Development of Chemical Isotope Labeling Liquid Chromatography Mass Spectrometry
on Quadrupole-Orbitrap Mass Spectrometers for Metabolome Analysis

by

Chu-Fan Wang

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

Doctor of Philosophy

Department of Chemistry

University of Alberta

Abstract

Chemical isotope labeling liquid chromatography-mass spectrometry (CIL LC-MS) is a powerful technique in metabolomics, offering extensive coverage of the metabolome. This method involves dividing metabolites into sub-groups based on their functional groups and subjecting them to specific labeling reactions. By employing well-designed labeling reagents, CIL LC-MS enables improved LC separation and MS detection, the use of differential isotope labeling further allows for accurate relative quantification. To date, a four-channel labeling protocol has been developed in our lab, i.e., profiling amine-/phenol-, carboxyl-, hydroxyl-, and carbonyl-sub-metabolomes, achieving a remarkable 86%-95% coverage of non-lipid metabolites in the databases of MCID, HMDB, KEGG, YMDB and ECMDB. While quadrupole time-of-flight (QTOF) MS has been commonly utilized in CIL LC-MS, this thesis research focuses on implementing CIL LC-MS analysis on Q-Orbitrap MS, with an emphasis on maximizing instrumental performance for metabolomics.

Chapter 2 compared the effects of different instrument types (TOF MS vs. Orbitrap MS) on metabolite detectability, considering detection sensitivity, dynamic range, ionization efficiency, and ion transportation efficiency. Instrument type was found to significantly impact metabolite detection in CIL LC-MS, many common peak pairs were detected from the two instruments; however, there were a significant number of unique

peak pairs detected in each type of instrument, necessitating careful comparison of data generated using different instruments.

Chapter 3 introduced a segmented spectrum scan method using Orbitrap MS in CIL LC-MS to enhance metabolite detection efficiency. During the LC data acquisition, the full m/z range was divided into multiple segments with the scanning of each segment to produce multiple narrow-range spectra. This approach realized a substantial increase in detectable metabolites while maintaining accurate relative quantification and precise peak ratio measurements.

Chapter 4 evaluated the performance of the segment-scan-assisted CIL LC-MS across different sample types such as feces, urine, blood, cell extracts, saliva, etc. A 120-m/z segment scan was established as a routine approach to further improve the overall MS detection of dansyl-labeled metabolites on Q-Orbitrap MS.

Chapter 5 explored the fragmentation patterns of dansyl-labeled amine-/phenol-containing metabolites using high-energy collisional dissociation (HCD) on quadrupole-Orbitrap MS, with a focus on MS/MS-based spectral elucidation to gain insights into the structure of dansyl-labeled metabolites.

Overall, this contributes to the optimization and understanding of CIL LC-MS analysis on Q-Orbitrap MS for metabolomics, advancing the capabilities of metabolite detection and identification in complex biological samples.

Preface

A version of Chapter 2 was published as: Chu-Fan Wang, and Liang Li. "Instrument-type effects on chemical isotope labeling LC-MS metabolome analysis: Quadrupole time-of-flight MS vs. Orbitrap MS." *Analytica Chimica Acta* 1226 (2022): 340255. I was responsible for the experimental design, data collection and analysis, as well as the manuscript writing. Professor Liang Li supervised the project and edited the manuscript.

A version of Chapter 3 was published as: Chu-Fan Wang, and Liang Li. "Segment Scan Mass Spectral Acquisition for Increasing the Metabolite Detectability in Chemical Isotope Labeling Liquid Chromatography–Mass Spectrometry Metabolome Analysis." *Analytical Chemistry* 94.33 (2022): 11650-11658. I was responsible for the experimental design, data collection and analysis, as well as the manuscript writing. Professor Liang Li supervised the project and edited the manuscript.

Chapter 4 has been submitted to *Analytica Chimica Acta* and it is under peer review. I was responsible for the experimental design, data collection and analysis, as well as the manuscript writing. Professor Liang Li supervised the project and edited the manuscript.

Acknowledgements

I would like to express my deepest gratitude and appreciation to the individuals who have supported and guided me throughout my research journey.

First and foremost, I am sincerely grateful to Dr. Liang Li. Without his exceptional guidance, unwavering support, and invaluable expertise, I could not have discovered the beauty of the world of metabolomics. His insightful feedback, encouragement, and commitment to excellence have played a crucial role in shaping the direction and quality of this research.

I would like to take this opportunity to express my heartfelt gratitude to my supervisory committee members, Dr. John Vederas and Dr. Ran Zhao, for their invaluable insights and constructive feedback that have greatly enriched the quality of my thesis. I am sincerely appreciative of their guidance and support throughout my research journey. Additionally, I am grateful to my external examiner, Dr. Lekha Sleno, for her valuable contribution in both participating in my oral defense and meticulously reviewing my thesis. I am deeply thankful to Dr. Mariusz Klobukowski for graciously accepting the role of the non-examining chair for my oral examination. I also wish to acknowledge the presence of Dr. Shira Joudan at my oral defense and extend my gratitude for her attendance.

I would like to acknowledge the members of the Li research group whom I worked with during my Ph.D. program: Yingwen Wang, Zhan Cheng, Carlos Canez Quijada, Vi Tran, Fernanda Monteiro Queiroz, Cyrene Jonah Catenza, Sicheng Quan, and Michal

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACN	Acetonitrile
AGC	Automatic gain control
CE	Capillary electrophoresis
CID	Collision-induced dissociation
CIL	Chemical isotope labeling
DC	Direct current
DDA	Data-dependent acquisition
DI	Direct injection
Dns	Dansylation
DnsHz	Dansylhydrazine
DmPA	4-(Dimethylamino) phenacyl
ECMDB	E. Coli Metabolome Database
ESI	Electrospray ionization
FTICR	Fourier-transform ion cyclotron resonance
FTMS	Fourier-transform mass spectrometry
GC	Gas chromatography
HCD	High-energy collisional dissociation
HILIC	Hydrophilic interaction liquid chromatography
HMDB	Human Metabolome Database
LC	Liquid chromatography
MCID	MyCompoundID
MCP	Microchannel plate
MeOH	Methanol
MS	Mass spectrometry
m/z	Mass-to-charge ratio
NMR	Nuclear magnetic resonance
OH	Hydroxyl group
Q	Quadrupole
RF	Radio frequency
RP	Reversed-phase
RSD	Relative standard deviation
RT	Retention time
SIL	Stable isotope label
S/N	Signal-to-noise ratio
TOF	Time-of-Flight
TQ	Triple quadrupole
USS	Universal serum standard
UUS	Universal urine standard
YMDB	Yeast Metabolome Database

Chapter 1

Introduction

1.1 Introduction to metabolomics

Metabolites are small molecules that undergo physical and chemical transformations in various biochemical processes within living organisms.[1–3] These processes, known as metabolism, enable organisms to sustain life (via nutrient acquisition, energy generation, toxics decomposition, etc.) and reproduce themselves.[4,5] The diversity of metabolites is immense, ranging from simple molecules such as amino acids, sugars, and fatty acids to more complex compounds like hormones, vitamins, and polyphenols. The levels and concentrations of metabolites can vary depending on factors such as environmental conditions, cellular state, and metabolic activity.[6] With their unique characteristics, metabolites offer a closer look into the phenotype compared to the other two fundamental components of living organisms, i.e., genes[7,8] and proteins.[9,10] Therefore, metabolite profiling, which involves the identification and quantification of a wide range of metabolites in biological samples, can provide valuable insights into the physiological state of a given biological system at a given time. This research, known as metabolomics, has emerged as a powerful tool for disease biomarker discovery, drug development, nutritional research, and other biological studies.

The term "metabolome" refers to the complete set of metabolites present within a biological system, including endogenous and exogenous molecules.[11,12] The ultimate

objective of metabolome profiling has been to collect qualitative and quantitative information for as many metabolites as possible while ensuring efficient sample throughput. From qualitative aspects, higher coverage of metabolome profiling can provide a more complete picture of the metabolic pathways and their interconnectedness, increase the chances of identifying disease-specific or condition-specific metabolites that can be used for in-depth diagnostic and prognosis biomarker discovery,[13,14] and enhance the accuracy and reliability of system-level analysis where metabolomic data is integrated with other omics data such as genomics and proteomics to understand the complex interactions and regulatory networks within biological systems. From quantitative aspects, due to tight metabolic homeostasis in biological systems such as humans, only small changes in metabolite concentrations are detectable in comparative samples with noticeable phenotype differences. Therefore, accurate quantification of metabolites is essential for obtaining meaningful and reproducible results in metabolomic studies.[15,16]

Typically, a routine metabolome analysis includes several key steps as follows:

i) sample preparation: Depending on the type of metabolome study, various sample types such as blood, urine, tissue/cell extracts, saliva, etc., may be collected. The samples are then subjected to processes such as homogenization and extraction to obtain the metabolites of interest.

ii) sample analysis: This step requires the use of a sensitive analytical platform with a high dynamic range to achieve comprehensive coverage of the metabolome while utilizing only small sample volumes.[17]

iii) data analysis: After sample analysis, sophisticated data analysis methods will be employed to extract useful information from the obtained data. This includes data processing, metabolite identification, statistical analysis, and pathway analysis. Computational tools and software are used to handle the large datasets generated from metabolome analysis to unravel the roles of metabolites in different biological processes.

The dynamic nature of the metabolome makes it highly sensitive to changes in physiological and environmental conditions, while at the same time, the unknown size of the overall metabolome, and the great diversity and more than 10^7-fold concentration[18] differences of metabolites also post a challenge for comprehensive metabolome characterization. With developments in analytical technologies, numerous efforts have been made in the past decades to develop new methods for metabolome analysis.[19,20]

1.2 Analytical technologies for metabolome profiling

The two most common detection platforms employed in metabolome analysis are nuclear magnetic resonance (NMR) and mass spectrometry (MS). NMR and MS offer distinct advantages in metabolome analysis due to their complementary capabilities. NMR provides structural information and allows for the quantification of a wide range of metabolites with excellent specificity and reproducibility, while it suffers from relatively

low sensitivity and requires a larger sample amount.[21–23] Moreover, NMR may also struggle with metabolite identification, especially for complex bio-samples. On the other hand, MS offers high sensitivity, enabling the detection of low-abundance metabolites and the ability to analyze a large number of metabolites in a single run.[24,25] However, identification of metabolites based on MS/MS spectra elucidation can be ambiguous,[16] using MS alone for metabolome analysis also suffers from ion suppression/enhancement effect when it comes to analysis of complex bio-samples.[26]

Nagana Gowda et al.[27] used *ex vivo* ^{1}H NMR to simultaneously analyze oxidized and reduced forms of major coenzymes to study cellular redox reactions in various cellular functions. *ex vivo* ^{1}H NMR was able to offer precise quantitative information on studied coenzymes with high reproducibility. However, when dealing with untargeted metabolomics, NMR faces limitations in sensitivity, making it less suitable for comprehensive compound identification within complex mixtures. Additionally, its resolution capabilities do not match those of MS techniques. In untargeted metabolomics, where the aim is to explore and identify a wide range of metabolites, including unknown compounds, MS is generally favoured due to its superior sensitivity, resolution, and ability to cover a broader metabolite.

In untargeted MS-based metabolomics, the integration of separation techniques, such as capillary electrophoresis (CE), gas chromatography (GC), and liquid chromatography (LC), with mass spectrometry is a valuable approach to further improve

the performance of metabolome analysis. The presence of separation techniques not only reduces sample complexity to improve MS detection but also provides accurate quantitative results. Among them, liquid chromatography-mass spectrometry (LC-MS) is particularly versatile and can be used for a wide range of metabolites, including polar and non-polar ones.[28,29] In contrast, GC[30,31] is usually for studying volatile and semi-volatile metabolites while CE[32,33] is only for highly polar and charged metabolites. In LC, the sample is dissolved in a liquid mobile phase and passed through a stationary phase, the separation of analytes is achieved based on their interactions with the stationary phase. The separated analytes are subsequently ionized and analyzed by the mass spectrometer.

1.2.1 Conventional LC-MS metabolomics

In conventional LC-MS metabolomics, achieving comprehensive metabolome coverage often requires the use of multiple LC-MS conditions. This approach involves employing different LC methods to separate metabolites with varying chemical properties and utilizing different ionization modes in MS for optimal detection.

During LC-MS analysis, reversed-phase liquid chromatography (RPLC) is employed to separate hydrophobic metabolites. On the other hand, hydrophilic interaction liquid chromatography (HILIC) is utilized to separate more polar compounds. In MS analysis, two different ionization modes, positive and negative, are typically used to enhance the detection of analytes with varying proton affinities. Positive ionization mode is suitable for detecting metabolites with higher proton affinity, while negative ionization

mode is effective for those with lower proton affinity. By employing both ionization modes, a wider range of metabolites can be detected and analyzed. Moreover, it is crucial to select a sensitive platform with a high dynamic range to meet the requirements of detecting metabolites with great differences in concentrations. This ensures that both low-abundance metabolites and those present at higher concentrations can be accurately detected and quantified.

Considering the different LC separation modes and ionization modes, for a single sample, four parallel LC-MS runs are typically performed, each utilizing a different instrumental setup to accommodate the specific LC separation mode (RPLC or HILIC) and ionization mode (positive or negative). Despite employing these strategies, detecting certain metabolites remains challenging due to factors such as their extremely low concentrations or poor ionization efficiency. Additionally, multiplying LC-MS runs, along with the subsequent data analysis, often results in increased costs and time requirements.

In conventional LC-MS, accurate quantitative analysis relies on the use of stable isotope labeled (SIL) internal standards.[34,35] These internal standards serve several purposes, including correcting for analyte losses during sample preparation, accounting for variability in the volume of sample introduced into the instrument, and compensating for matrix effects. However, SIL internal standards can be prohibitively expensive, and their accessibility is limited. Consequently, a relatively new approach has been developed in the field of metabolomics, i.e., chemical isotope labeling (CIL) LC-MS.

1.2.2 Chemical isotope labeling LC-MS metabolomics

CIL LC-MS offers an alternative approach for metabolomic analysis. In principle, a pair of rationally designed differential isotope labeling reagents (e.g., H/D, $^{12}C/^{13}C$, and $^{14}N/^{15}N$) are employed to react with the metabolites in comparative samples (e.g., individual samples are labeled with the light reagent while a pooled or control sample is labeled with the heavy reagent), followed by mixing and LC-MS analysis.[36,37] In CIL LC-MS, each metabolite in the individual sample corresponds to a CIL internal standard in the pooled sample. By comparing the peak area ratios between the light and heavy isotopically labeled metabolites, quantitative information can be obtained, providing insights into the relative abundance changes of metabolites across different samples.

The introduction of mass tags through isotope labeling brings additional advantages: Firstly, it changes the chemical properties of labeled metabolites, thereby improving the LC separation and ionization efficiency. For example, ^{12}C-/^{13}C-dansylation (Dns) is used to modify amine-/phenol-containing metabolites, the introduced dansyl moiety enhances the hydrophobicity of the labeled metabolites, enabling efficient LC separation using a simple RPLC setup (no extra HILIC separation needed).[38] Additionally, the tertiary amine of the introduced dansyl group serves as a protonated site to improve the ionization process (i.e., only positive ion mode is required for MS detection). Taken together, a higher coverage of metabolome profiling can be realized using CIL LC-MS with a simpler LC-MS configuration compared to the conventional LC-MS. Furthermore, the presence of

mass tags facilitates data processing and analysis. Unlike conventional LC-MS, using differential isotope labeling LC-MS, metabolites are detected as peak pairs, while chemical and background noises are detected as singlet peaks. As a result, noise peaks detected from different types of instruments can be eliminated when only the peak pairs from the metabolites are compared, allowing for more accurate metabolite identification.

With these points in mind, it is crucial to design a rational labeling reagent in CIL LC-MS. An ideal CIL reagent should meet several requirements: i) The reagent should be capable of reacting with a wide range of metabolites, considering the diverse nature of metabolites. Encouragingly, a previous study[39] has provided valuable insights into the functional groups present in a wide range of metabolites. By filtering out lipids, inorganic species, and hydrocarbons that are usually not targeted for analysis by CIL LC-MS, the study revealed that a significant proportion (86%-96%) of metabolites in various metabolite databases (such as MCID, HMDB, KEGG, YMDB, and ECMDB) contain one or more of the five functional groups as follows: amine, phenol, hydroxyl, carbonyl, and carboxyl. In this case, rational labeling reagents that specifically target these functional groups can be found or designed. ii) The CIL reaction should exhibit sufficient and consistent labeling efficiency. It is important to ensure that the labeling reagent efficiently labels the target metabolites, including low-abundance metabolites. Insufficient labeling efficiency may lead to incomplete or partial labeling, resulting in information loss regarding the detected metabolome. In addition, accurate quantification relies on consistent

labeling efficiency, as inconsistent labeling can introduce biases and affect the reliability of quantitative results. iii) The CIL reaction should minimize or eliminate the formation of reaction byproducts that may interfere with LC separation and suppress MS detection. Unintended byproducts can complicate the chromatographic separation, leading to peak broadening, co-elution, or interference with metabolite detection. iv) Labeled metabolites should be stable for sample preparation, storage, and analysis. Any chemical degradation or losses could lead to distorted results. v) High-quality MS/MS spectra can be acquired based on the labeling chemistry. Metabolite identification is normally finished by matching accurate mass and retention time (RT) to a standard library. However, some metabolite standards may not be commercially available, in this case, MS/MS spectra interpretation can provide structural information which helps identify unknown metabolites.

Based on those requirements, a 4-channel CIL LC-MS method has been developed in our lab. Four reactions are used to derivatize metabolites from different sub-metabolomes: Dns for amines/phenols,[38] 4-(Dimethylamino) phenacyl bromide (DmPA) derivatization for carboxylic acids,[40] base-activated Dns for hydroxyls[41] and dansylhydrazine (DnsHz) derivatization for carbonyls.[42] It is demonstrated that a 10- to 1000-fold increase in MS detection sensitivity has been achieved with precise quantification by using this strategy.

1.3 Instrument-type effect on CIL LC-MS metabolomics

In CIL LC-MS, the detection ability of MS remains a limiting factor in achieving comprehensive metabolome profiling since different types of mass spectrometers exhibit variations in detection sensitivity, mass resolution, dynamic range, and other performance parameters. The choice of MS instrument can impact the overall detection capability and lead to different metabolome analysis results.

1.3.1 Mass spectrometry

As shown in Figure 1.1, there are three main components in a mass spectrometer, namely, ion source, mass analyzer, and detector.

Figure 1.1 Basic components of a mass spectrometer.

In CIL LC-MS, electrospray ionization (ESI)[43,44] is commonly employed, labeled-metabolites are ionized in the ion source chamber. In principle, eluent containing the labeling metabolites from LC passes through an ESI capillary applied with a high voltage, at the end of the capillary tip, LC flow forms a Taylor cone which consists of a highly charged surface that emits tiny droplets. The formed droplets then move towards the MS inlet with solvent molecules evaporating due to the high temperature and low pressure within the ionization chamber, leading to the formation of smaller droplets. At the same time, ions are formed through one or more mechanisms, including protonation and adduct

formation. ESI is considered a soft ionization technique, meaning it generates predominantly intact molecular ions with minimal fragmentation, which facilitates subsequent identification based on the mass of intact ionized metabolites detected.

Ionized labeled-metabolites are then subjected to a mass analyzer where the ions are separated based on their mass-to-charge ratio (m/z). Multiple mass analyzers are commercially available such as quadrupole,[45] time-of-flight (TOF),[46] Orbitrap,[47] and Fourier-transform ion cyclotron resonance (FTICR).[48]

i) A quadrupole mass analyzer consists of four parallel metal rods arranged in a rectangular configuration (see Figure 1.2). The operation of the quadrupole is based on a combination of radio frequency (RF) and direct current (DC) voltages applied to the rods. Briefly, the RF voltage creates an oscillating electric field, while the DC voltage creates a static electric field. These fields work together to control the transmission or exclusion of ions based on their m/z. With the simple configuration and relatively low cost, quadrupole has been widely used in MS, furthermore its ability to selectively transmit or exclude ions based on m/z also allows it to function as an ion filter in tandem MS where multiple mass analyzers are combined to enable MS/MS capabilities.
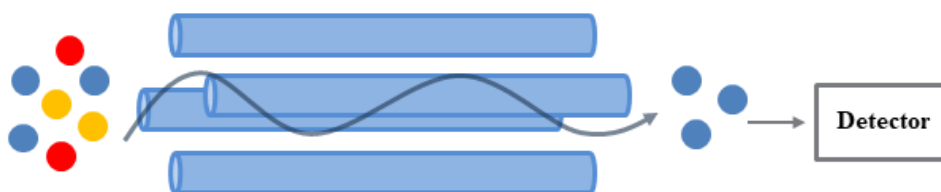
Figure 1.2 Schematic of a quadrupole mass analyzer.

ii) In a TOF mass analyzer, as shown in Figure 3, ions are accelerated into a flight tube by applying an electric field. The acceleration causes ions of different m/z ratios to gain an identical kinetic energy. Ions with higher m/z ratios travel slower, while ions with lower m/z ratios travel faster, enabling the separation of ions based on m/z. The flight tube is typically evacuated to reduce collisions with gas molecules, ensuring that the ions travel unimpeded. At the end of the flight tube, there is a microchannel plate (MCP) detector that measures the arrival time of ions.
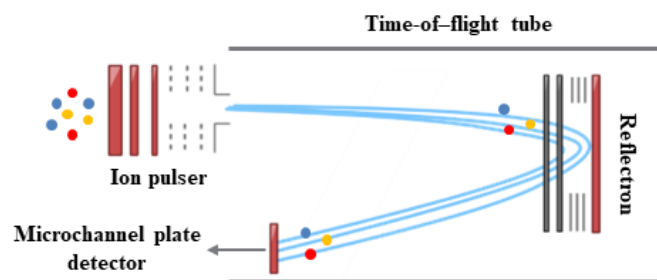
Figure 1.3 Schematic of a time-of-flight mass analyzer.

iii) An Orbitrap mass analyzer consists of two primary components: a central spindle-like electrode and an outer electrode, as shown in Figure 4. A voltage is applied between these electrodes to generate an electrostatic field. When ions are introduced into the Orbitrap, they are trapped and confined within the electrostatic field. The ions oscillate along the central spindle electrode due to the forces exerted by the electric field. The outer electrodes surrounding the central electrode act as receiver plates for detecting the image current resulting from the axial oscillations of the trapped ions. Ions with different m/z are detected and analyzed based on their frequencies of oscillations. In Orbitrap MS, the

number of ions to be analyzed in a single MS scan is fixed, there is another ion-trapping

device called C-Trap to temporarily store the ions before mass analysis.
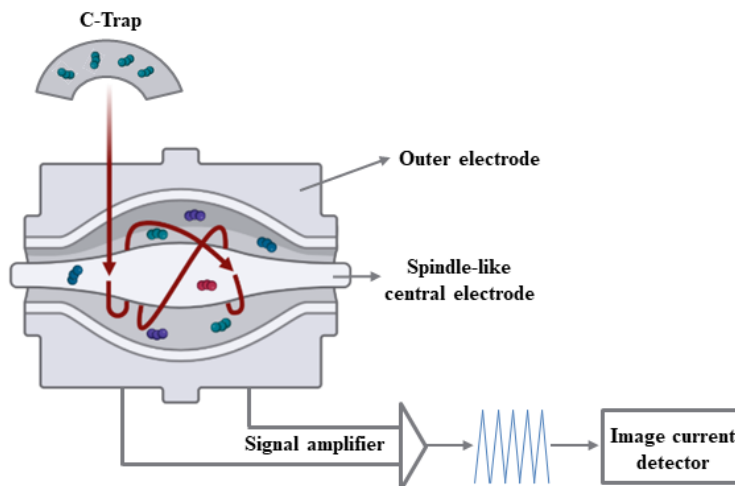
Figure 1.4 Schematic of an Orbitrap mass analyzer.

