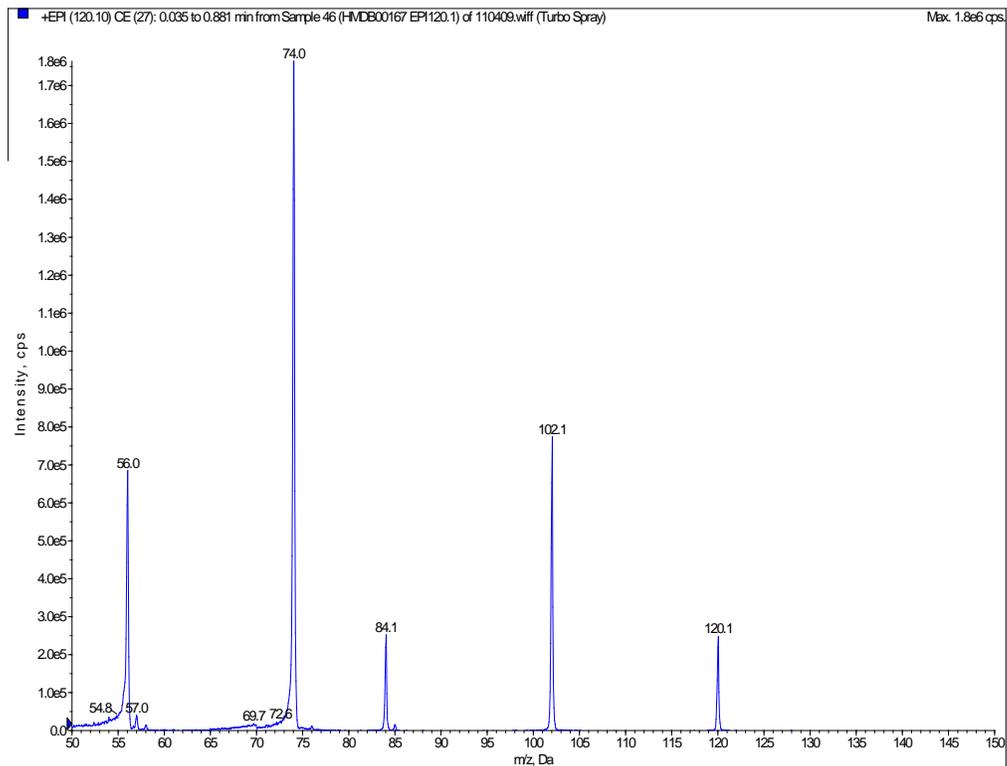
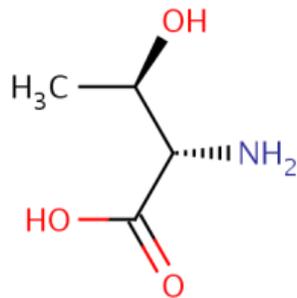
HMDB00159
L-Phenylalanine (2-amino-3-phenyl-propanoic acid)

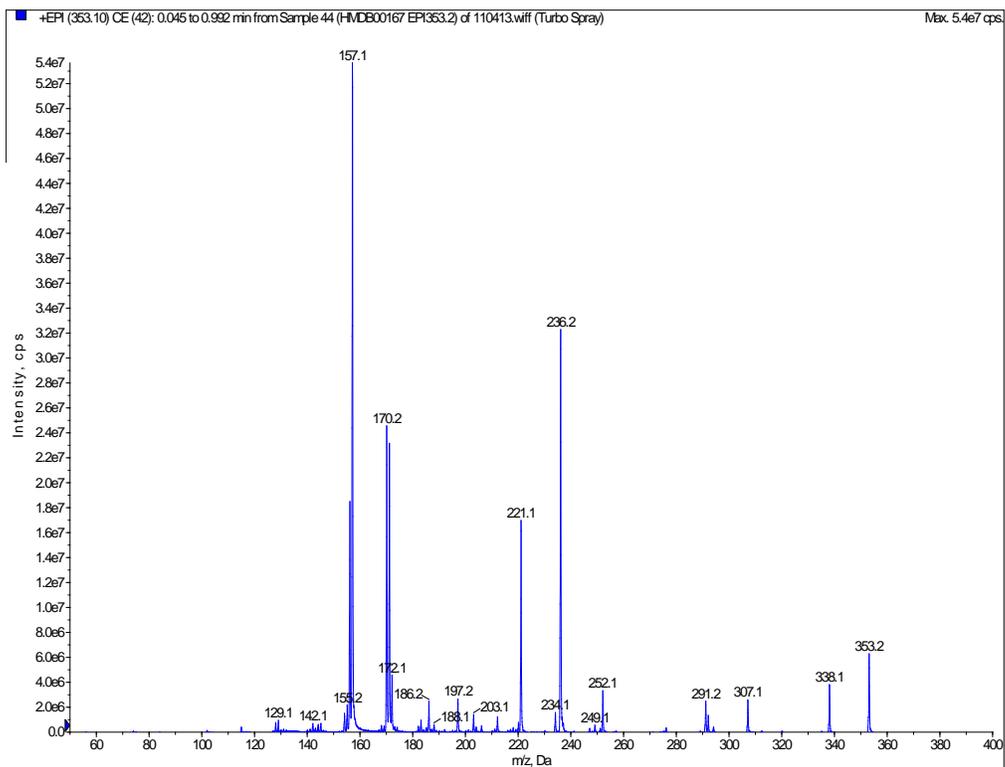
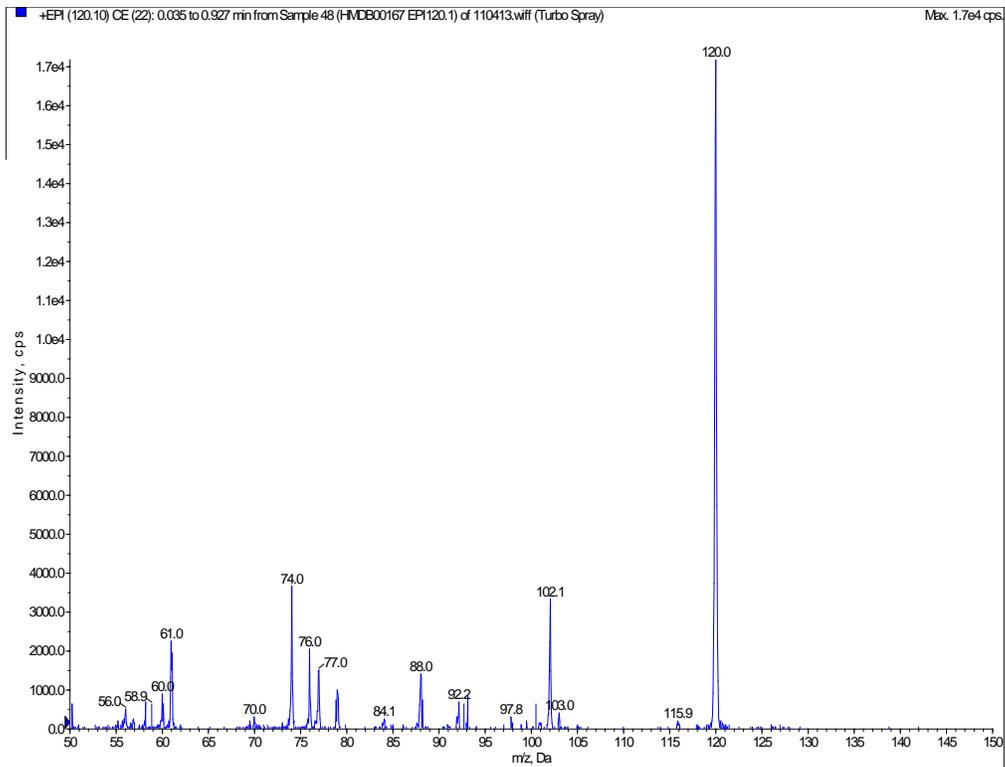




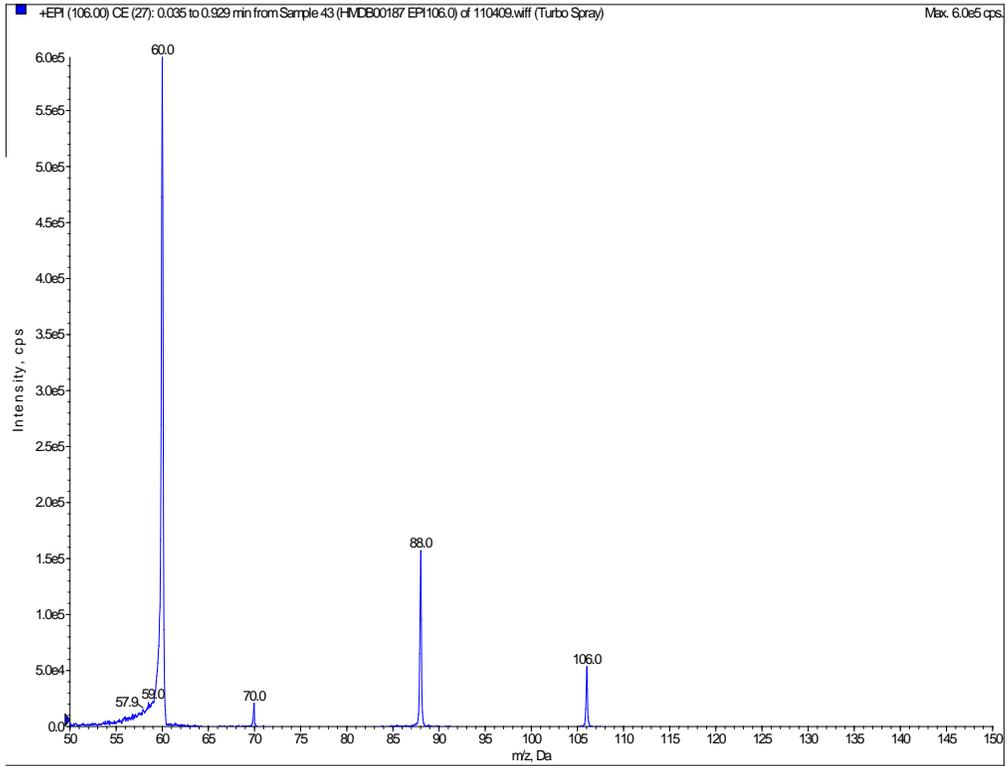
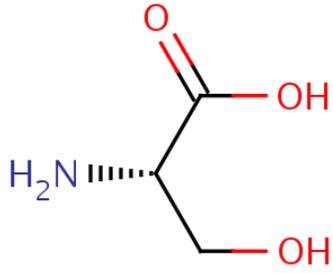
HMDB00167

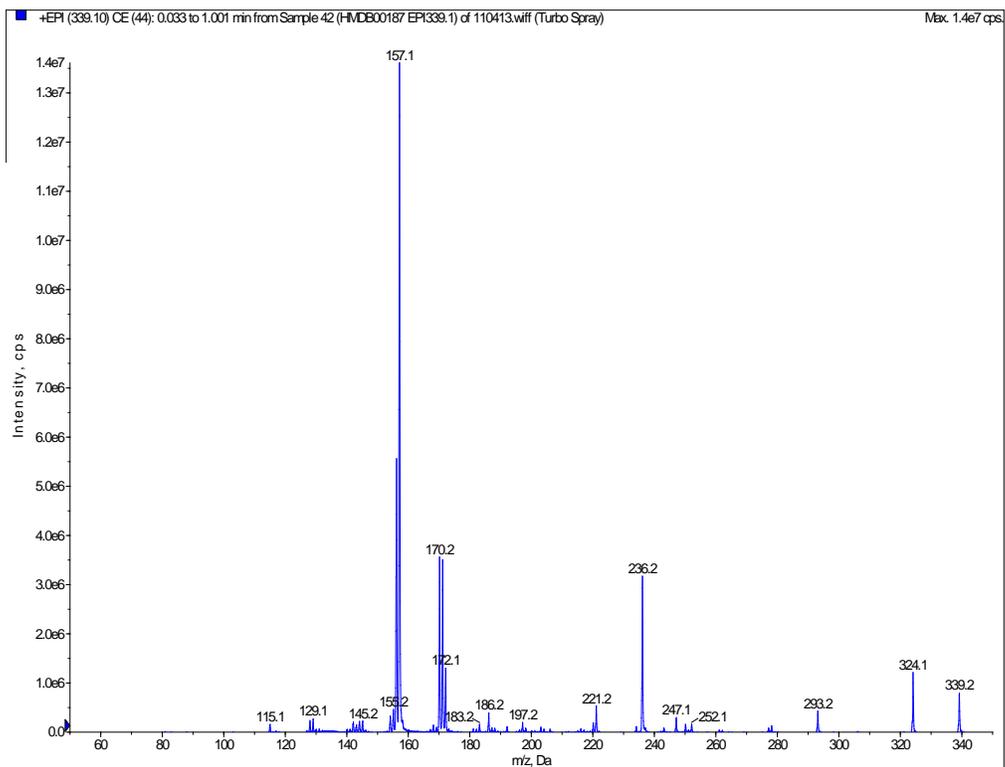
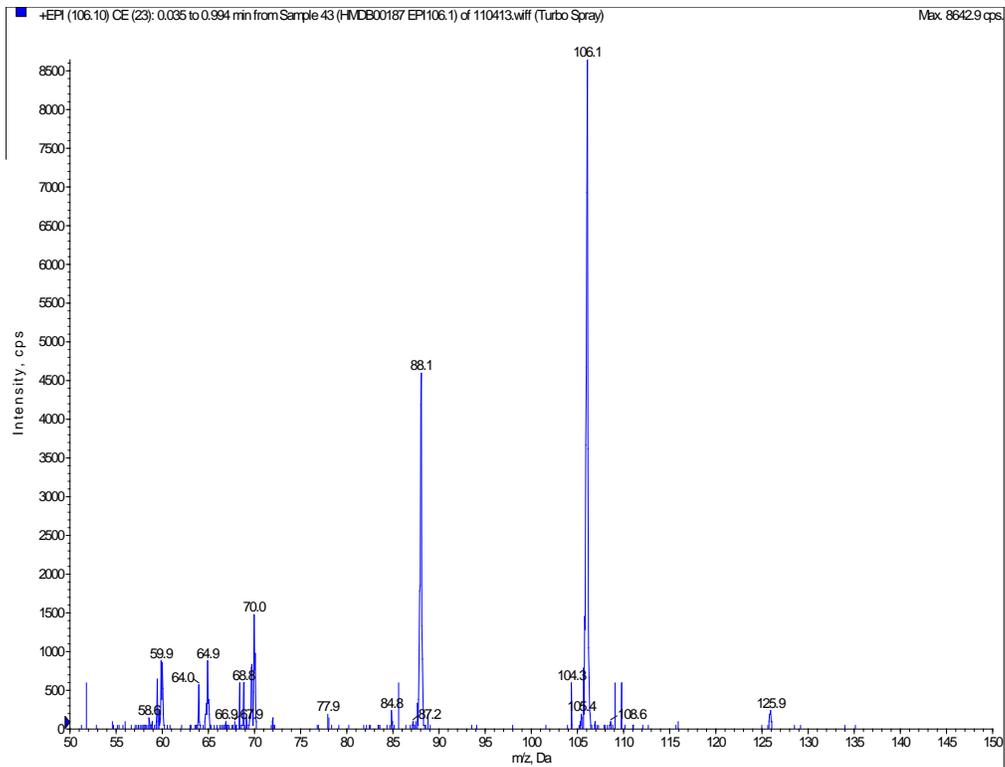
L-Threonine (2-amino-3-hydroxy-butanoic acid)





HMDB00187
L-Serine (2-amino-3-hydroxy-propanoic acid)

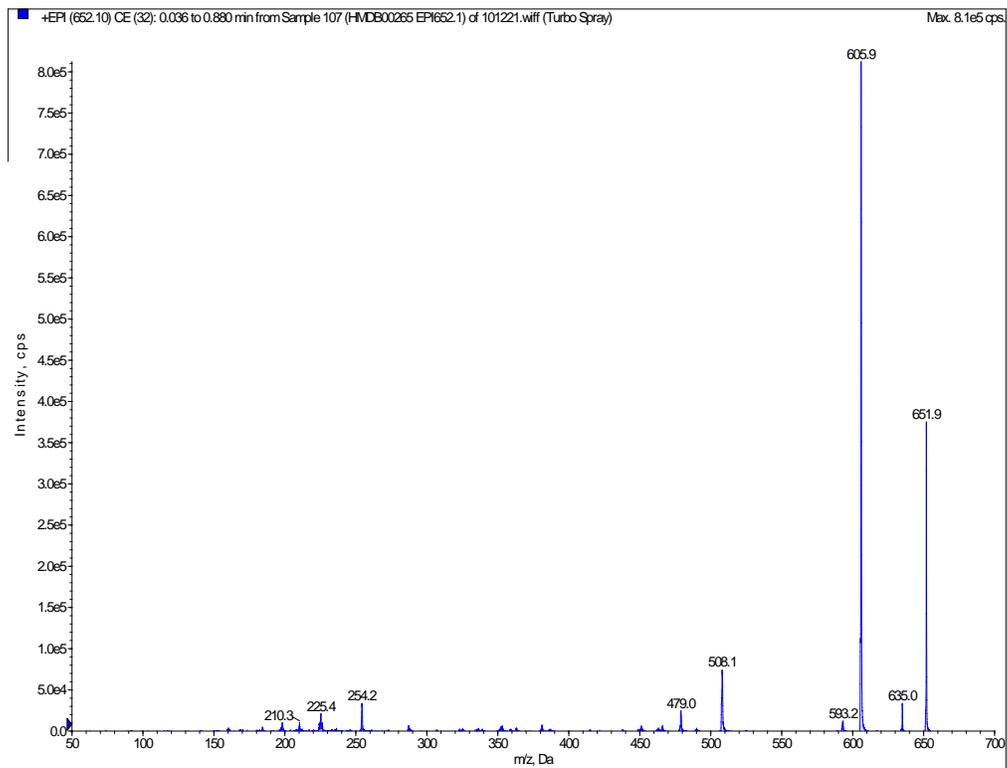
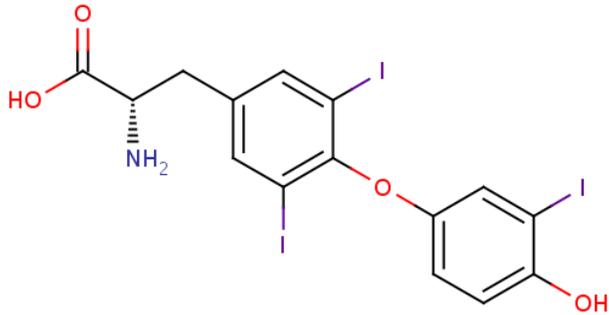


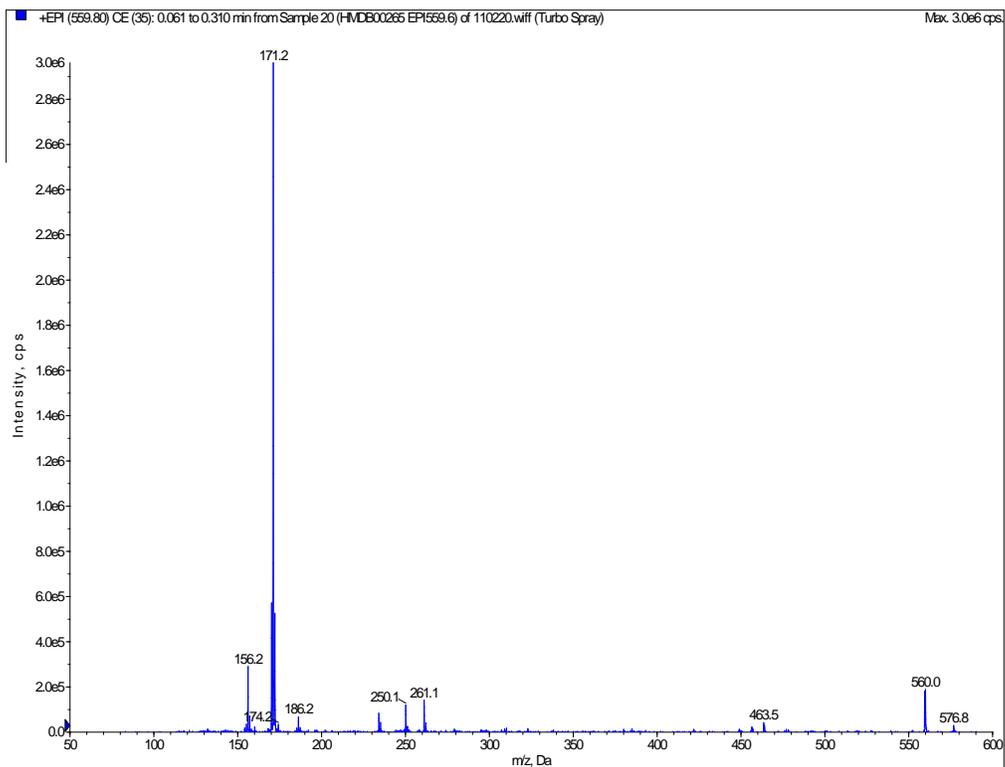
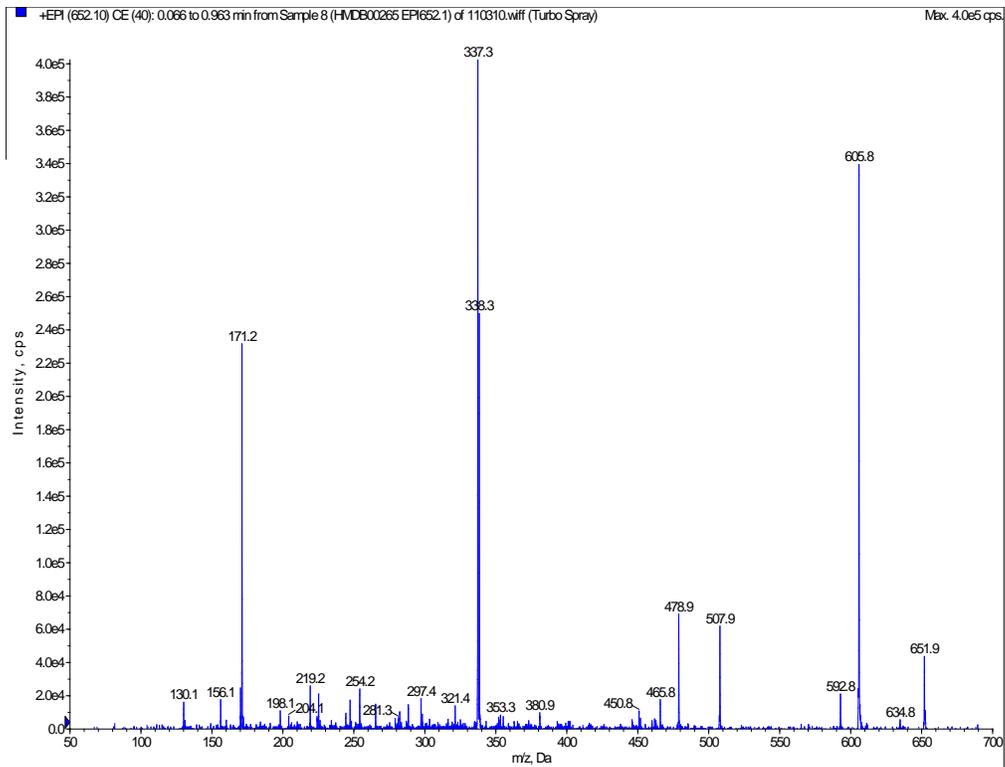


HMDB00265

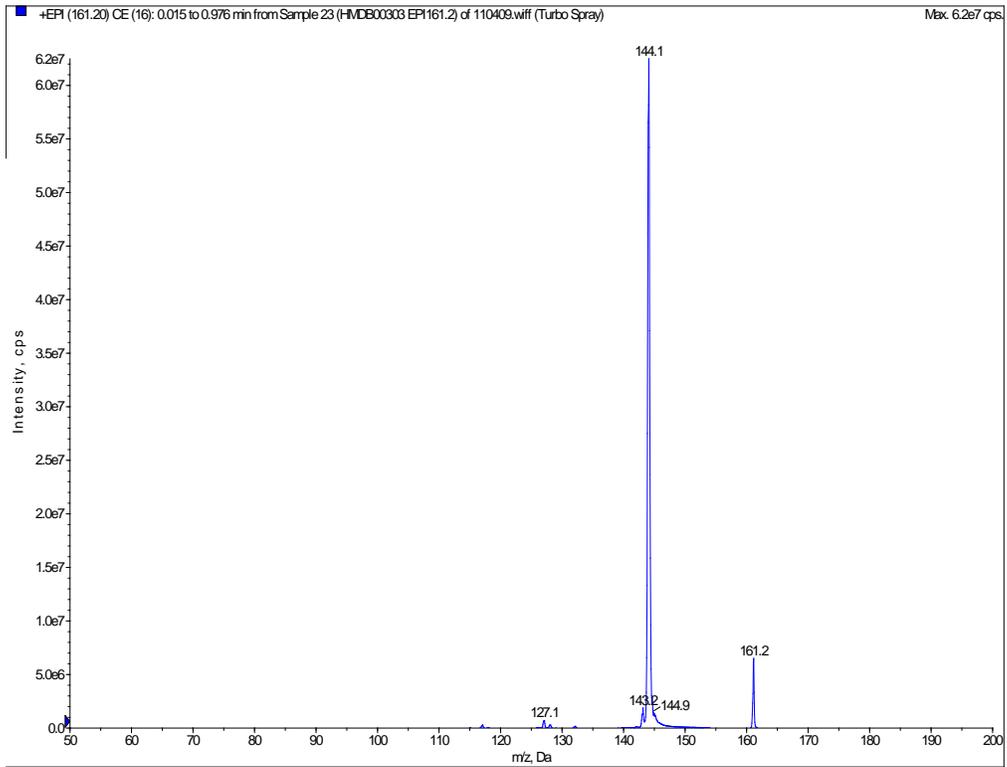
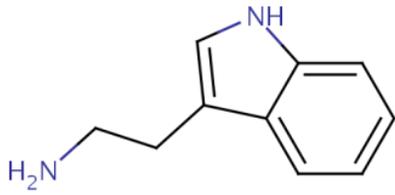
Liothyronine

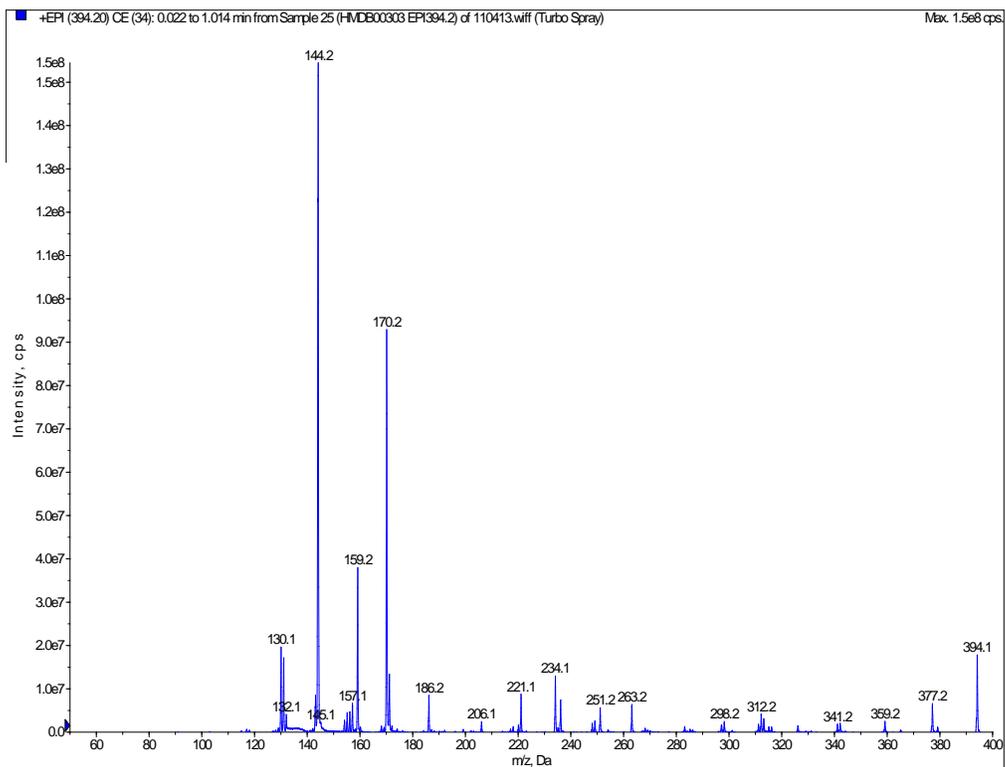
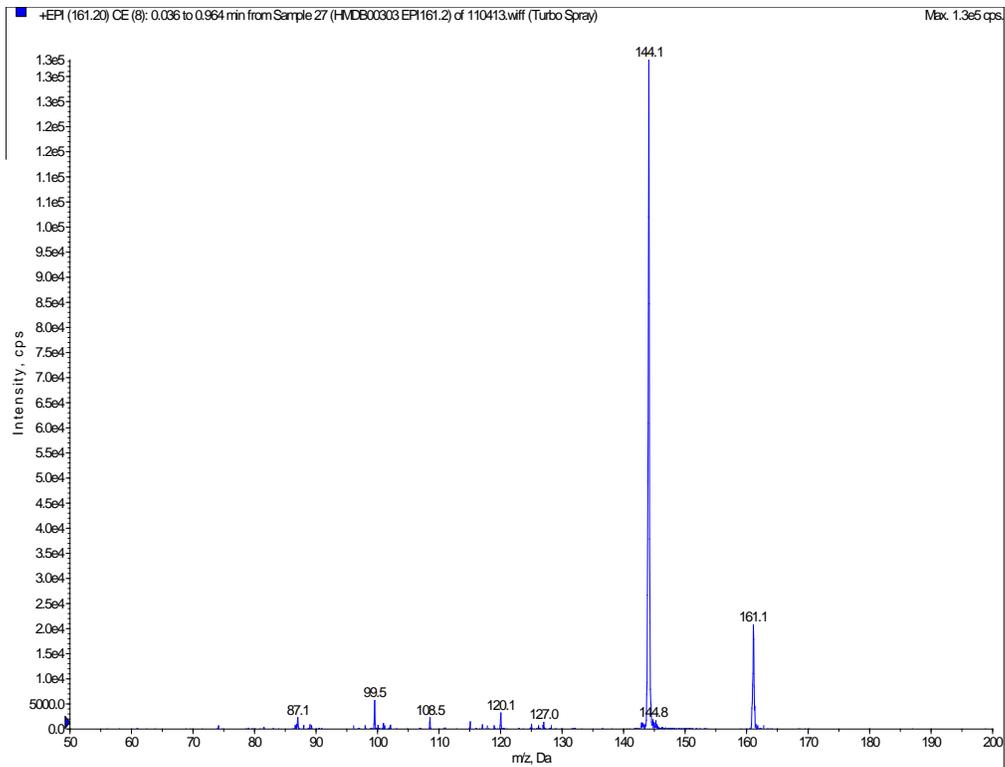
(2-amino-3-[4-(4-hydroxy-3-iodo-phenoxy)-3,5-diiodo-phenyl]-propanoic acid)





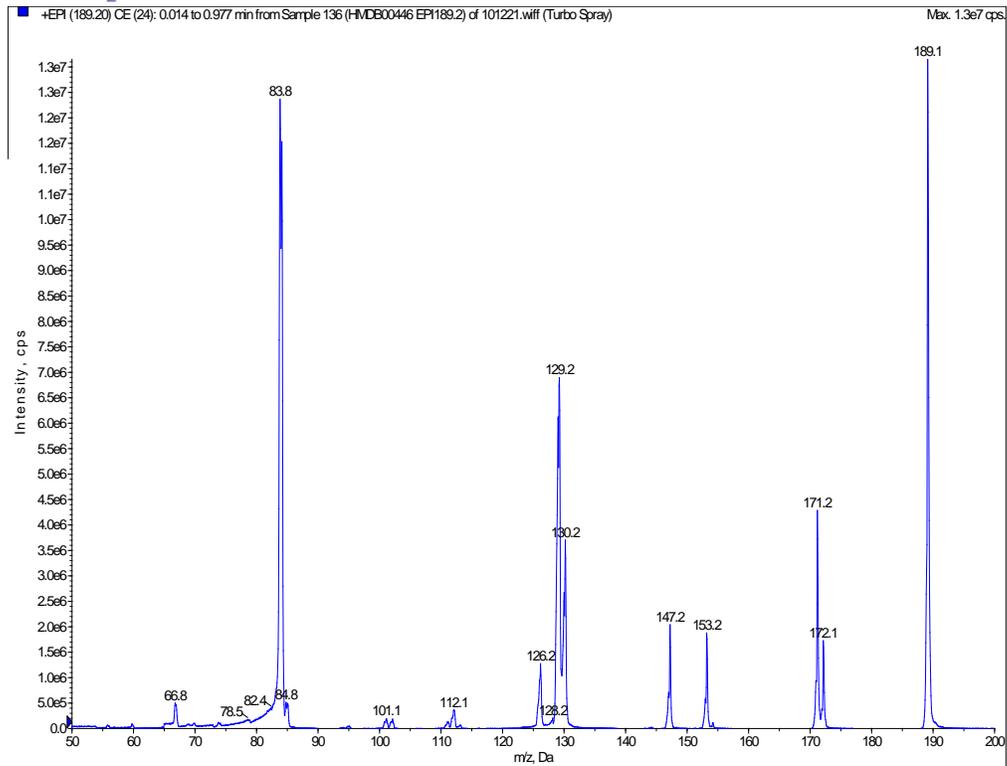
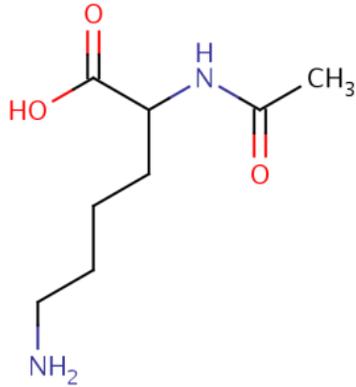
HMDB00303
Tryptamine (2-(1H-indol-3-yl)-ethylamine)

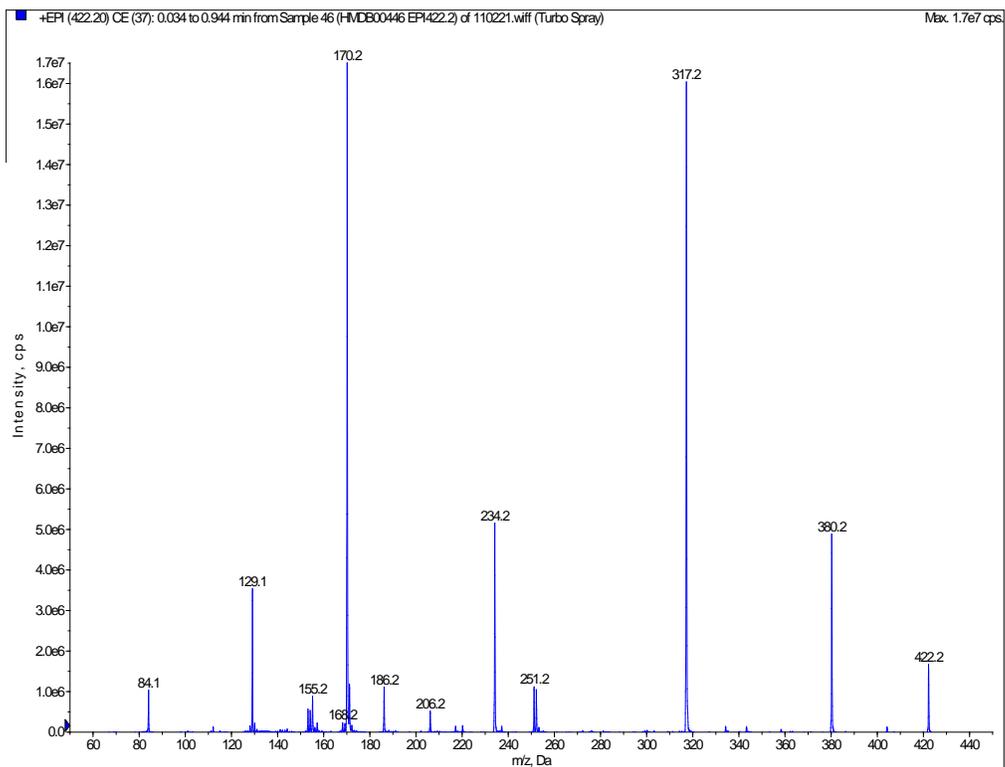
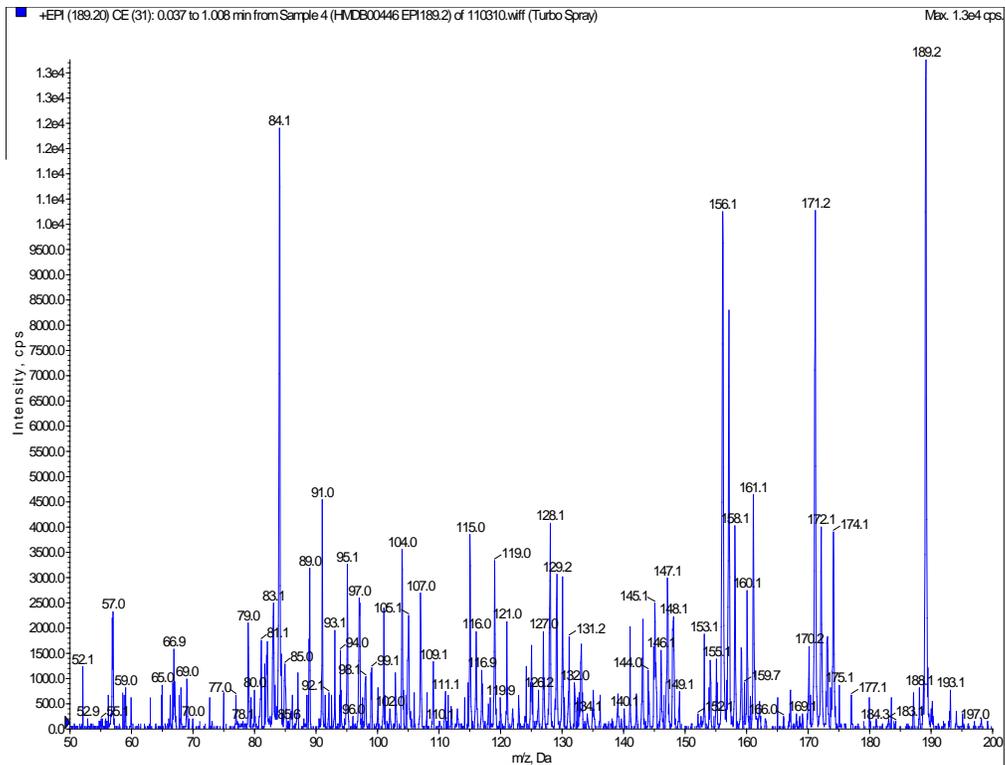




HMDB00446

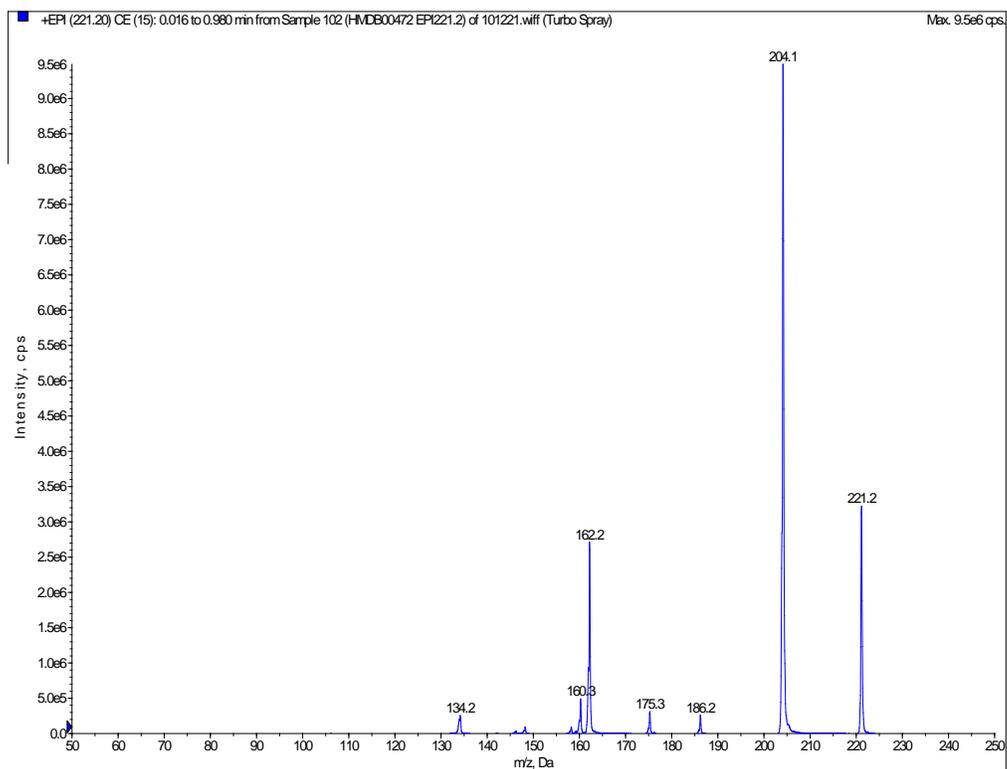
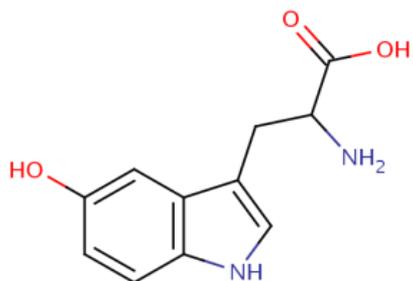
N-Alpha-acetyllysine (2-acetamido-6-aminohexanoic acid)

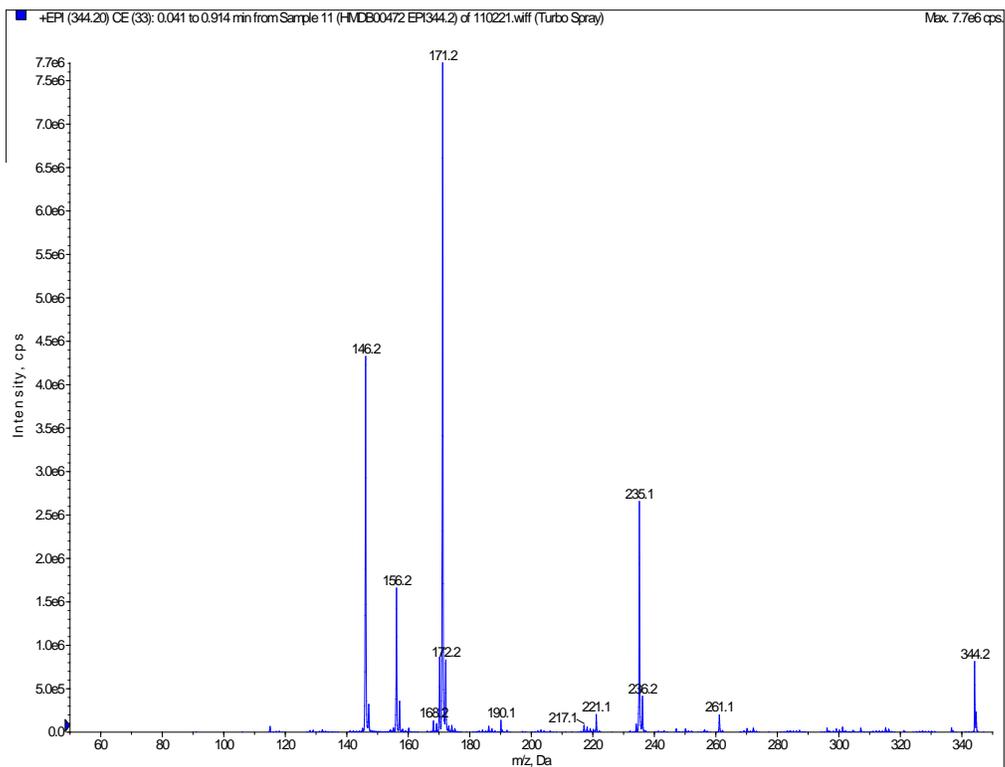
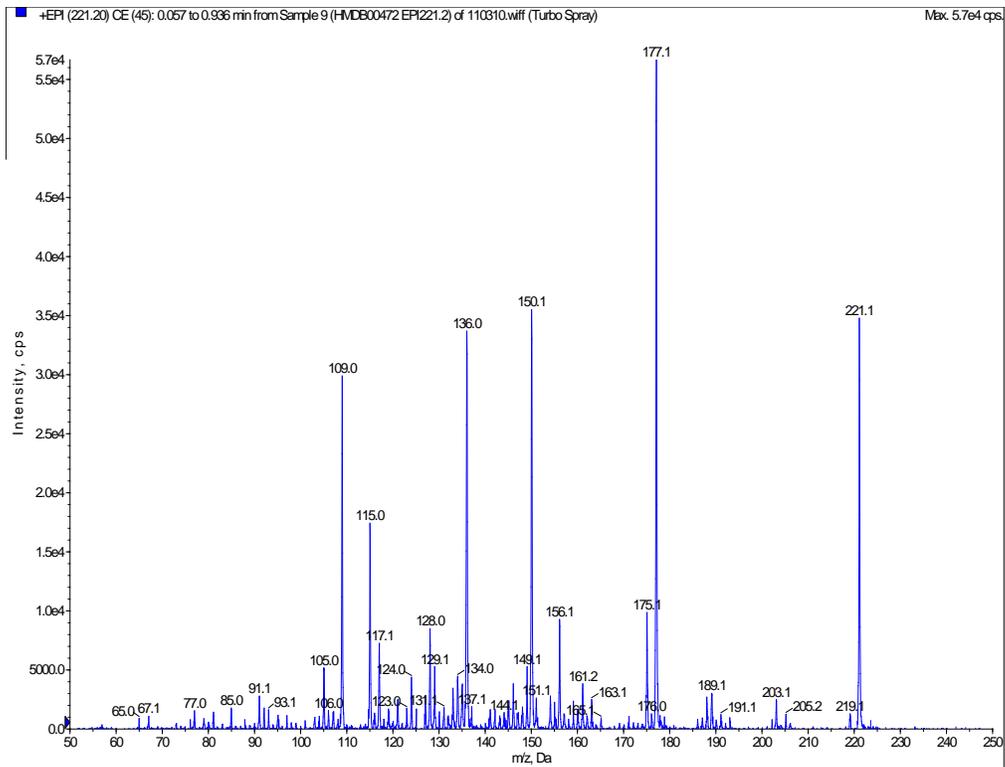




HMDB00472

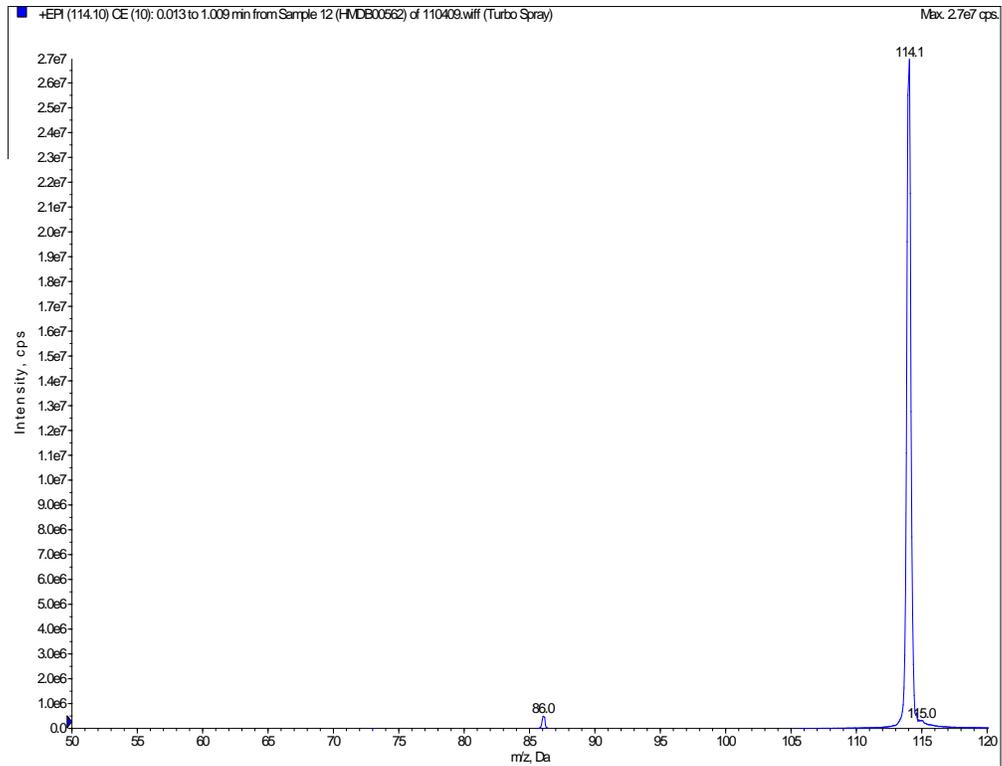
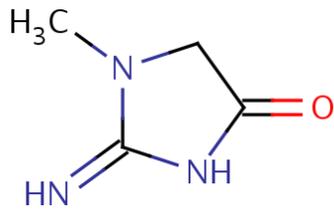
5-Hydroxy-L-tryptophan (2-amino-3-(5-hydroxy-1H-indol-3-yl)propanoic acid)

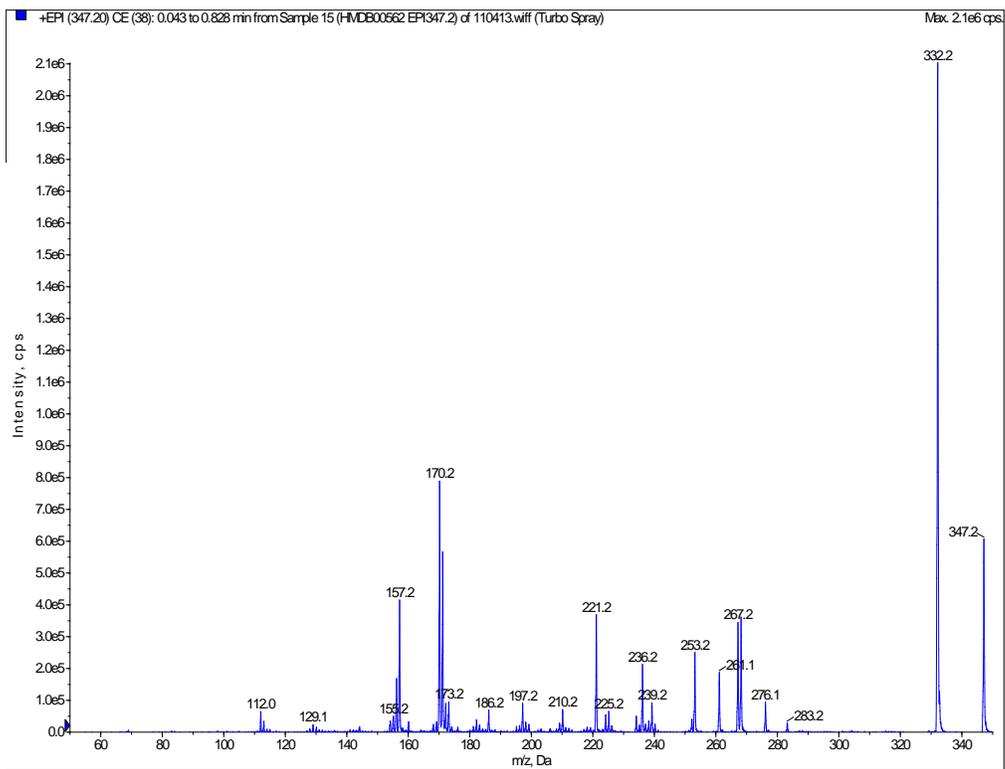
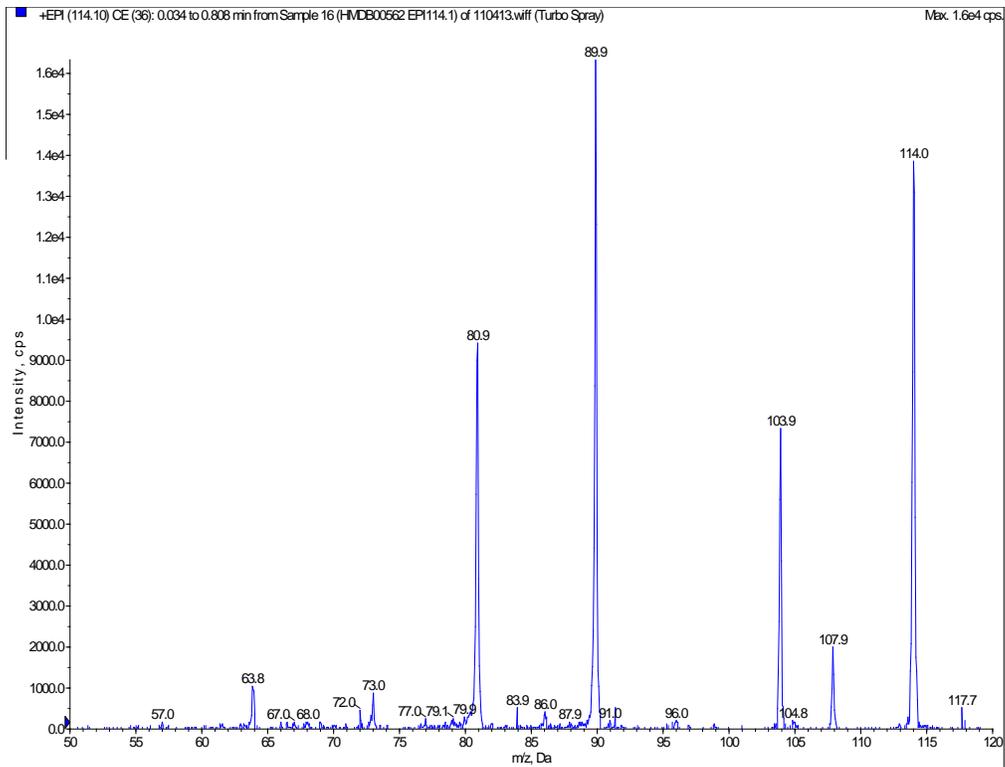