Unlike single-stage MS, tandem MS[49] involves two stages of mass analysis,

allowing for the characterization of compounds by generating fragments with more

structural information. In tandem MS/MS, precursor ions selected in the first stage MS

(MS1) are subjected to a collision cell where they collide with neutral gas to induce

fragmentation, the generated fragment ions are analyzed in the second stage of MS. The

most popular configuration of tandem MS are triple quadrupoles (TQ), quadrupole-TOF

(Q-TOF), and quadrupole-Orbitrap (Q-Orbitrap). TQ MS is a widely used technique for

quantitative analysis in targeted-metabolomics due to its high sensitivity, selectivity,

quantification accuracy and low cost. However, TQ MS is not a good option for untargeted

metabolomic because of its low resolution and mass accuracy. In non-targeted

metabolomics, as mentioned in section 1, the ultimate objective of metabolome profiling

is to collect qualitative and quantitative information for as many metabolites as possible. Compared to TQ MS, Q-TOF MS and Q-Orbitrap MS offer higher resolution and mass accuracy while retaining the capability for informative fragmentation at the same time. On this account, most recent studies related to LC-MS based non-targeted metabolomics were mainly performed on these two platforms.[50–57]

In previous studies, Q-TOF MS has been extensively used for CIL LC-MS.[18,39,41,42,58,59] However, considering that different types of mass spectrometers have different detection sensitivity, mass resolution, dynamic range, etc., it would be useful to evaluate the effects of MS instrument type on the detectability of labeled metabolites from these aspects when adapting the established CIL LC-MS method on Q-Orbitrap MS. Moreover, understanding the effects of using different instruments can shed light on the variations in detected metabolites, which may arise when similar studies are conducted. By comprehending these differences, researchers can potentially explain the discrepancies in metabolome data generated from various experiments. This knowledge not only enhances our understanding of metabolomics data but also offers valuable insights to develop improved methods for enhancing overall metabolite detectability.

1.3.2 In the aspect of sensitivity

Both Q-TOF MS and Q-Orbitrap MS offer high detection sensitivity. In TOF MS, mass spectral recording speed can reach as high as 10,000 spectra per second, which is the highest among all the other mass analyzers. While in real-world applications, such high

spectral acquisition speed is not necessary, in this case, spectra averaging is carried out where multiple individual mass spectra are combined to maintain the final scan speed in a proper range. In this process, the in-spectrum signal-to-noise ratio (S/N) is enhanced by reducing random noise, resulting in a higher detection sensitivity. Saying the number of mass spectra combined during the averaging procedures is N, the final scan speed is roughly equal to the original scan speed divided by N. Moreover, it has been reported that the in-spectrum S/N is positively proportional to the root squared of N. In TOF MS, ions transmitted by quadrupole will immediately gain kinetic energy from the ion pulser and get extracted into the flight tube (see Fig 1.3). In contrast, in Orbitrap MS, ions will be temporarily stored in the C-Trap before being introduced into the Orbitrap (see Figure 1.4), the presence of the C-Trap provides a means to accumulate ions and improve the detection sensitivity. By accumulating the ions for a longer duration, the detection sensitivity could be further enhanced while the overall scan speed will be sacrificed at the same time.

However, in metabolomics, frequent sampling is often necessary to capture small changes in metabolite concentrations between comparative samples with noticeable phenotype differences.[60,61] It leads to limited sample volume, whether obtained through invasive methods (such as blood or cerebrospinal fluid collection) or non-invasive methods (such as urine or saliva collection). In this case, the reliable detection of metabolites present at extremely low abundances in the collected samples heavily depends on the sensitivity of the tandem MS to be used. In other words, the detection of those metabolites is prone to

sensitivity changes among different tandem MS platforms. For example, in CIL LC-MS, a peak pair could be detected by an MS when its intensity was above the detection limit whilst it could be missed by another MS when its intensity was below the detection limit, bringing up inconsistent metabolome analysis results generated using different instruments.

1.3.3 In the aspect of dynamic range

In CIL LC-MS, an equalized ionization efficiency can be achieved among different labeled metabolites, which is accomplished through consistent labeling efficiency and structure normalization during the labeling process. This ensures that the signal intensities of those metabolites are more directly comparable to their original concentrations in biological samples, which vary by several orders of magnitude. As discussed above, the sensitivity of an instrument determines the lower limit of detection for labeled metabolites, while dynamic range represents the ability for an instrument to capture and measure signals across a wide range of intensity levels. An ideal dynamic range in CIL LC-MS should be able to cover the whole metabolome to be profiled. However, in real-world applications, although both Q-TOF MS and Q-Orbitrap MS offers a relatively wide dynamic range, limitation unique to each instrument is still available, leading to different MS detections of metabolites.

In Q-TOF MS, the in-spectrum dynamic range is always limited by the MCP detector. As shown in Figure 1.5, an MCP detector[62] consists of a Pb-glass material with millions of microchannels fused. When ions enter the microchannels of the MCP detector,

they undergo a process where multiple secondary electrons are generated each time they hit the channel walls. This process repeats several times until the ions leave the channels, resulting in the generation of thousands of output electrons that are then detected. Since the total number of channels of an MCP detected is constant, for those ionized metabolites with high abundance, their signals can reach saturation in TOF MS, resulting in signal distortion and potential loss of quantitative accuracy.

The in-spectrum dynamic range of Orbitrap MS is generally limited due to the finite capabilities of its detection system. In Orbitrap MS, ion detection relies on the measurement of the image current generated by ions as they orbit in the electrostatic field of the Orbitrap analyzer. The image current is then converted into a mass spectrum. The limitation arises from the maximum current that can be measured by the ion detector. When ions with high abundance are present in the mass spectrum, they can generate a strong image current, leading to accurate and reliable detection of these ions. However, when ions with low abundance are present, their image current is weaker, and they might be overshadowed by the noise in the detection system. As a result, low-abundance ions may not be detected or detected with poor signal-to-noise ratio, reducing the overall dynamic range of the mass spectrum.
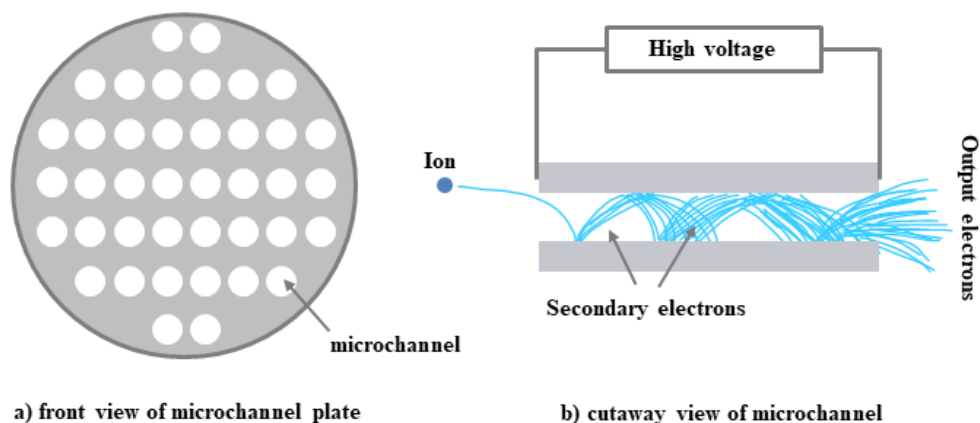
Figure 1.5 Schematic of a microchannel plate detector.

1.3.4 In the aspect of resolution

In MS, resolution is typically expressed as the separation between two adjacent peaks (Δm) to the mass of the (second) peak (M), i.e., Δm/M. It represents the ability to distinct peaks for ions that differ slightly in mass. The usage of high-resolution MS provides enhanced mass accuracy and improved spectral quality. In CIL LC-MS-based metabolomics, the peak pair detectability among different platforms is dependent on the resolution as well since more peaks can be resolved under a higher resolution in theory. When it comes to TOF MS vs Orbitrap MS, the former is normally operated a certain range of resolution, while the latter offers more options in resolution.

1.3.5 In the aspect of MS/MS capability

In tandem MS, precursor ions are selected and accelerated in the collision cell (typically a quadrupole or hexapole) to collide with neutral gas. The collision-induced energy causes the bonds within the precursor ions to break, forming a series of product

ions. These product ions carry information about the structure and composition of the precursor ions.

The specific fragmentation pathways and mechanisms that occur in the collision cell are referred to as collision-induced dissociation (CID),[63] which is governed by gas-phase chemistry. CID is widely used in quadrupole-based instruments, such as TQ MS and Q-TOF MS to produce informative fragmentation patterns. For example, with b- and y-type fragment ions in peptide fragmentation.[64,65] Therefore, it is suitable for targeted metabolome analysis and quantification.

Higher-energy collisional dissociation (HCD)[66] is a dissociation technique developed on Orbitrap MS. In earlier designs of ion-trap instruments like the Orbitrap MS, CID was performed within a linear ion trap. This method, also known as in-trap CID, employed a resonant excitation technique, requiring ions to be accelerated relatively slowly within the trap to gain energy. It contrasted with the typical beam-type CID used in TQ or Q-TOF MS instruments. In more recent designs, the linear ion trap has been replaced by a quadrupole-based collision cell, aligning it with the configuration used in TQ and Q-TOF MS instruments. To distinguish this newer collision technique from the previous CID method used in early Thermo Fisher designs, the term "HCD" was introduced for the collision cell in later models. Although HCD has the same configuration compared to CID in TQ/QTOF, it has been reported that HCD can improve the fragmentation in the low mass region compared to CID.[67]

In LC-MS-based metabolomics, accurate metabolite identification is obtained by matching experimental accurate mass and retention time against a standard library. However, due to the great diversity in metabolites, not all metabolites have an available standard. In this case, high-quality experimental MS/MS spectra serve as complimentary information for metabolite identification and structure validation. It is important to acknowledge that different tandem MS platforms can produce different results for the same molecule since different molecular dissociation techniques are employed. Therefore, choosing the appropriate tandem MS platform and optimizing the experimental strategy is crucial for obtaining the best results in metabolomics analysis.

In addition to considering the quality of MS/MS spectra, the acquisition speed of MS/MS is another crucial factor to be taken into account when aiming for high-throughput non-targeted MS/MS spectral collection. In terms of TOF resolution, there is generally little difference between the MS and MS/MS modes. This offers an advantage over the Orbitrap analyzer, where the resolution in MS/MS mode is often compromised for the sake of speed. However, in the context of targeted MS/MS-based metabolite identification, the Orbitrap MS is more suitable than TOF MS, particularly for low-intensity precursor ions. The presence of the C-Trap enables the accumulation of precursor ions, enhancing the overall sensitivity of MS1 and resulting in the generation of high-quality MS/MS spectra.

1.4 Scope of the thesis

My research aims to implement CIL LC-MS analysis on Q-Orbitrap MS, with a focus on maximizing the instrumental performance for metabolomics.

In Chapter 2, we examined the effects of instrument type on the detectability of true metabolites with a focus on the comparison of TOF MS vs. Orbitrap MS, regarding four main factors: a) detection sensitivity, b) detection dynamic range, c) ionization efficiency, and d) ion transportation efficiency.

Following the conclusion obtained in Chapter 2, a segmented spectrum scan method, where the full m/z range was divided into multiple segments with the scanning of each segment to produce multiple narrow-range spectra, was developed using Orbitrap MS in CIL LC-MS for improving the metabolite detection efficiency in (Chapter 3). The improvement in the dynamic range of Orbitrap MS could reduce the difference in peak pairs detection observed. A remarkable increase in detectable metabolites was achieved with good relative quantification accuracy and peak ratio measurement precision compared to conventional full scan method.

In Chapter 4, the performance of segment-scan-assisted CIL LC-MS among different types of samples was evaluated. An optimal 120-m/z segment spectral acquisition method was developed as a routine approach for CIL LC-MS metabolome analysis on Q-Orbitrap MS.

In Chapter 5, we conducted a comprehensive investigation into the fragmentation patterns of dansyl-labeled amine-/phenol-containing metabolites using HCD on Q-Orbitrap MS. More informative MS/MS spectra that encompassed structural information related to both the dansyl group and the metabolite moiety were collected and used for metabolite identification and structural analysis.

Chapter 2

Instrument-type Effects on Chemical Isotope Labeling LC-MS

Metabolome Analysis: Quadrupole Time-of-flight MS vs. Orbitrap MS

2.1 Introduction

Metabolomics aims to profile all the small molecules that are chemically transformed during the metabolism in a biological system. However, these small molecules or metabolites show a great variety in their chemical structures and concentrations on both temporal and spatial scales. Thus, it is a major challenge to detect and quantify all the metabolites using conventional analytical methods, such as liquid chromatography-mass spectrometry (LC-MS), due to the limited detectability and dynamic range.[68] To address this challenge, especially to increase the profiling coverage of the metabolome, many MS-based methods have been developed.[69,70] Of those, an enabling strategy is to use differential chemical isotope labeling (CIL) of metabolites coupled with LC-MS.[71]

In CIL LC-MS, metabolites are classified into different sub-metabolomes depending on the functional groups present in molecules (e.g., amine, hydroxyl, etc.).[71–78] Then for each submetabolome, a pair of labeling reagents with differential isotopes (e.g., ^{12}C- and ^{13}C-reagents) are selected to react with the common functional group of metabolites in comparative samples (e.g., an individual sample vs. a pooled sample), followed by mixing the labeled samples for LC-MS analysis. Finally, by merging the submetabolome analysis results, higher coverage of the whole metabolome profiling can

be realized.[39] With the proper selection of the reagents, CIL can improve the LC separation and the ionization efficiency of metabolites.[71] Moreover, the configuration of the LC-MS setup does not need to be changed for routine metabolome analysis, as a result of altering the physical properties of metabolites after labeling. For example, ionic metabolites can be converted into relatively hydrophobic derivatives after reacting with a reagent containing a hydrophobic moiety. All these labeled metabolites can be retained on reversed-phase (RP) LC. If the reagent contains more readily positive chargeable groups, the labeled metabolites can be detected in positive ion mode alone.

While using one LC-MS condition (i.e., RPLC and positive ion detection) is operationally more convenient for metabolite detection and quantification, the use of a proper LC-MS configuration becomes critical for analyzing as many metabolites as possible. In our previous studies, quadrupole time-of-flight (QTOF) MS has been extensively used for CIL LC-MS. However, considering that different types of mass spectrometers have different detection sensitivity, mass resolution, dynamic range, etc., it would be useful to examine the effects of MS instrument type on the detectability of labeled metabolites. For example, understanding the possible differences in detected metabolites using different instruments may help explain the variations of metabolome data generated from similar studies.[79] More importantly, understanding the differences may guide us to develop better ways to improve the overall metabolite detectability by optimizing the hardware configurations.

In this work, two popular platforms, LC-QTOF-MS and LC-Orbitrap-MS, were selected to analyze the same biological samples following the same CIL LC-MS analysis workflow, and then the performance of these two instruments was compared. We used ^{12}C-/^{13}C-dansylation LC-MS to analyze the amine/phenol submetabolome of human urine and serum samples. One unique aspect of this comparison was the use of peak pairs detected as a gauge of analytical performance. Unlike conventional LC-MS, using differential isotope labeling LC-MS, metabolites are detected as peak pairs, while chemical and background noises are detected as singlet peaks.[59] As a result, noise peaks detected from different types of instruments can be eliminated when only the peak pairs from the metabolites are compared.

2.2 Experimental

2.2.1 Chemicals and reagents

All the chemicals and reagents used in this work, unless otherwise stated, were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). For the differential isotope labeling reagents, ^{12}C$_2$-dansyl chloride (DnsCl) and ^{13}C$_2$-DnsCl were from Nova Medical Testing Inc. (NovaMT) (www.novamt.com). LC-MS grade water, methanol, acetonitrile (ACN) were purchased from Honeywell (Edmonton, AB, Canada), and LC-MS grade formic acid was purchased from Thermo Fisher Scientific (Hamton, NH, USA).

2.2.2 Sample preparation

The urine and serum samples used in this study were the universal urine standard (UUS) and the universal serum standard (USS) from NovaMT. They are the pooled urine or serum generated by mixing the same amount of aliquots of 20 healthy individuals. The urine sample was diluted 4 times in LC-MS grade water before labeling. Proteins in serum were removed by protein precipitation using the following procedure. In a microcentrifuge tube, 30 μL of USS was mixed with 90 μL of precooled LC-MS grade methanol, vortexed, spun down and then kept at -20 °C for 1 hr. After that, the mixture was centrifuged at >10000 rpm for 10 min. 90 μL of supernatant was taken from the centrifuged mixture, transferred into a new microcentrifuge tube and dried using a SpeedVac. Finally, the dried serum sample was dissolved in 25 μL of LC-MS grade water, vortexed and spun down.

2.2.3 Dansylation labeling

In this work, dansylation LC-MS was used to analyze the amine/phenol submetabolome. The prepared urine and serum samples as described above were labeled following the labeling protocol published previously.[38] Briefly, in a microcentrifuge tube, 25 μL of the prepared sample was mixed with 12.5 μL of 250 mM sodium carbonate/sodium bicarbonate buffer. The solution was vortexed, spun down and mixed with 12.5 μL of LC-MS grade ACN and 25 μL of freshly prepared DnsCl (18 mg/mL in ACN, for light labeling) or $^{13}C_2$-DnsCl (18 mg/mL in ACN, for heavy labeling), followed by incubation at 40 °C for 45 min. Then 5 μL of sodium hydroxide solution (250 mM) was

added into the incubated mixture, vortexed, spun down and incubated at 40 °C for another

10 min to quench the labeling reaction. Finally, the solution was mixed with 25 μL of

formic acid (425 mM in 1:1 ACN/water) to neutralize the excess sodium hydroxide.

2.2.4 LC-MS analysis

The $^{12}C_2$- and $^{13}C_2$-labeled samples were mixed in equal volume and centrifuged at

12000 rpm for 10 min prior to LC-MS analysis. Two LC-MS platforms, a quadrupole-

Orbitrap mass spectrometer (Q-Exactive, Thermo-Fisher Scientific, Waltham, MA, USA)

coupled with an ultra-high-performance LC system (Vanquish, Thermo-Fisher Scientific,

Waltham, MA, USA), and a QTOF mass spectrometer (maXis II, Bruker, Billerica, MA,

USA) coupled with an ultra-high-performance LC system (Dionex UltiMate 3000,

Thermo-Fisher Scientific, Waltham, MA, USA). Two LC systems were equipped with the

same Agilent Zorbax Eclipse Plus C18 column (2.1 mm× 100 mm, 1.8 μm particle size, 95

Å pore size) for separation, and used the same LC separation condition as follows: mobile

phase A was 0.1% (*v/v*) LC-MS grade formic acid in the same grade water, and mobile

phase B was 0.1% (*v/v*) LC-MS grade formic acid in the same grade ACN. The gradient

elution profile was: t = 0.0 min 25% B, t = 10.0 min 99% B, t = 13.0 min 99% B, t = 13.1

min 25% B, t = 16.0 min 25% B. The flow rate was 400 μL/min, and the column

temperature was kept at 40 °C. There were slight differences in chromatographic peak

shapes of some metabolites using the Vanquish and Ultimate 3000 LC systems. However,

in CIL LC-MS, metabolites are detected as peak pairs, the slight difference in peak shape

would affect neither metabolite identification results nor relative quantification results, because the shape of chromatographic peaks would have the same effect on both light (^{12}C-labeled) peak and heavy (^{13}C-labeled) peak.

All mass spectra were acquired in positive ion mode with the m/z range from 200 to 1000 at a scan rate of 1 Hz for both MS. Orbitrap MS was operated at two different resolutions: 60k which is close to that of QTOF MS (80k, reported by Brucker) and a higher resolution at 120k. The MS parameters used for QTOF MS were: ion mode – positive; dry temperature - 230 °C; dry gas - 8 L/min; capillary voltage – 4.5 kV; nebulizer - 1.0 bar; endplate offset - 500 V; spectra rate - 1.0 Hz. The parameters for Orbitrap MS were: ion mode - positive; sheath gas flow rate - 60 a.u.; aux gas flow rate - 30 a.u.; sweep gas flow rate - 2 a.u.; spray voltage - 3.5 kV; capillary temperature – 350 °C; vaporizer temperature – 350 °C; AGC target - 1e6; maximum inject time - 120 ms (60k resolution), 200 ms (120k resolution).

The mass resolutions of the two instruments were measured for a number of sodium formate cluster ions in the m/z range from 200 to 1000 with different signal intensities (up to 50 times of intensity difference). In the experiment, sodium formate solution (0.1 mM) was directly infused into the ESI source at a flow rate of 0.1 mL/min. The ion source was operated in the positive ion mode and the related parameters were set to be the same as those used for LC-MS analysis. 30 continuous MS scans were collected. The mean values and the standard deviations of mass resolutions were calculated based on 10 continuous

MS spectra. Table 2.1 shows the measured resolutions obtained from the maXis II QTOF MS and the Orbitrap MS. The resolutions for QTOF ranged from ~24k at m/z 226.95 to ~41k at m/z 702.86 for the high abundance ions, while the resolutions were from ~21k at m/z 294.93 to ~31k at m/z 736.85 for low abundance ions. For the Orbitrap operated at 60k resolution, the measured resolutons ranged from ~67k at m/z 226.95 to ~33k at m/z 906.82 for high intensity ions. For the low abundance ions, the resolution ranged from ~57k at m/z 294.93 to ~32k at m/z 940.82. For the Orbitrap operated at 120k resolution, the measured resolutions ranged from ~132k at m/z 226.95 to ~64k at m/z 906.82 for the high abundance ions and ranged from ~115k at m/z 294.93 to ~64k at m/z 940.82 for the low abundance ions. The reduction of mass resolution for the low abundance ions was not as significant in Orbitrap as in QTOF. In summary, the actual resolutions achievable by these two instruments exhibited significant variability compared to the values reported by the vendors, primarily dependent on both the intensities and the m/z of the detected ions.

2.2.6 Metabolite identification

The detected peak pairs were identified or matched by using IsoMS based on a three-tiered approach.[39] The first-tier (positive identification) was obtained by matching accurate mass and retention time (RT) to a labeled standard library (CIL library) composed of over 1600 endogenous metabolites,[82] the second-tier (putative identification with high confidence level) was completed based on searching accurate mass and RT against a linked identity library (LI library) composed of over 7000 pathway-related metabolites with

predicted RT,[39] and the third-tier results (putative match) was identified by matching

accurate mass to the MyCompoundID library (MCID library).[83] The peak pairs with the

same RT and m/z values within the defined search tolerances gave the same list of positive

or putative identities for QTOF and Orbitrap. This was also true for comparing the

identification results between 60k and 120k in the Orbitrap datasets.

Table 2.1 Signal intensities and mass resolutions of sodium formate cluster ions obtained
by QTOF MS and Orbitrap MS. Data are presented as mean ±S.D. (n=10).

	m/z	QTOF MS		Orbitrap MS (60,000)		Orbitrap MS (120,000)	
		Intensity $(\times 10^5)$	Resolution	Intensity $(\times 10^6)$	Resolution	Intensity $(\times 10^6)$	Resolution
High abundance ions	226.95	7.2±0.1	24206±185	104.6±0.2	67002±89	40.9±0.1	132802±331
	362.92	20.3±0.1	32534±193	32.8±0.1	52739±262	13.1±0.1	105206±496
	430.91	33.2±0.3	38860±26	56.8±0.3	47872±124	22.6±0.1	94739±339
	498.90	18.8±0.2	37011±571	39.0±0.1	44972±126	17.3±0.1	90072±47
	566.89	23.5±0.2	39938±273	34.1±0.1	42106±81	20.0±0.2	84206±533
	634.87	19.5±0.2	38905±681	28.9±0.2	39539±94	20.4±0.1	79939±402
	702.86	24.1±0.1	40730±113	31.5±0.4	37836±46	22.4±0.2	75506±163
	770.84	16.6±0.1	39042±65	20.9±0.2	36170±48	19.0±0.2	71872±249
	838.83	12.1±0.2	37214±191	17.9±0.1	34672±47	17.9±0.1	68939±385
	906.82	6.6±0.3	33073±210	11.1±0.1	33039±94	13.5±0.2	66472±43
Low abundance ions	294.93	1.90±0.11	20603±177	7.14±0.23	57306±134	4.10±0.32	114709±226
	381.29	1.53±0.04	22062±237	1.52±0.12	49338±168	0.24±0.03	92403±1024
	472.90	1.13±0.04	22912±344	2.13±0.03	44271±205	0.25±0.02	82868±654
	540.89	2.06±0.03	24102±332	2.74±0.18	42038±171	0.35±0.04	77868±1237
	608.87	3.07±0.04	25221±196	3.04±0.06	39206±244	0.45±0.02	75538±569
	668.87	4.67±0.12	28931±769	5.23±0.15	38139±169	8.23±0.34	77072±169
	736.85	6.96±0.09	31212±246	5.77±0.04	36417±126	8.31±0.13	72606±81
	804.84	5.63±0.08	30282±341	4.86±0.08	34806±163	7.97±0.11	70339±33
	872.83	3.93±0.12	29205±145	3.52±0.09	33306±424	7.36±0.22	67472±249
	940.82	3.00±0.43	27682±399	3.47±0.14	31739±235	6.43±0.26	63971±46

2.2.6 Metabolite identification

The detected peak pairs were identified or matched by using IsoMS based on a three-tiered approach.[39] The first-tier (positive identification) was obtained by matching accurate mass and retention time (RT) to a labeled standard library (CIL library) composed of over 1600 endogenous metabolites,[82] the second-tier (putative identification with high confidence level) was completed based on searching accurate mass and RT against a linked identity library (LI library) composed of over 7000 pathway-related metabolites with predicted RT,[39] and the third-tier results (putative match) was identified by matching accurate mass to the MyCompoundID library (MCID library).[83] The peak pairs with the same RT and m/z values within the defined search tolerances gave the same list of positive or putative identities for QTOF and Orbitrap. This was also true for comparing the identification results between 60k and 120k in the Orbitrap datasets.