HMDB00562

Creatinine (2-amino-1,5-dihydro-1-methyl-4H-imidazol-4-one)

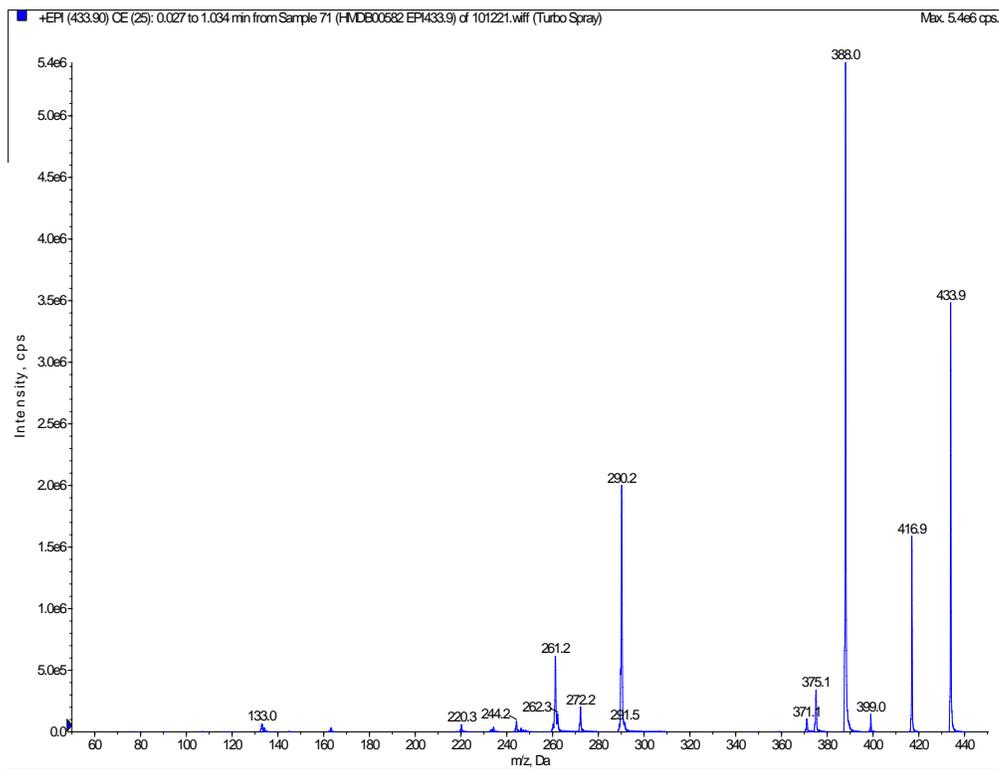
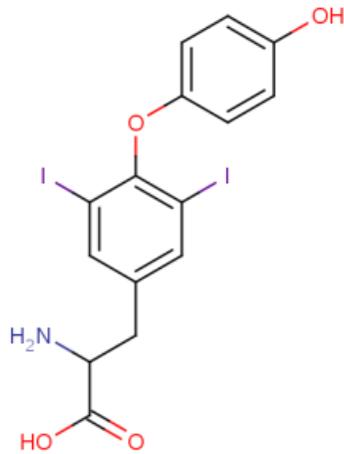


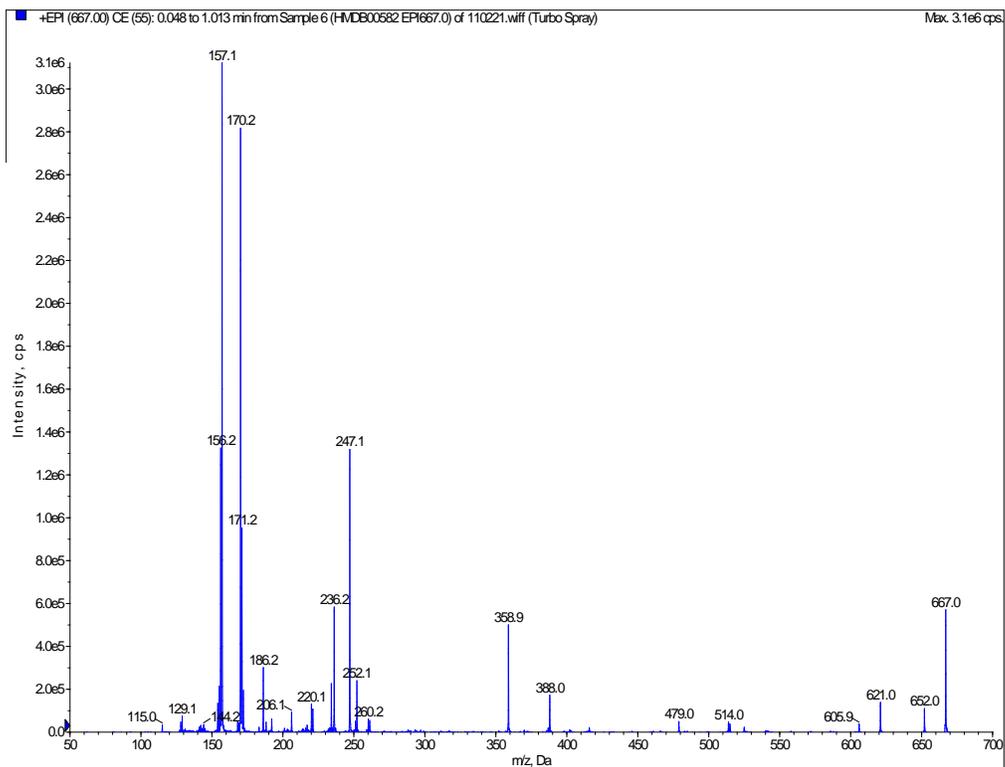
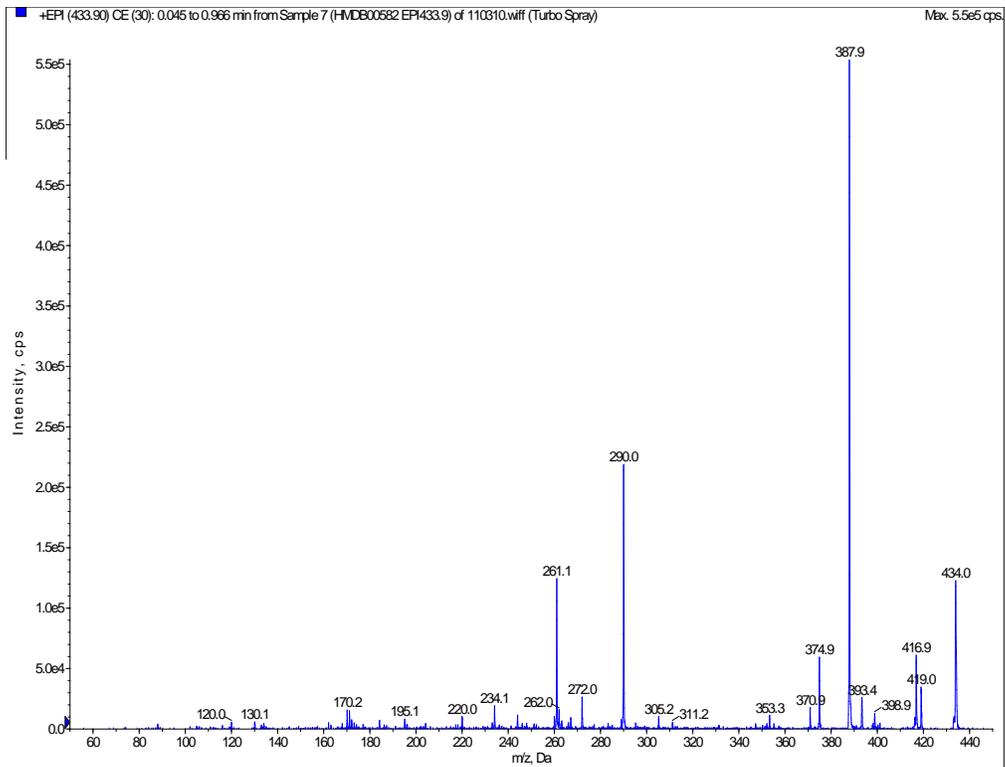


HMDB00582

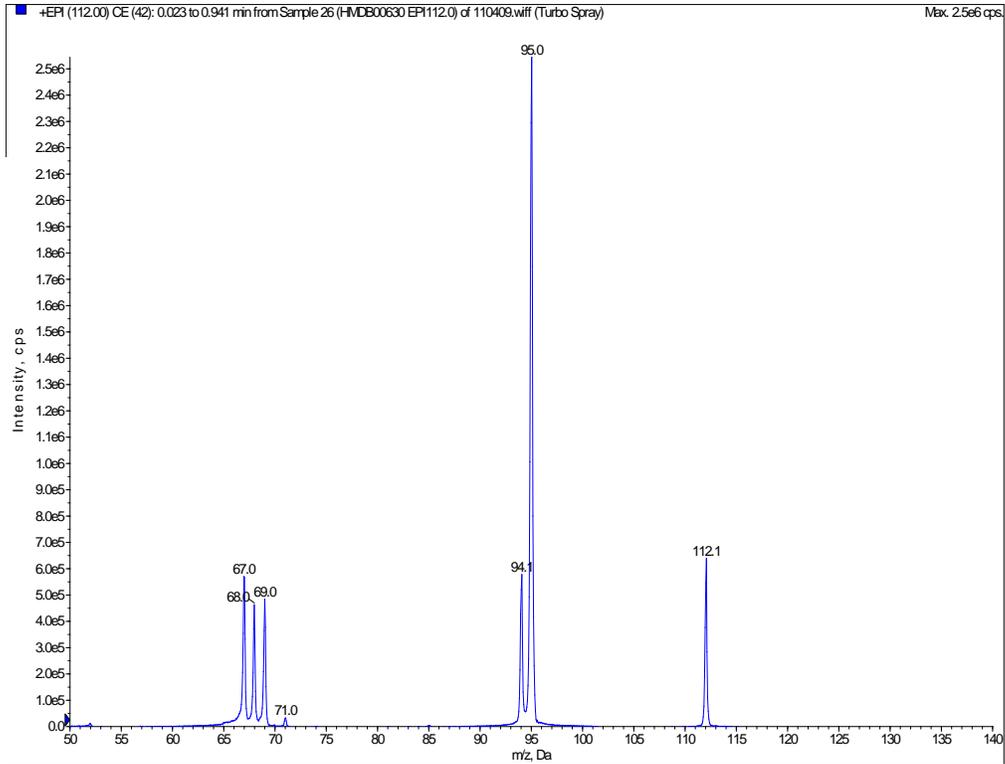
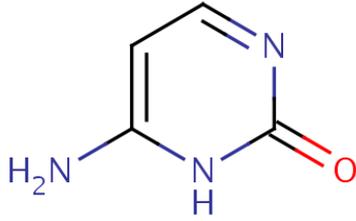
3,5-Diiodothyronine

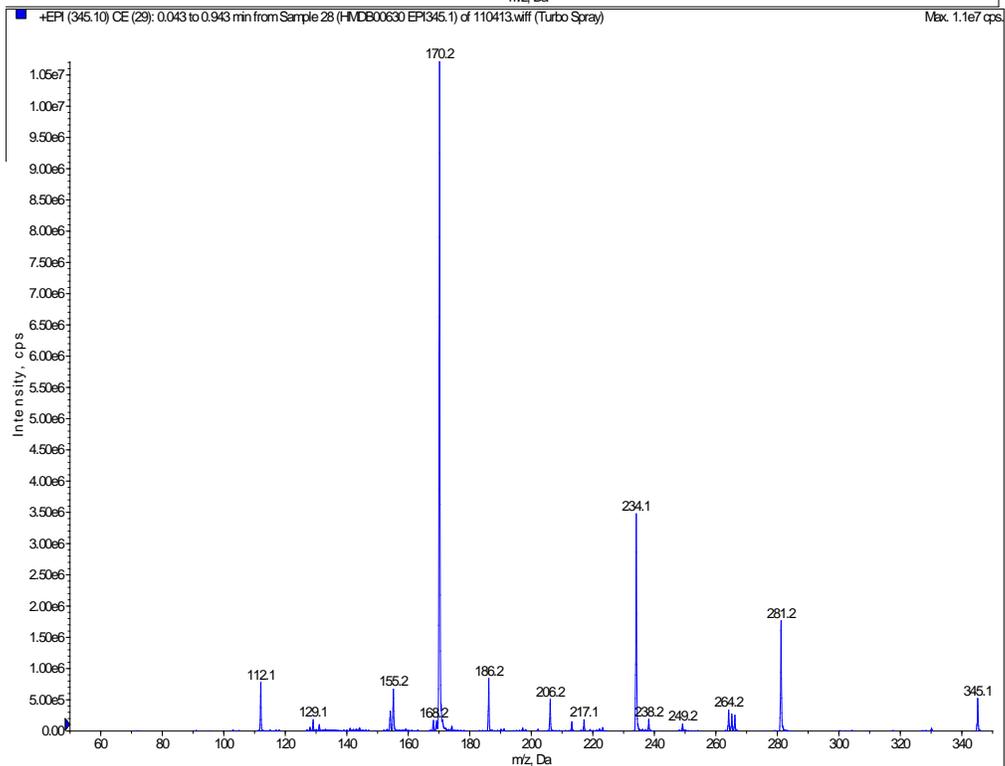
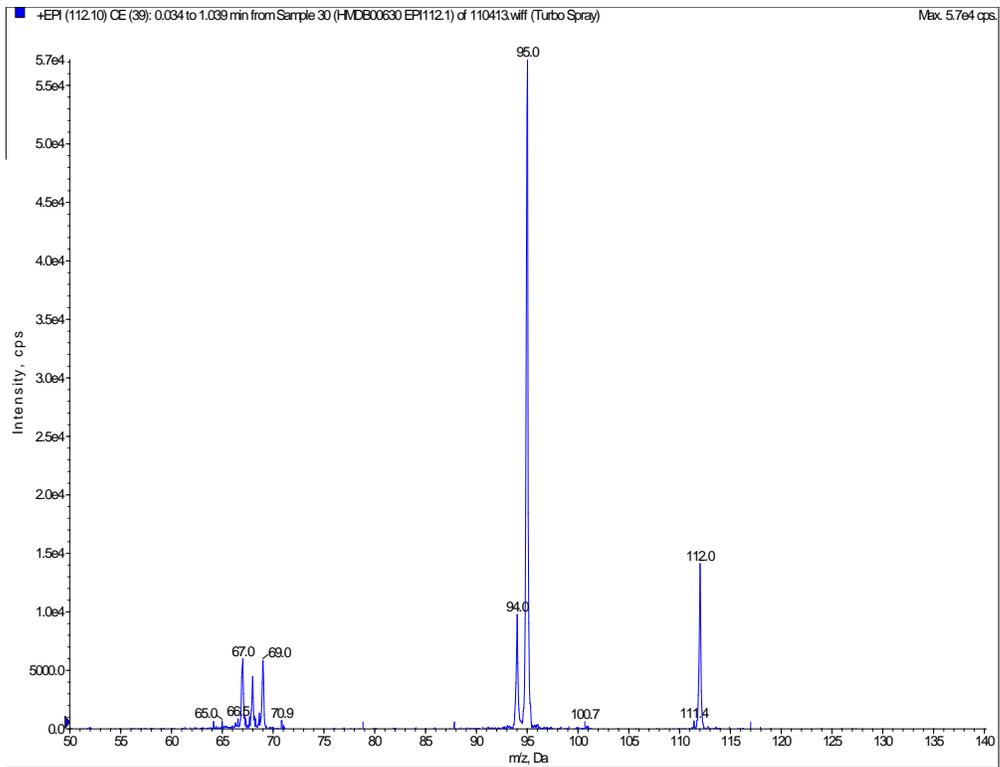
(2-amino-3-[4-(4-hydroxyphenoxy)-3,5-diiodo-phenyl]-propanoic acid)



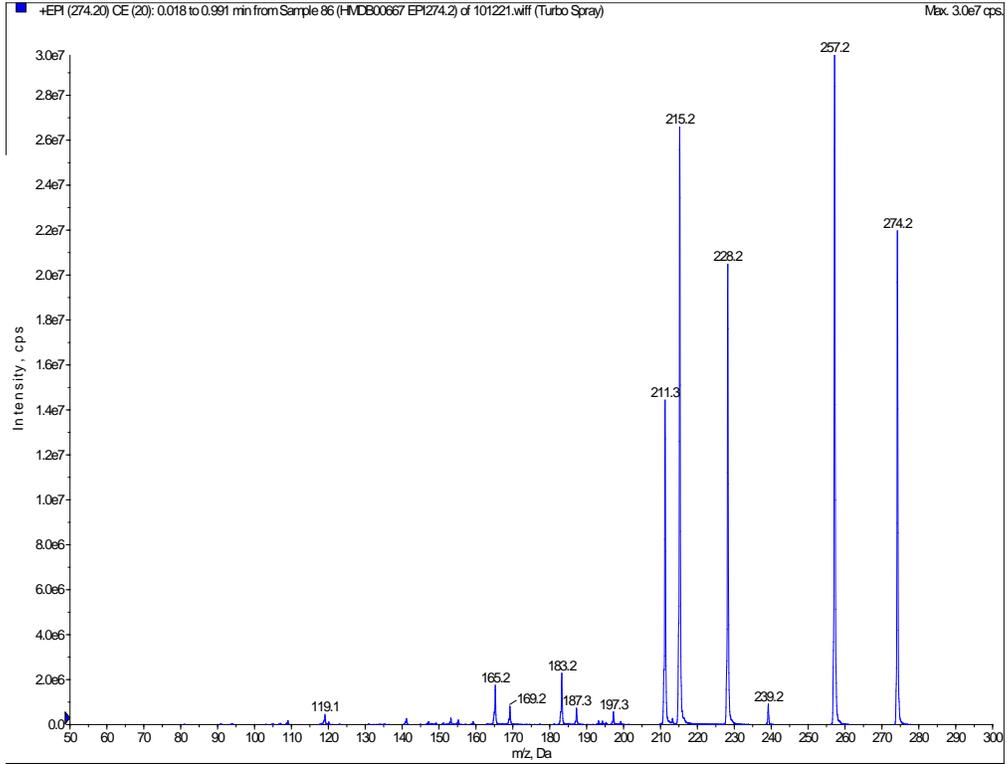
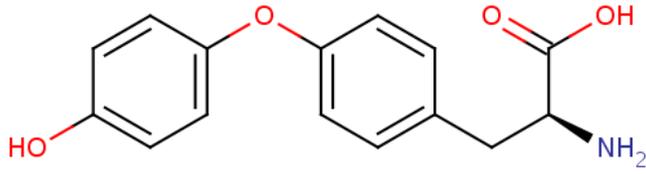


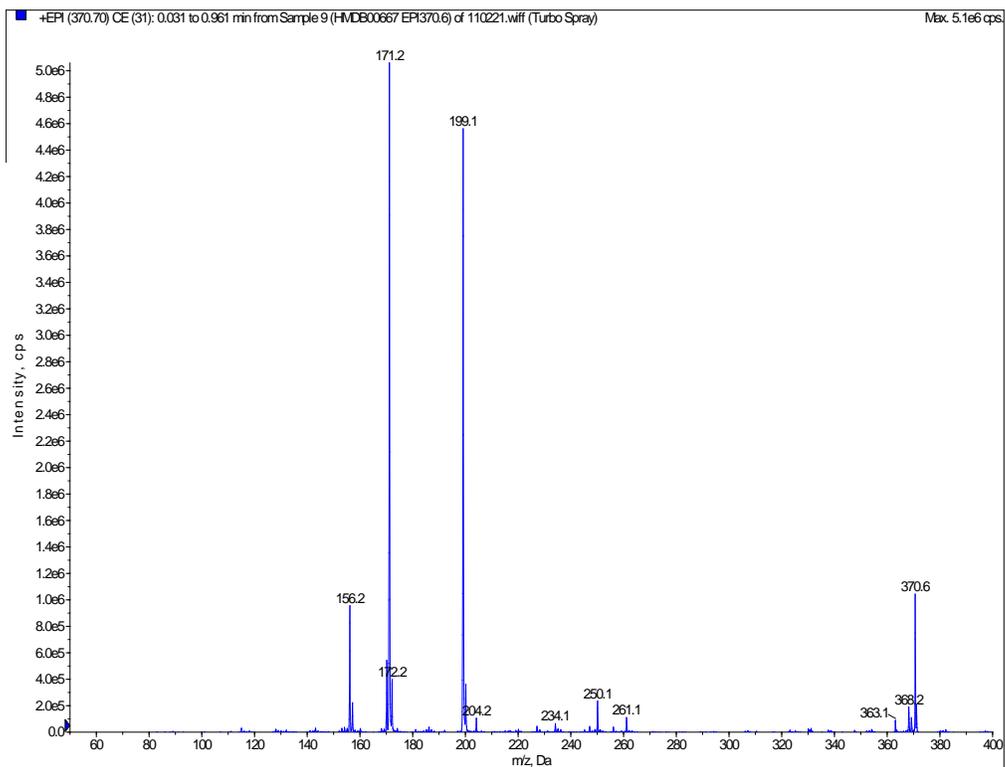
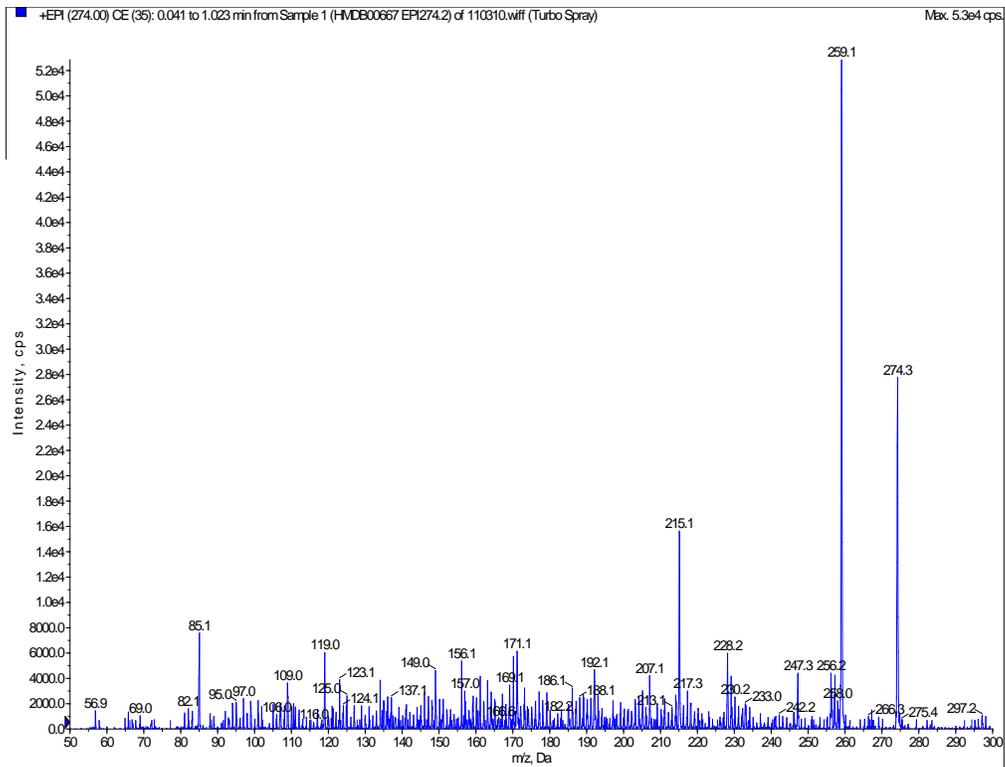
HMDB00630
Cytosine (4-amino-2(1H)-pyrimidinone)





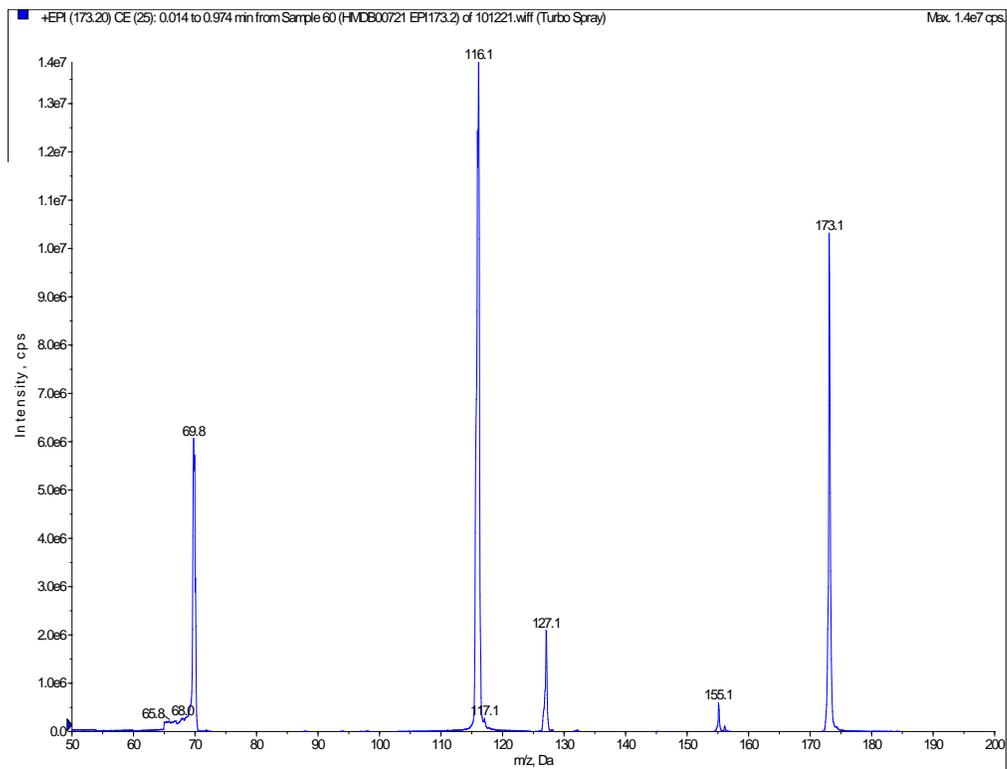
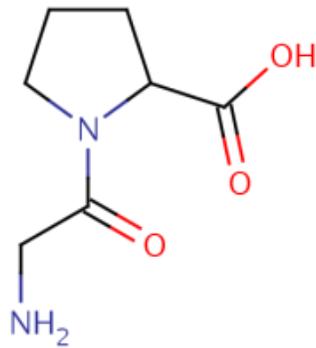
HMDB00667
L-Thyronine (O-(4-hydroxyphenyl)-L-Tyrosine)

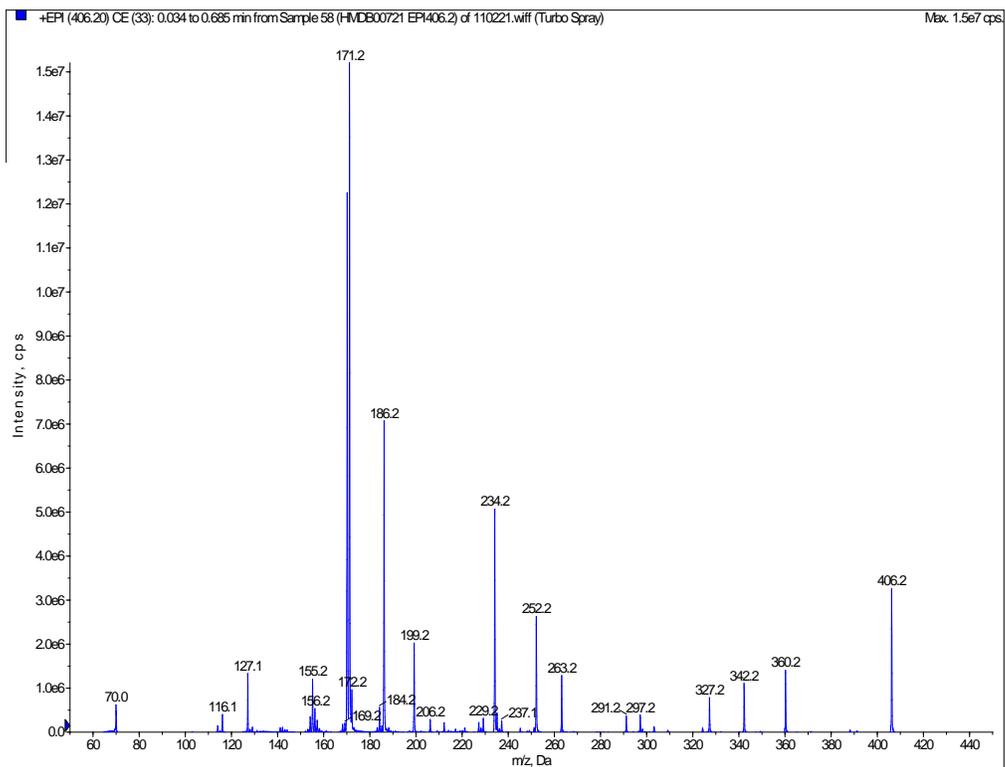
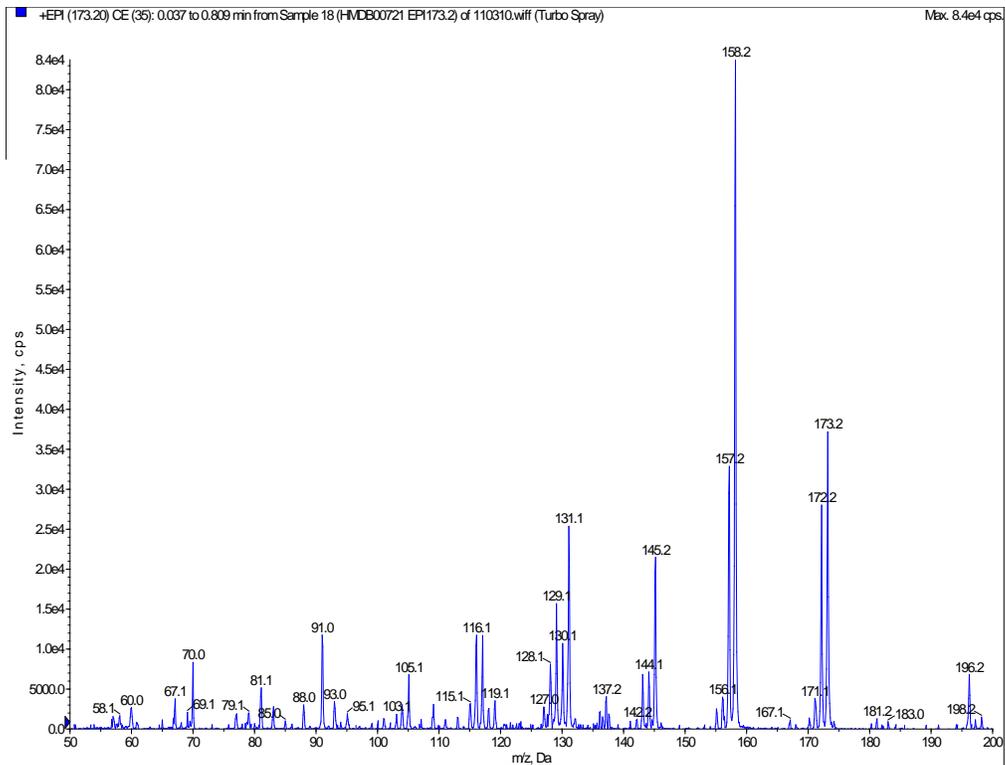




HMDB00721

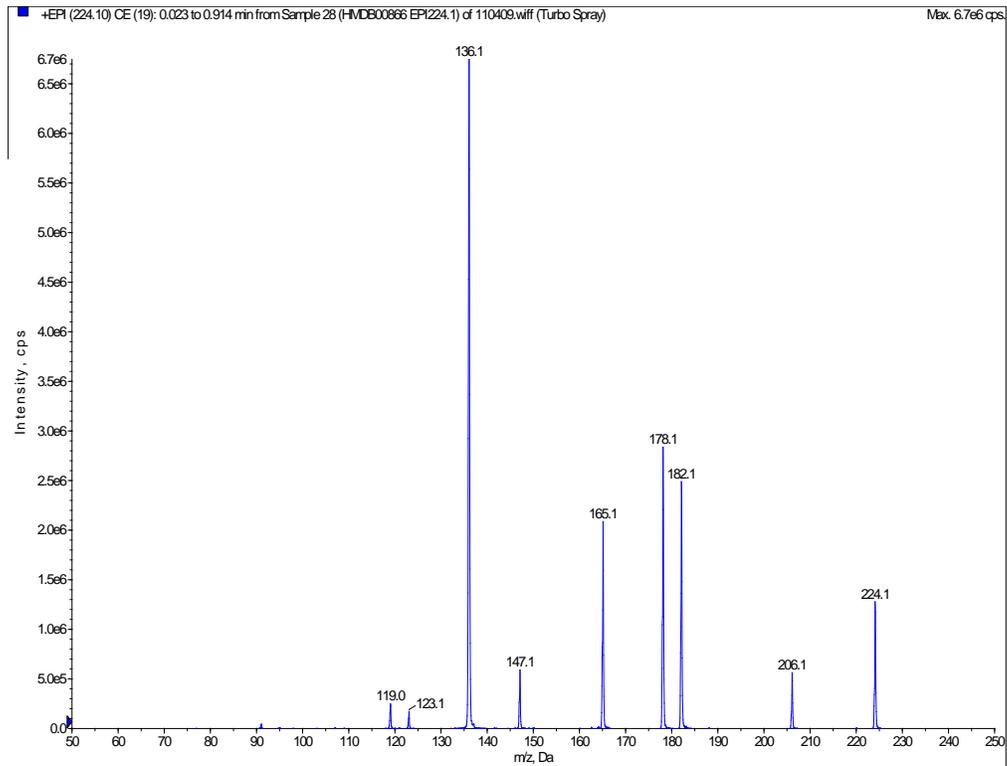
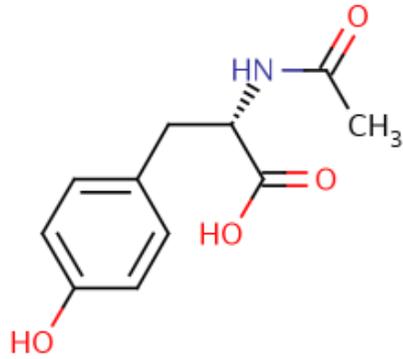
Glycylproline (1-(2-aminoacetyl)pyrrolidine-2-carboxylic acid)

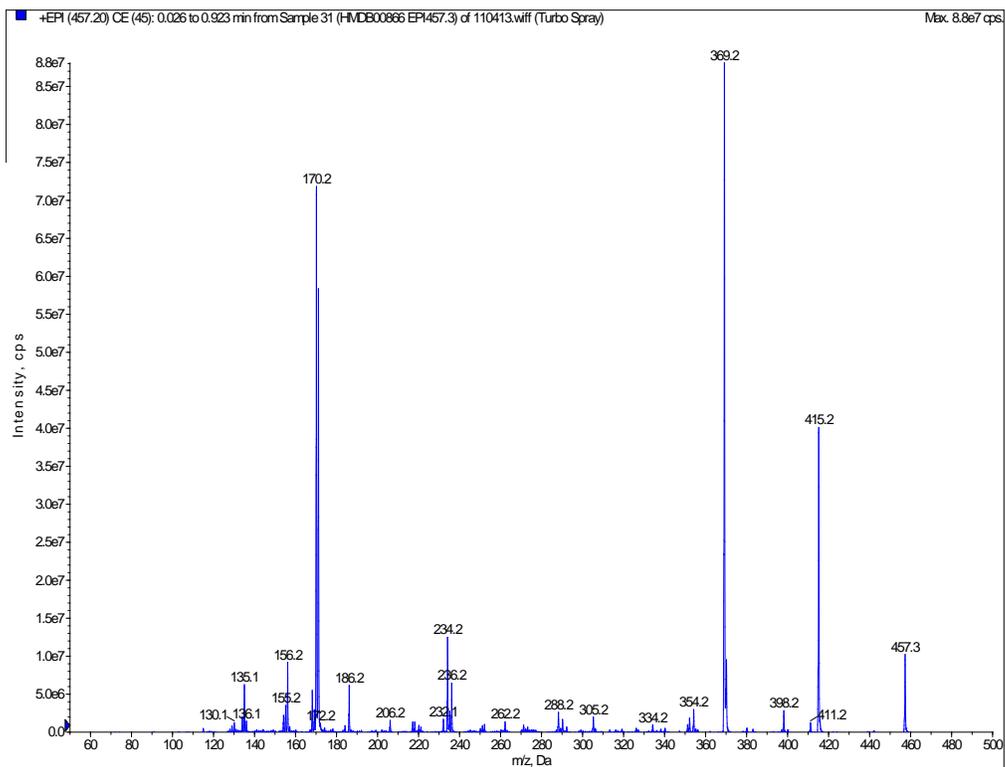
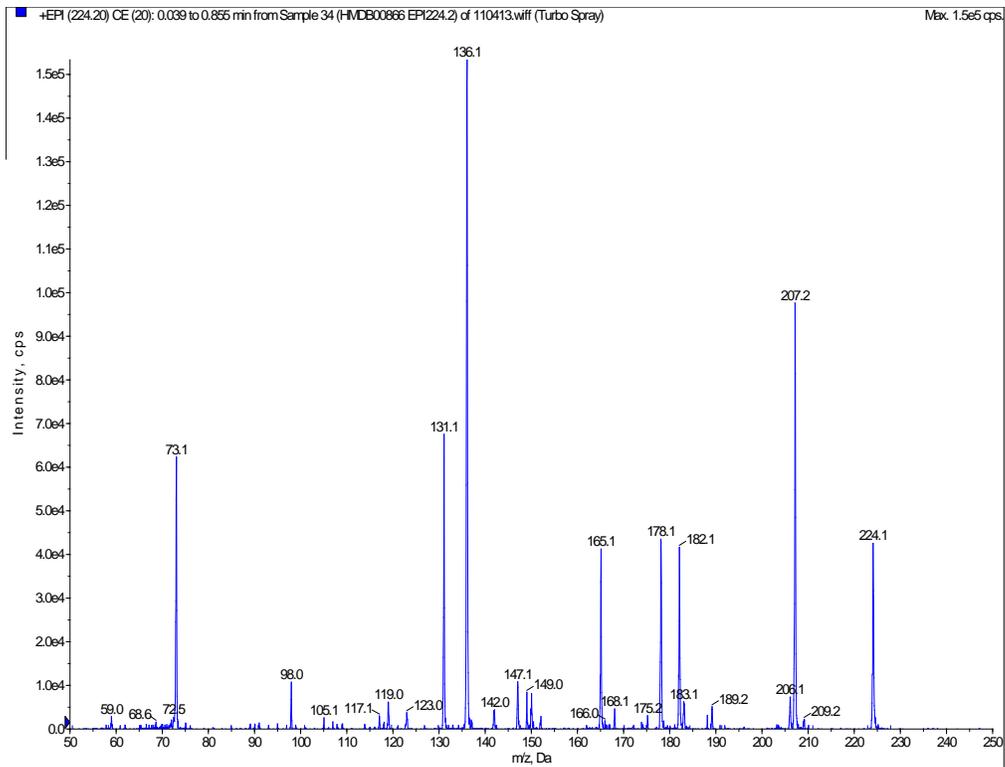




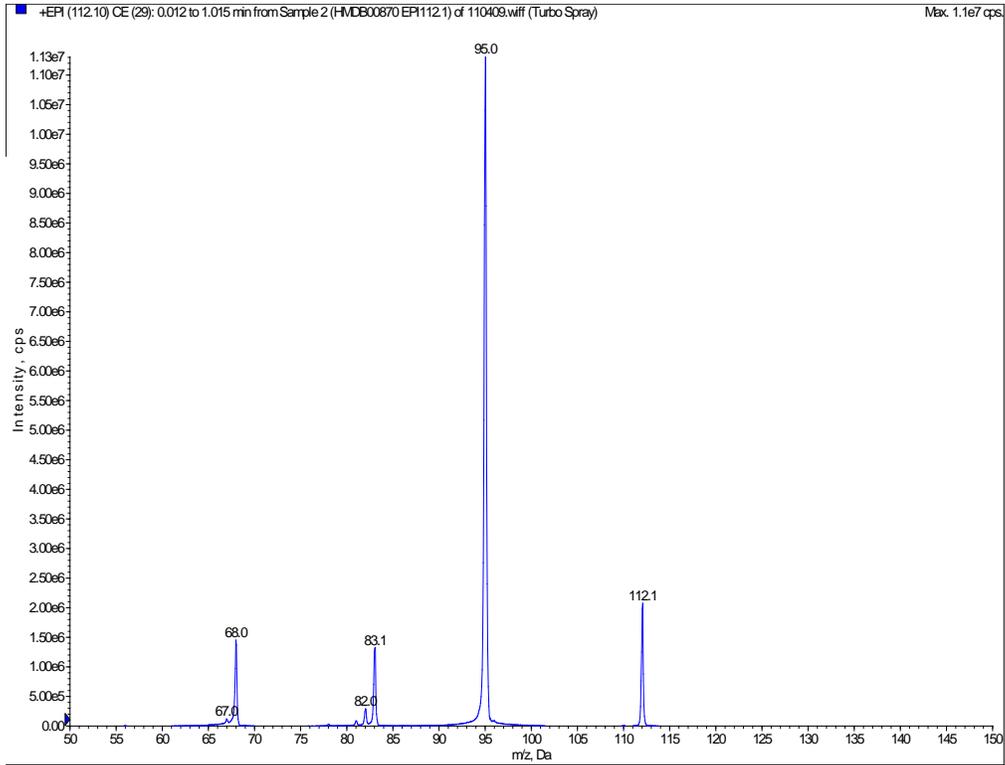
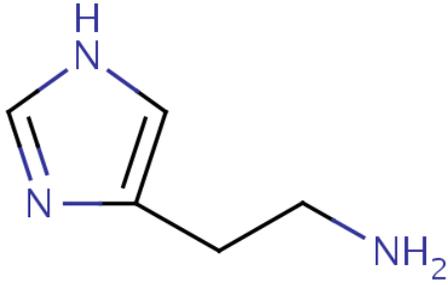
HMDB00866

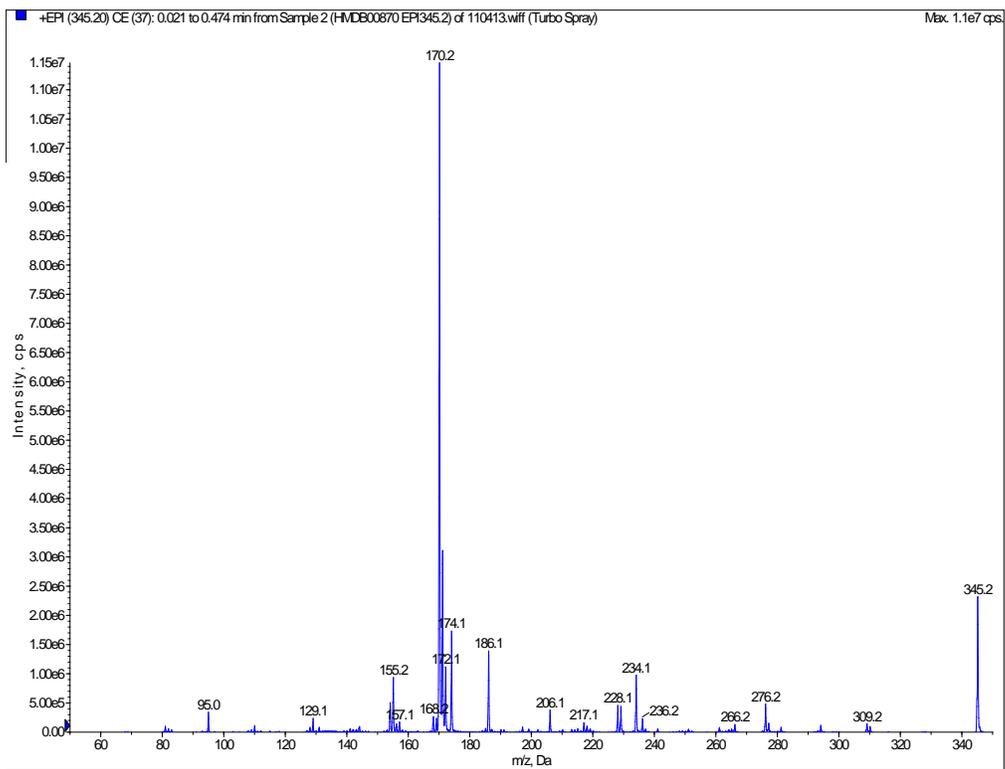
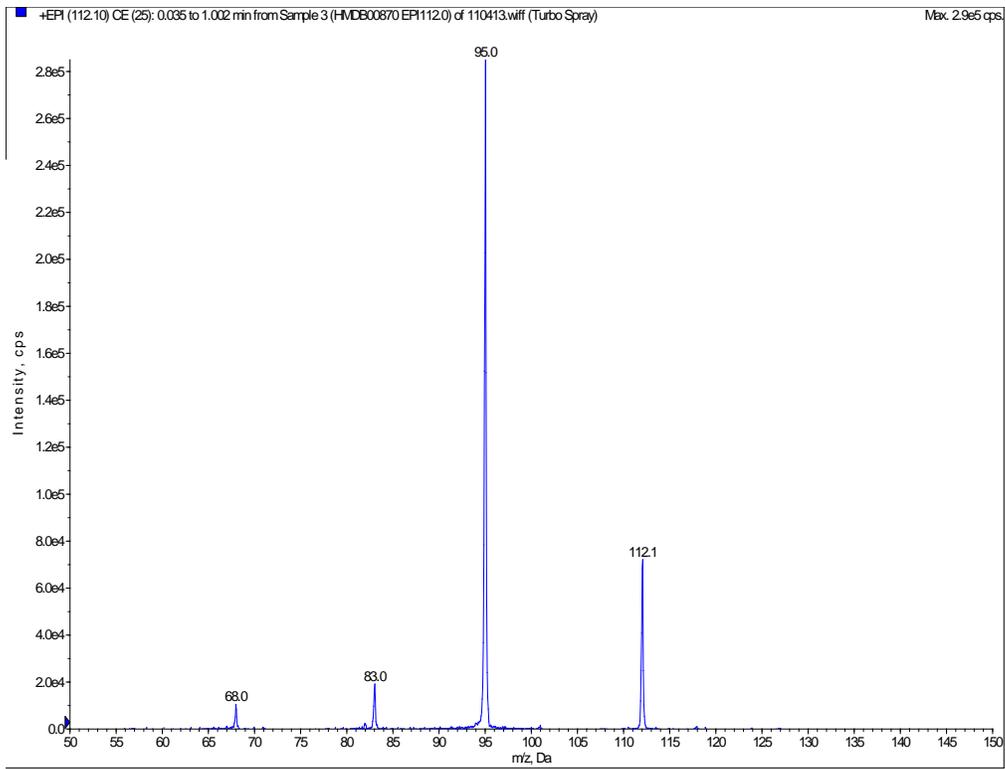
N-Acetyl-L-tyrosine (2-acetylamino-3-(4-hydroxyphenyl)-propanoic acid)





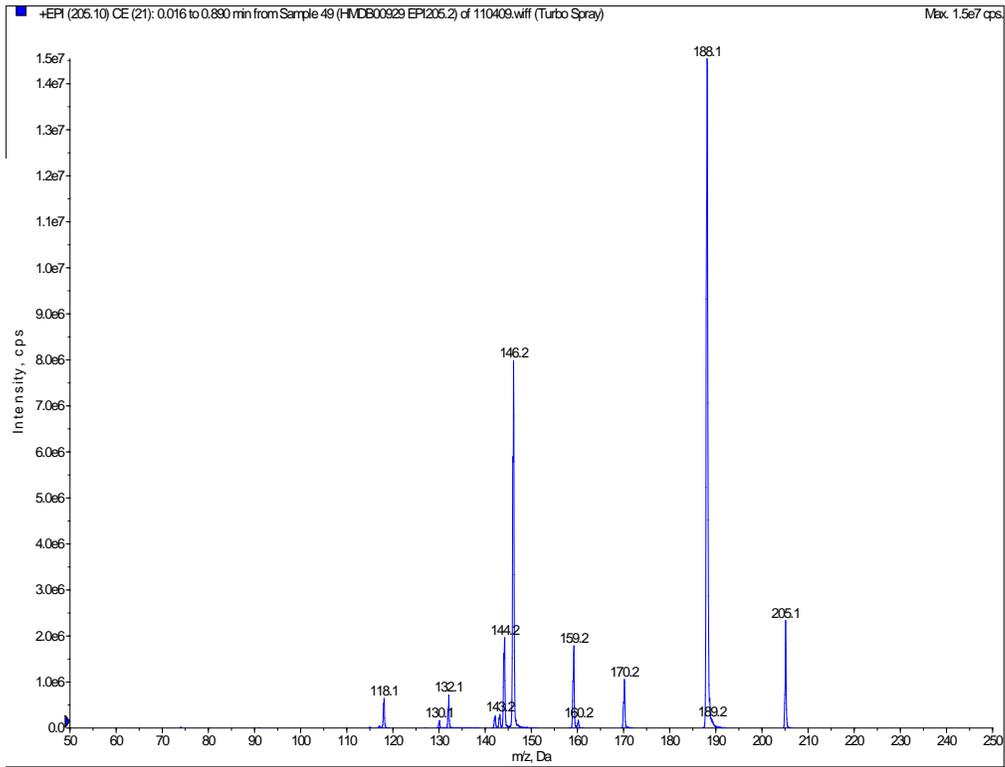
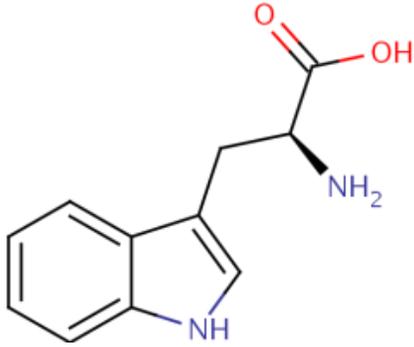
HMDB00870
Histamine (2-(3H-imidazol-4-yl)ethanamine)