2.3 Results and discussion

2.3.1 Urine and serum metabolome analysis

Due to the limited dynamic range of a mass spectrometer, for the co-eluting metabolites with huge concentration differences, some ionized metabolites with low abundance may not be able to be detected. Increasing the injection volume into LC while not causing signal distortion or sample carryover issues, is a common way to maximize the detectable ionized metabolites in LC-MS.[18] On this account, optimization of injection

volume was carried out for both LC-MS platforms using labeled urine and serum samples and the results are summarized in Figure 2.1.

Figure 2.1a displays a plot of the peak pair number detected from labeled urine and serum samples as a function of injection volume on the LC-QTOF-MS platform, illustrating that the maximum number of detected peak pairs could be reached using an injection volume of 6 µL for both labeled samples. Similarly, as illustrated in Figures 2.1b and 2.1c, the optimal injection volume was 4 µL for the LC-Orbitrap-MS platform operated at the resolution of 60k and 120k; this optimal volume was found for both labeled samples. It was also observed that all the relative standard deviations of the detected peak pair number from triplicate injections were below 1.5%, indicating injection replicates of the same sample gave excellent reproducibility on the number of peak pairs detected.

Figure 2.2 shows the base-peak ion chromatograms obtained from 6 µL injection of the dansyl-labeled urine and serum samples in LC-QTOF-MS and 4 µL injection of the same samples in LC-Orbitrap-MS. The retention times of both chromatograms were calibrated using a set of RT calibrants.[58] Many chromatographic peaks were detected across the entire gradient time window, illustrating the metabolome complexity of these samples and good detectability of the methods used.
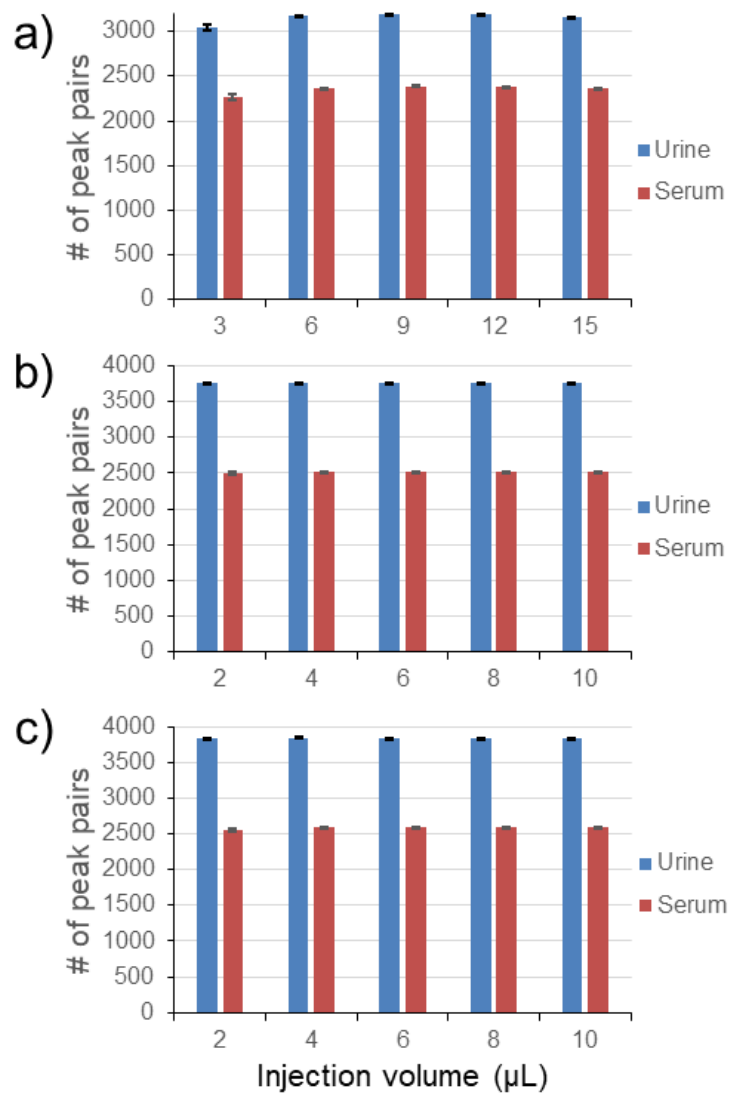
Figure 2.1 Number of peak pairs detected from labeled urine and serum samples as a function of injection volume using a) LC-QTOF-MS operated at a resolution around 80 k, b) LC-Orbitrap-MS operated at a resolution of 60 k, and c) LC-Orbitrap-MS operated at a resolution of 120 k. Data are presented as mean ± S.D. (n = 3).
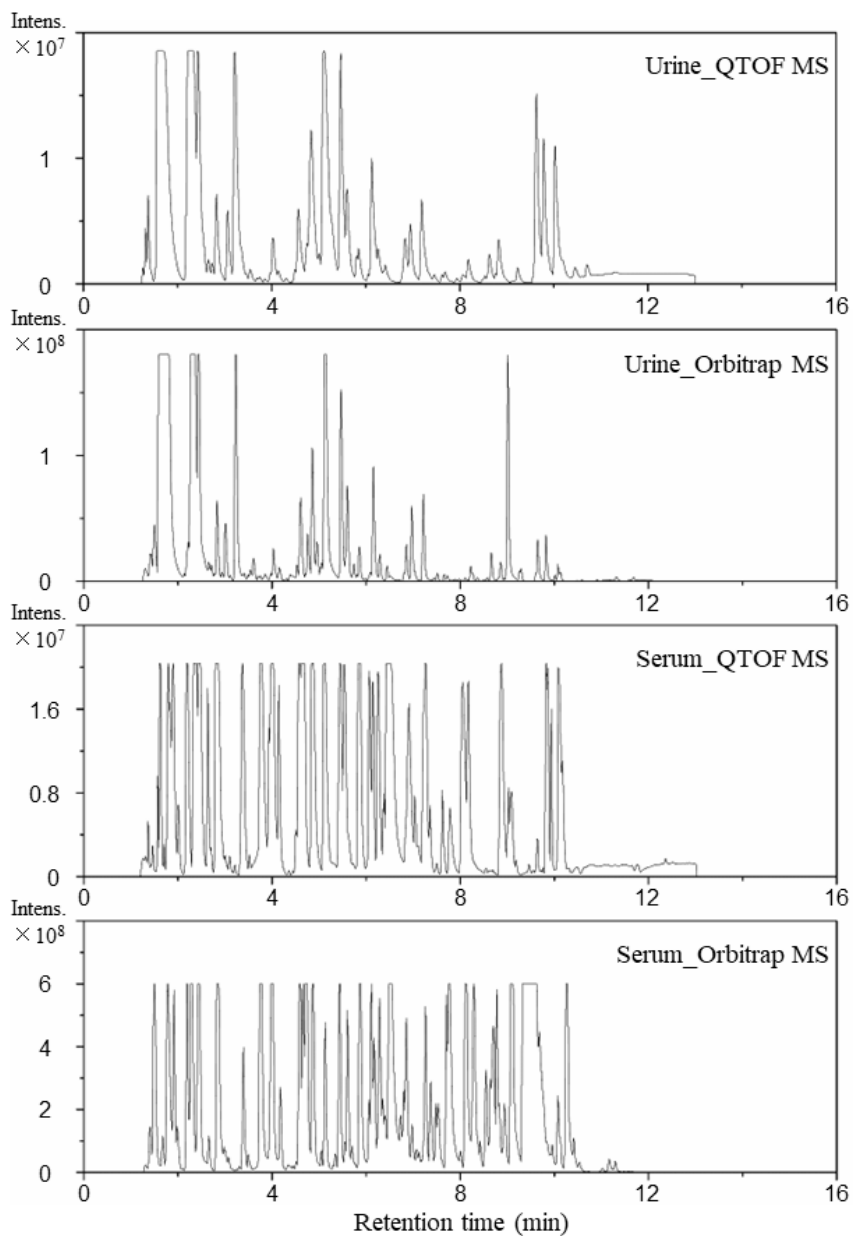
Figure 2.2 Base-peak ion chromatograms obtained from 6 μL injection of the dansyl-labeled urine and serum samples in LC-QTOF-MS and 4 μL injection of the same samples in LC-Orbitrap-MS (a saturated intensity limit was manually set in the chromatograms collected by Orbitrap MS for a fair comparison).

Table 2.2 Number of peak pairs or metabolites detected and identified (or matched) from dansyl-labeled urine and serum samples by different LC-MS platforms operated under the optimal sample injection conditions.

Sample type	Peak pair type	QTOF MS	Orbitrap MS (res. of 60k)	Orbitrap MS (res. of 120k)
Human urine	Total peak pairs	3172	3765	3839
	Tier 1	145	175	175
	Tier 2	352	605	655
	Tier 3	2321	2774	2896
	IDs and matches	2818 (88.9%)	3554 (94.4%)	3726 (97.0%)
Human serum	Total peak pairs	2370	2518	2592
	Tier 1	148	179	176
	Tier 2	307	326	336
	Tier 3	1778	1840	1942
	IDs and matches	2233 (94.2%)	2345 (93.1%)	2454 (94.7%)

The numbers of peak pairs detected and identified (or matched) from urine and serum samples by using the two LC-MS platforms under the optimal injection conditions are summarized in Table 2.2. Overall, more peak pairs were detected by LC-Orbitrap-MS in both urine and serum samples, compared to those detected by LC-QTOF-MS. For urine samples, 3172 peak pairs were detected in total by LC-QTOF-MS. Among them, a total of 2818 peak pairs (88.9%) were identified or matched, including 145 peak pairs positively identified as tier-1 metabolites, 352 peak pairs putatively identified with high confidence as tier-2 metabolites, and 2321 peak pairs matched as tier-3 metabolites. Compared to the above results, more metabolites were identified and matched in tiers 1, 2 and 3 based on the peak pairs detected by LC-Orbitrap-MS. Similar observations were also found in identification results from the serum samples. As shown in Figure 2.1, the injection volume

that Orbitrap MS required to reach the maximum number of peak pairs was lower than QTOF MS, indicating that Orbitrap MS has higher overall sensitivity than QTOF MS.

The numbers of peak pairs detected and identified (or matched) from urine and serum samples by using the two LC-MS platforms under the optimal injection conditions are summarized in Table 2.2. Overall, more peak pairs were detected by LC-Orbitrap-MS in both urine and serum samples, compared to those detected by LC-QTOF-MS. For urine samples, 3172 peak pairs were detected in total by LC-QTOF-MS. Among them, a total of 2818 peak pairs (88.9%) were identified or matched, including 145 peak pairs positively identified as tier-1 metabolites, 352 peak pairs putatively identified with high confidence as tier-2 metabolites, and 2321 peak pairs matched as tier-3 metabolites. Compared to the above results, more metabolites were identified and matched in tiers 1, 2 and 3 based on the peak pairs detected by LC-Orbitrap-MS. Similar observations were also found in identification results from the serum samples. An analysis of overlapping peak pairs between urine and plasma metabolomes was carried out. Only 29.5% - 34.3% of serum peak pairs were also detected in urine samples using difference LC-MS platforms, illustrating the significant difference in metabolome composition between these two sample types. Therefore, Therefore, the selection of the sample type in metabolomics studies should be carefully done, depending on the specific research objectives. As shown in Figure 2.1, the injection volume that Orbitrap MS required to reach the maximum number of peak pairs was lower than QTOF MS, indicating that Orbitrap MS has higher

overall sensitivity than QTOF MS.

Figure 2.3 shows the Venn diagram of the numbers of the peak pairs or metabolites detected in urine or serum by using the two LC-MS platforms. For the comparison, peak pairs with the same accurate mass and calibrated RT within a tolerance window (±5 ppm for accurate mass and ±60 s for calibrated RT) detected in the two platforms were regarded as the same. For the following discussion, we refer peak pairs or metabolites detected by using LC-QTOF-MS as QTOF peak pairs. Orbitrap peak pairs are referred to the peak pairs detected by using LC-Orbitrap-MS with a resolution of 60k. Likewise, HR-Orbitrap peak pairs represent the peak pairs detected in the same way, but at a higher resolution of 120k.
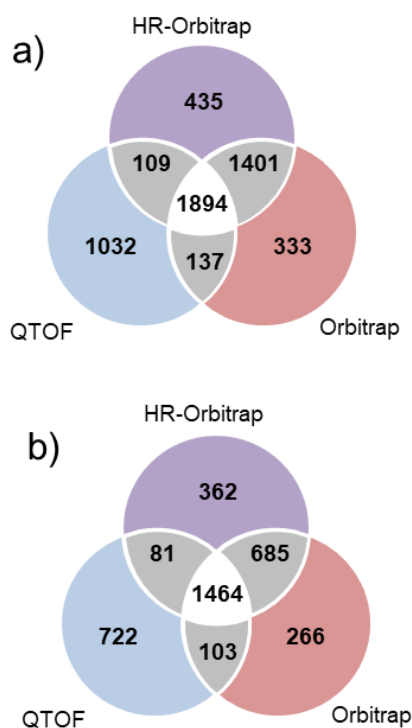
Figure 2.3 Venn diagrams showing the number of the peak pairs or metabolites detected from a) urine samples and b) serum samples by using different LC-MS instruments.

2.3.2 Comparison of QTOF and Orbitrap MS performance at a similar resolution

As illustrated in Figure 2.3, when comparing the QTOF and Orbitrap peak pairs, there were 1141 and 1734 unique peak pairs detected by QTOF from urine and serum samples, respectively. At the same time, 803 and 951 unique peak pairs were detected by Orbitrap from urine and serum samples, respectively. To further reveal the differences in the metabolites detected by the two LC-MS platforms at a similar resolution, Figures 2.4 and Supplemental Figure 2.6 show the distribution of QTOF peak pairs and Orbitrap peak pairs detected at different ranges of signal intensities from urine and serum samples, respectively. For urine and serum, metabolites with relatively low abundances account for a large proportion of all the detected peak pairs. Moreover, ignoring the peak pairs with the lowest intensities, there is a trend that the number of peak pairs or metabolites decreases dramatically as the intensity increases within the instrument detection dynamic range, illustrating the wide distribution of the concentrations of metabolites within a single biological sample and also proving that the CIL LC-MS method increased the metabolome coverage significantly, especially for those metabolites with low concentrations.

Figure 2.4 Number of peak pairs detected from urine samples at different ranges of ion counts using a) LC-Orbitrap-MS at 60k and b) LC-QTOF-MS.

As shown in Figure 2.4, for urine samples, more Orbitrap peak pairs were found in each range of signal intensity compared to the QTOF peak pairs, especially for those with relatively low intensities. On average, higher S/N values were observed in Orbitrap than QTOF, although some individual metabolites might be detected with higher S/N in one instrument over the other (see below). Higher S/N can be achieved by increasing the signal intensity or lowering the background noise.

There are possibly three main sources contributing to the differences of S/N observed. One is related to the generation of ions. Although electrospray ionization (ESI) was used in both platforms, the spray tip and ion source designs from the two manufacturers are different. These design differences can result in different ionization efficiencies of the same metabolite and different extents of ion suppression or enhancement of the metabolite with co-eluting metabolites and/or background chemicals.[84,85] Another factor is related the ion transport process.[86-88] The interfaces used to transport the ions from the ion source to the vacuum system of the mass spectrometers and the ion guides used to move the ions to the mass analyzers are very different. These can contribute to the difference in ion detection efficiency. Finally, the ion detection mechanism for TOF involving time-of-flight separation and multichannel plate detector is very different from that of Orbitrap with an image current detector. The detector responses of metabolite ions and associated noise/background ions may be different. In addition, the highest signal counts are limited by detector saturation in TOF, while they are limited by the trap capacity in Orbitrap. While we cannot readily measure the extent of the contributions of the above three factors for the observed S/N's of the QTOF and Orbitrap peak pairs, on average, the Orbitrap MS achieved a better detection sensitivity and a wider dynamic range, compared to QTOF MS for the analysis of dansyl-labeled urine and serum samples.
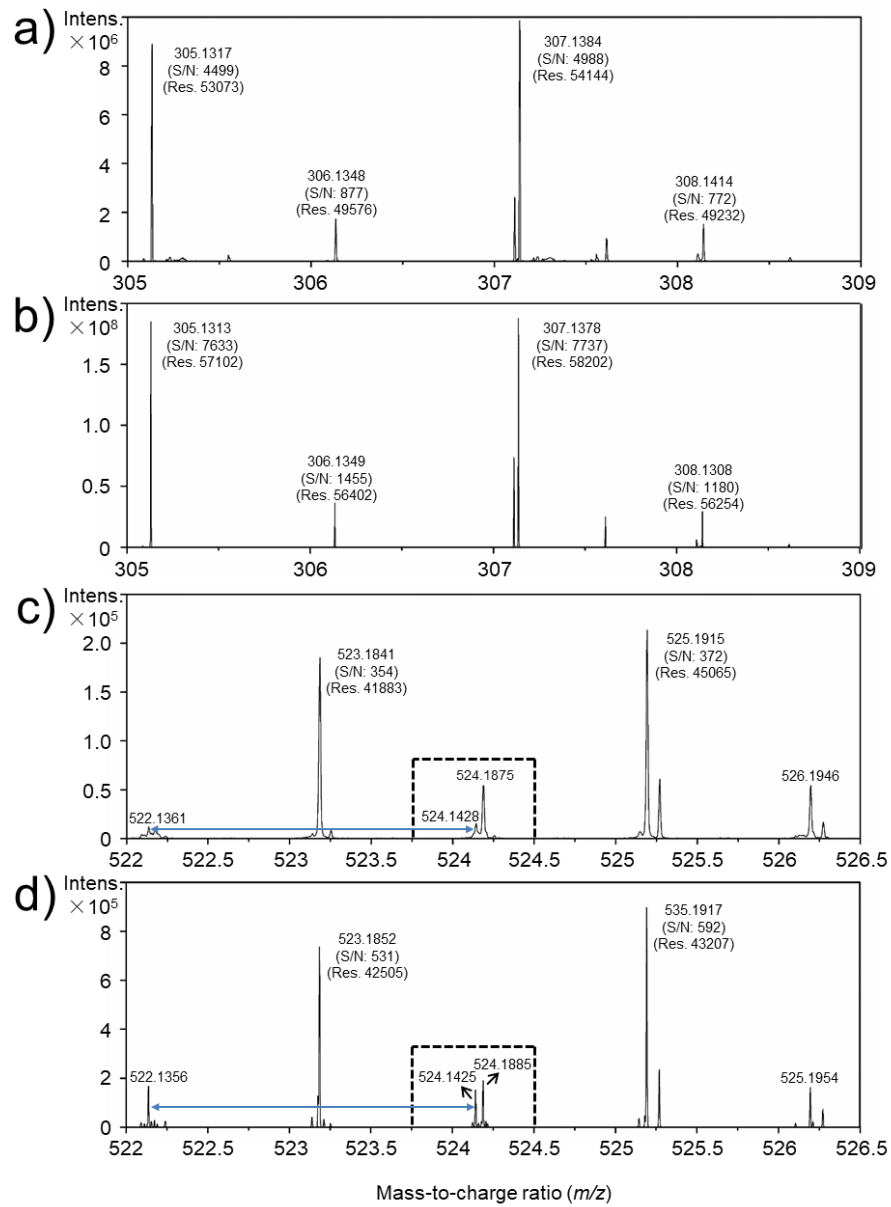
Figure 2.5 Expanded mass spectra of a peak pair detected within the m/z range of 305 to 309 by both a) QTOF MS and b) Orbitrap MS. Expanded mass spectra of two peak pairs detected within the m/z range of 522 to 526.5 by both c) QTOF MS and b) Orbitrap MS.

We have examined the differences of peak pair detection by the two platforms in more detail. Figure 2.5 shows a representative case of a peak pair with higher S/N values observed in Orbitrap (Figure 2.5b), compared to QTOF (Figure 2.5a). As shown in Figure 2.5, the actual difference in resolution was not as pronounced as that we found using sodium formate solution. It appeared that the complexity of the sample may also have an influence on the observed resolution. For most peak pairs detected in the m/z range of 300 to 700 m/z, the two mass spectrometers were operated at a similar resolution according to the collected spectra. Figure 2.4 and Figure 2.6 show that the lower abundance peak pairs account for the differences in the percentages of the Orbitrap peak pairs vs. the QTOF peak pairs. All three factors described above may contribute to the differences. For example, lower signals are more susceptible to ion suppression and ion loss during the transport process can have a larger effect on low-intensity ions.

Figure 2.5c and 2.5d shows the expanded mass spectra of two peak pairs detected by both QTOF MS and Orbitrap MS. While the peak pair at m/z 523.1841 and 525.1915 in QTOF has a similar pattern to those at m/z 523.1852 and 525.1917 in Orbitrap, the peak pair at m/z 522.1361 and 524.1428 in QTOF has much lower S/N than those at m/z 522.1356 and 524.1425 in Orbitrap.

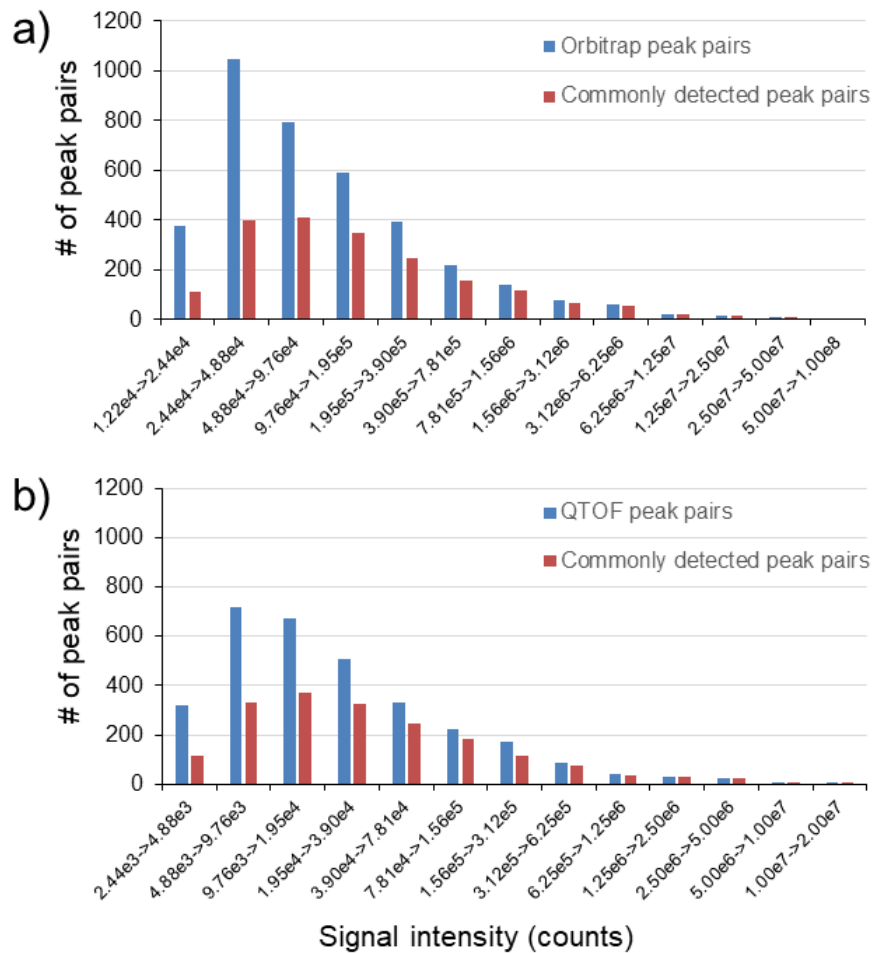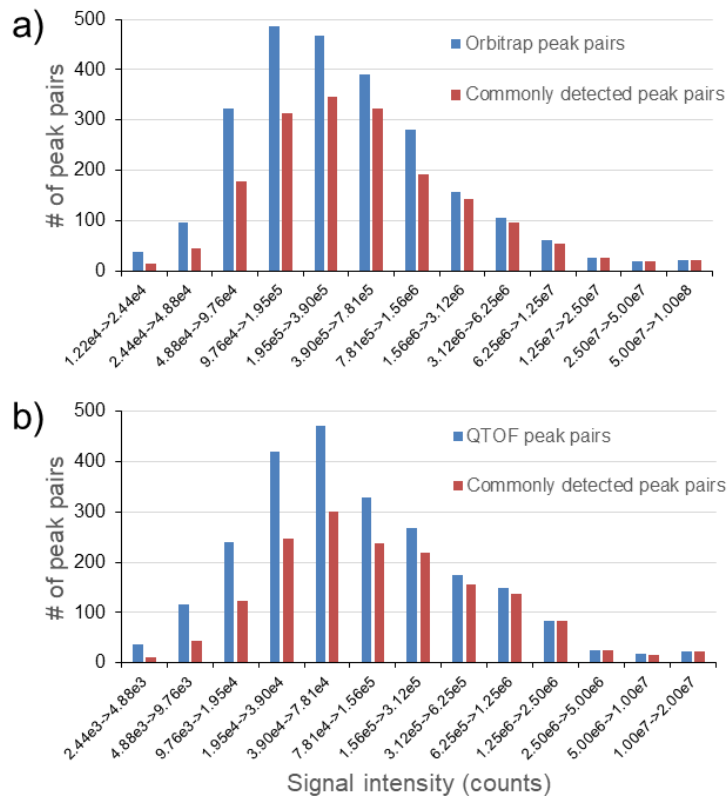**Figure 2.6** Number of peak pairs detected from serum samples at different ranges of ion counts using a) LC-Orbitrap-MS and b) LC-QTOF-MS operated at a similar resolution.

In some cases, a peak pair could be detected by an MS when its intensity was above the detection limit whilst it could be missed by another MS when its intensity was below the detection limit. This is the main reason why most peak pairs unique to one instrument were found with relatively low abundance. As an example, Figure 2.7a shows the unique QTOF peak pair at m/z 353.1171 and m/z 355.1239 that was not detected in Orbitrap (Figure 2.7b), while Figure 2.7d shows the unique Orbitrap peak pair at m/z 421.1219 and m/z 423.1289 that was not detected in QTOF (Figure 2.7c).
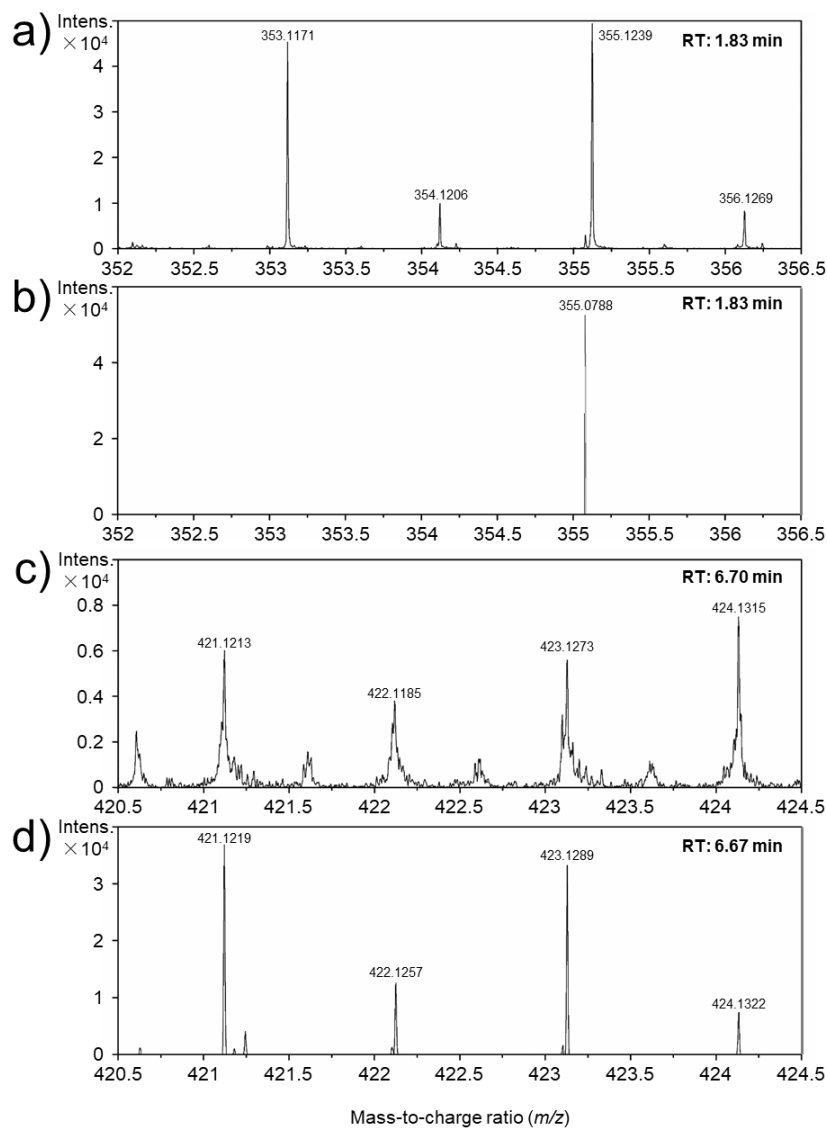
Figure 2.7 Expanded mass spectra of a peak pair uniquely detected by (a) QTOF MS, but not by (b) Orbitrap MS. Expanded mass spectra of a peak pair not detected by (c) QTOF MS with background noises, but clearly detected by (d) Orbitrap MS. Mass spectra presented here were collected at the apex of corresponding chromatographic peak.

We have compared the detectability of metabolites that were positively identified using the CIL standard library in the QTOF and Orbitrap platforms. Figure 2.8 shows the

Venn diagram of the numbers of identified metabolites in the two platforms. Over 90% of the tier-1 metabolites in urine or serum samples were commonly identified by QTOF and Orbitrap, in despite of the tier-1 isomers present in the identification results. Most of the endogenous metabolites in our CIL library have relatively high concentrations in biological samples such as urine and serum. Therefore, they are more likely to be detected, even if one instrument may reduce the signal intensity of a metabolite to some extent; however, the signal may still be above the detection limit. We anticipate that, when the CIL standard library is expanded to include more metabolites that may not be present in high concentrations in biological samples in the future, the detectability difference of identified metabolites from different LC-MS platforms will likely be increased. For example, the unique QTOF peak pair shown in Figure 2.7a was positively identified as homoserine, and the unique Orbitrap peak pair shown in Figure 2.7d was positively identified as indoleacrylic acid. Based on the metabolites identified in this work, we did not observe any specific trends of one instrument having a better chance of detecting a certain type of metabolites.
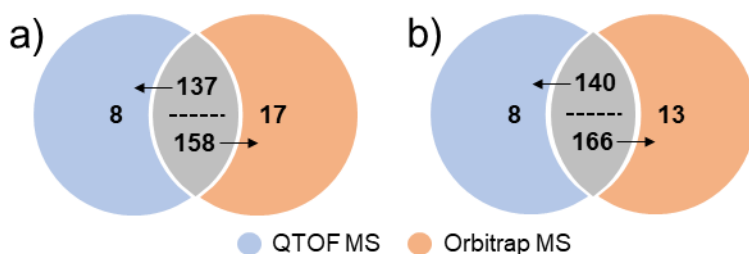
Figure 2.8 Venn diagrams showing the number of tier-1 metabolites detected from a) urine samples and b) serum samples by using different LC-MS instruments. Isomers in the

identification results were considered as the same metabolite during the comparison, leading to slightly different numbers of common peak pairs.

Taken together, the above analyses indicate that, while many common peak pairs could be detected from QTOF and Orbitrap, a significant number of peak pairs were detected uniquely in each platform. Thus, if we combined the two datasets produced from the two platforms, the overall number of peak pairs detected could be increased. For example, for the urine samples, the total number of peak pairs detected by the two platforms was 4906, compared to 3172 using QTOF alone (i.e., 54.7% increase) or 3765 using Orbitrap alone (i.e., 30.3% increase). Similarly, for the serum samples, QTOF detected 2370 peak pairs and Orbitrap detected 2518 peak pairs, and the combined number was 3321. This work also suggests that for generating more reproducible metabolome results in a metabolomics study, using one platform, either QTOF or Orbitrap, to analyze all the comparative samples within the study is critical. In order to increase the number of commonly detectable metabolites from the two platforms, thereby making the metabolome datasets generated from different platforms more comparable, we believe that more studies are needed to better understand the instrumental factors contributing to the differences. In this regard, we examined the mass resolution or resolving power of the Orbitrap to see if mass resolution plays any roles.

2.3.3 Evaluation of Orbitrap MS performance at different resolutions for metabolome analysis

The Venn diagrams in Figure 2.3 show that 544 unique HR-Orbitrap peak pairs and 470 unique Orbitrap peak pairs were detected from the urine samples, while 443 unique HR-Orbitrap peak pairs and 369 unique Orbitrap peak pairs were detected from the serum samples. The number of unique peak pairs and the percentage of unique peak pairs to the total number of peak pairs detected are much lower in the Orbitrap vs. HR-Orbitrap comparison than those in the Orbitrap vs. QTOF comparison. Data acquisition at higher resolution may require the use of a lower spectral acquisition rate in Orbitrap. In this work, a built-in function in the Orbitrap software, Microscan,[88–90] was used to maintain an average spectral collection speed of 1 Hz on the LC-Orbitrap-MS platform operated at different resolutions. In principle, a defined number of microscans (i.e., complete MS scans generated originally from the Orbitrap MS) were averaged to export a new scan and be written into the raw data. Thus, the new scan speed was modified by the Microscan and approximately equal to the original scan speed divided by the defined Microscan value. During the microscans averaging, the signal-to-noise ratio of the new spectra also increased proportionally to the square root of the Microscan value. In other words, for the Orbitrap MS used in this work, the signal-to-noise ratio was sacrificed for a higher resolution data acquisition. Using this type of setup, we compared the performance of the Orbitrap MS at different resolutions for metabolome analysis.

Figure 2.9 and Figure 2.11 show the distribution of Orbitrap peak pairs and HR-Orbitrap peak pairs detected at different ranges of signal intensity from urine and serum samples, respectively. Orbitrap peak pairs had a similar distribution as a function of the signal intensity compared to HR-Orbitrap peak pairs due to the same LC-MS configuration. However, it was also observed that some unique peak pairs were detected at specific resolutions. As discussed above, sensitivity dropped as the Orbitrap MS was switched to a higher resolution. In this case, fewer peak pairs could be detected at the high resolution, resulting in the presence of those unique peak pairs shown in Figures 2.9a and 2.11a. At the same time, the higher resolution also allowed the detection of newly resolved peak pairs, resulting in the presence of those unique peak pairs shown in Figures 2.9c and 2.11b. Figure 2.10 shows an example of a newly resolved peak pair detected by the Orbitrap MS operated at a higher resolution (m/z 473.1905 and 475.1972). Thus, unique peak pairs were detected in the Orbitrap MS operated at either 60k or 120k resolution. As a result, the overall peak pair number detected or the metabolome coverage was increased if we merged the two datasets. In the case of urine samples, the total peak pair number was found to be 4309, representing a 14.4% increase over the 3765 peak pairs detected at 60k and an 11.4% increase over the 3869 peak pairs detected at 120k. For the serum samples, the combined number was 2961, compared to 2518 at 60k and 2592 at 120k.
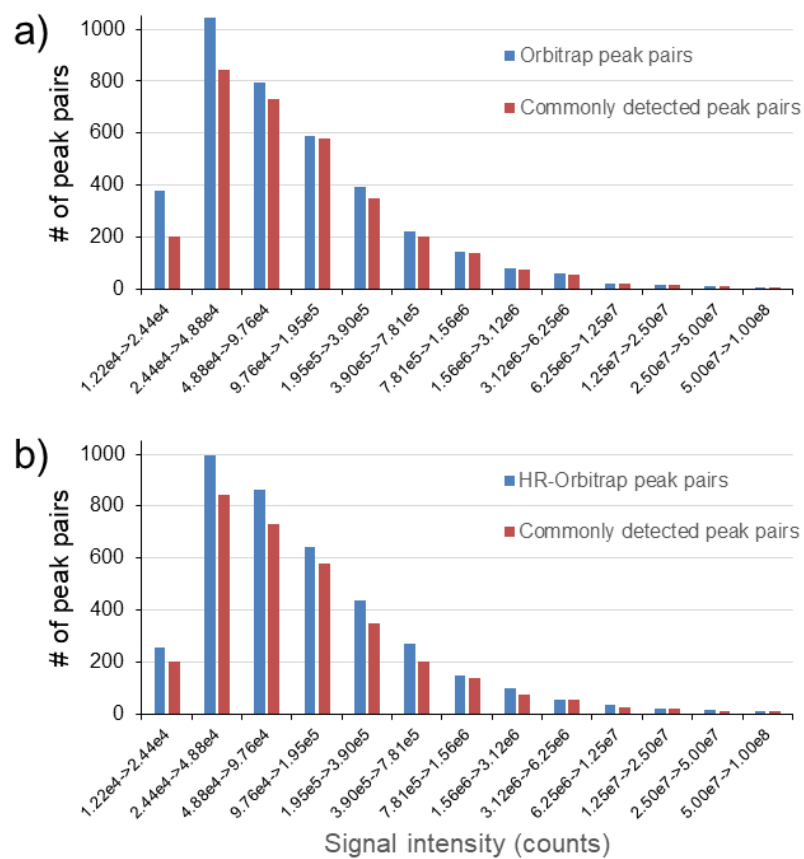
Figure 2.9 Number of peak pairs detected from urine samples at different ranges of ion counts using a) LC-Orbitrap-MS operated at the resolution of 60k and b) LC-Orbitrap-MS operated at the resolution of 120k.
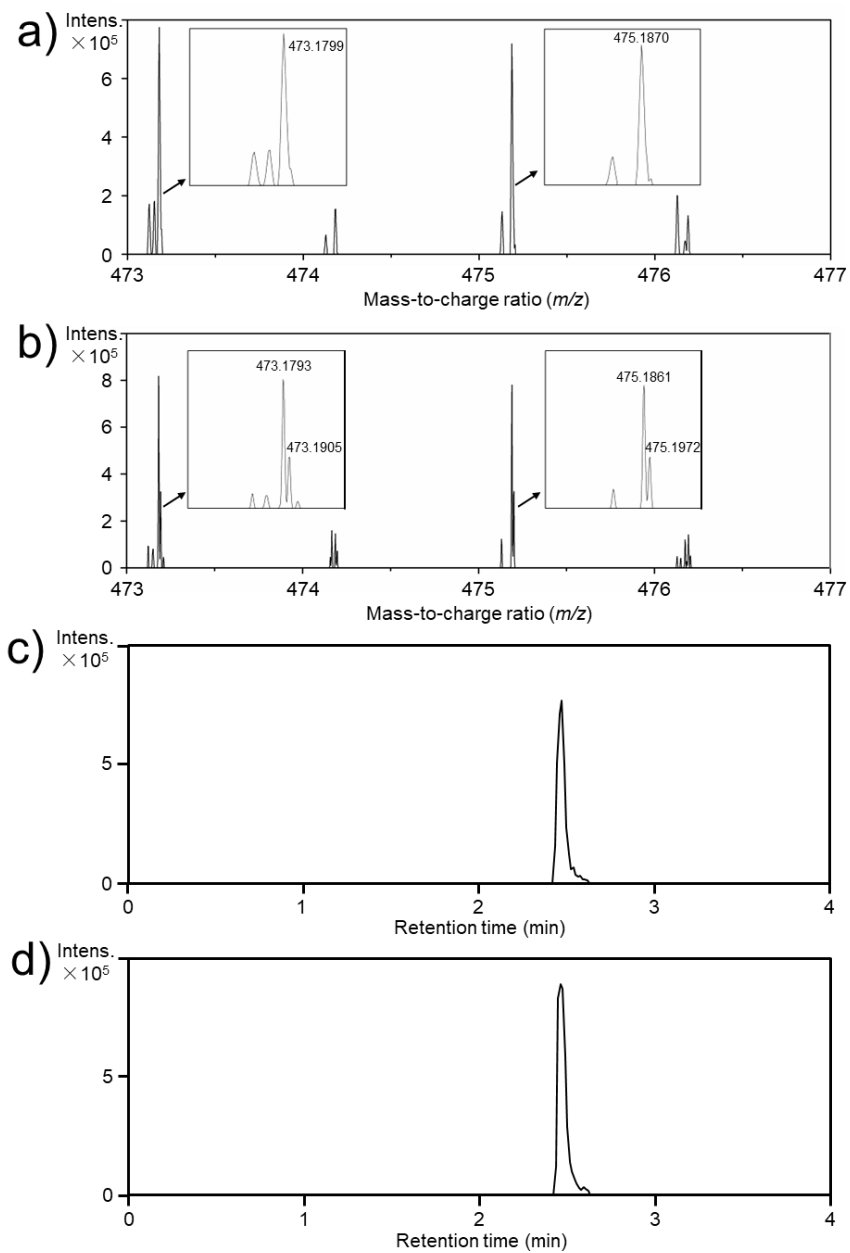
Figure 2.10 Expanded spectra of a peak pair (m/z 473.1905 and m/z 475.1972, tier-3 metabolite) not resolved by the Orbitrap MS operated at the resolution of (a) 60k, but resolved at the resolution of (b) 120k. Extracted ion chromatograms of (c) m/z 473.1905 and (d) m/z 475.1972.
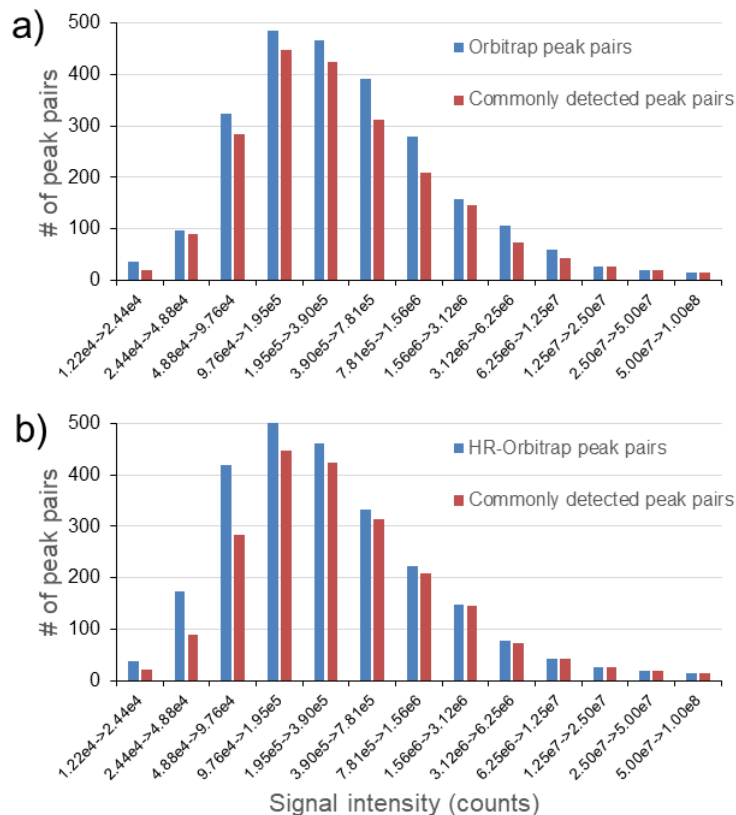
Figure 2.11 Number of peak pairs detected from serum samples at different ranges of ion counts using a) LC-Orbitrap-MS operated at the resolution of 60k and b) LC- Orbitrap-MS operated at the resolution of 120k.

2.4 Conclusions

In this work, the instrument performance of LC-QTOF-MS and LC-Orbitrap-MS for CIL LC-MS-based metabolome analysis has been evaluated. QTOF MS and Orbitrap MS showed different detectability of peak pairs, due to a combination of possible sources of variations including in a) detection sensitivity of MS, b) detection dynamic range of MS, c) ionization efficiency from the use of different spray tip designs, and d) ion transportation efficiency from the use of different ion source designs of the two mass spectrometers. The

combined datasets obtained from QTOF and Orbitrap could provide a significantly higher coverage of the serum or urine metabolome. The performance of LC-Orbitrap-MS operated at different resolutions for CIL metabolome analysis was also evaluated. Higher resolution increased the possibility of resolving overlapped peaks, but at the expense of decreased sensitivity. As a result, the number of peak pairs detected at 120k resolution was similar to that at 60k resolution. However, combining the two datasets obtained at 60k and 120k resolutions could lead to a significant increase in the overall number of peak pairs detected in urine or serum samples. The significant differences in peak pair detection by different instruments, as demonstrated in this work, also suggest that, in order to detect and quantify the common metabolites in different samples, it is important to use the same type of the MS platform. We believe that future research is needed to not only understand the differences in instrument performance, but also significantly reduce the differences with a goal of detecting most, if not all, metabolites commonly by different LC-MS platforms.

Chapter 3

Segment Scan Mass Spectral Acquisition for Increasing Metabolite Detectability in Chemical Isotope Labeling LC-MS Metabolome Analysis

3.1 Introduction

Chemical composition of the metabolome of biological systems is very complex, as evident from large numbers of compound entries in various metabolome databases.[91] More importantly, analytical techniques have been advanced rapidly, allowing the detection of many compounds not detectable using less sensitive methods in the past.[92] On the other hand, due to tight metabolic homeostasis in biological systems, only small changes of metabolite concentrations are detectable in comparative samples with noticeable phenotype differences, requiring the use of a large number of samples to reveal the statistically significant changes.[93–95] Therefore, metabolome analysis faces the challenge of achieving high coverage to detect as many metabolites as possible, while maintaining high sample throughput to analyze many comparative samples.

To meet the challenge, mass spectrometric detection needs to be optimized to analyze metabolites as efficiently as possible. In the area of trapped-ion high-resolution mass spectrometry (MS), either Fourier transform ion-cyclotron resonance (FT-ICR) or Orbitrap, due to the limit of the number of ions that can be trapped, it is not possible to purposely saturate the signals of high-abundant ions in order to detect the co-existing low-

abundant ions, which is more commonly done in other mass analyzers such as quadrupole time-of-flight (QTOF) MS.[18,96] To increase the dynamic range of ion detection in FT-ICR and Orbitrap, a strategy of collecting multiple segmented spectra separately and then stitching them into a full spectrum has been developed.[97,98] Ions within a pre-defined narrow mass range are selected using an ion selector (e.g., a quadrupole device) and entered into the ICR cell or Orbitrap for mass analysis with high resolution and accuracy. The segmented-spectrum ions composed of a smaller number of different ions (i.e., simpler composition) can be trapped in a much larger overall number than that of the same narrow range using the full spectrum ions, increasing the chance of detecting lower abundant ions.

There are several studies showing the significant benefits gained by using segment scan over full scan, particularly in the areas of petroleomics[97,99] and metabolomics,[98,100,101] although extra time may be required to perform multiple segment scans. Petrochemicals such as crude oils represent one of the most complex sample types. Direct injection (DI) ionization that produces continuous ions over a long period is often used, in combination with segment scan, for petrolome analysis using FT-ICR and Orbitrap MS. One report showed the use of repeated sample injections for LC-FTICR-MS-based petrolome analysis with each injection focusing on one segment of a narrow mass range.[102] DI-MS with segment scan has also been used for metabolome analysis with improved coverage compared to full scan in various metabolomics applications.[98,100,101] In addition, segment scan has been shown to increase the accuracy of isotopic ion abundance measurement in

targeted metabolic flux analysis (e.g., by reducing space charge effect) with separate scans of two segments, instead of one full scan, in LC-Orbitrap-MS.[103] However, to our knowledge, there is no report of detailed study of segment scan mass spectral acquisition for increasing metabolomic coverage in LC-MS-based metabolome analysis. This is possibly due to the use of LC for separating a complex sample into simpler co-eluting mixtures, negating the need for using the segment scan approach. In conventional LC-MS, many metabolites cannot be ionized with high efficiencies using electrospray ionization (ESI) or retained and separated effectively in a reversed-phase or other modes of LC. Thus, the mass spectra of co-eluting mixtures at a given retention time are relatively simple.