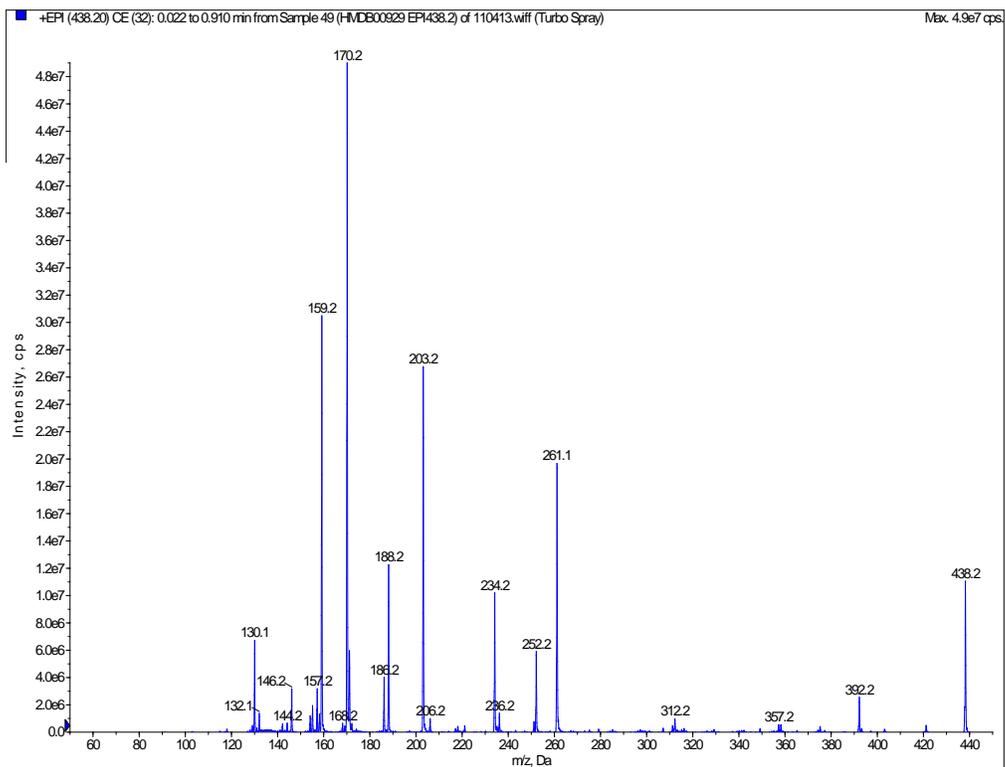
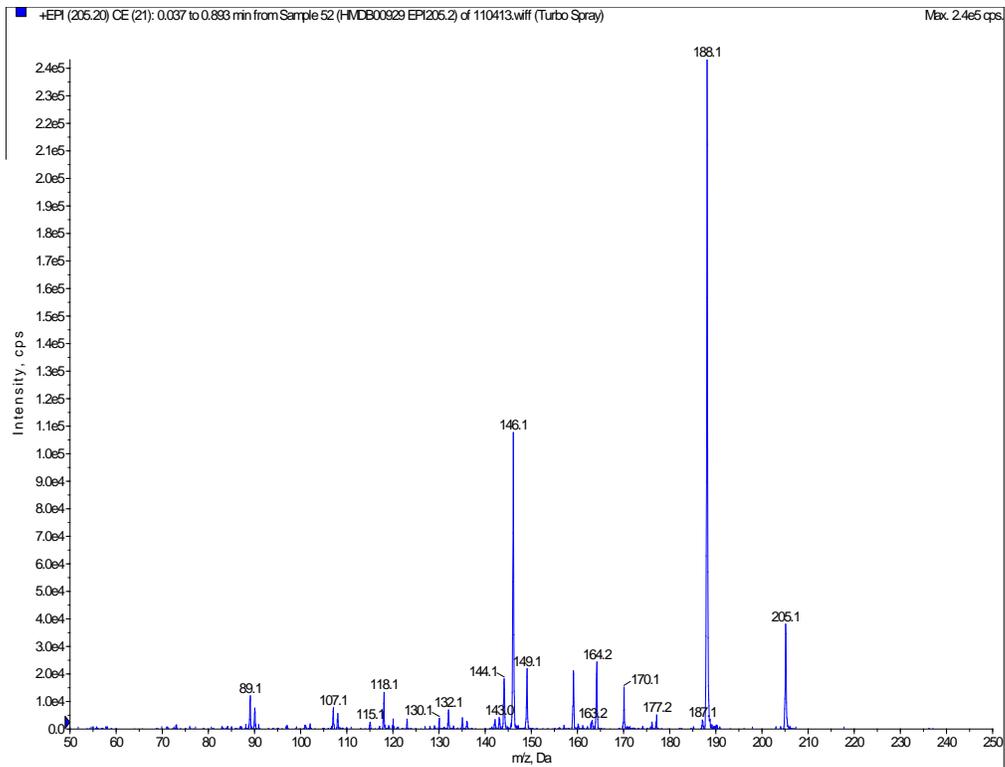




HMDB00929

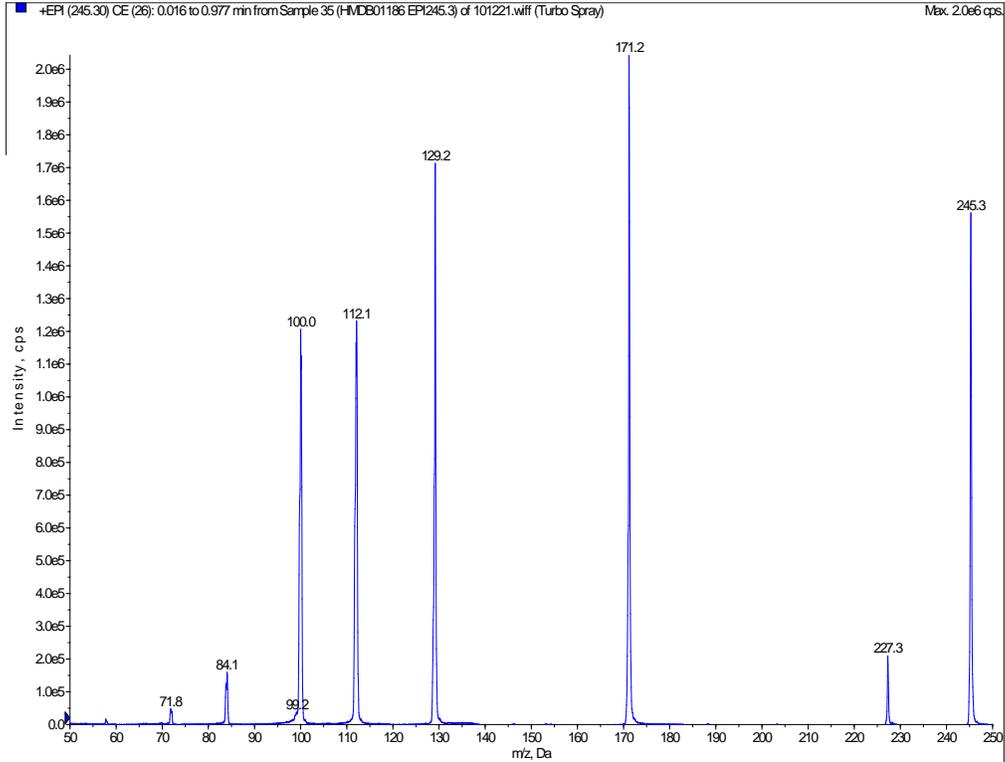
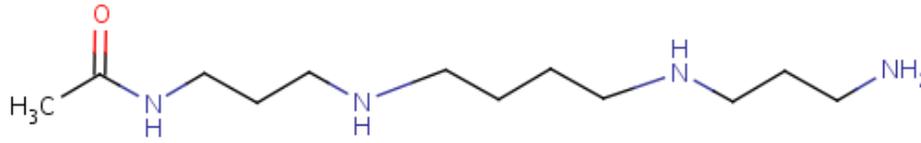
L-Tryptophan ((2S)-2-amino-3-(1H-indol-3-yl)propanoic acid)

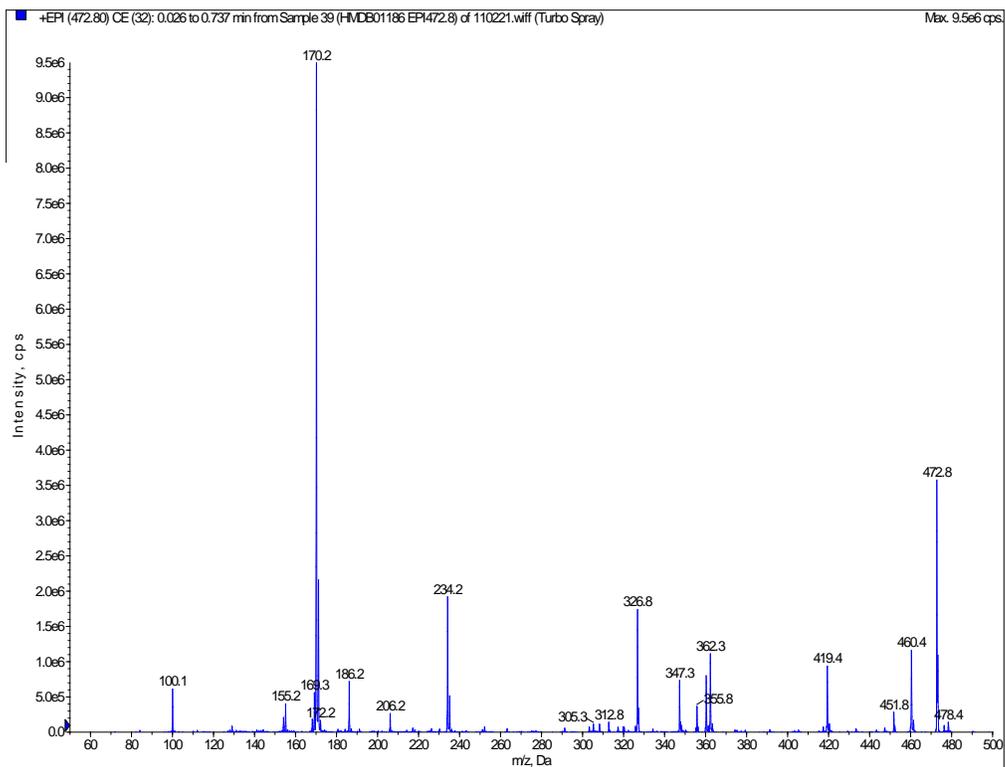
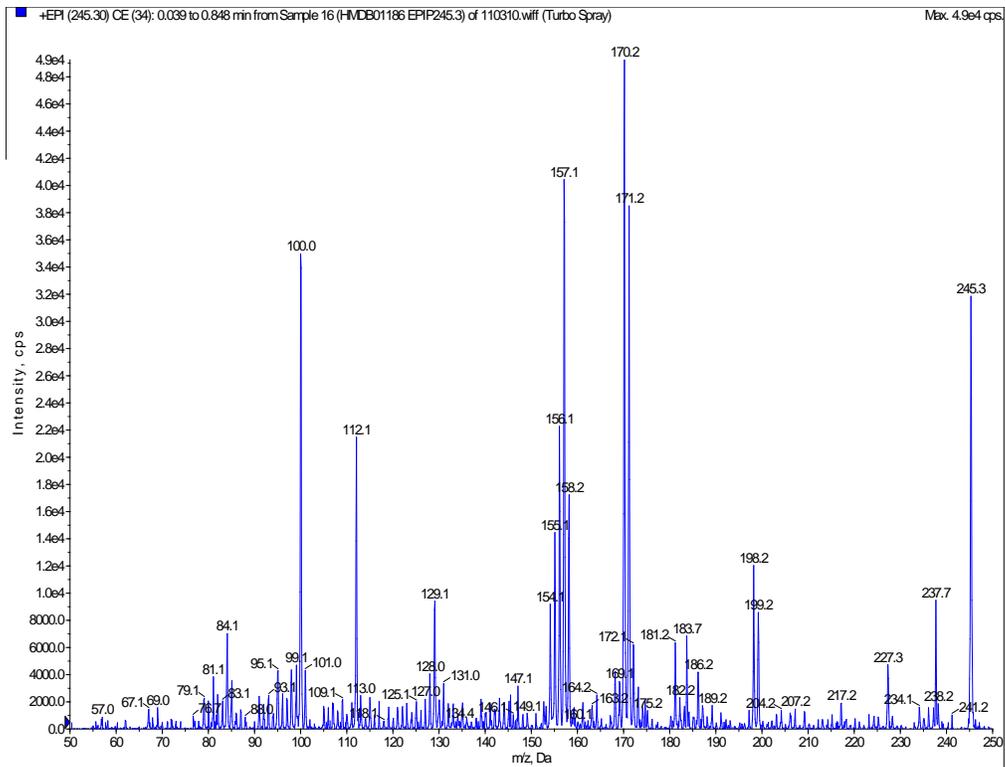




HMDB01186

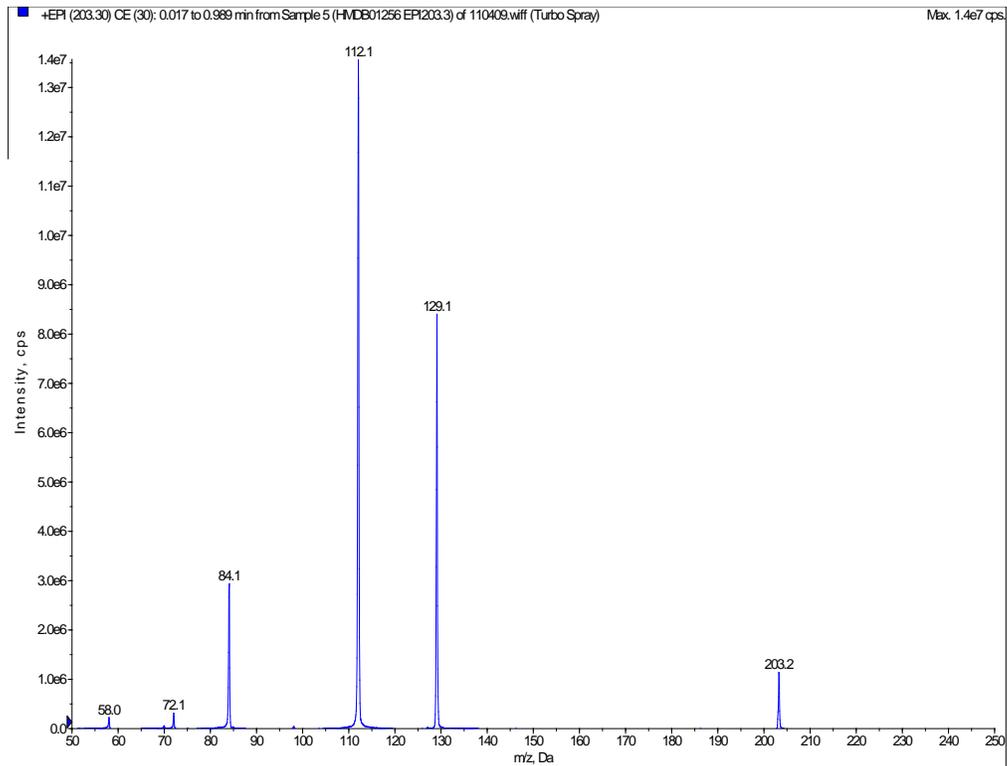
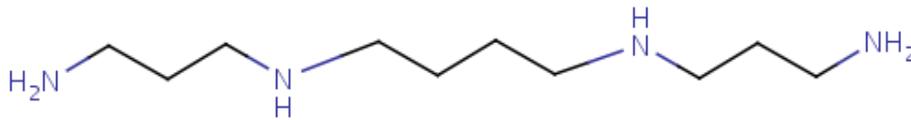
N1-Acetylspermine (N-[3-[4-(3-aminopropylamino)butylamino]propyl]acetamide)

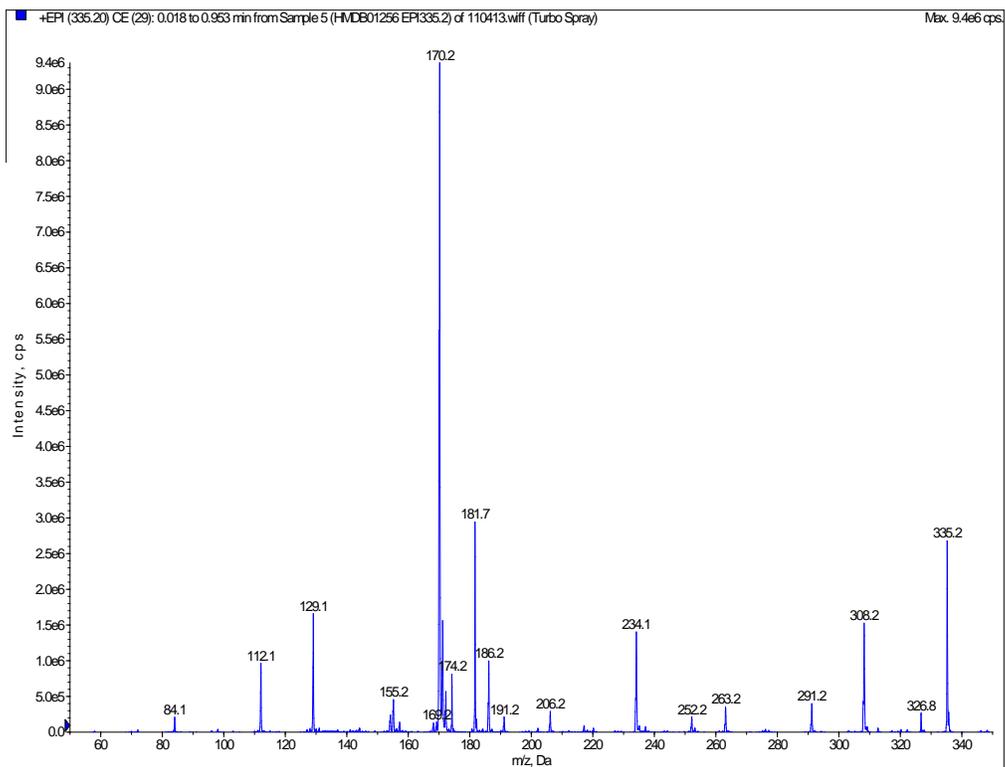
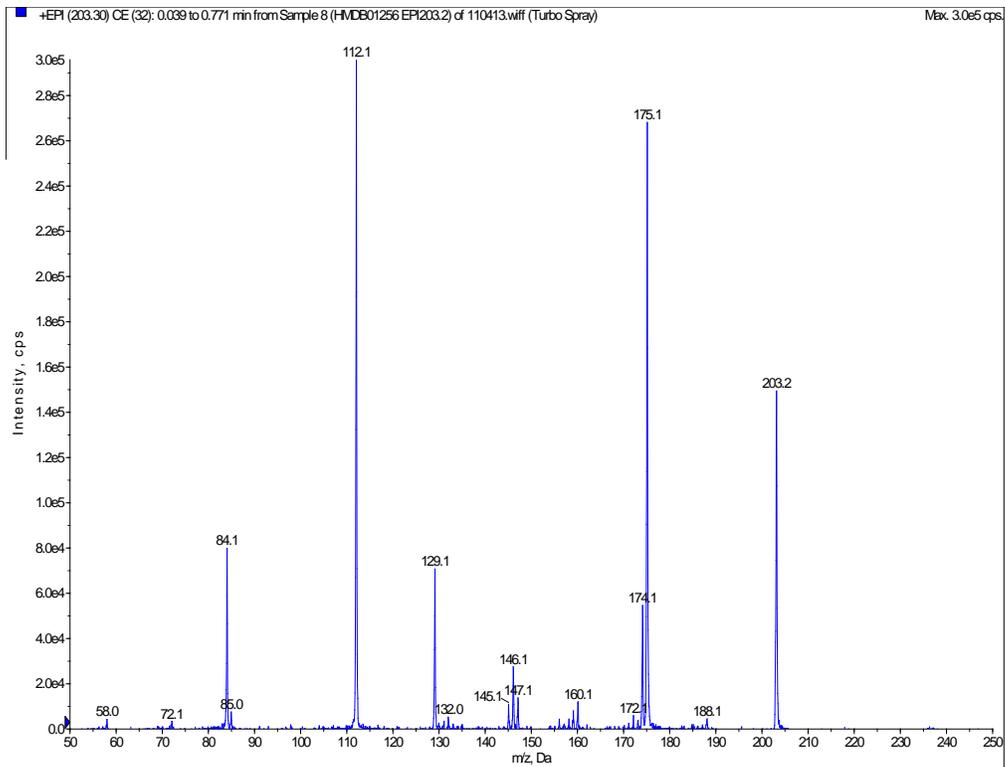




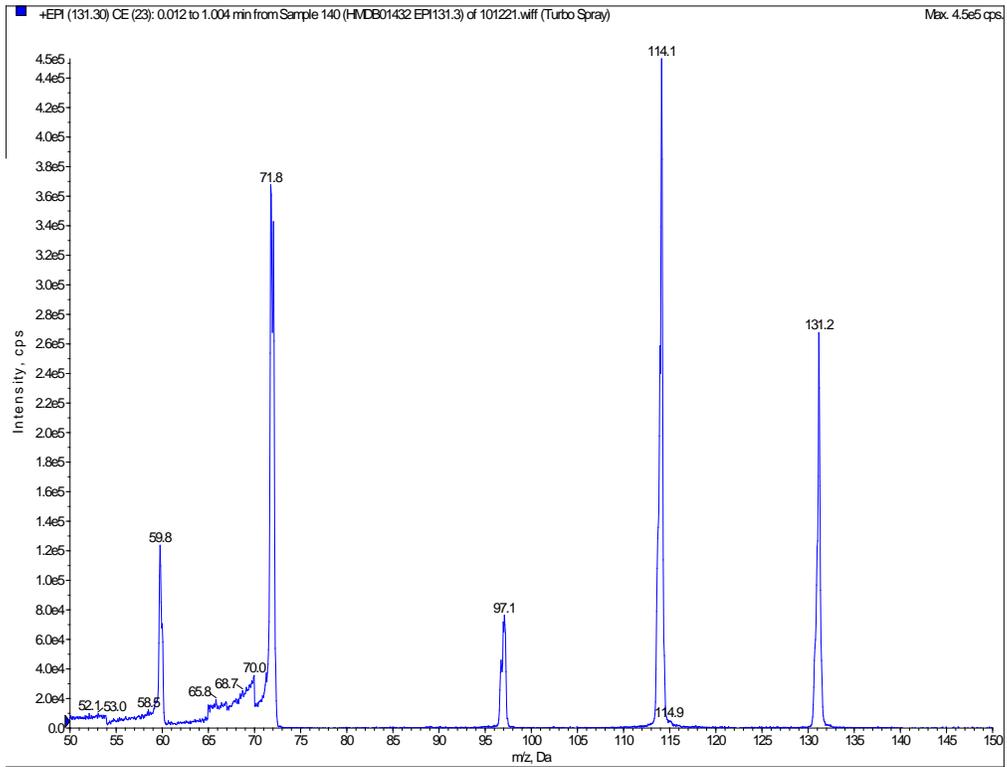
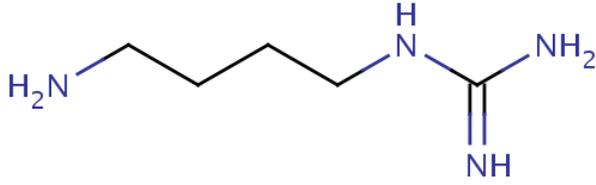
HMDB01256

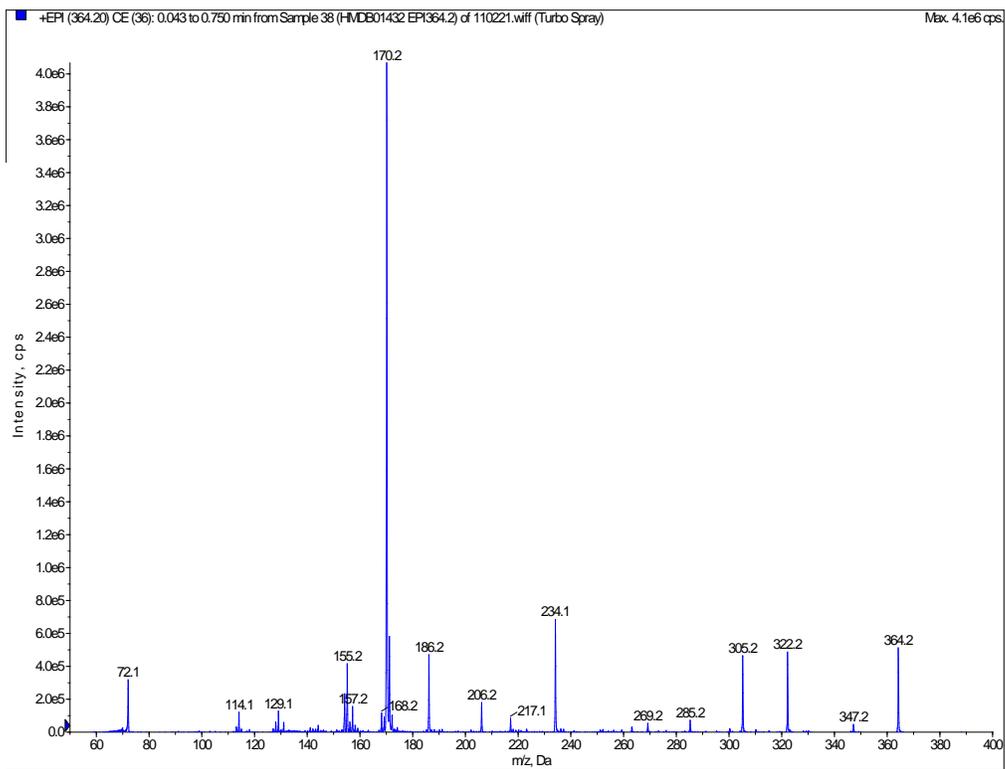
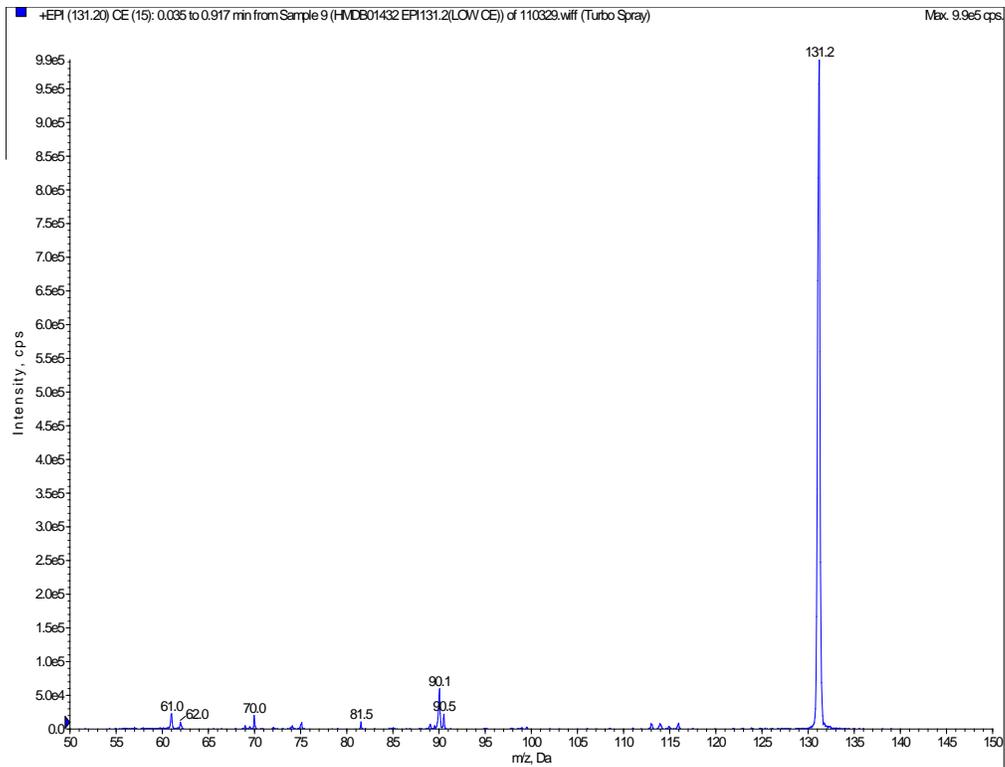
Spermine (N,N'-bis(3-aminopropyl)butane-1,4-diamine)