However, chemical derivatization of metabolites allows the detection of many more metabolites.[71,104–106] For example, in high-performance chemical isotope labeling (CIL) LC-MS, rationally designed chemical labeling reagents are used to derivatize metabolites to enhance LC separation and ESI efficiency simultaneously, resulting in much improved metabolite detectability, compared to conventional label-free LC-MS methods.[38] Using differential isotope labeling (e.g., ^{12}C-reagent labeling of a sample and ^{13}C-reagent labeling of a pooled sample, followed by mixing and LC-MS), accuracy relative quantification of a labeled metabolite in comparative samples can be performed. In this study, we demonstrated a significant increase in the number of metabolites detectable using segment scan Orbitrap MS for chemical isotope labeling (CIL) LC-MS metabolome analysis. One unique aspect of CIL LC-MS is that the labeled metabolites are detected in peak pairs,[59]

while the noise and background ions as well as non-derivatized metabolites are detected in singlet peaks, allowing direct comparison of the number of true metabolites detected in segment scan vs. full scan.

3.2 Experimental section

3.2.1 Segment scan workflow

Figure 3.1 shows the schematic of the workflow for full scan and segment scan mass spectral data acquisition in Orbitrap for CIL LC-MS. Samples are differentially labeled using ^{12}C- or ^{13}C-dansylation reagents. In this work, we used human urine samples for the study, as the urine metabolome is very complex and its analysis would benefit greatly from an optimized workflow. Urine metabolomics is also widely used for biomarker discovery research[107] and thus the developed method reported herein can be directly used for real-world applications. Dansyl labeling targets the analysis of the amine/phenol submetabolome,[108] while other labeling methods can be used for analyzing other chemical-group-based submetabolome (e.g., acid, hydroxyl, and carbonyl).[39] The ^{12}C-/^{13}C-dansyl labeled urine was injected in triplicate (n=3). Dansylated metabolites are more readily ionized as protonated molecules with little other adduct ions formed which are much more stable than their unlabeled counterparts, resulting in simpler spectra and much less in-source or skimmer region fragmentation.[38] Moreover, the labeled metabolites were detected in peak pairs in mass spectra, enabling the use of proper algorithms and software such as IsoMS (a python-based software developed in our lab for CIL LC-MS

data processing) to filter out the redundant peak pairs (i.e., any residual adduct ions, dimers, multimers, heteromers, and fragment ions) to retain only $[M+H]^+$ for a metabolite.[109] Thus, the number of peak pairs detected represented a list of putative labeled metabolites. Noise and chemical backgrounds were detected as singlet peaks and filtered out. This approach of counting peak pairs provided a more accurate assessment of the metabolomic coverage, compared to counting the number of features; many features detected in Orbitrap were not originated from metabolites in samples.

As shown in Figure 3.1, the conventional full scan data acquisition scans the whole m/z range for small molecule detection (i.e., m/z 220 to 1000). The low-mass cut-off of m/z 220 was used, because dansyl labeling increases the molecular ion mass of each metabolite by 234 Da. The advantage of using this cut-off is that the low-mass ions from salts, solvents, impurities and other background molecules, which are often detected in high abundance in conventional or label-free LC-MS, are not trapped, thereby increasing trapping and subsequent detection of analyte ions. In the segment scan data acquisition, the full scan range was divided into multiple segments of narrow mass range with small overlaps (10 m/z) in adjacent segments. Various sizes of segments were examined in this study to determine the optimal conditions in terms of the metabolomic coverage and the analysis time (see below). Data processing for segment scan involved the conversion of the raw mass spectra data into peak lists using msConvert. A sub-list of peaks within a segment mass range was processed using IsoMS[59] to pick the peak pairs, filter redundant peak pairs,

calculate peak intensity ratios, group and align peak pairs from multiple sample runs in the same manner as those used in full scan data acquisition (Figure 1). After peak intensity ratio calculation, peak pairs with distorted ratio (below 0.1 or above 10) will be removed. The peak pairs attributed to the same metabolites from multiple LC-MS runs were aligned based on the retention time and accurate mass. A Zero-filling program[81] was used to retrieve the intensity ratios for the missing peak pairs in the aligned files. Finally, the sub-lists of peak pairs generated by IsoMS were merged and any duplicate peak pairs found in the overlapped m/z ranges were removed by retaining the higher intensity peak pairs.
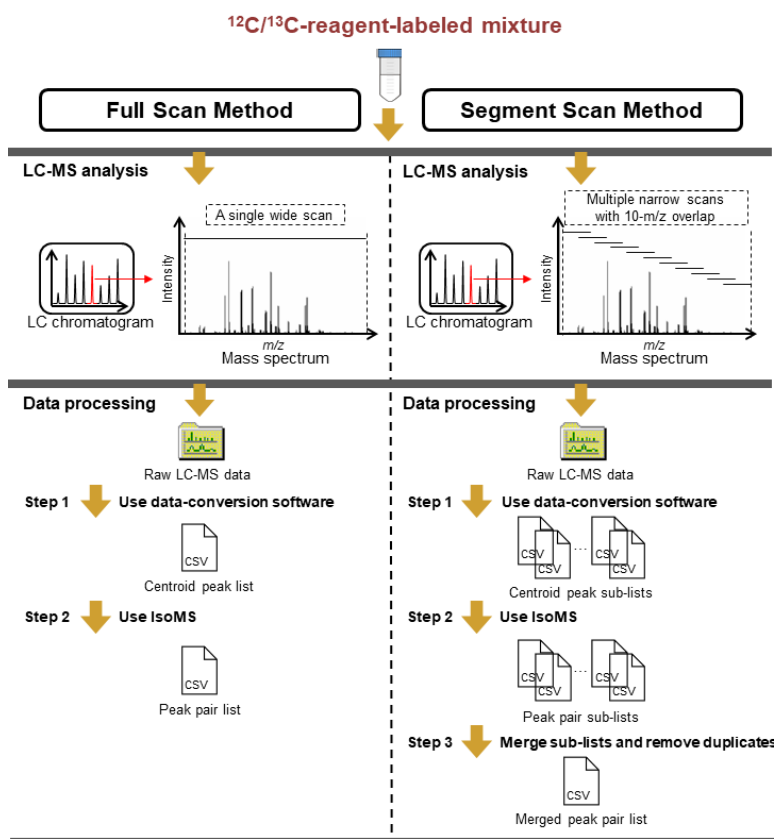

Figure 3.1 Workflow for analyzing the labeled samples in CIL LC−MS using the full scan and segment scan methods.

3.2.2 Dansylation labeling

The urine sample used in this study was the universal urine standard (UUS) from NovaMT, which is a pooled urine prepared by mixing the same amount of aliquots of 20 healthy individuals. The urine sample was diluted 4 times in LC-MS grade water and labeled following the labeling protocol published previously.[39] Briefly, 25 μL of the diluted urine sample was mixed with 12.5 μL of 250 mM sodium carbonate/sodium bicarbonate buffer in a 0.5-mL microcentrifuge tube. The mixture was then vortexed, spun down, and added with 12.5 μL of LC-MS grade ACN and 25 μL of freshly prepared DnsCl (18 mg/mL in ACN, for light labeling) or ^{13}C-DnsCl (18 mg/mL in ACN, for heavy labeling), followed by incubation at 40 °C for 45 min. Next, the incubated mixture was added with 5 μL of sodium hydroxide solution (250 mM), vortexed, spun down, and sit at 40 °C for another 10 min to quench the labeling reaction. The quenched solution was mixed with 25 μL of formic acid (425 mM in 50:50 LC-MS grade ACN/water) to neutralize the excess sodium hydroxide. Finally, the ^{12}C- and ^{13}C-reagent-labeled urine samples were mixed in equal volume and centrifuged at 12000 rpm for at least 10 min to precipitate any proteins present in the mixture and the supernatant was ready for LC-MS analysis.

3.2.3 LC-MS

A quadrupole-Orbitrap mass spectrometer (Q-Exactive HF, Thermo-Fisher Scientific, Waltham, MA, USA) coupled with an ultra-high-performance LC system (Vanquish, Thermo-Fisher Scientific, Waltham, MA, USA) was employed for LC-MS

analysis. The ^{12}C-/^{13}C-reagent-labeled urine sample was injected onto an Agilent Zorbax

Eclipse Plus C18 column (2.1 mm× 100 mm, 1.8 μm particle size, 95 Å pore size). The LC

condition used is listed as follows: mobile phase A - 0.1% (v/v) LC-MS grade formic acid

in the same grade water; mobile phase B - 0.1% (v/v) LC-MS grade formic acid in the same

grade ACN; injection volume - 8 μL; flow rate - 400 μL/min; column temperature – 40 °C,

and a 16-min gradient (t = 0.0 min 25% B, t = 10.0 min 99% B, t = 13.0 min 99% B, t =

13.1 min 25% B, t = 16.0 min 25% B).

The mass spectrometer was equipped with an electrospray ionization (ESI) source,

using the conditions listed as follows: ion mode - positive; sheath gas flow rate – 60 a.u.;

aux gas flow rate – 30 a.u.; sweep gas flow rate – 2 a.u.; spray voltage – 3.5 kV; capillary

temperature – 350 °C; vaporizer temperature – 350 °C. Two different scan modes were

used as follows: full scan or a conventional MS1 data collection mode using a wide m/z

range (i.e., 220 to 1000) and segment scan or a collection of multiple narrow and

overlapping isolation windows with varying scan ranges (see Results). All mass spectra

were acquired at a scan rate of 1 Hz. Two resolutions (i.e., 60,000 and 120,000) were used

for analyzing the labeled urine mixture.

3.2.4 Metabolite identification

The detected peak pairs were identified or matched by using IsoMS based on a

three-tiered approach. The first-tier (positive identification) was obtained by matching

accurate mass and retention time (RT) to a labeled standard library (CIL library) composed

of over 1600 endogenous metabolites, the second-tier (putative identification with high confidence level) was completed based on searching accurate mass and RT against a linked identity library (LI library) composed of over 7000 pathway-related metabolites with predicted RT, and the third-tier results (putative match) was identified by matching accurate mass to the MyCompoundID library (MCID library).

3.3 Results and discussion

3.3.1 Ion capacity and mass accuracy

For mass analyzers involving trapping ions in space, such as Orbitrap and FT-ICR, the detection dynamic range is proportional to the total number of ions that can be trapped, otherwise known as the ion capacity of the trap or ICR cell. In Orbitrap, ions transferred from the quadrupole are temporally stored and an automatic gain control (AGC) system is used to control the ion capacity. Ion number near or at the capacity may lead to space-charge effects which can adversely affect mass measurement accuracy, as well as the accuracy of measuring the ion counts or quantification.[100,110] Therefore, in this work, optimization of ion capacity was carried out first to maximize the dynamic range while not causing adverse effects.

The Orbitrap instrument was initially calibrated using the Thermo ESI positive ion calibration solution. Sodium formate solution was analyzed with three different AGC values (1e6, 5e5, and 1e5) and two different resolutions (60k and 120k). A series of stable cluster ions with known m/z values can be generated from sodium formate, which can be

used to assess mass accuracy. For each AGC value, spectra acquisition (in triplicate) was carried out within a series of 30-m/z isolation windows centered at m/z 226.95, 430.91, 566.88, 702.86, and 838.83 (i.e., the m/z of sodium formate cluster ions). The results obtained are summarized in Figure 3.2A for 60k resolution and Figure 3.2B for 120k resolution. No obvious mass accuracy loss was found as the AGC value increased from 1e5 to 1e6. The values of mass accuracy were distributed randomly over the five selected isolation windows collected at two different resolutions. For this particular instrument, the isolation windows centered at m/z 430.91 gave somewhat higher mass errors than the other ranges of m/z values, but still of less than 1.5 ppm.

Since the composition of sodium formate solution is relatively simple with a few dominant cluster ions shown within the isolation windows, our next set of experiments used the more complex urine samples for the optimization of AGC value. We focused the isolation windows centered at m/z 339.10, 323.11, 399.14, 279.12, and 324.59, corresponding to molecular ions of five labeled metabolites, dansyl-serine, dansyl-alanine, dansyl-phenylalanine, dansyl-dimethylamine and dansyl-tyrosine, respectively. As demonstrated in Figure 3.2C and 3.2D, similar results to the sodium formate solution were obtained even from a much more complex sample, indicating that the space-charge effect was not so strong in Orbitrap MS when using a narrow isolation window for spectral acquisition. With external calibration, the mass accuracy with errors of below 1.6 ppm, as shown in Figure 3.2D is acceptable for the CIL LC-MS method for peak alignment and

metabolite identification. In the following work, the AGC target was set to 1e6 in order to achieve the highest detection dynamic range for the segment scan method.
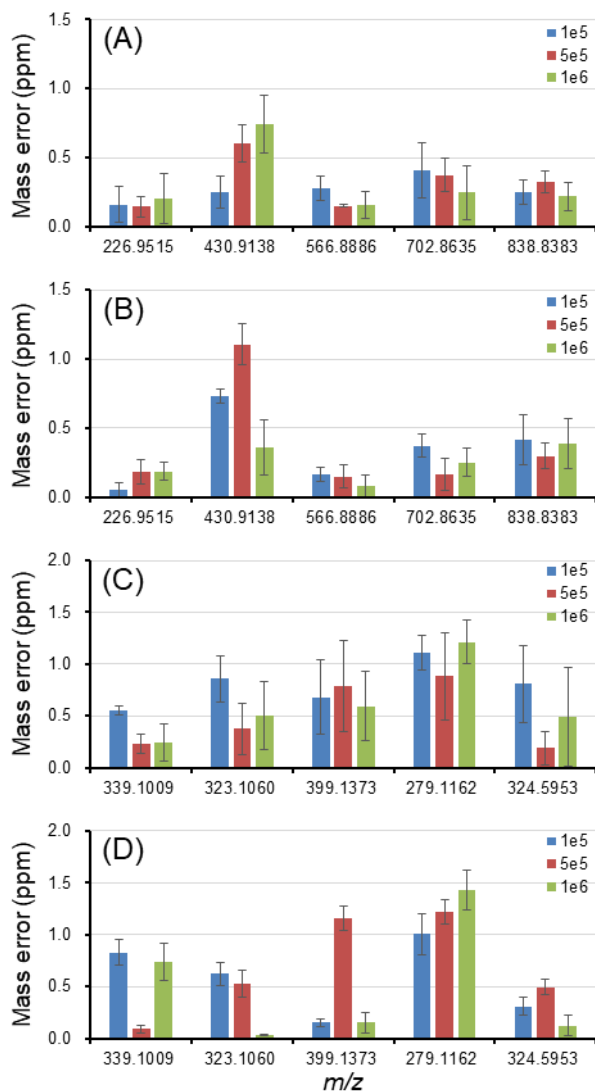

Figure 3.2 Mass measurement errors of five sodium formate cluster ions obtained at a mass resolution of (A) 60,000 and (B) 120,000 and five dansyl-labeled metabolites in urine obtained at a mass resolution of (C) 60,000 and (D) 120,000 with different AGC values. The error bars show the standard deviation of triplicate measurements.

3.3.2 Segment width and peak pairs detected

In segment scan, using a smaller mass range or narrower segment should increase the overall number of analyzed ions, but at the expense of analysis speed. For LC detection, a prolonged analysis would reduce the sample throughput significantly, which is not desirable for metabolomics studies of a large number of samples. To evaluate how segment width affects the number of peak pairs detected, labeled urine sample was analyzed at 60k resolution using a series of isolation windows centered at m/z 339.10, 323.11, and 399.14, with widths increasing from 30-m/z to 780-m/z (780-m/z was used in the full scan mode). For each segment, the number of detected peak pairs per 30-m/z was obtained. As demonstrated in Figure 3.3A, for the spectra collected within the 30-m/z range centered at three m/z values, the peak pair number decreased as the segment width was increased. For example, the percentage increase was 92, 80, and 74% for isolation centered at m/z 339.10 with the segment width of 30-, 40-, and 60-m/z, respectively (Figure 3.3A). Similarly, as Figure 3.3B shows, for segment scan at 120k resolution, the percentage increase was found to be 78, 68, and 62% for isolation centered at m/z 339.10 with the segment width of 30-, 40-, and 60-m/z, respectively.
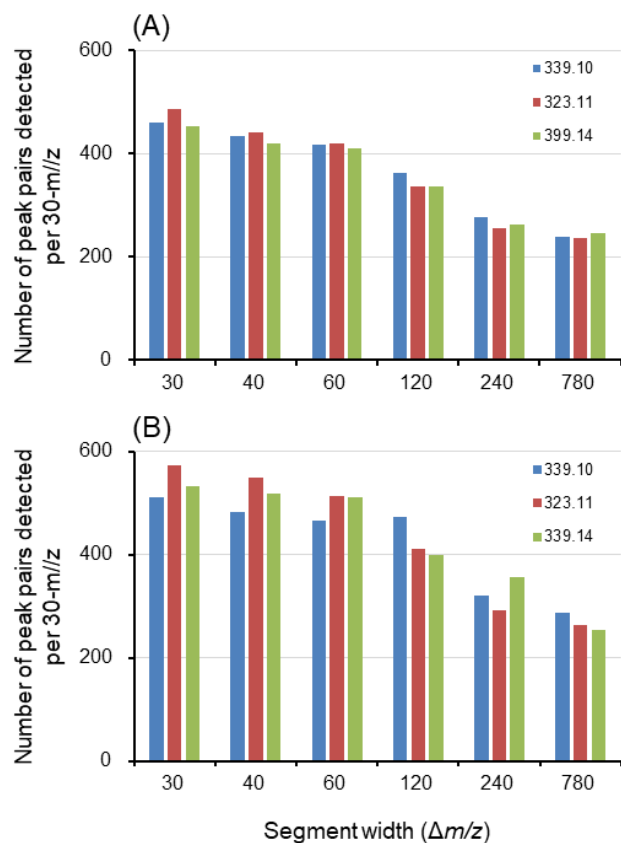
Figure 3.3 Average number of peak pairs detected per 30 m/z window at a mass resolution of (A) 60,000 and (B) 120,000 from dansyl-labeled urine as a function of segment width used in the segment scan method. The peak pair numbers presented in the figure refer to the number of peak pair detected within a specific 30-m/z range during the 16-min LC gradient.

In the case of using the segment width of 120-m/z, the percentage increase depended on the m/z ranges analyzed with the largest gain (51%) when the isolation was centered at m/z 339.10 and the smallest gain (37%) at m/z 399.14. Using the segment width of 240-m/z, the peak pair number increase was relatively small over the full scan mode. It

appears that a limit of gain was reached at the segment width of 60-m/z. This limit is most likely governed by the complexity of the co-eluting metabolites after LC separation. Reducing the segment width does not increase the dynamic range of detection, if the composition of the trapped ions is relatively simple. Thus, the segment scan method benefits the most for analyzing very complex samples, as eloquently demonstrated in DI-MS analysis of petrochemicals and metabolome samples.[98–101] In this regard, when chemical derivatization is used to improve metabolite detectability in CIL LC-MS, many more metabolites from urine become ionizable. As a result, a significant increase in the number of peak pairs detected can be achieved when segment scan is used, even when LC is used to reduce the number of co-eluting metabolites from the urine sample.

3.3.3 Segment number and analysis time

Compared to DI-MS, applying the segment scan method to LC-MS requires the consideration of maintaining the chromatographic peak shape integrity as well as the overall gradient analysis time. A sufficient number of mass spectra needs to be collected to define a fast-rising chromatographic peak in ultra-high-performance LC separation for accurate quantification. Our previous studies have shown an optimal mass spectral acquisition rate at 1 Hz in CIL LC-MS.[59,111] For the Orbitrap used in this study, when the AGC target is set to 1e6, the highest scan speed that the instrument can reach is 7 scans per second at 60k resolution and 3 scans per second at 120k resolution. In segment scan, the overall scan time is distributed to each m/z isolation window when the AGC value is

reached. As a result, the scan speed for each segment is approximately equal to the overall scan speed divided by segment number. In order to maintain an average spectral acquisition speed of 1 Hz, every single LC-MS analysis can only accommodate at most 7 and 3 segments at the resolution of 60k and 120k, respectively. If segment number exceeds the limit (7 or 3), additional injection(s) of the same sample is needed, which would prolong the overall analysis time for a sample and require more samples. The latter may not be possible in the analysis of samples of limited amounts.

We examined the correlation of segment width, number of segments and the overall data acquisition time. Table 3.1 shows the results at 60k and 120k resolutions. In all cases, a 10-m/z overlap between adjacent segments was used to eliminate the edge effect.[112,113] At 60k resolution, the use of segment width of 120-m/z required 7 segments to cover the full scan range. In this case, only one sample injection was needed, allowing direct comparison with the full scan LC-MS experiment. However, if the segment width was reduced to 60-m/z, 15 segments were required, necessitating the use of three injections to cover the full scan m/z range. At a segment width of 30-m/z, 38 segments were used with six injections. At 120k resolution, even setting the segment width at 140-m/z, 6 segments were needed, requiring two injections. The use of a 60-m/z segment width required 15 segments and five injections.

The last column in Table 3.1 lists the relative number of peak pairs (in % increase relative to full scan) in each condition. Compared to the full scan method, the segment scan

method was able to detect 56, 64, and 82% more peak pairs when using a segment width of 120-, 60-, and 30-m/z at 60k resolution. Considering that the total data acquisition time for these widths was 16 (one injection), 48 (three repeat injections), and 96 min (six repeat injections), respectively, the segment width of 120-m/z should be considered as an optimal compromise, as it could detect 56% more peak pairs while using the same analysis time as the conventional full scan method. At 120k resolution, 74, 81, and 89% increase in the detected peak pairs was achieved using the segment scan method with a segment width of 120-, 60-, and 30-m/z, respectively. The peak pair number increase was higher using 120k resolution, compared to 60k resolution, if the same data acquisition time was used. If two injections per sample were used, operating the Orbitrap at 120k with the segment width of 140-m/z performed better than at 60k resolution. However, the moderate-higher increase of 73% with two injections at 120k vs. 56% with one injection at 60k may not justify the use of two injections per sample, particularly if many samples are analyzed in a metabolomics project.

Taken together, the results in Table 3.1 show that the segment scan method reaches an optimal compromise between the gain in the number of detected peak pairs and the overall analysis time, when the instrument is operated at 60k resolution with the segment width of 120-m/z. It is remarkable to see a net gain of 56% more peak pairs detected, compared to the full scan method, using the exactly same analysis time (e.g., increasing from 3765 pairs in full scan to 5867 pairs in segment scan). The results in Table 1 also

show that, if achieving higher coverage of detection is more important than the analysis speed, multiple repeat sample injections can increase the peak pair number from 3843 in full scan to 7273 in segment scan (89% gain) when the Orbitrap was operated at 120k with the segment with of 60-m/z.

Table 3.1 Relations of segment width, number of segments, total data acquisition time, and number of peak pairs detected at 60,000 or 120,000 resolution for the segment scan method.

Resolution	Segment width (m/z)*	Number of segments	Data acquisition time for each sample (min)**	Total number of peak pairs detected	Percent increase relative to full scan (%)
60,000	780 (full scan)	1	16	3765	-
	120	7	16	5867	56
	60	15	48	6170	64
	30	38	96	6863	82
120,000	780 (full scan)	1	16	3843	-
	140	6	32	6673	74
	95	9	48	6955	81
	60	15	80	7274	89

*Adjacent segments had an overlap of 10-m/z.

**16-min LC-MS analysis per injection (e.g., 48-min acquisition time required 3 repeat injections).

3.3.4 Detection dynamic range

We took a closer look at the major reasons behind the very significant increase in the number of peak pairs detected. Figure 3.4A shows the expanded full scan spectrum obtained at 60k resolution, while Figure 3.4B shows the same region of the spectrum obtained using the optimal segment scan method. The m/z 295.1101 ions were the dominant ions in full scan, masking the analysis of some low abundance ions (Figure 3.4A inset). In contrast, some low abundance ions were clearly detected in segment scan (Figure 3.4B inset). Figure 3.4C shows a plot of the number of peak pairs detected as a function of

ion signal intensity (counts) in multiple pins. In both 60k and 120k resolutions, more peak pairs were detected in segment scan than full scan in all intensity ranges. However, higher proportions of gains were observed for peak pairs with lower signal intensities. These results are in line with what one would expect from segment scan, i.e., the increased detection dynamic range in segment scan allows more low abundance ions to be detected, compared to full scan.
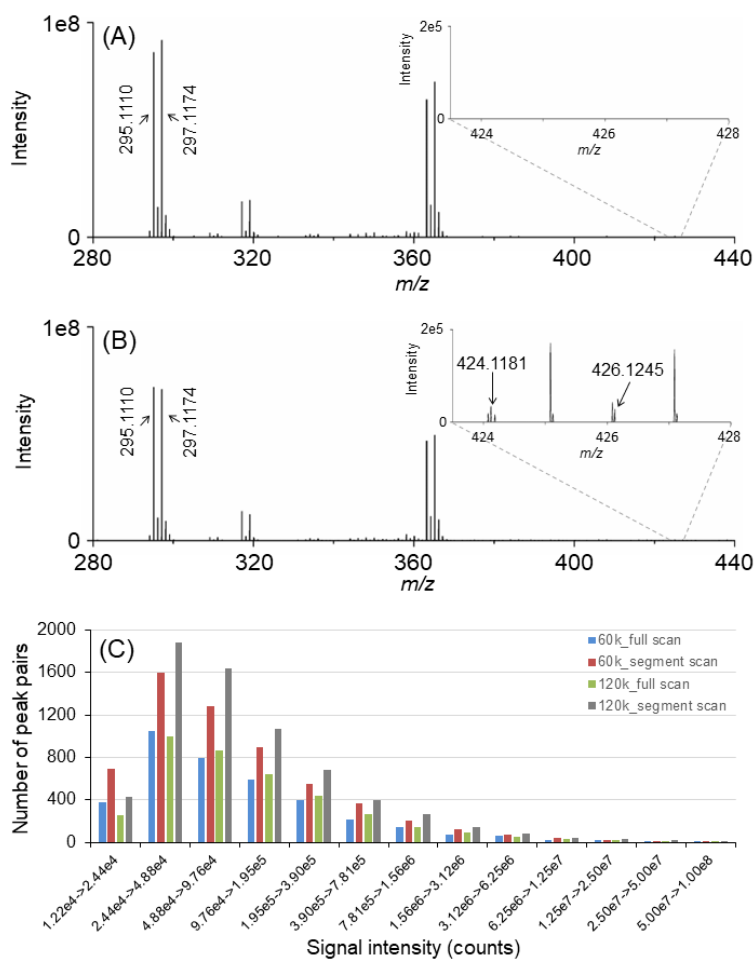

Figure 3.4 Expanded mass spectra of dansyl labeled urine collected at the retention time of 2.38 min using (A) the conventional full scan method and (B) the optimal segment scan

method. (C) Number of peak pairs detected as a function of ion signal intensity in different bins.

It should be noted that dansyl labeling of metabolites has the effect of reducing the ionization efficiency differences of metabolites with vastly different structures, as the structure of the dansyl moiety determines the ionization efficiency of a dansyl labeled metabolite.[82] Because of the similarity of ionization efficiencies of labeled metabolites, the ion intensities reflect much more closely the metabolite concentrations, i.e., the labeled metabolites detected with low ion intensities are the ones with low concentrations in the sample. For example, the peak pair with m/z 295.1110 in Figure 4A was identified as ethanolamine, and the low intensity peak pair with m/z 424.1181 was identified as glycyl-aspartate. Based on the relative intensity of these two peak pairs, we could estimate that ethanolamine was around 2000 times more concentrated than glycyl-aspartate in the sample. Thus, the segment scan method allows the detection of low concentration metabolites better than the full scan method.

3.3.5 Metabolite comparison

The Venn diagrams in Figure 3.5 show the number of peak pairs detected using the two different scan methods. In comparison, peak pairs obtained by two different scan methods but with an accurate mass error below 5 ppm and a calibrated RT error below 60 s were regarded as the same. As illustrated in Figure 3.5, at both resolutions, the peak pairs detected using segment scan could cover more than 93% of those detected using full scan.

Therefore, there is a net gain of many newly detected peak pairs using segment scan in CIL LC-MS, resulting in a significant improvement in metabolomic coverage.
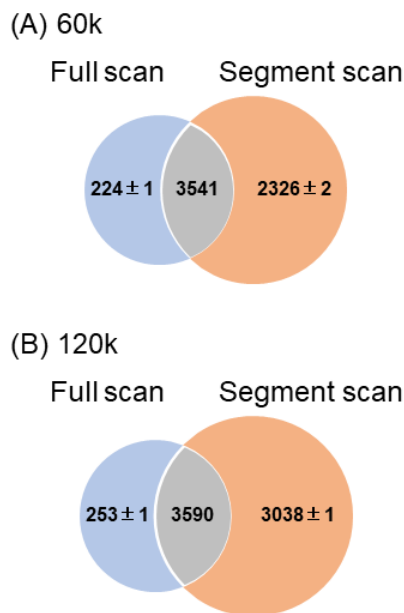

Figure 3.5 Venn diagrams showing the number of metabolites detected using the conventional full scan method and the optimal segment scan method at a mass resolution of (A) 60,000 and (B) 120,000.

The peak pairs detected were identified using the three-tiered approach.[114] The distribution of the numbers of identified or matched peak pairs is shown in Table 2. Out of 3765 peak pairs detected using the full scan method at 60k resolution, 3520 peak pairs (93.5%) could be identified or matched. Among them, 157 peak pairs were positively identified as tier-1 metabolites, 605 peak pairs were putatively identified with high confidence as tier-2 metabolites, and 2759 peak pairs were matched as tier-3 metabolites. Compared to the above results, more peak pairs were detected using segment scan, most of

which were the tier-3 metabolites. Similar observations were also found in the comparison of identification results obtained at 120k resolution (Table 2). The results shown in Table 2 indicate that most of the newly detected labeled metabolites using segment scan have not been positively identified using the current databases. It is clear that more development work is needed to expand the CIL standard library in the future.

Table 3.2 Comparison of metabolite identification results between the optimal segment scan method* and the full scan method collected at 60,000 or 120,000 resolution.

Resolution		Full scan method	Optimized segment scan method
60,000	Total peak pair number	3765	5867
	Tier 1 ID number	157	158
	Tier 2 ID number	604	712
	Tier 3 matched number	2759	4705
	Number of IDs and matches	3520 (93.5%)	5575 (95.0%)
120,000	Total peak pair number	3843	6673
	Tier 1 ID number	159	163
	Tier 2 ID number	654	774
	Tier 3 matched number	2792	5340
	Number of IDs and matches	3605 (93.8%)	6277 (94.1%)

*The optimal segment width at 60,000 and 120,000 resolution is 120 m/z and 140 m/z, respectively.

3.3.6 Peak ratio measurement

In metabolomics, in addition to achieving as high coverage as possible, accurate quantification of the metabolic changes among comparative samples is also critical. Using the segment scan method, the absolute intensities of detected peaks may be affected during segmented scans and stitching processes. The observed edge effect with much reduced signal intensities for ions with m/z closer to the cutoff boundaries is an example. Even

73

using 10-m/z overlap to eliminate the edge effect, duplicate peaks can have different intensities in the two segments. Figure 3.6A and Figure 3.6B show an example of the same peak pair detected in two adjacent segments. The absolute intensities are very different, making the process of stitching segment spectra into a full spectrum more challenging, which may affect the absolute quantification accuracy.

CIL LC-MS can overcome this quantification challenge. While isotope standards are commonly used as a reference for accurate quantification of targeted metabolites, conventional or label-free LC-MS for untargeted metabolome analysis is semi-quantitative, as it does not use the isotope references. However, in CIL LC-MS, a pooled sample with almost the same composition (metabolite types and concentrations) as those of the individual samples of the same type (e.g., a pooled urine from individual urine samples) is labeled with a heavy isotope reagent to form a reference for all light isotope labeled individual samples.[114] Peak intensity ratios of the differentially labeled metabolites are measured to reflect the relative abundance differences among different samples. In this work, we analyzed the labeled urine samples with a theoretical ratio of 1.0 between the light and heavy labeled metabolites. The peak ratios of the same peak pairs in Figure 3.6A and Figure 3.6B are similar and close to 1.0. Thus, we can accurately perform relative quantification based on the peak ratio value, while the absolute intensity may be affected in the stitching process. In fact, the segment scan method used in CIL LC-MS does not stitch the segment spectra into a full spectrum. Instead, we generate the peak pair sub-lists

of individual segments and then combine them into a full list with the removal of duplicate peak pairs in adjacent 10-m/z overlap regions.

We examined the global distributions of the peak ratios of all peak pairs detected in full scan or segment scan. Figure 6C shows the number of peak pairs as a function of peak area ratios in multiple bins. The percentage of peak pairs with the ratios in the range of 0.8-1.25 (±25% from 1.0) was 87% in full scan and 81% in segment scan. Thus, the measurement accuracy did not change much in segment scan, compared to full scan. Figure 6D shows the number of peak pairs as a function of the relative standard deviation (RSD) of peak area ratios from triplicate measurements. The percentage of peak pairs with RSD<10% was 89% in full scan and 88% in segment scan. The precision of the measurement in segment scan did not change much either, compared to full scan.
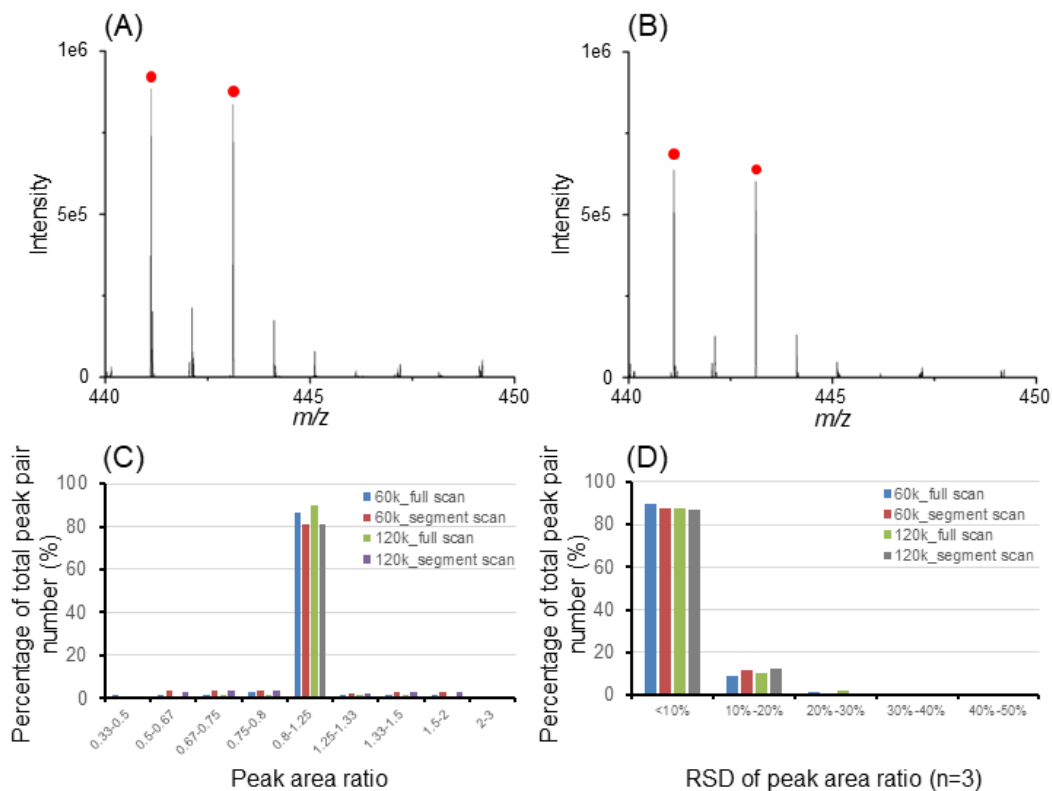
Figure 3.6 Expanded mass spectra of dansyl labeled urine collected at the retention time of 3.04 min at the overlapped regions of two adjacent segments: (A) segment scan from m/z 330 to m/z 450 and (B) segment scan from m/z 440 to m/z 560. (C) Number of peak pairs detected as a function of peak area ratio from the 1:1 mix of ^{13}C-/^{12}C-dansyl-labeled urine. (D) Number of peak pairs detected as a function of relative standard deviation (RSD) of the measured peak ratios (n=3).

3.4 Conclusions

We have developed a segment scan method (i.e., 120-m/z segments at 60,000 resolution and 140-m/z segments at 120,000 resolution) in Orbitrap MS for improving metabolite detection efficiency in CIL LC-MS-based metabolome analysis. Compared to

the conventional full scan method, segment scan allowed the detection of many more metabolites. For example, in a 16-min LC-MS analysis of dansyl labeled human urine samples, 5867 peak pairs or metabolites (not features) were detected in segment scan, compared to 3765 peak pairs detectable in full scan, representing a 56% gain. This remarkable increase in detectable metabolites indicates the complexity of chemical compositions of biofluids such as urine. The LC elutes, even after efficient separation (e.g., UHPLC in reversed-phase mode), can still contain many co-eluting components. Segment scan MS offers a means of improving the detectability of these co-eluting metabolites, thereby increasing the overall metabolomic coverage, while maintaining a similar sample analysis speed, for global metabolomics studies. To implement this method for routine CIL LC−MS metabolome analysis of samples with different metabolomic complexity, there will be a need to optimize the segment number and analysis time to maximize the performance for a specific type of sample. We expect that a higher gain in the number of detectable peak pairs can be achieved when using

the segment scan method to analyze a more complex sample, compared to the conventional full scan method. Once the segment number and analysis time are optimized for a specific type of sample, it should become part of the standard operating procedure for metabolome analysis of this type of sample.

Chapter 4

Unraveling the Potential of Segment Scan Mass Spectral Acquisition for Chemical Isotope Labeling LC-MS-based Metabolome Analysis: Performance Assessment across Different Types of Biological Samples

4.1 Introduction

Metabolomics seeks to profile a wide range of metabolites with accurate quantitative and qualitative information.[115,116] Depending on the research objectives, different types of biological samples are required, with invasive and non-invasive collection methods involved. Invasive methods provide direct access to specific compartments or tissues in a biological system, samples like blood/plasma, cerebrospinal fluid, and tissues are collected by penetration or extraction. On the other hand, non-invasive method is less invasive, posing minimal risk to participants. Samples collected in this way such as feces, urine, sweat, and saliva may have lower metabolite coverage, especially those organ-/tissue-specific metabolites. Regardless of which sample collection method to be used, the sample amount is always limited. However, within a certain type of biological sample, metabolites have a wide range of physicochemical properties and concentrations.[117,118] Moreover, the structures and concentrations of metabolites are highly sensitive to various interconnected environmental factors.[6] This challenge underscores the importance of employing sensitive and robust analytical techniques compatible with different sample types for metabolomics research.

To meet these requirements, a high-performance chemical isotope labeling technique has been developed in our lab for liquid chromatography-mass spectrometry (LC-MS) metabolome analysis.[38,40–42] Compared to the conventional LC-MS method, a well-designed chemical labeling reagent is used to derivatize metabolites to improve both LC separation and ionization efficiency, allowing for more metabolites to be detected. Moreover, the usage of differential isotope labeling (e.g., ^{12}C-reagent vs. ^{13}C-reagent) allows accurate relative quantification.[111] In principle, individual samples are labeled with the ^{12}C-reagent and then mixed with a ^{13}C-reagent-labeled pooled sample. For each ^{12}C-labeled metabolites, there is always a corresponding ^{13}C-labeled metabolite available, serving as an internal standard during the LC-MS analysis.

In CIL LC-MS, the enhancement in metabolite detectability calls for optimization of the mass spectrometric parameters to capture the full dynamic range of metabolite signals. This becomes particularly important in Fourier-transform mass spectrometry (FTMS) like Orbitrap MS and Fourier-transform ion cyclotron resonance (FT-ICR), where the MS detection is normally limited by the finite dynamic range of image current detector.[119] To improve the overall in-spectrum dynamic range, a strategy using multiple narrow windows for spectra collection was developed in FTMS.[98,120] Instead of using a wide scan range, ions collected in a narrower isolation window tend to show less diversity in intensity difference, in this case, for those ions with lower abundance, they get more chance to get detected.

In our previous work,[121] we have applied this strategy of optimizing mass spectrometric detection in CIL LC-MS-based metabolome analysis using Orbitrap MS. This approach has been demonstrated to improve the overall metabolite detection efficiency from human urine samples. However, to establish this method as a routine approach for CIL LC-MS metabolome analysis and meet the specific requirements of different sample types, further optimization of parameters is necessary. In this work, we evaluated the performance of the segment scan method in CIL LC-MS analysis across various common sample types, including feces, urine, plasma, cell extracts, and saliva. A segment scan-assisted CIL LC-Orbitrap MS method has been developed as a routine assay for metabolome analysis.

4.2 Experimental

4.2.1 Chemicals and reagents

All the chemicals and reagents used in this work, unless otherwise stated, were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). HPLC grade water was purchased from Honeywell (Charlotte, NC, USA). ^{12}C- and ^{13}C-dansyl chloride were from Nova Medical Testing Inc. (NovaMT) (www.novamt.com).

4.2.2 Workflow

Figure 4.1 shows the schematic of the workflow for this study. In addition to urine, the performance of segment scan method in CIL LC-MS for another four major sample types (i.e., feces, blood, cell lines, and saliva) was evaluated. For each type of biological

sample, it was split into two aliquots, followed by ^{12}C-labeling and ^{13}C-labeling. The light-labeled sample and the heavy-labeled sample were then mixed in a 1:1 ratio to obtain a ^{12}C-/^{13}C-labeled mixture. The purpose of this mixing was to ensure that, in theory, the concentrations of the ^{12}C-labeled metabolites were equivalent to those of the ^{13}C-labeled metabolites in each type of biological sample. In CIL LC-MS, different from chemical noise or other backgrounds, every labeled metabolite will show up as a peak pair, therefore, the performance of segment scan method can be evaluated based on the enhancement in detected peak pairs compared to those obtained using conventional full scan method. In segment-scan-assisted CIL LC-MS analysis, the full scan range was divided into multiple isolation windows with a 10-m/z overlap. For a given mass scan range (e.g., in CIL LC-MS, a 220-1000 m/z range has been used to accommodate the mass shifts caused by isotopic labeling while minimizing the low-mass ions attributed to salts, solvents, impurities and other background molecules), various segment widths were examined to find the optimal balance between analysis performance and the additional analysis time required. The choice of segment width affected the number of segments used and consequently impacted the total analysis time for LC-MS.

Data acquired using segment scan method was processed following the procedure as described in section 3.2.1. Briefly, data processing was conducted in segments. Peaks in each segment/isolation window were extracted from the raw data using msConvert[80], and then processed by IsoMS[59] through several key steps including peak pairs picking,

redundant peak pairs filtering, peak intensity calculation, peak pairs alignment from multiple sample runs in the same way as those used in the full scan where data was acquired in a wide scan/window. After peak alignment, a zero-filling program[81] was used to retrieve the intensity ratios (light peak/heavy peak) for the missing peak pairs in the aligned files. Finally, the sub-lists of peak pairs generated from IsoMS were merged and any duplicate peaks found in the overlapped m/z range were removed.

4.2.3 Fecal sample collection and preparation

All animal experiments were conducted according to a protocol that was approved by the Ethics Approval Board of the University of Alberta, Edmonton, Canada. A total of 40 healthy female C57BL/6 mice were housed in groups under standard conditions, fed a standard chow, and allowed access to sterile water. A spot feces collection method was used to avoid any potential urinal contamination in this work, briefly, the mouse was lifted from the cage by holding its tail, leading to a quick defection due to stress. Fresh fecal pellets were collected into a 1.5-mL Eppendorf tube. For further sample processing, a fecal pellet was transferred to a new 1.5-mL Eppendorf tube, followed by a six-run subsequential solvent extraction. In each run, 100 μL of solvent was added into the tube, and mixed with the feces by vortexing for 5 min to generate a homogenized suspension. The suspension was then centrifuged at 12,000g for 10 min to get the supernatant. For the first three runs, water was used as the solvent, while ACN was used for the remaining three runs. Supernatants obtained from the extraction process were combined, mixed sufficiently,

dried down using a SpeedVac, and re-suspended in 100 μL of H$_2$O/ACN (50:50 vol%) for

later chemical isotope labeling.
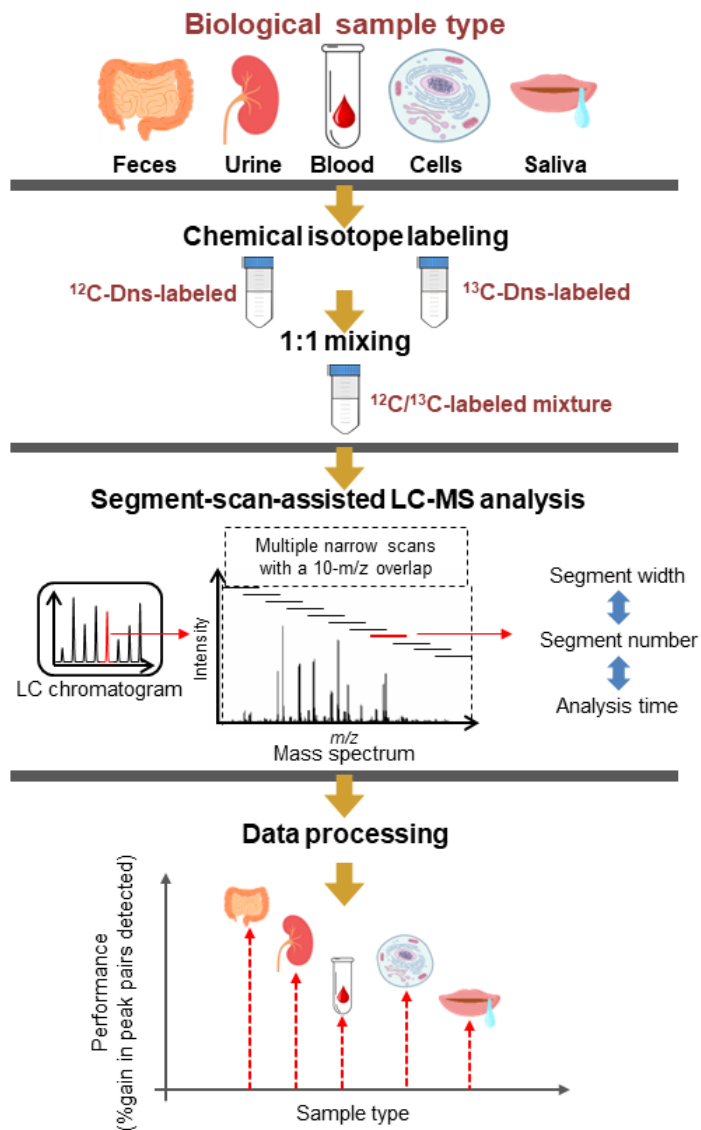
Figure 4.1 Workflow of performance comparison of segment scan method in CIL LC-MS

for different types of samples.

4.2.4 Serum sample preparation

A universal serum standard (USS) from NovaMT was used for this work, which is a pooled serum by combining an equal number of aliquots derived from 20 healthy individuals. A protein removal process was conducted before chemical isotope labeling. In a 500-μL Eppendorf tube, 60 μL USS was mixed with 180 μL MeOH adequately. The mixture was then placed in a -20 °C freezer for 1 hr to precipitate the protein in the USS, followed by centrifugation at 12,000g for 10 min. The supernatant was then dried down using a SpeedVac and resuspended in 50 μL of water.

4.2.5 Cell extract sample preparation

Cell culture: Cell lines A549 were used in this work. The cells were cultured on a 2D surface in Gibco™ Ham's F-12K (Kaighn's) medium (Fisher Scientific, Hamton, NH, USA). The culture was incubated at 37 °C in a humidified atmosphere with 5% CO_2.

Cell harvest: Initially, the growth medium was removed, and the cell cultures were washed three times with cold phosphate buffer saline. Subsequently, 1 mL of cold MeOH was added to quench the metabolites. The cells were detached by scraping and transferred into 1.5-mL Eppendorf tubes, followed by a MeOH removal process using a SpeedVac.

Cell lysis: A freeze-thaw-cycle method was employed for cell lysis. In principle, 200 μL MeOH/H_2O (50:50 vol%) was added into cell pellets. The Eppendorf tube was placed in liquid nitrogen for 1 min, and then thawed in water for 1 min (i.e., a freeze-thaw cycle). After five cycles, the tube was centrifuged at 16,000g for 10 min to obtain the

supernatant, which was then transferred to a new 1.5-mL Eppendorf tube, dried down using a SpeedVac, and resuspended in water to get cell line extracts.