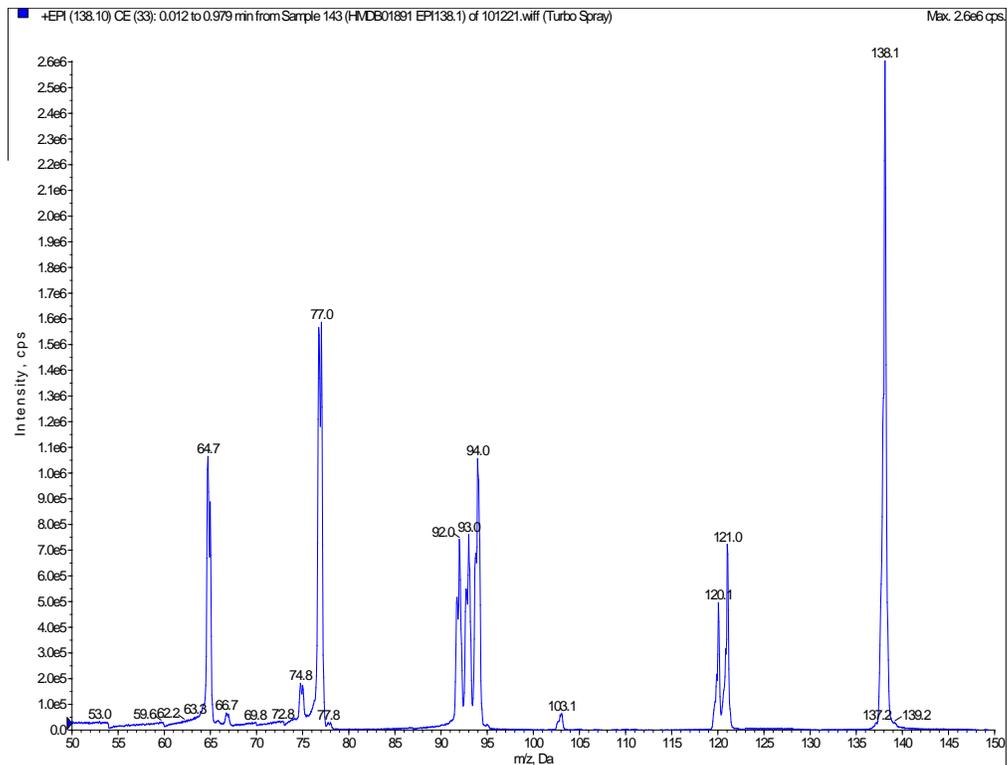
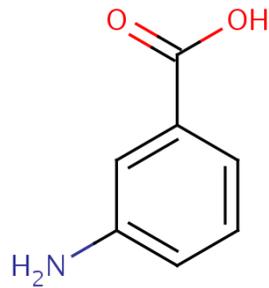


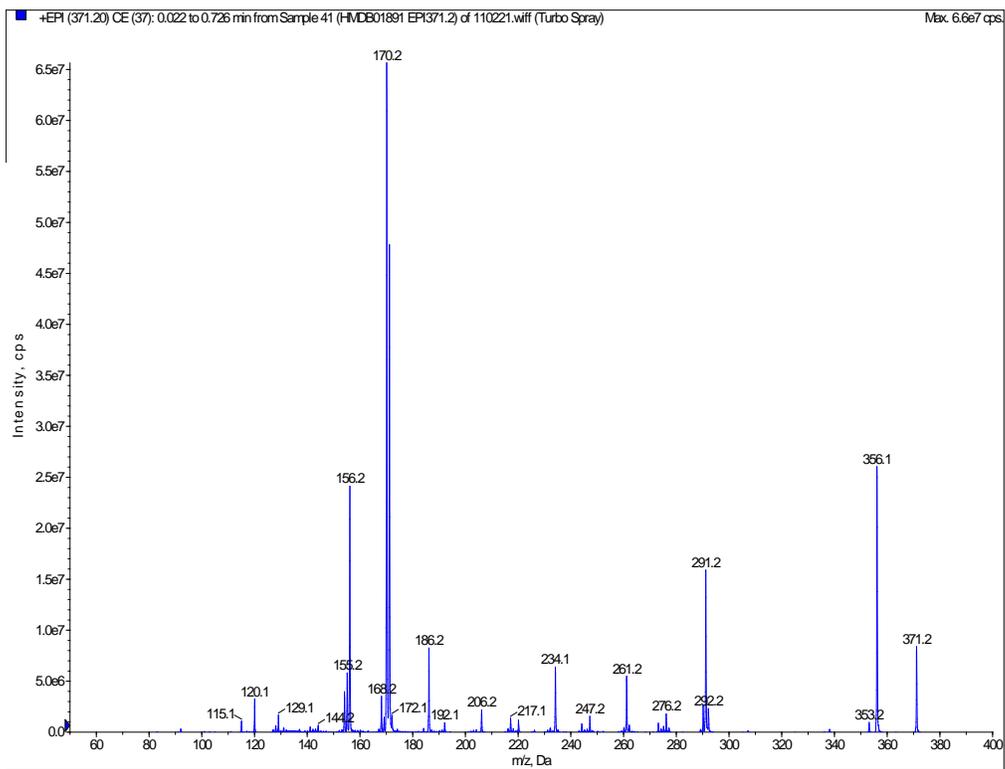
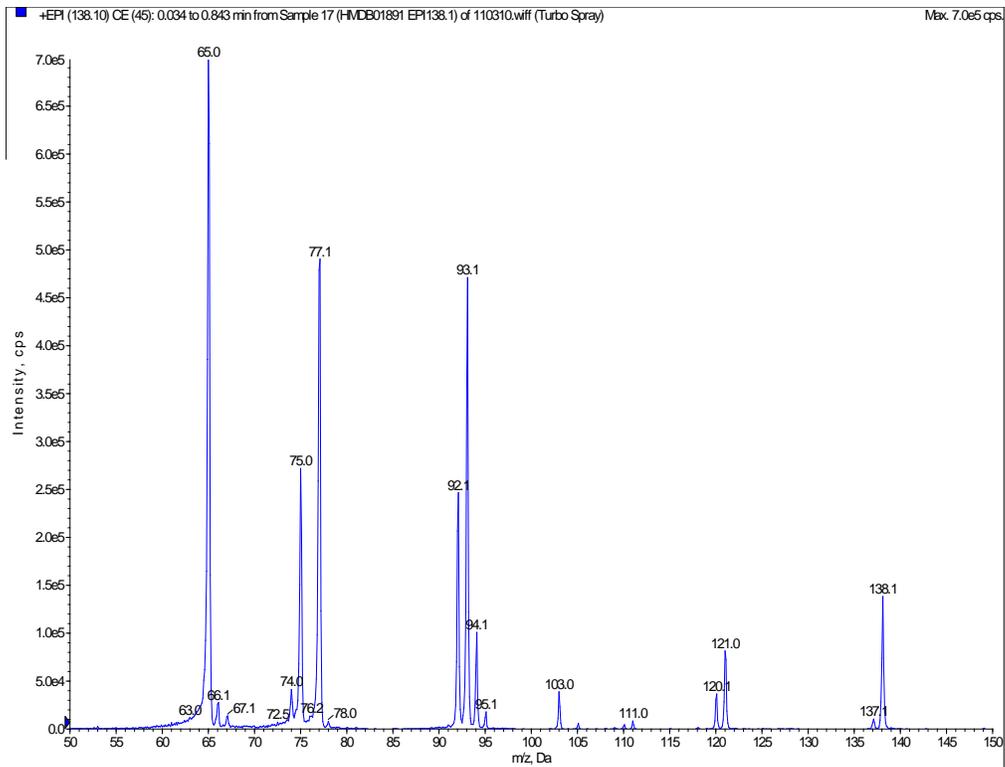
HMDB01432
Agmatine (2-(4-aminobutyl)guanidine)





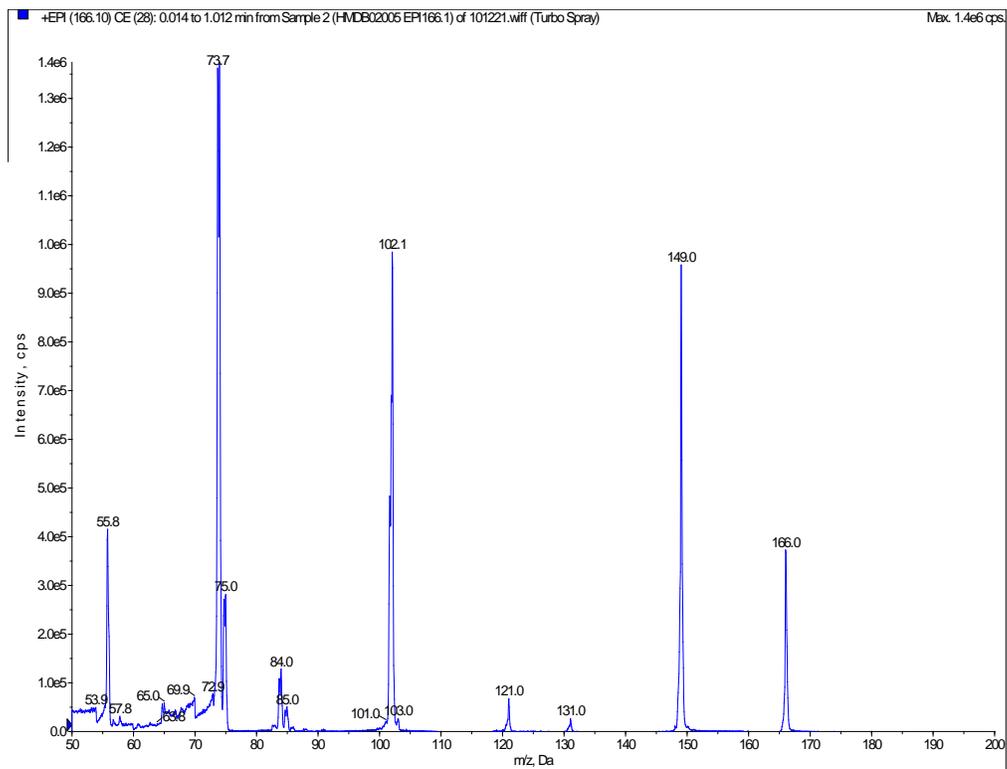
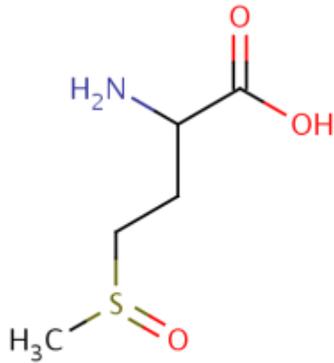
HMDB01891
m-Aminobenzoic acid (3-aminobenzoic acid)

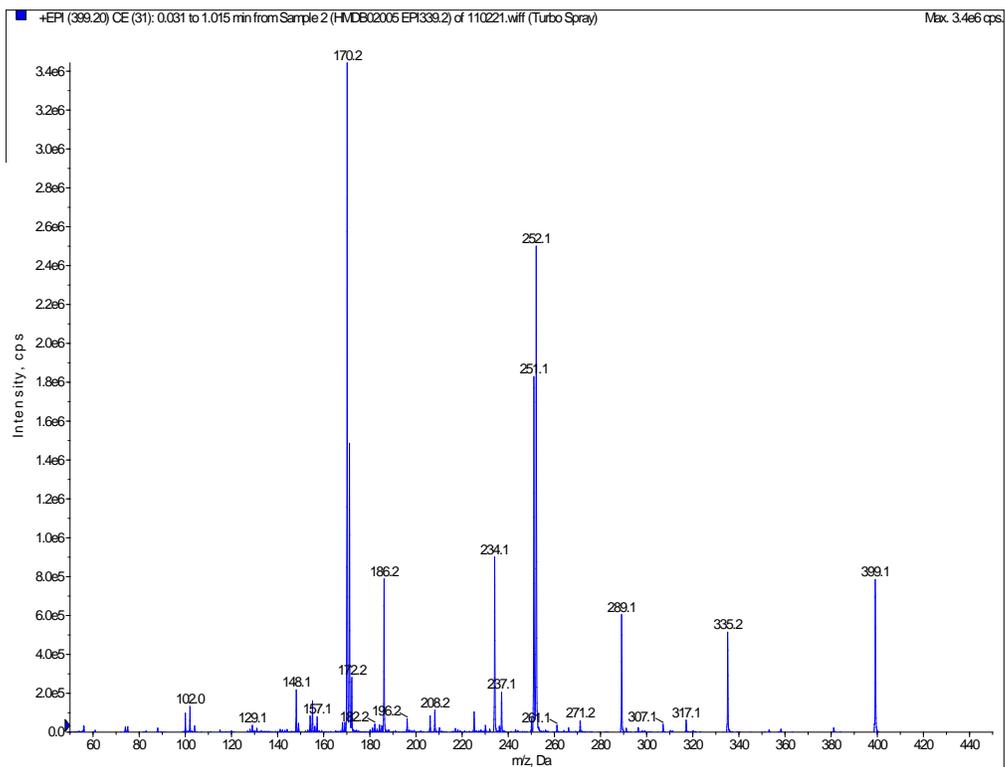
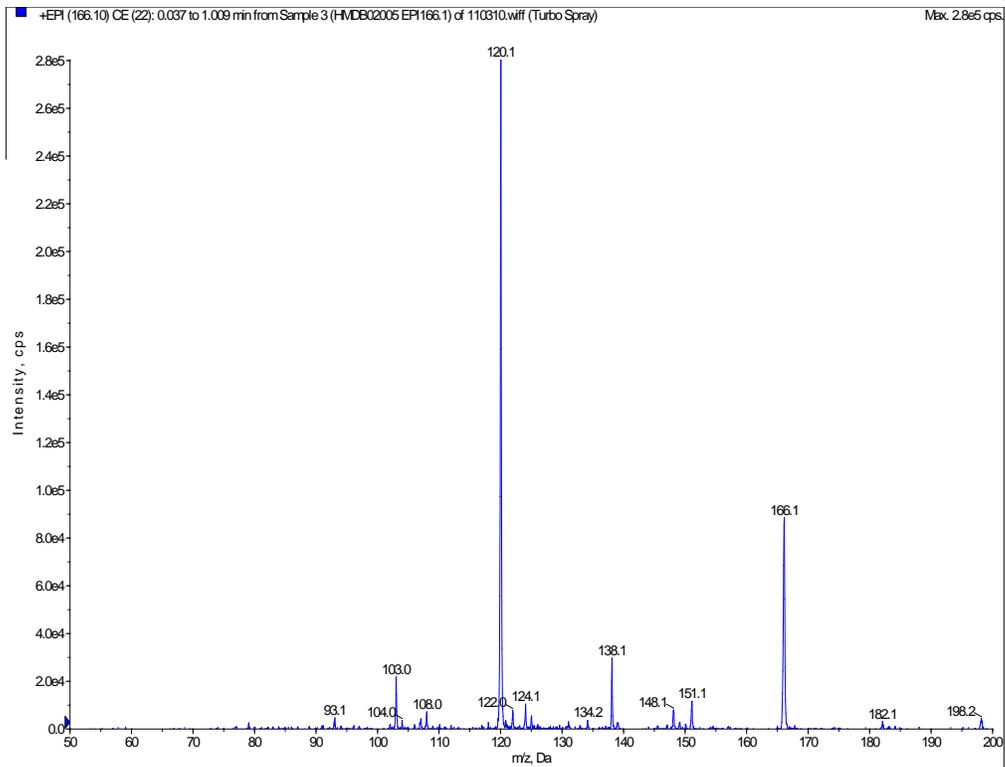




HMDB02005

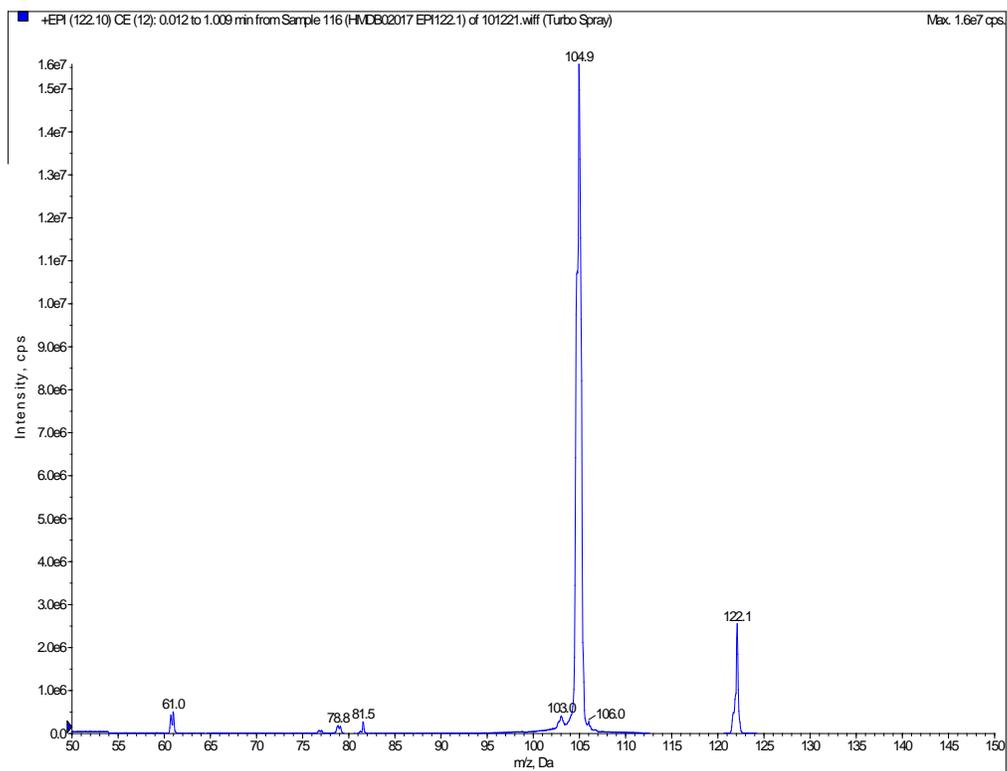
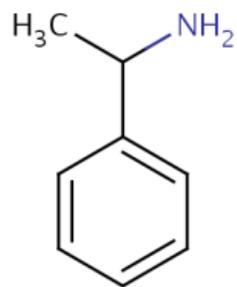
Methionine sulfoxide (2-amino-4-methylsulfinyl-butanoic acid)

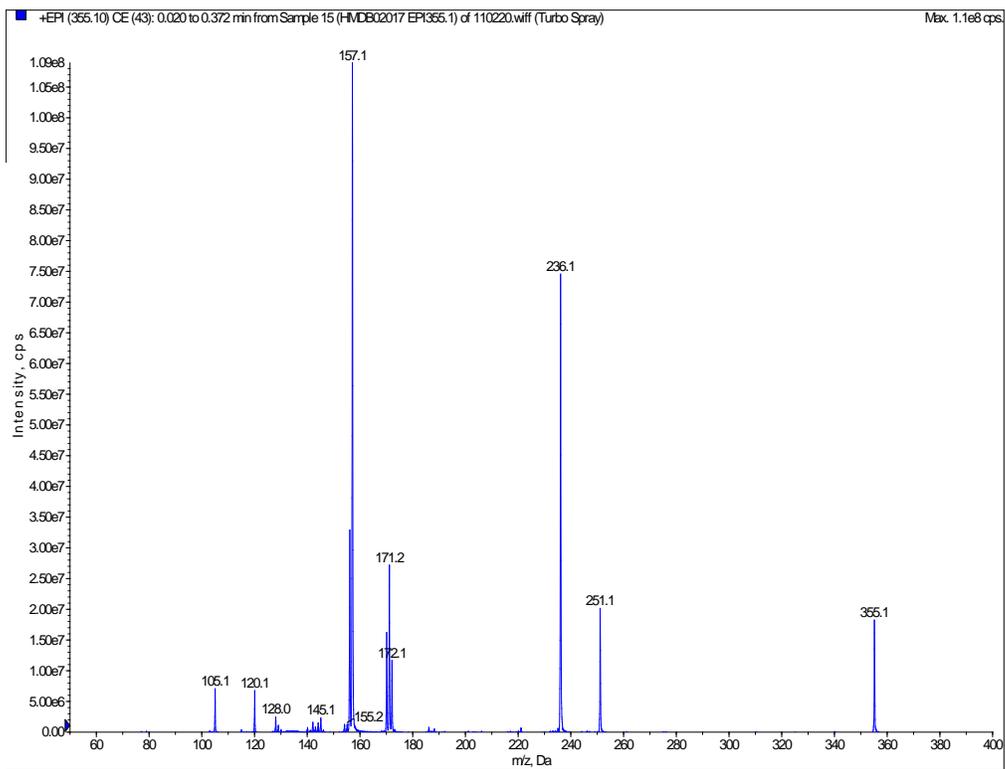
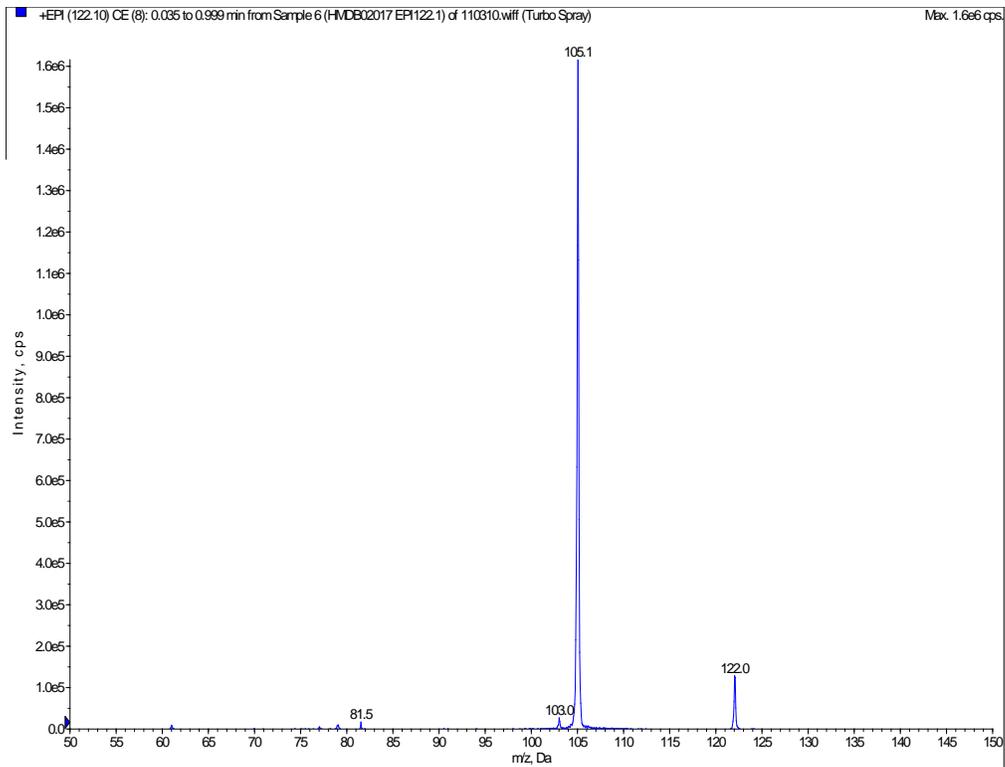




HMDB02017

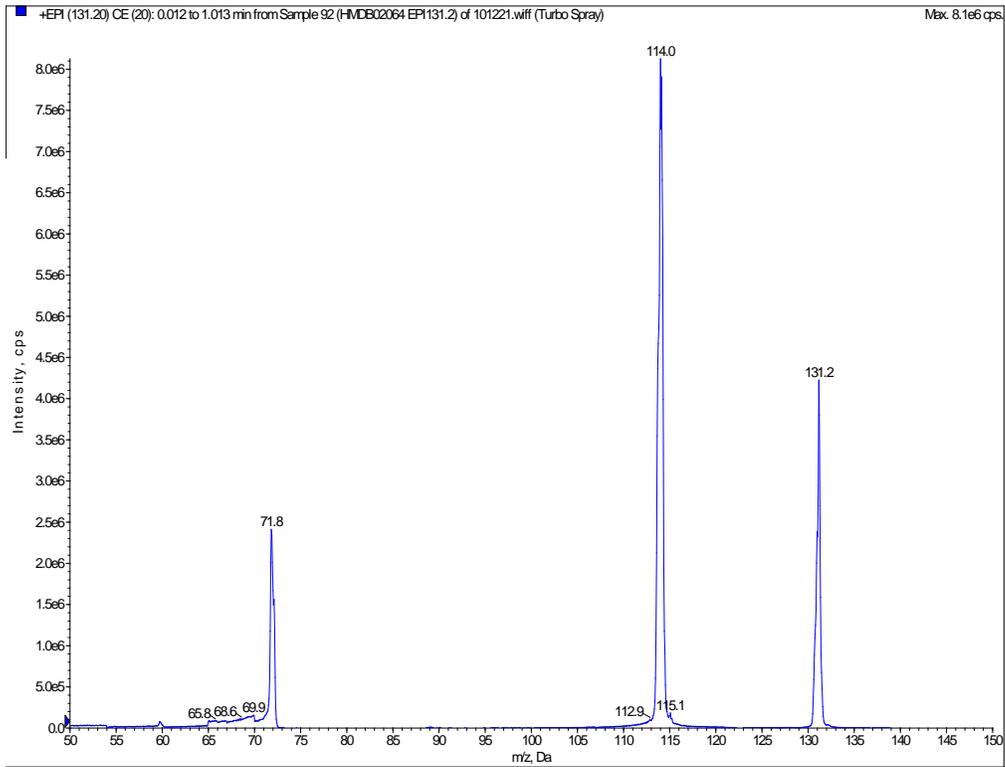
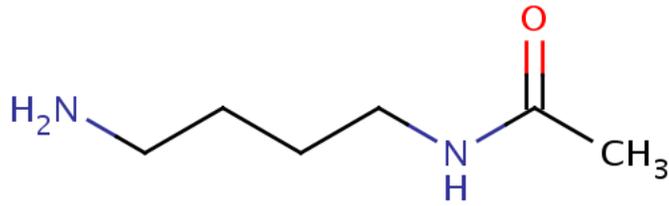
1-Phenylethylamine (1-phenylethanamine)

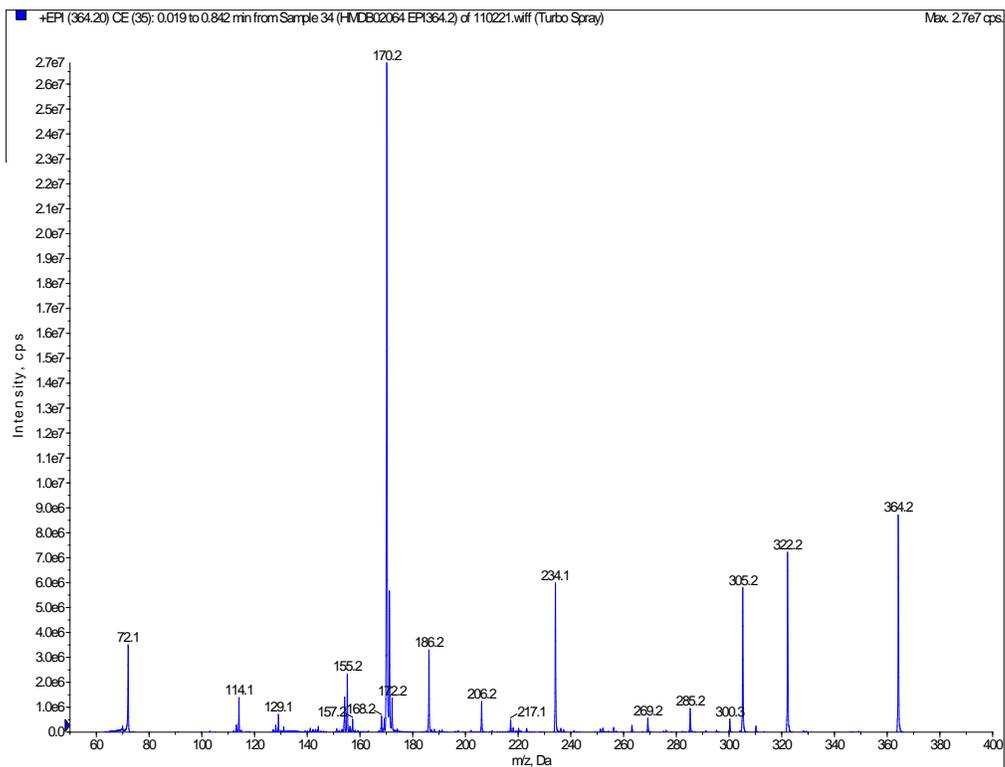
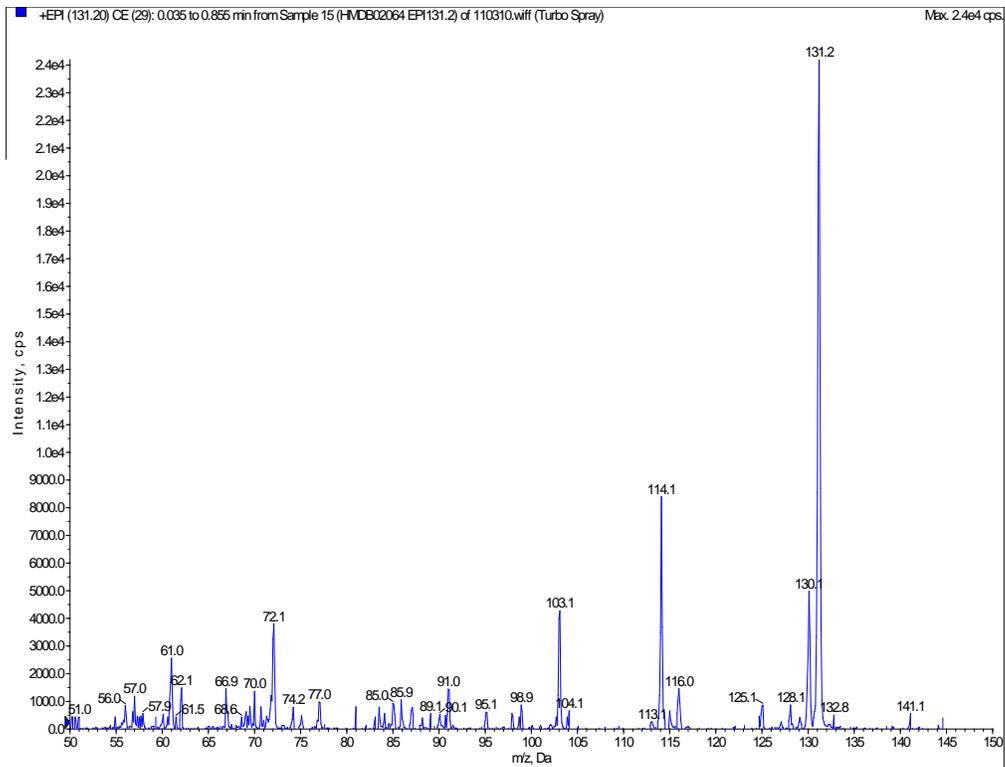




HMDB02064

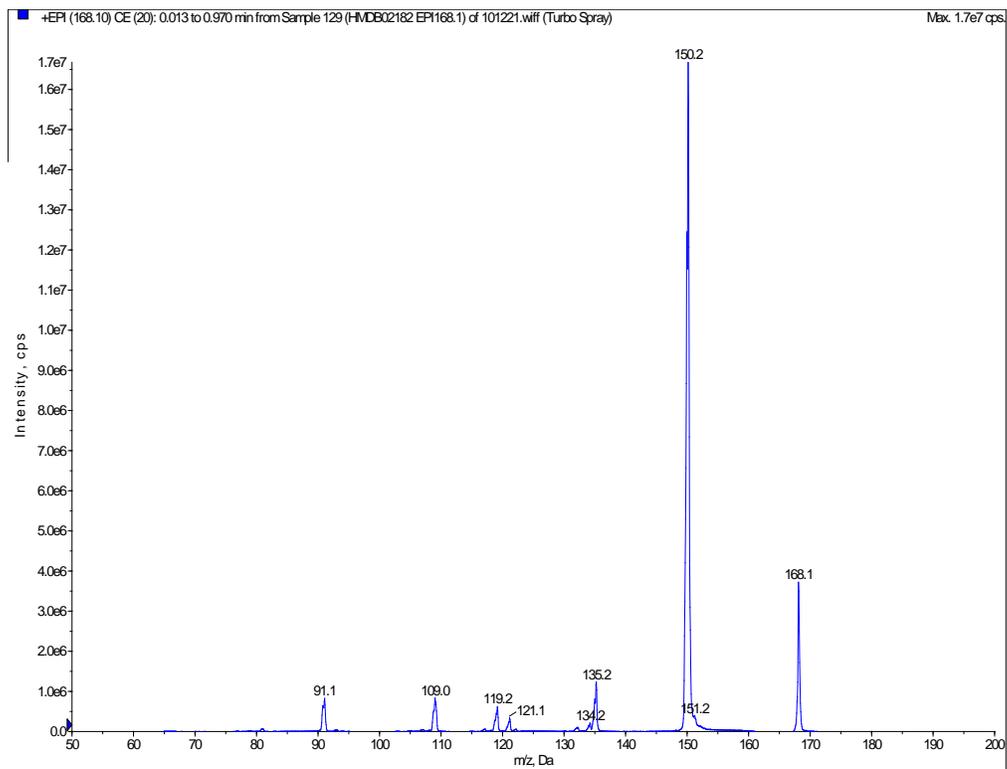
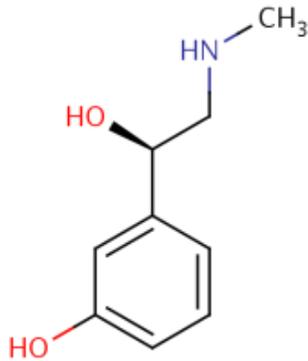
N-Acetylputrescine (N-(4-aminobutyl)acetamide)

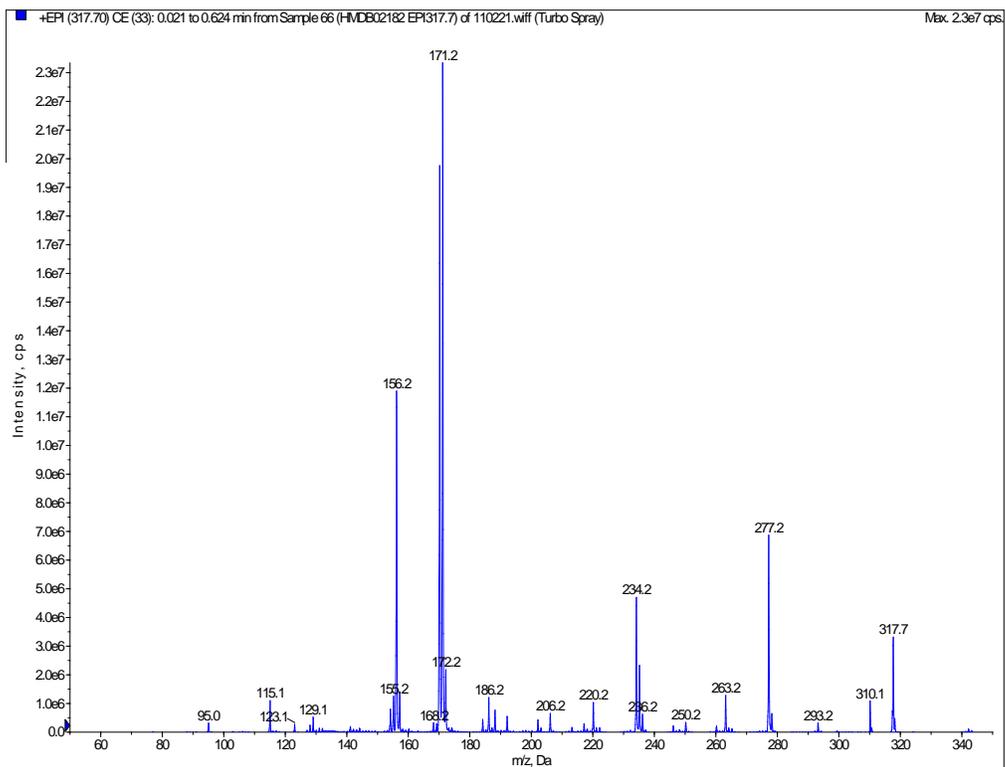
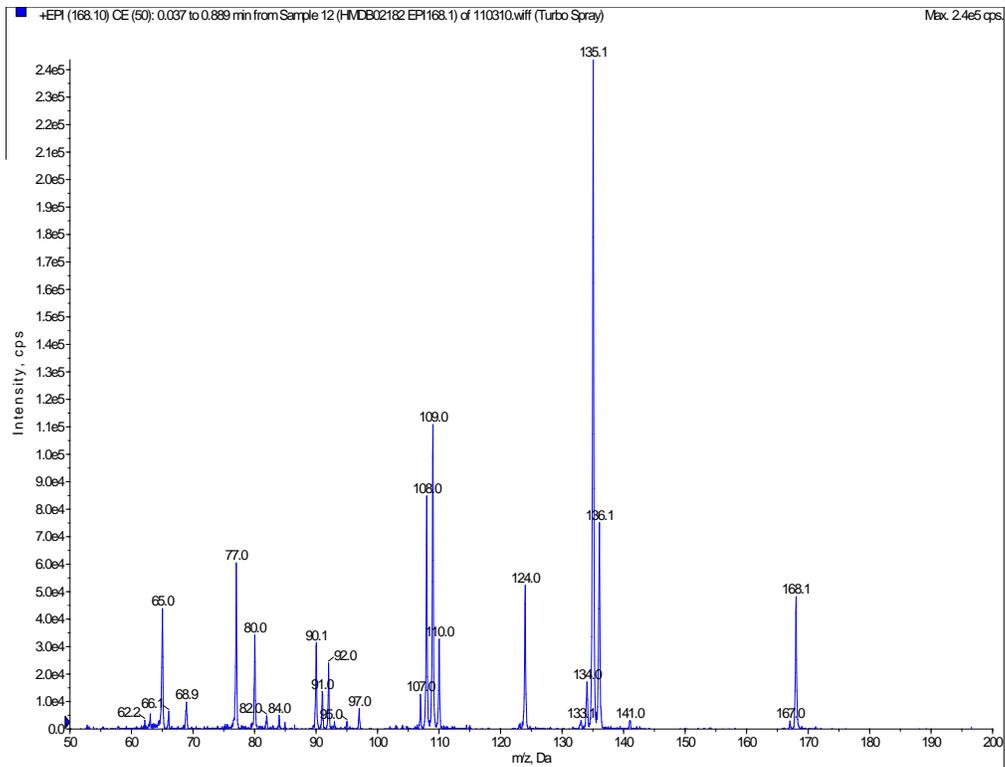




HMDB02182

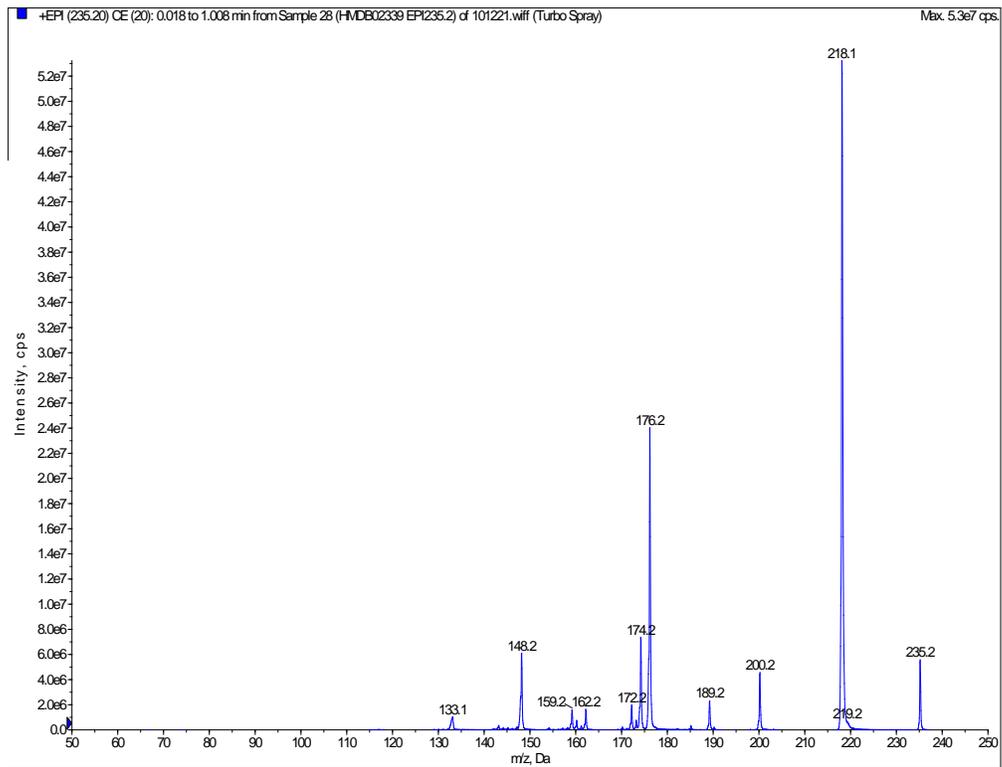
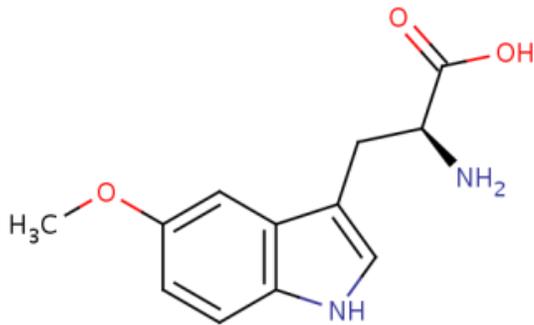
Phenylephrine (3-[(1R)-1-hydroxy-2-methylaminoethyl]phenol)

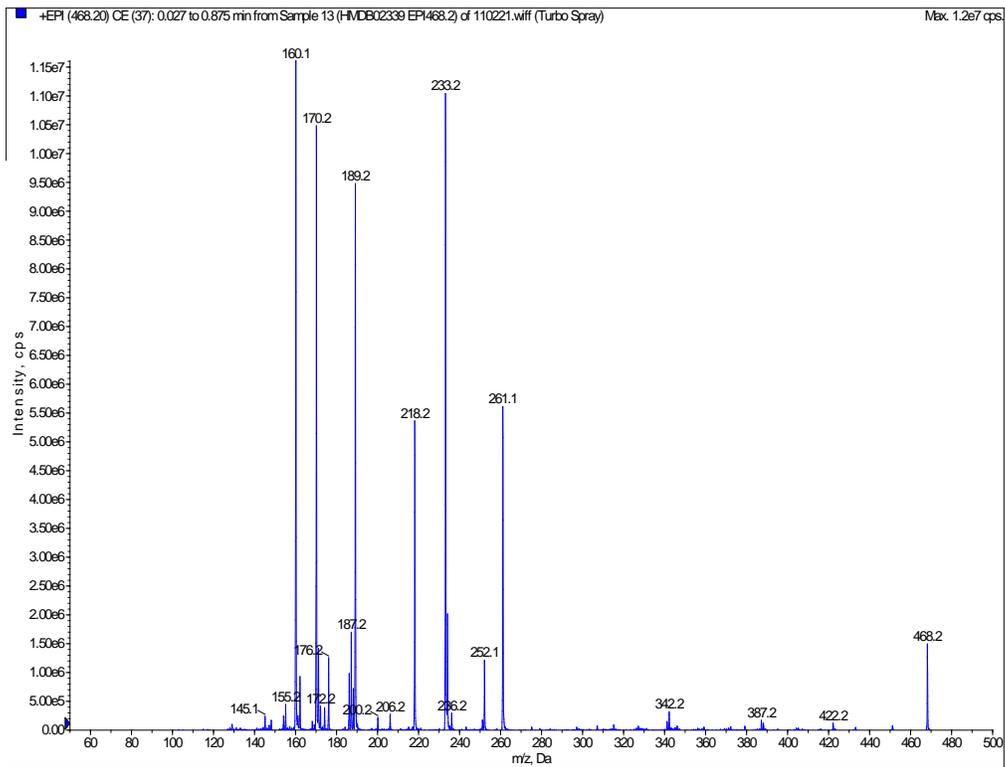
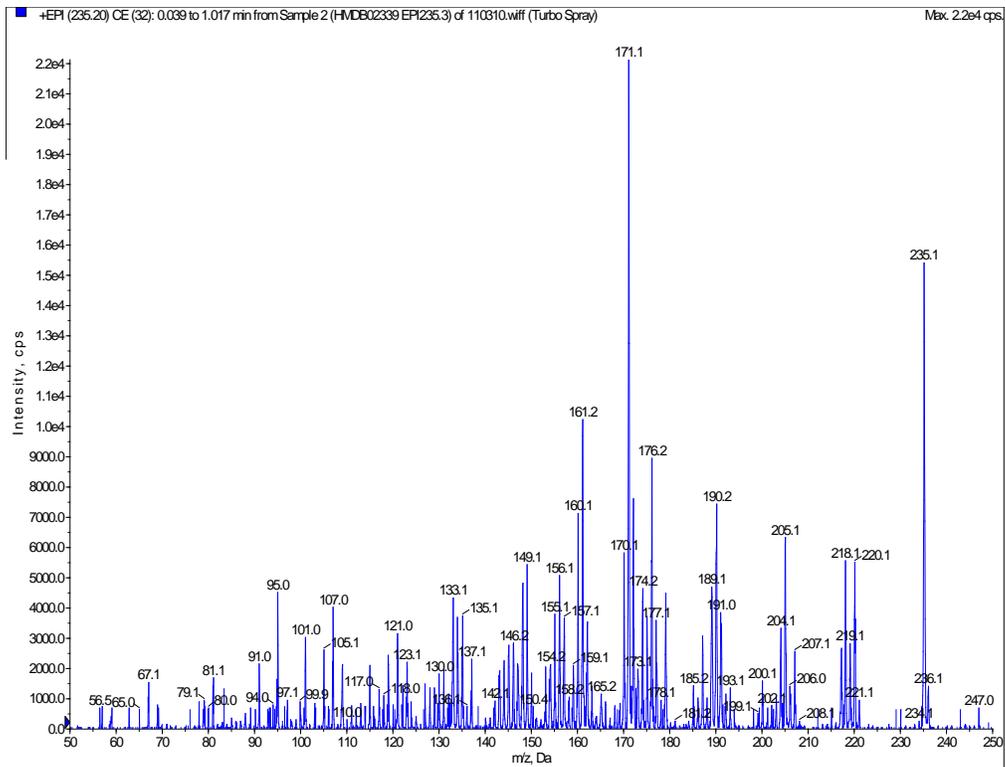




HMDB02339

5-Methoxytryptophan ((2S)-2-amino-3-(5-methoxy-1H-indol-3-yl)propanoic acid)

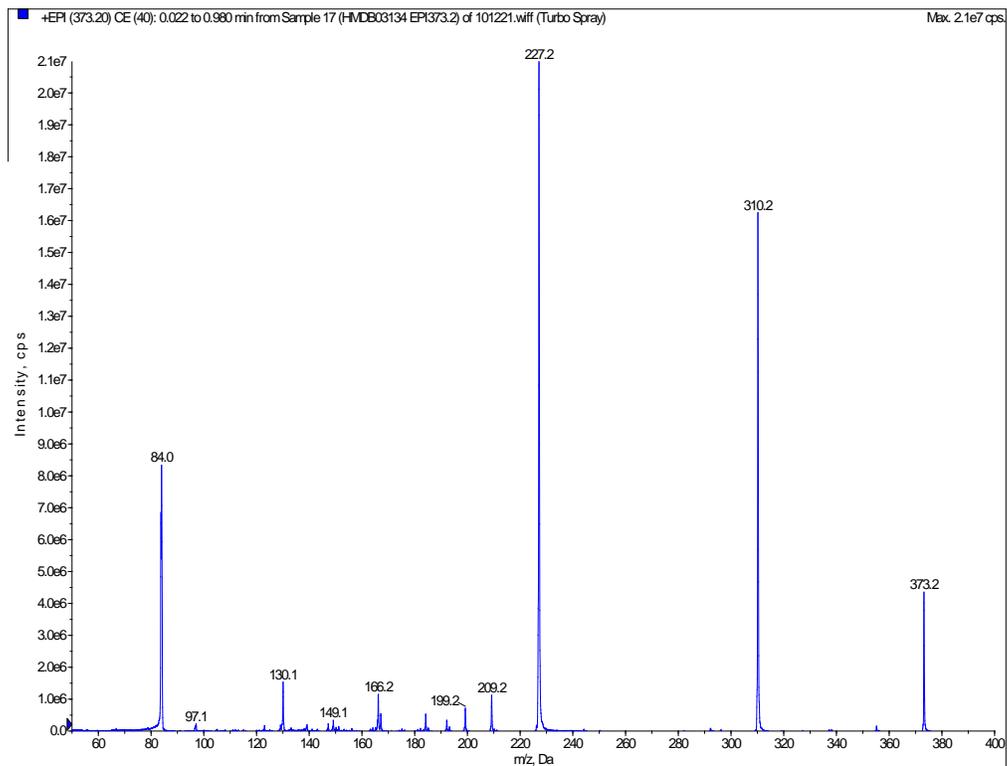
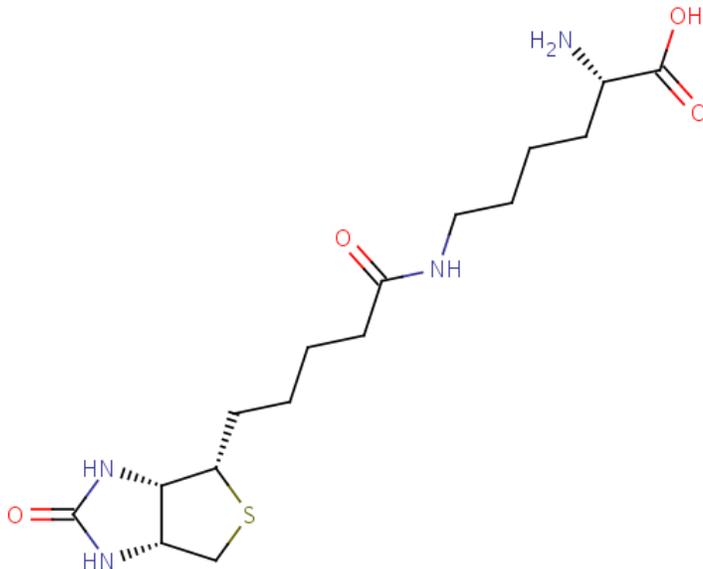