4.2.6 Saliva sample preparation

In this study, human saliva samples were collected in accordance with prevailing human research ethics guidelines. Approximately 1 mL of saliva was collected from the donor in a 1.5-mL tube during the afternoon, after a mouth rinse with water. The collected saliva samples underwent a protein precipitation process similar to the one described in section 4.2.3. 240 μL of MeOH was added into 80 μL saliva to precipitate protein. After centrifugation, the supernatant was dried down and resuspended in 50 μL water.

4.2.7 Chemical isotope labeling

The labeling protocol was identical for all the prepared biological samples, following the same procedures described in Chapter 3.

4.2.8 LC-MS analysis

A quadrupole-Orbitrap mass spectrometer (Q-Exactive HF, Thermo-Fisher Scientific, Waltham, MA, USA) coupled with an ultra-high-performance LC system (Vanquish, Thermo-Fisher Scientific, Waltham, MA, USA) was employed for LC-MS analysis. The ^{12}C-/^{13}C-reagent-labeled sample mixtures were injected onto an Agilent Zorbax Eclipse Plus C18 column (2.1 mm× 100 mm, 1.8 μm particle size, 95 Å pore size). The LC conditions used for both MS and MS/MS were the same as those reported in Chapter 2.

The mass spectrometer was equipped with an electrospray ionization (ESI) source, using the conditions listed as follows: ion mode - positive; sheath gas flow rate – 60 a.u.; aux gas flow rate – 30 a.u.; sweep gas flow rate – 2 a.u.; spray voltage – 3.5 kV; capillary temperature – 350 $^{\circ}$C; vaporizer temperature – 350 $^{\circ}$C. Two different scan modes were used as follows: full scan or a conventional MS1 data collection mode using a wide m/z range (i.e., 220 to 1000) and segment scan or a collection of multiple narrow and overlapping isolation windows with varying scan ranges (see Results). All mass spectra were acquired at a scan rate of 1 Hz. Resolution was set to 60,000 for MS detection.

4.2.9 Metabolite identification

The detected peak pairs were identified or matched using IsoMS with a three-tiered approach[82], providing varying levels of confidence and specificity based on the available libraries and matching criteria. In the first tier (positive identification), accurate mass and retention time (RT) were compared to a labeled standard library known as the CIL library. This library comprises over 1600 endogenous metabolites. For the second tier (putative identification with a high confidence level), accurate mass and RT were cross-referenced against a linked identity library referred to as the LI library. This library consists of over 7000 pathway-related metabolites with predicted RT values. In the third tier (putative match), accurate mass was compared to the MyCompoundID library (MCID library)[83] to identify potential matches.

4.3 Results and discussions

4.3.1 Metabolic complexity among different sample types

Figure 4.2 shows the peak pairs number detected under the optimal injection volume from five types of samples using the convention full-scan-based LC-MS method. The peak pair number reflects the extent of metabolite coverage in a sample. A higher peak pair number suggests a broader representation of metabolites, indicating a more complex sample composition. Conversely, a lower peak pair number may indicate limited metabolite coverage and a less complex sample. As shown in Figure 4.2, based on the number of peak pairs detected, feces and urine samples show the highest complexity in composition, followed by serum and cell extract samples, the lowest sample complexity was found in human saliva metabolome.
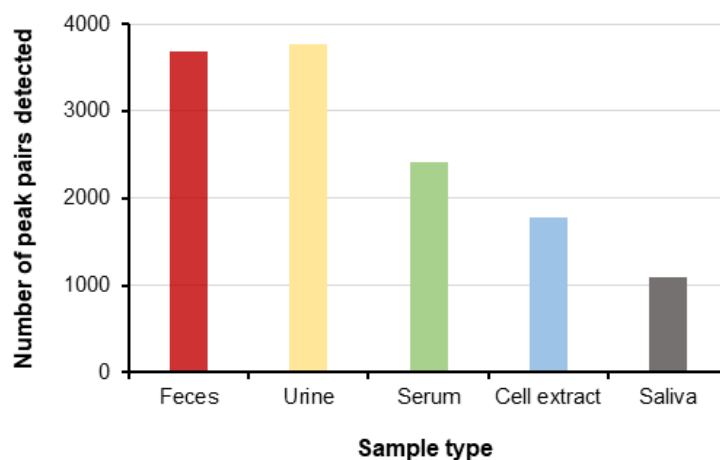
Figure 4.2 Number of peak pairs detected from five types of samples using the conventional full-scan-based LC-MS method.

These observations are consistent with those found in our previous study.[122–125] Within biological systems, blood assumes a crucial function as the principal transporter of

diverse substances throughout the body.[12] It transports molecules, dissolved gases, nutrients, and metabolic wastes to every tissue and organ. This transportation ensures that the necessary substances reach the cells and tissues that require them for proper functioning. As a result, the concentration of metabolites in cells and tissues is generally lower compared to their levels in the blood, as they are continuously consumed. Saliva, a biofluid produced by the salivary glands, is believed to receive certain metabolites from the blood through the infusion process in salivary glands.[126,127] Thus, saliva is a good alternative biofluid for disease diagnosis and prognosis when blood sampling is not feasible. However, the complexity of the saliva metabolome is generally lower compared to the metabolomes of cells/tissues or serum/plasma due to the infusion process in salivary glands. In the case of urine metabolome, since urine is generated in kidneys where soluble metabolic waste and excess water are extracted from the blood, metabolites show a higher concentration distribution in urine compared to those in blood[128], allowing for more metabolites to be detected in CIL LC-MS. Fecal metabolome differs from the other four metabolomes as it comprises downstream metabolic waste following gastrointestinal ingestion and digestion. The microbiome[129,130] is also a component of the fecal metabolome, with the microbiome playing a role in the gastrointestinal tract. Consequently, the fecal metabolome exhibits an even wider range of metabolites.

4.3.2 Evaluation of segment scan performance on different sample types

In segment scan, a higher overall detection dynamic range can be achieved by employing a narrower segment for spectra acquisition to cover the metabolome being analyzed, while the overall analysis time is also increased with a higher requirement of overall segment scan, leading to poor sample throughput. Considering the different metabolic complexity in different biological samples, a proper segment width should be used to maximize the performance of segment scan method in the metabolomics of a specific sample type.

In LC-MS-based metabolomics, acquiring a sufficient number of mass scans is crucial to obtain chromatographic peaks with integrity, ensuring accurate quantification. Our previous work[59,111] has demonstrated that a mass spectral acquisition speed of over 1 Hz is necessary for CIL LC-MS. For the Orbitrap MS employed in this work, the overall scan speed is around 7 spectra per second when it is operated at a resolution of 60,000. Therefore, a single LC-MS run can only accommodate up to 7 segments in order to maintain a good LC result, in this case, the total number of segments should be (close to) a multiple of 7 when one selects a proper segment width. In this work, the optimal performance of segment scan method was determined based on three segment widths, 120, 60, and 30 m/z, with 7, 15, and 38 segments involved, respectively. As shown in Figure 4.3, for urine and fecal samples, a significant improvement in metabolite detectability, around 64% and 56% increase in detected peak pairs, was obtained when the mass spectral

acquisition mode was changed from full scan (780 m/z) to a 120-m/z segment scan method. A further improvement was observed when a narrower segment width was used, 72% and 94% more peak pairs were detected from fecal samples within 60-m/z and 30-m/z segments, 64% and 82% more peak pairs were detected from urine sample within 60-m/z and 30-m/z segment. Similar observations were also obtained in serum and cell extract samples while with relatively lower MS detection improvements. 17%, 17%, and 36% more peak pairs were detected from serum samples within 120-m/z, 60-m/z, and 30-m/z segments, respectively, and 31%, 39%, and 53% more peak pairs were detected from cell extracts within 120-m/z, 60-m/z, and 30-m/z segments, respectively. Notably, there was only a minor increase (approximately 5% more detected peak pairs) in the results of the saliva metabolome analysis when comparing the 120-m/z and 60-m/z segment scan methods to the conventional full scan method. A more substantial enhancement in metabolites detection (25% more detected peak pairs) was observed within 30-m/z segments, although requiring a 96-min LC-MS analysis for a single sample.
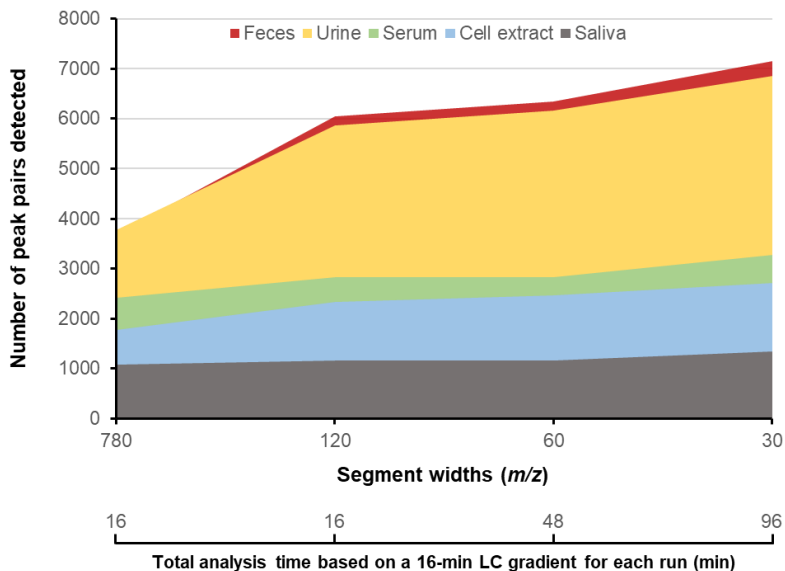
Figure 4.3 2-D area chart showing the peak pairs number detected from five types of samples using segment-scan-assisted CIL LC-MS methods with different segment sizes (Results present for urine samples were obtained in Chapter 3).

Overall, although the results revealing the highest coverage of metabolome profiling were obtained using the 30-m/z segment scan method, the total LC-MS analysis time for a single sample was 96 min based on a 16-min LC gradient (i.e., three injections were needed), resulting in low sample throughput for real-world applications. Alternatively, the 120-m/z segment method offered satisfactory results without requiring additional analysis time compared to the full scan method. Figure 4.4 shows the global intensity distribution of the detected peak pairs from different samples using the 120-m/z segment scan method vs. the conventional full scan method. In CIL, the overall ionization efficiency of the labeled metabolites is equalized with the presence of dansyl group. The intensities of those detected metabolites can reflect their original abundances in the biological samples.

In Figure 4.4, it was observed that the majority of metabolites in all four sample types exhibited relatively low abundances, resulting in low intensities of the detected peak pairs. In the case of saliva samples, the limited diversity of metabolites contributed to a reduced number of low-abundance compounds present in the metabolome. Consequently, the performance of the segment scan method was not as high as that in analyzing more complex samples, as the improved in-spectrum dynamic range primarily focused on ions with low intensities.

In summary, the 120-m/z segment scan method can be established as a routine approach for CIL LC-MS, effectively enhancing the detection efficiency of metabolites in Q-Orbitrap MS. This method demonstrates compatibility with various sample types, particularly those with high or medium metabolic complexity. Although, for less complex samples with limited metabolite diversity such as saliva and sweat, the practicality of the 120-m/z segment scan method is not as high as that of more complex samples, the overall metabolome coverage could still be improved by employing the 30-m/z segment scan method as an alternative strategy.

4.3.3 Comparison of metabolite identification

As shown in Table 4.1, in the fecal metabolome, a total of 3,682 peak pairs were detected using the conventional full scan method. Out of these, 3,426 peak pairs (93.0%) could be identified or matched. Among the identified peak pairs, 266 were classified as tier-1 metabolites, 539 as tier-2 metabolites, and 2,621 as tier-3 metabolites. When

comparing the results to the optimal segment scan method, it was observed that most of the

newly detected peak pairs using the segment scan method belonged to the tier-3 metabolites.

This distribution pattern was consistent across other metabolomes as well. Although these

newly detected peak pairs could not be positively identified at the current stage, they still

hold value in statistical analysis for biomarker discovery, pathway analysis, and other

biological studies. These unidentified peak pairs can provide valuable information and

contribute to the overall analysis. If any of these peaks become of interest, further specific

structural analysis can be conducted using MS/MS spectra elucidation to determine their

identities.

.3.4 Peak pair ratio assessment

Accurate quantification ensures that the reported metabolite concentrations reflect

the true physiological conditions, enabling researchers to draw meaningful and accurate

conclusions about biochemical pathways, metabolic interactions, and the impact of

external factors on metabolism. In CIL LS-MS-based metabolomics, accurate relative

quantification is achieved based on the light-/heavy-labeled peak area ratio. In this work,

we analyzed the global distribution of peak ratios for all peak pairs detected in both full

scan and segment scan methods obtained from various sample types. The theoretical peak

area ratio of a detected peak pair should be equal to 1 since two aliquots of biological

samples were light- and heavy-labeled following the same protocol and mixed in a 1:1 ratio

in volume. In Figure 4.5, we plotted the number of peak pairs against peak area ratios,

categorized into multiple bins. Among different types of samples, approximately 81%-90% of peak pairs in the full scan and 72%-84% in the segment scan fell within the range of 0.8-1.25 (±25% from 1.0). The slight reduction in measurement accuracy observed in the segment scan can be attributed to the inclusion of a greater number of low-intensity peaks. Furthermore, Figure 4.6 illustrates the number of peak pairs plotted against the relative standard deviation (RSD) of peak area ratios, obtained from triplicate measurements. Among different types of samples, it was found that about 82%-96% of peak pairs in the full scan and 76%-92% in the segment scan exhibited an RSD of less than 10%. Hence, the precision of the measurement in the segment scan remained relatively stable compared to the full scan.
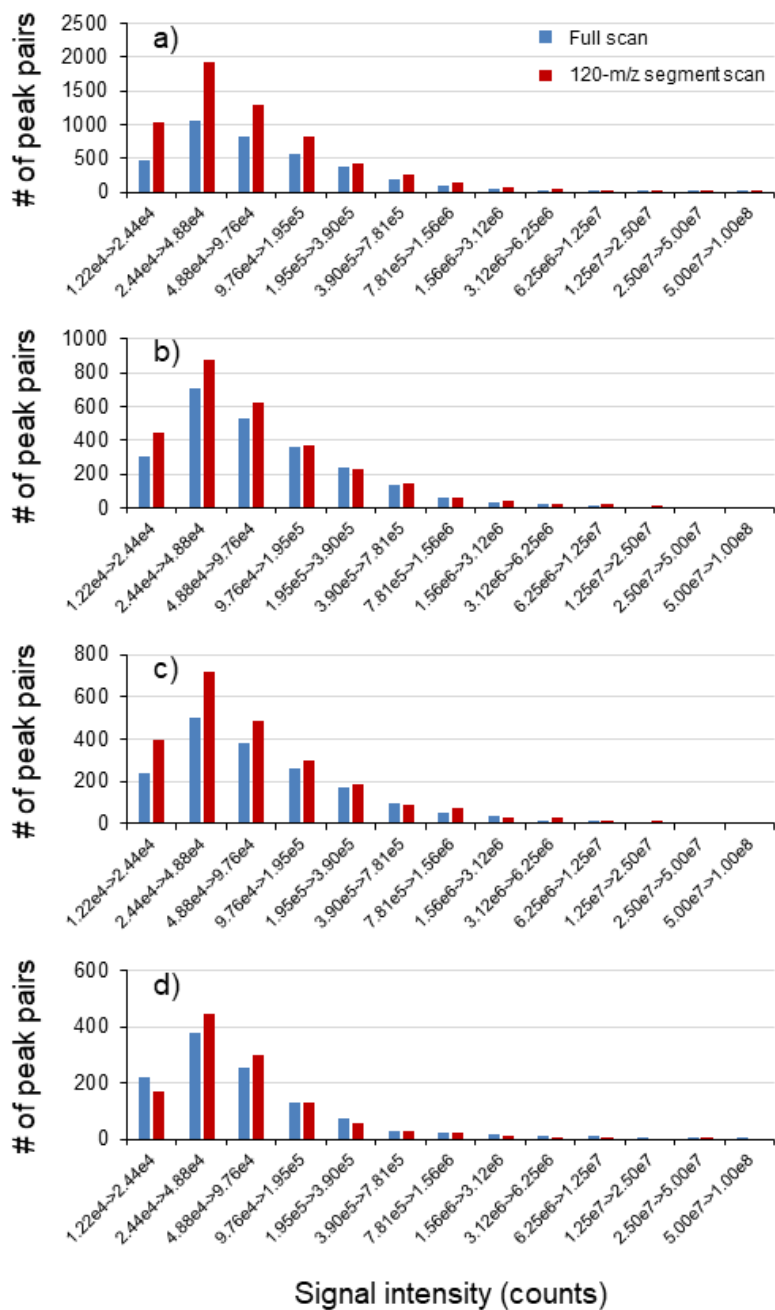
Figure 4.4 Number of peak pairs detected as a function of ion signal intensity in different

bins using 120-m/z segment scan vs. full scan from a) fecal samples, b) serum samples, c)

cell extracts, and d) saliva samples.

Table 4.1 Comparison of metabolite identification results between optimal segment scan method and full scan method collected at 60,000 resolution.

Sample type	Peak pair type	Full scan method	Optimized segment scan method
Feces	Detected in total	3682	6048
	Tier 1	266	265
	Tier 2	539	842
	Tier 3	2621	4607
	IDs and matches	3426 (93.0%)	5714 (94.5%)
Urine	Detected in total	3765	5867
	Tier 1	157	158
	Tier 2	604	712
	Tier 3	2759	4705
	IDs and matches	3520 (93.5%)	5575 (95.0%)
Serum	Detected in total	2414	2836
	Tier 1	156	166
	Tier 2	250	300
	Tier 3	1800	2167
	IDs and matches	2206 (91.4%)	2633 (92.8%)
Cell extract	Detected in total	1779	2335
	Tier 1	109	100
	Tier 2	195	263
	Tier 3	1329	1810
	IDs and matches	1633 (91.8%)	2173 (93.1%)
Saliva	Detected in total	1071	1164
	Tier 1	80	107
	Tier 2	113	111
	Tier 3	742	761
	IDs and matches	935 (87.3%)	979 (84.1%)
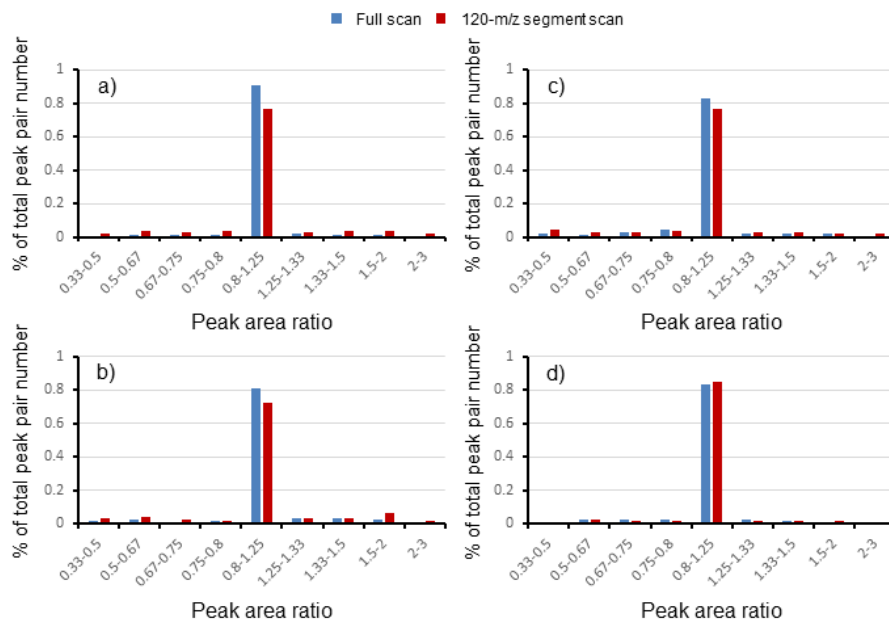
Figure 4.5 Percentage of the total number of peak pairs detected as a function of peak area ratio from a 1:1 mix of ^{13}C/^{12}C dansyl-labeled a) fecal mixture, b) serum mixture, c) cell mixture, and d) saliva mixture.
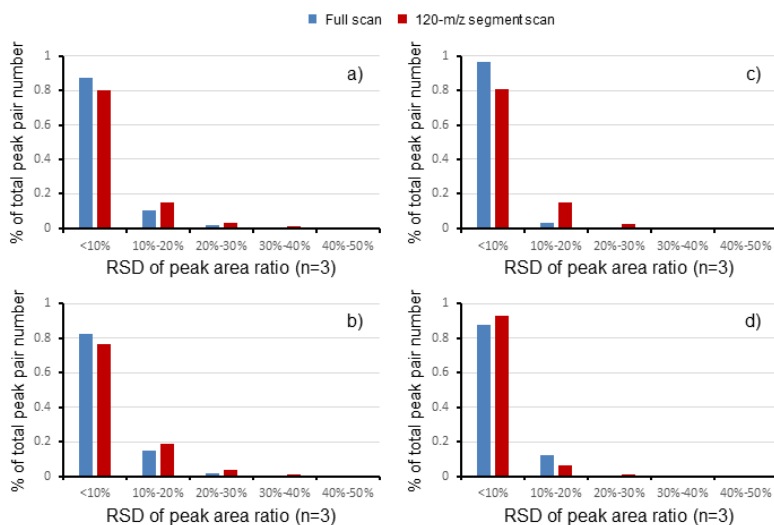
Figure 4.6 Percentage of the total number of peak pairs detected as a function of RSD value of the measured peak area ratio (n=3) from a 1:1 mix of ^{13}C/^{12}C dansyl-labeled a) fecal mixture, b) serum mixture, c) cell mixture, and d) saliva mixture.

4.4 Conclusions

We evaluated the performance of segment-scan-assisted CIL LC-MS among different types of samples. The segment scan method demonstrated higher performance in samples with higher metabolic complexity, such as feces and urine, which contain a wider variety of metabolites. In these samples, up to 94% more peak pairs were detected compared to conventional full scan methods. For less complex samples with limited metabolite diversity, like saliva, only around 25% more peak pairs were detected compared to full scan. Overall, a 120-m/z segment scan method was validated as a routine approach for CIL LC-Q-Orbitrap MS, providing enhanced detection efficiency for metabolites. The method proved to be compatible with various sample types, particularly those exhibiting high or medium metabolic complexity. However, it is important to acknowledge that for less complex samples with limited metabolite diversity, such as saliva and sweat, the overall performance of the optimal segment method in terms of peak pairs detection can still be improved, although the extend of improvement is not as high as that in analyzing more complex samples. In such cases, the benefits of enhanced peak pairs detection should be weighed against the total LC-MS analysis time, considering the additional time required for data processing.

Chapter 5

Fragmentation of Dansyl-labeled Amine- and Phenol-containing Metabolites for MS/MS-based Structure Elucidation Using Higher Energy Collisional Dissociation

5.1 Introduction

Chemical isotope labeling (CIL) liquid chromatography-mass spectrometry (LC-MS) is a powerful technique in metabolomics with high coverage of metabolome profiling.[131,132] In CIL LC-MS, metabolites are divided into different sub-groups based on the functional groups they contain, followed by a labeling reaction specific to the functional group. With a rational design of the labeling reagents, significant improvement in LC separation and MS detection can be achieved. Moreover, the usage of differential isotope labeling creates references (i.e., a pooled sample labeled with a heavy reagent) for each metabolite in the individual samples labeled with a light reagent, allowing for accurate relative quantification.[111,133] To date, a four-channel labeling protocol has been developed in our lab for CIL LC-MS metabolome analysis, i.e., ^{12}C-/^{13}C-dansylation for amine-/phenol-containing sub-metabolome,[38] ^{12}C-/^{13}C-dimethylaminophenacyl (DmPA) bromide for carboxyl-containing sub-metabolome,[42] base-activated ^{12}C-/^{13}C-dansylation for hydroxyl-containing sub-metabolome,[41] and ^{12}C-/^{13}C-dansylhydrazine (DnsHz) for carbonyl-containing metabolome.[40] The combination of these four sub-metabolomes can achieve a total 86%-95% metabolite coverage[39] based on metabolome database search such

as MyCompoundID library (MCID), Human Metabolome Database (HMDB), Yeast Metabolome Database (YMDB), and E. Coli Metabolome Database (ECMDB).