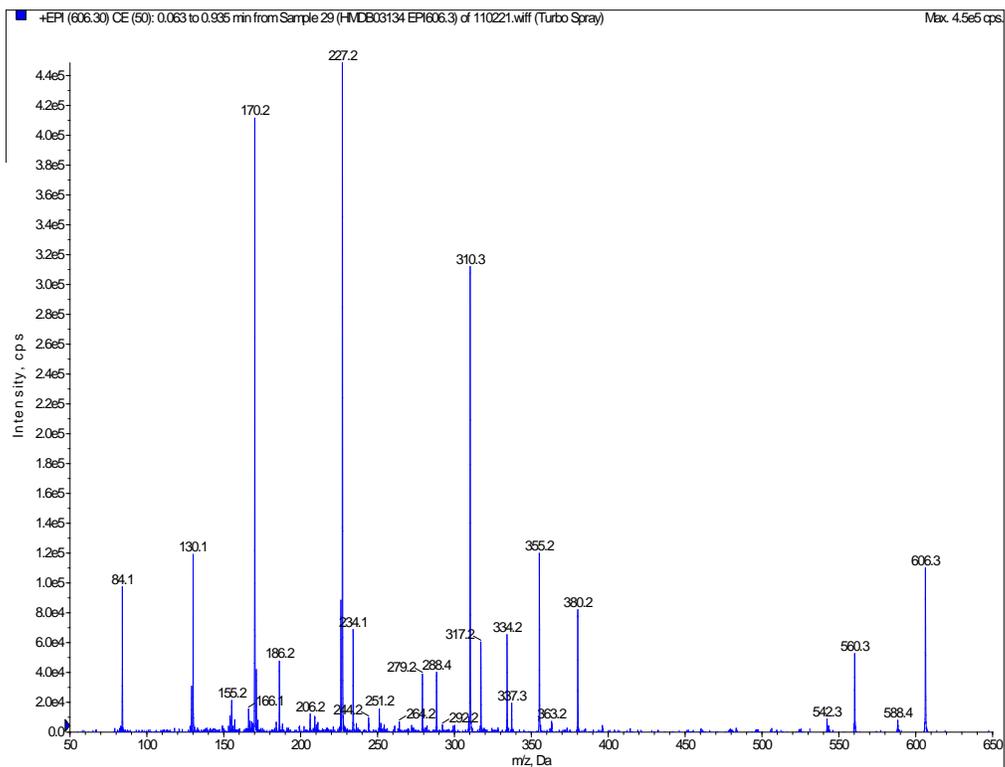
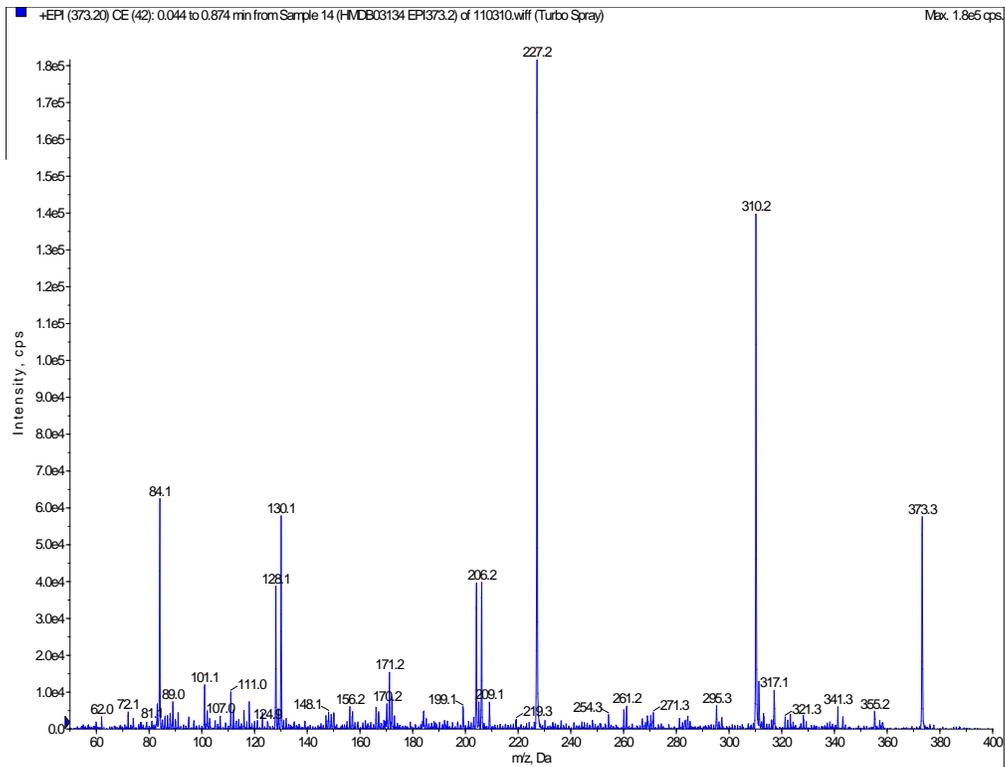


HMDB03134

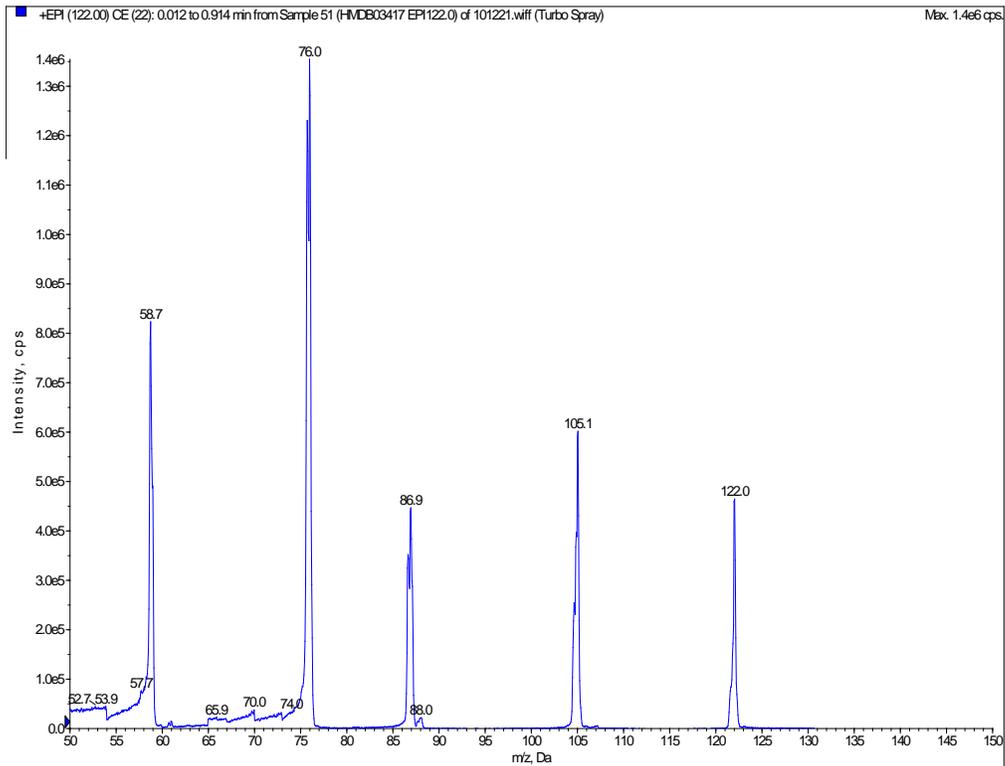
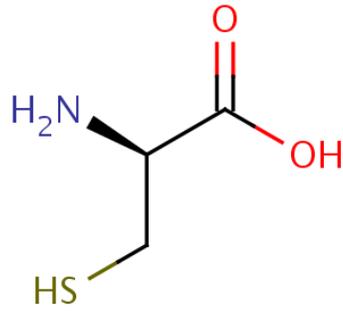
Biocytin

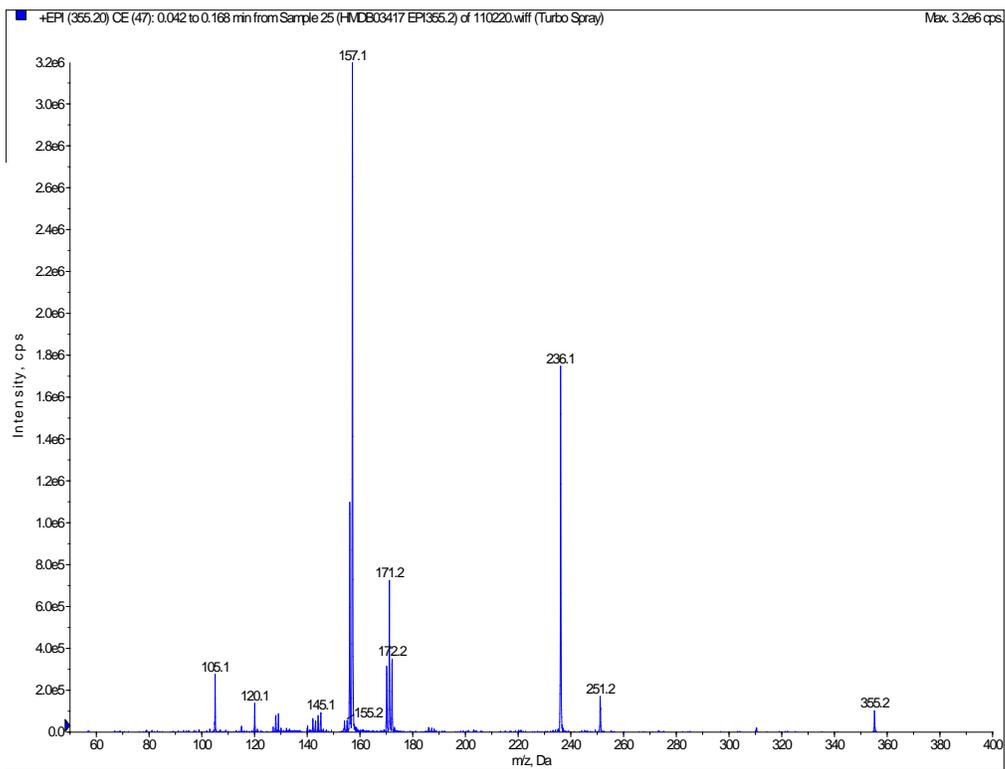
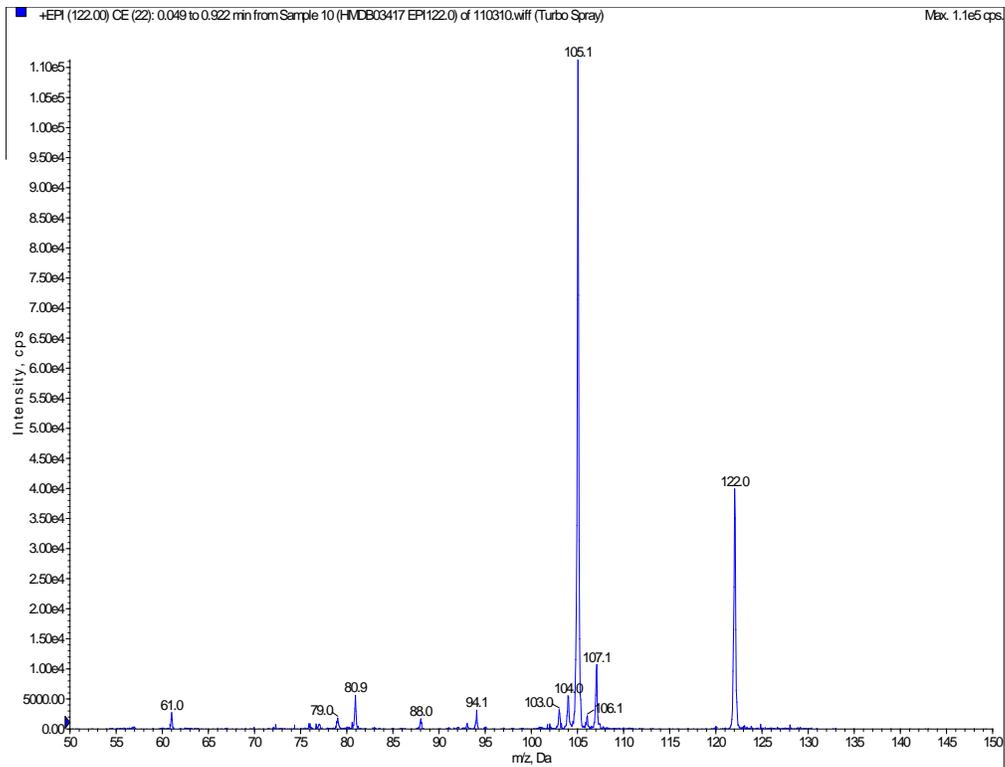
((2S)-2-amino-6-[5-[(1S,2S,5R)-7-oxo-3-thia-6,8-diazabicyclo[3.3.0]oct-2-yl]pentano-ylamino]hexanoic acid)





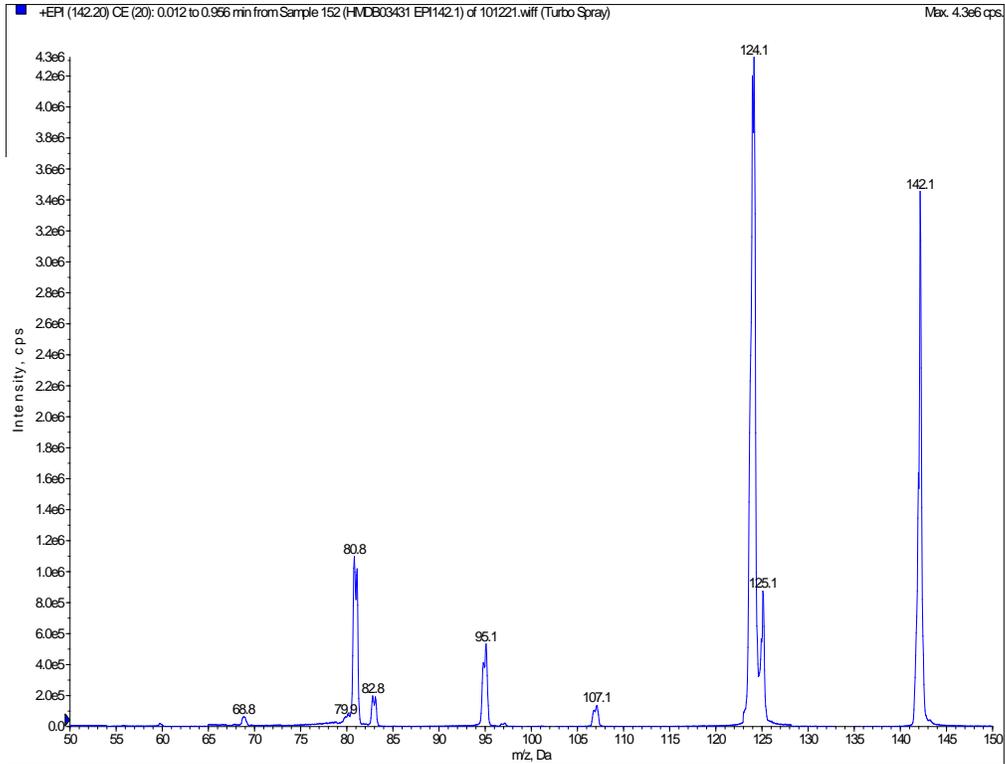
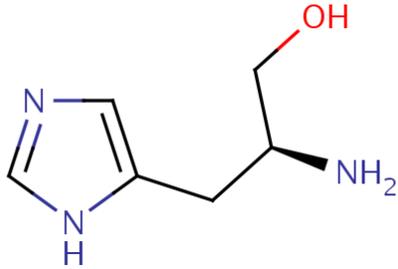
HMDB03417
D-Cysteine ((2S)-2-amino-3-sulfanyl-propanoic acid)

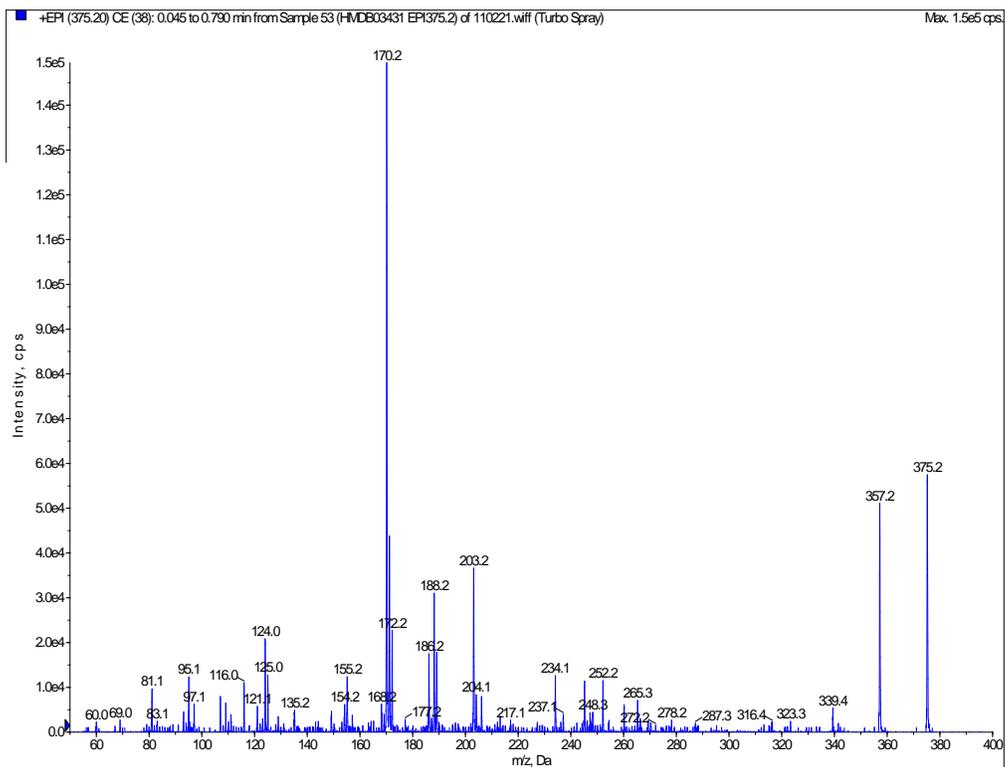
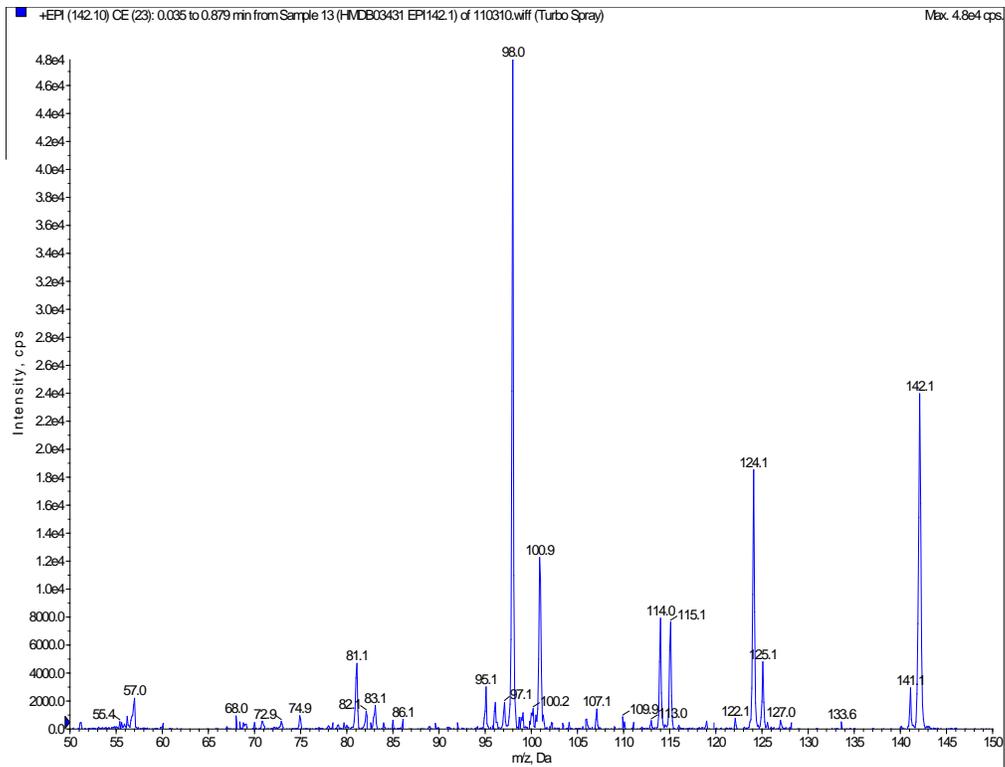




HMDB03431

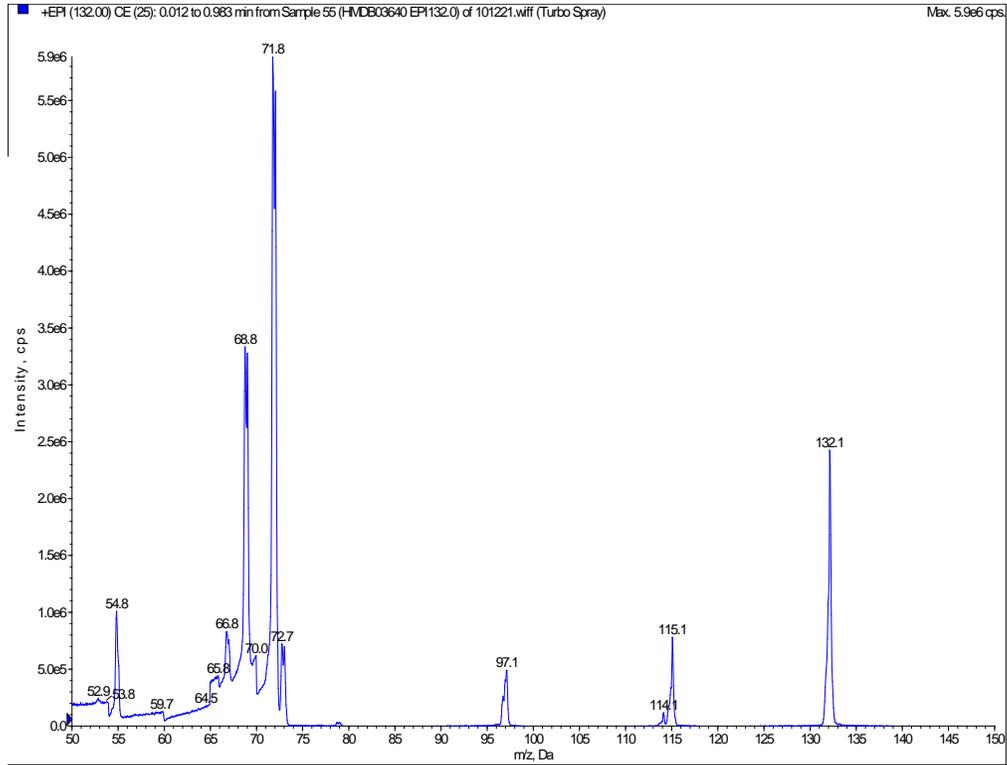
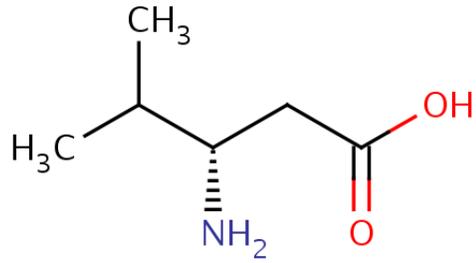
L-Histidinol ((2S)-2-amino-3-(3H-imidazol-4-yl)propan-1-ol)

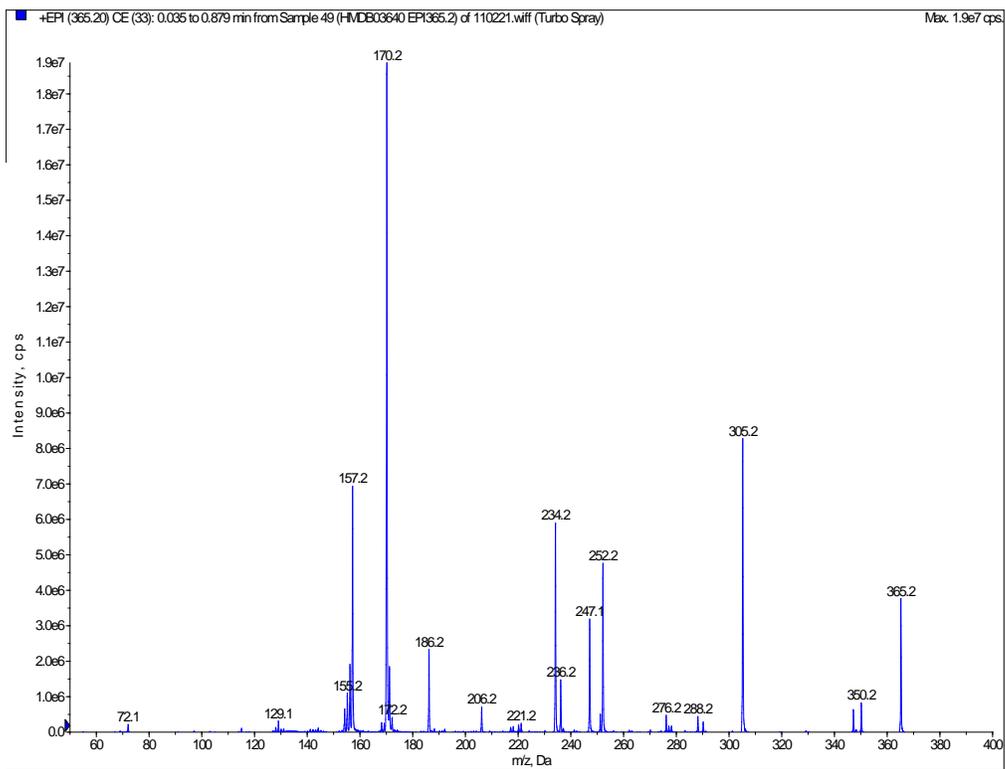
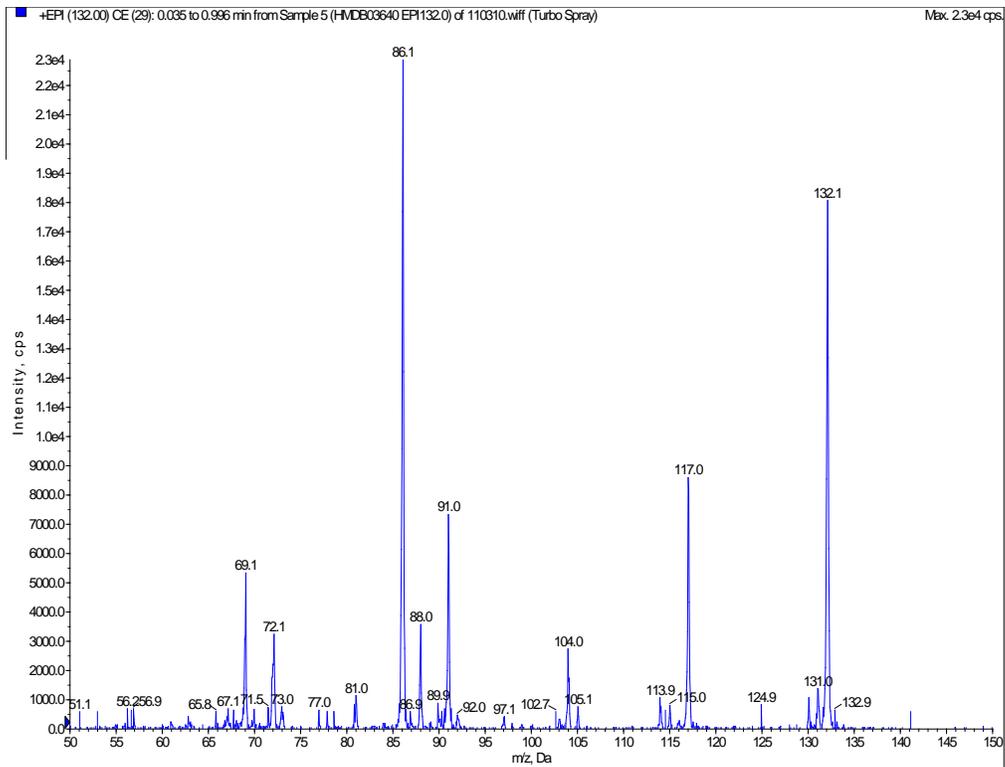




HMDB03640

Beta-Leucine ((3S)-3-amino-4-methyl-pentanoic acid)





HMDB03966

Selenomethionine (2-amino-4-methylselanyl-butanoic acid)