In addition to high metabolome coverage and relative quantification accuracy, obtaining reliable metabolite identification results is another crucial aspect of metabolomics research. In CIL LC-MS, a unique quantitative aspect is that metabolites are present as peak pairs (^{12}C-labeld metabolites and ^{13}C-labeld metabolites), making them distinct from other noises or chemical backgrounds in MS detection. The detected peak pairs are identified based on a three-tier approach[82,83]: In the first tier (positive identification), accurate mass and retention time (RT) were compared to a labeled standard library known as the CIL library. This library comprises over 1600 endogenous metabolites. For the second tier (putative identification with a high confidence level), accurate mass and RT were cross-referenced against a linked identity library referred to as the LI library. This library consists of over 7000 pathway-related metabolites with predicted RT values. In the third tier (putative match), accurate mass was compared to the MyCompoundID library (MCID library) to identify potential matches. For the tier-2 and tier-3 metabolites, only putative identification or match can be obtained due to the absence of reference standards, in such cases, high-quality MS/MS spectra can offer valuable complementary information for structure analysis.

Our research objective is to understand the fragmentation patterns of known compounds labeled with a dansyl moiety in order to facilitate the interpretation of MS/MS

spectra of unknown metabolites in two aspects. Firstly, for amine- and phenol-containing metabolites, we found a simple way to differentiate the amine-containing metabolites from the phenol-/hydroxyl-containing metabolites. Knowing which group an unknown metabolite belongs to can be helpful in accurate mass search against a metabolome database. Secondly, fragmentation patterns related to the metabolite moiety could serve as identifiers in MS/MS-based metabolite identification,[134] these patterns not only aided in corroborating the identities of Tier-2 metabolites but also elevated the confidence levels of the overall identification results. Higher energy collisional dissociation (HCD) on Orbitrap MS has shown promise in improving fragmentation in the low mass region, which could significantly enhance the quality and informativeness of the MS/MS spectra.[135,136] Therefore this technique has been employed in this work to obtain informative MS/MS spectra, which will be valuable for metabolite identification and structural analysis.

5.2 Experimental

5.2.1 Workflow

Figure 5.1 shows the schematic of the workflow for this work, which was divided into two steps. Firstly, specific compounds representing amine- and phenol-containing metabolites were selected as standards and labeled with dansyl chloride, followed by a targeted MS/MS spectra collection with collision energy varying from 10 to 50 eV. This variation in collision energy enables the generation of a wide range of fragment ions, facilitating a comprehensive study of the fragmentation patterns.
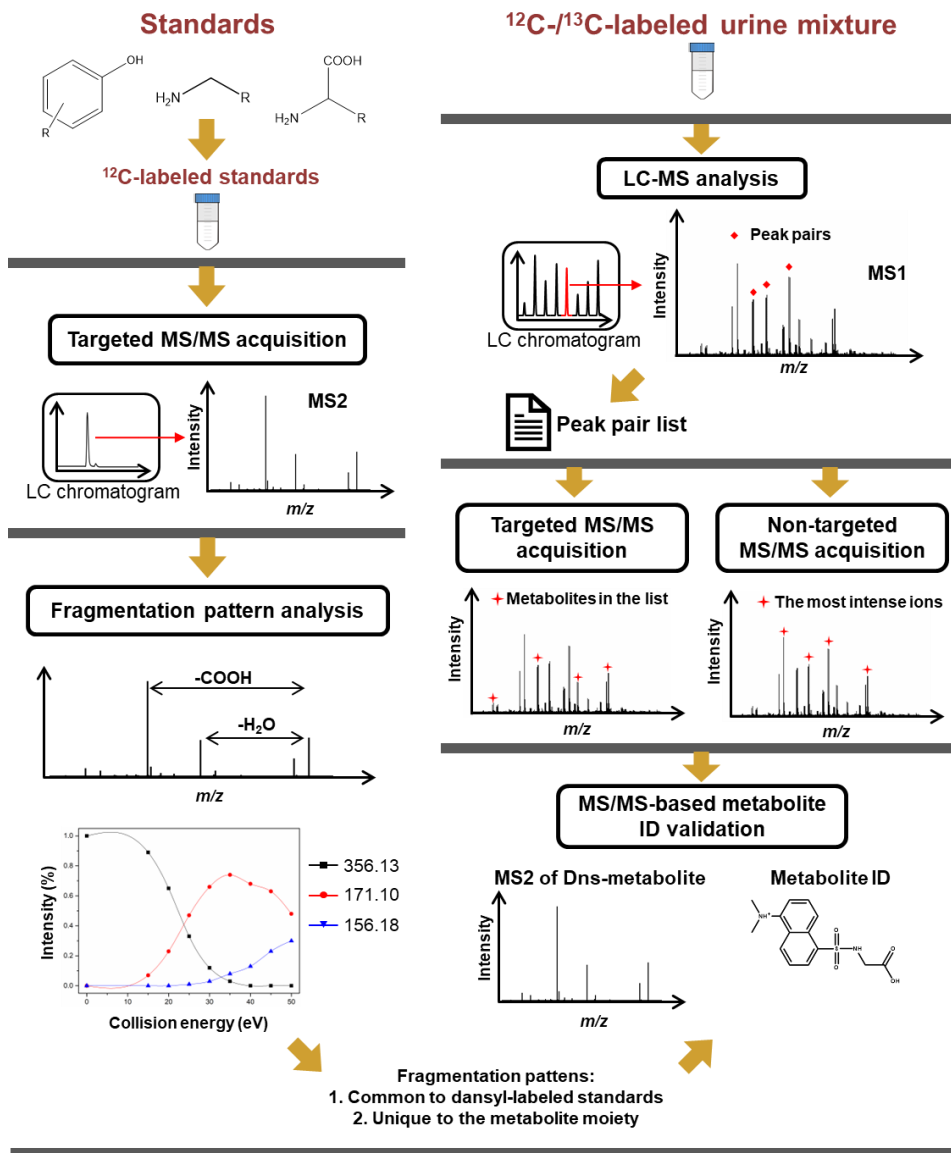
Figure 5.1 Workflow for fragmentation pattern analysis using dansyl-labeled standards and MS/MS-based ID validation for dansyl-labeled amine- and phenol-containing metabolites.

Secondly, high-quality MS/MS spectra of the dansyl-labeled metabolites in biological samples were collected using both targeted and non-targeted MS/MS. Targeted MS/MS is specifically directed to analyze and fragment pre-selected precursor ions of

interest (i.e., protonated ^{12}C-labeled metabolites), in this case, a routine CIL LC-MS analysis of a ^{12}C-/^{13}C-labeled urine mixture was performed to obtain the peak pair list which contains the corresponding peaks of the ^{12}C- and ^{13}C-labeled metabolites. In non-targeted MS/MS, a top N data-dependent acquisition (DDA) method[137] was employed, where the selection of precursor ions was based on their intensities in the MS1 scan. The top N most intense peaks were chosen for fragmentation, where N represents a predetermined number of ions. Non-targeted MS/MS shows less specificity than targeted MS/MS but it consumes less sample and setup time since no predefined targeted ions need to be determined. In this work, both of these two MS/MS methods were employed to increase the total MS/MS coverage of the metabolites of interest while maintaining a good spectral quality. After 8 runs of CIL LC-MS/MS, around 1200 and 2500 spectra were collected using targeted and non-targeted MS/MS methods, respectively.

Based on fragmentation pattern analysis of selected standard and tier-1 metabolites, some basic fragmentation types or rules related to dansyl-labeled metabolites were proposed. For tier-2 metabolites, which only have putative identification results with high confidence levels due to the lack of standards, the proposed rules were employed to validate their identification. In this work, all the dansyl-labeled metabolites, otherwise stated, refer to metabolites with a single dansyl tag.

5.2.2 Metabolite standards

Essential amino acids (alanine, glycine, isoleucine, leucine, proline, valine, phenylalanine, tryptophan, aspartic acid, glutamic acid, serine, threonine, methionine, asparagine, and glutamine), amine-containing compounds (dimethylamine, 1-naphthylamine, aniline, 6-aminocaproic acid), and phenol-containing compounds (4-ethylphenol, 4-hydroxyphenylacetic acid, and 4-hydroxybenzoic acid) were selected to represents amine- and phenol-containing metabolites for fragmentation pattern analysis. For each standard, a 100 μM solution was prepared for chemical isotope labeling.

5.2.3 Universal urine sample

A universal urine sample from NovaMT (www.novamt.com) was used in this work for MS/MS spectra acquisition. It is an equal mixture of human urine samples collected from 12 healthy individuals. The urine sample was diluted four times in water before chemical isotope labeling.

5.2.4 Dansylation labeling

A 25 μL aliquot of the prepared standard solutions or diluted urine sample was used for dansylation labeling following the protocol described in Chapter 2.

5.2.5 Instrumentation

LC-MS and LC-MS/MS analysis were performed on a quadrupole-Orbitrap mass spectrometer (Q-Exactive HF, Thermo-Fisher Scientific, Waltham, MA, USA) coupled with an ultra-high-performance LC system (Vanquish, Thermo-Fisher Scientific, Waltham,

MA, USA). Samples were injected onto an Agilent Zorbax Eclipse Plus C18 column (2.1 mm× 100 mm, 1.8 μm particle size, 95 Å pore size). The LC conditions used for both MS and MS/MS were the same as those reported in Chapter 2.

The mass spectrometer was equipped with an electrospray ionization (ESI) source, using the conditions listed as follows: ion mode - positive; sheath gas flow rate – 60 a.u.; aux gas flow rate – 30 a.u.; sweep gas flow rate – 2 a.u.; spray voltage – 3.5 kV; capillary temperature – 350 °C; vaporizer temperature – 350 °C. MS1 was collected with a mass from 220 to 1000 at a resolution of 60,000. For targeted MS/MS, MS2 was operated using the parameters listed as follows: resolution – 15,000; AGC target – 2e5; isolation window – 1 m/z; collision energy – stepped from 20 to 50 eV; precursor intensity threshold: 1e5. For non-targeted MS/MS, MS2 was operated using the parameters listed as follows: resolution – 15,000; AGC target – 2e5; isolation window – 1 m/z; top N – 8; precursor intensity threshold – 1e5; dynamic exclusion – 4 seconds.

5.3 Results and discussion

5.3.1 Fragmentation pattern analysis

For a dansyl-labeled metabolite, its MS/MS spectrum typically displays fragmentation related to both dansyl group and metabolite moiety. As shown in Figure 5.2, characteristic fragments related to the dansyl group[138–140] cut the MS/MS spectrum into two major regions. In the first region (from the precursor ion down to m/z 252),[141] neutral losses of functional groups are commonly observed, which are indicative of specific

substructures within the dansyl-labeled metabolites. In the second region (lower than m/z 157), fragments with lower m/z serve as valuable structural identifiers (i.e., fingerprint), giving structural insight into the metabolite moiety. Therefore, in the following discussion, fragmentation patterns of dansyl-labeled metabolites were analyzed from two main aspects: fragmentation related to the dansyl group and fragmentation in the neutral loss region and fingerprint region, based on the MS/MS spectra collected from standards and tier-1 metabolites.
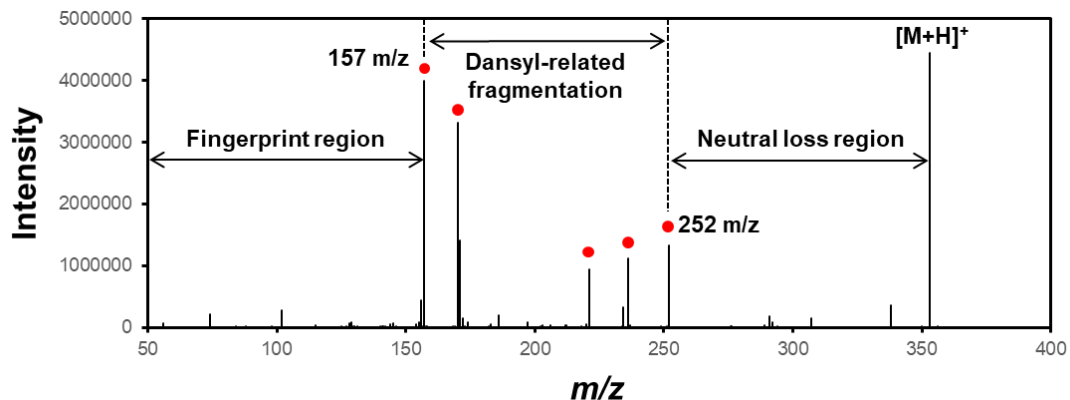
Figure 5.2 MS/MS spectrum of dansyl-labeled threonine at a stepped collision energy of 10-30-50 eV (●: fragments related to the dansyl group).

5.3.1.1 Fragmentation related to the dansyl group

Phenol-containing metabolites Figure 5.3a shows the MS/MS spectrum of dansyl-4-ethylphenol collected with stepped collision energies of 10-30-50 eV. Considering the simple structure, most of the fragments generated from dansyl-4-ethylphenol were attributed to the dansyl group. Three pairs of characteristic ions were observed, namely,

356.13/341.11 m/z, 292.17/277.15 m/z, and 171.10/156.08 m/z, with each pair representing a mass difference of 15.02 Da. The overall fragmentation pattern of dansyl-4-ethylphenol can be derived from these six fragment ions, as shown in Figure 5.3b, which can be further validated by a fragmentation pathway analysis based on these six ions (see Figure 5.4a). The mass difference of 15.02 Da indicated a loss of radical CH_3 group, with a charge-induced fragmentation process involved. For dansyl-labeled metabolites, the tertiary amine group serves as a protonation site during ionization, resulting in a change in the hybridization of nitrogen from sp^2 to sp^3. A more stable environment for the nitrogen atom can be achieved through the loss of a CH_3 radical, causing the hybridization state to revert to sp^2. In Figure 5.3b, there were another two types of charge remote fragmentations involved for dansyl-labeled phenol-containing metabolites, generating fragment pair 341.11/277.15 m/z with a neutral loss of SO_2, and a follow-up radical cation pair 171.10/156.08 m/z. The neutral loss of the SO_2 group merged the two separated conjugation systems (i.e., benzene ring and naphthalene ring) with a sp^2 hybridized oxygen, leading to a more stable resonance structure.[142] In addition to these three fragment pairs, it was noted that no neutral loss of ethyl was observed during the fragmentation, it may be because the long spatial distance between the #4 carbon within the benzene ring and the hydrogen in the ethyl chain hindered the remote hydrogen rearrangement as shown in Figure 5.3c. The six characteristic fragmentations related to the dansyl group are

commonly present in all dansyl-labeled phenol-containing metabolites regardless of their

chemical structures.
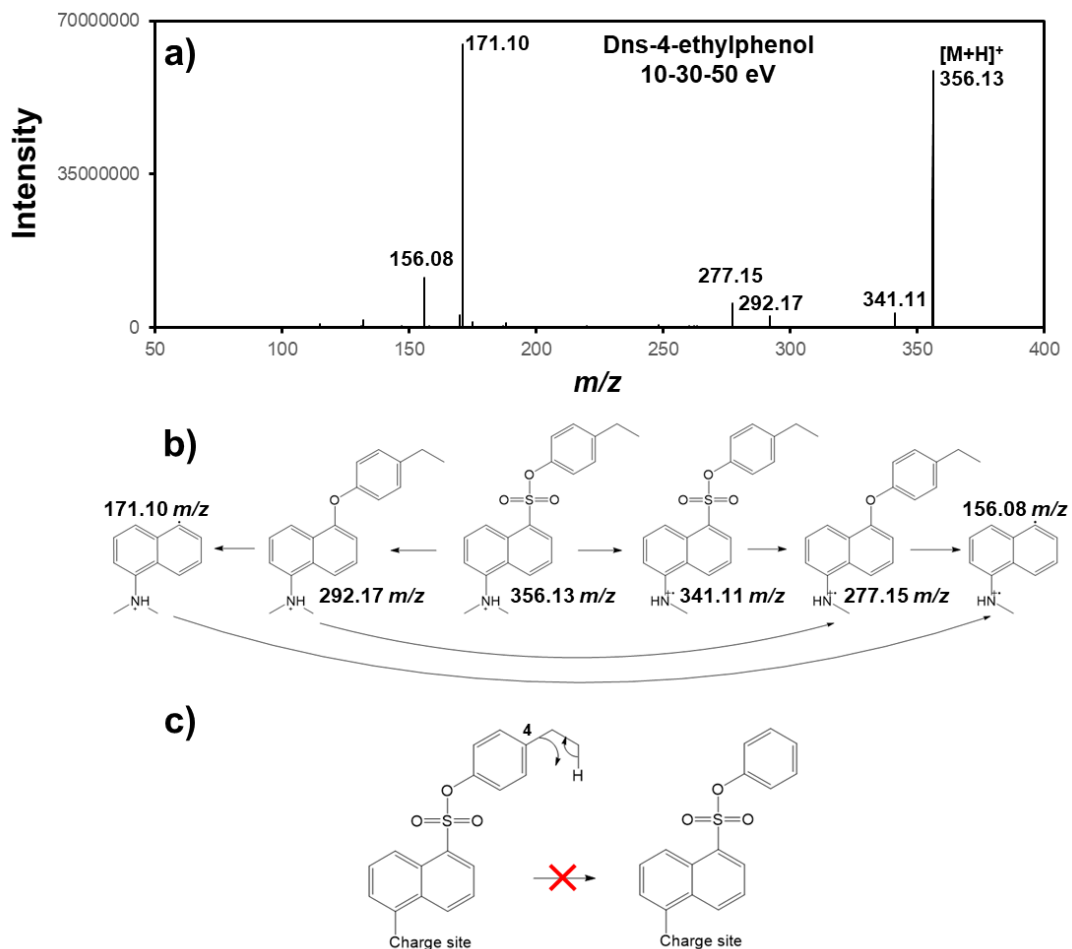
Figure 5.3 a) MS/MS spectrum of dansyl-labeled 4-ethylphenol collected with a stepped

collision energy of 10-30-50 eV, b) fragmentation patterns of the major fragments present

in the MS/MS spectrum, c) neutral loss of ethyl was not observed during the fragmentation.
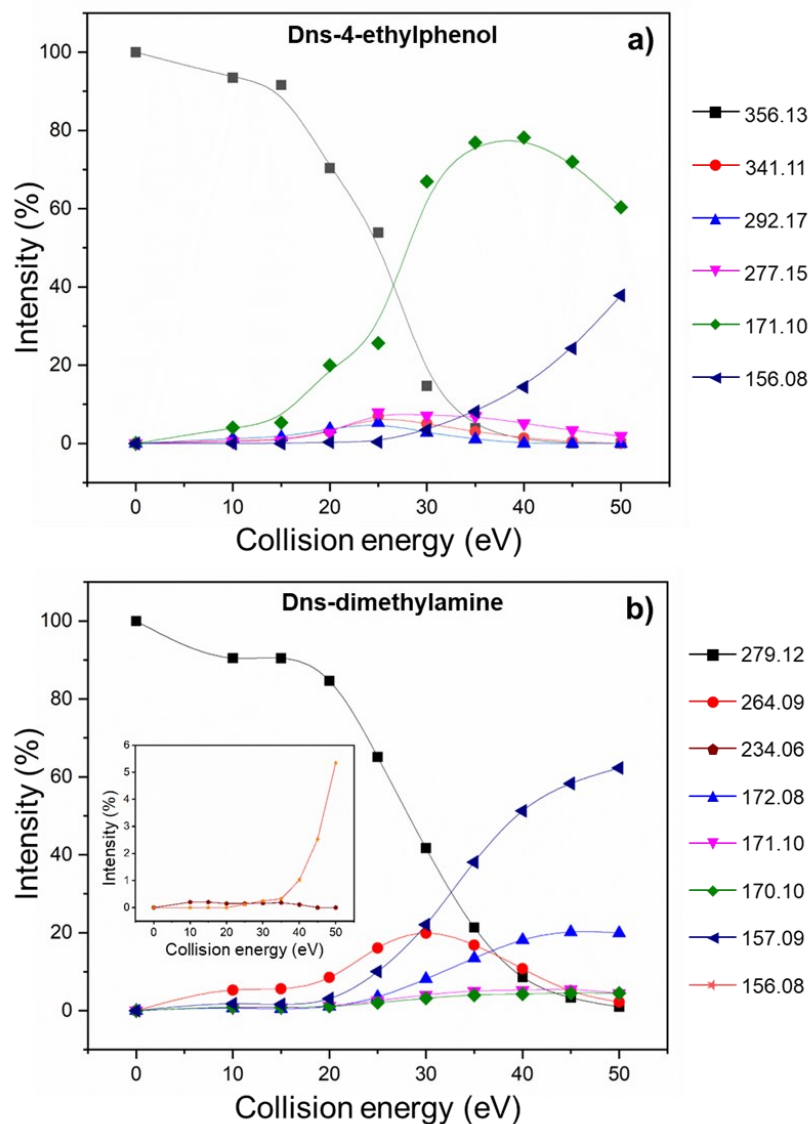
Figure 5.4 Fragment ions plotted as a function of HCD energy for a) Dns-4-ethylphenol and b) Dns-dimethylamine.

Amine-containing metabolites Compared to phenol-containing metabolites, amine-containing metabolites showed a different fragmentation pattern related to the dansyl group. As shown in Figure 5.5a, dansyl-dimethylamine was selected for fragmentation pattern analysis. Seven characteristic fragments were observed following

the fragmentation patterns shown in Figure 5.5b. The overall fragmentation pattern was validated through a fragmentation pathway analysis (see Figure 5.4b). For dansyl-labeled amine-containing metabolites, there were at least two amine groups serving as protonation sites within the molecule, further generating type-1 and type-2 precursor ions (see figure 5.5b). Since type-1 and type-2 ions had different protonation positions, they followed different pathways during the fragmentation. For type-1 ions, charge remote fragmentations were dominated, generating fragments at 264.09, 172.08, 171.10, 157.09, and 156.08 m/z. Among these fragments, ions at 157.09 m/z are particularly abundant in the MS/MS spectra. Additionally, fragments at 172.08 m/z, specific to amine-containing metabolites, were observed. Therefore, fragments 172.08 m/z can be used as an identifier for amine-containing metabolites. A possible mechanism involving charge-remote rearrangement was proposed to explain the formation of these fragments (Figure 5.5c). The loss of CH_3 radical discussed above was also found in the fragmentation of type-1 ions. In contrast, charge migration fragmentations were mainly involved for type-2 ions, forming fragments 234.06 and 170.10 m/z. In this pathway, no further loss of a CH_3 radical occurred due to the different charge position compared to type-1 ions. In addition to the fragmentations discussed above, another three types of fragments ions related to the dansyl group were also found in the amine-containing metabolites with more complex structures than dimethylamine.

m/z 236.06/251.08 fragments The mechanism of fragments 236.06/251.08 m/z involves a charge-remote beta-hydrogen rearrangement, as demonstrated by Figure 5.6c. Such fragmentations were commonly found in the MS/MS of amine-containing metabolites as long as they contain beta-hydrogen such as Dns-isoleucine (Figure 5.6a). Figure 5.6d shows an exception where there was not beta-hydrogen in Dns-glycine, therefore no fragments 236.06/251.08 m/z was found in its MS/MS spectrum (Figure 5.6b).

m/z 252.07 fragments As shown in Figure 5.7, metabolites containing both amine and hydroxyl/carboxylic groups often exhibited a charge-induced hydroxyl rearrangement, leading to the formation of fragment ions at 252.07 m/z. Notably, no further loss of radical CH_3 was observed for fragments 252.07 and 234.06 m/z because only type-2 protonated ions were involved in the process. While not universally present in all amine-containing metabolites, fragment pairs such as 236.06/251.08 m/z and 234.06/252.07 m/z can still serve as identifiers to distinguish amine-containing metabolites from phenol-containing metabolites.

Neutral loss of -SO$_2$ group Aromatic amines demonstrated a behavior similar to phenol-containing metabolites during fragmentation. In Figure 5.8, the MS/MS spectrum of Dns-3-aminobenzoic acid was studied, and it displayed a neutral loss of -SO$_2$ group, akin to phenol-containing metabolites. Additionally, fragments at 171.10 and 156.08 m/z were also prominent in the collected MS/MS spectra. However, in this case, ions at 252.07 m/z were not detected, which can be attributed to the absence of -SO$_2$ group. Furthermore,

the presence of a benzene ring in the structure of Dns-3-aminobenzoic acid spatially

hindered the charge-induced hydroxyl rearrangement related to the formation of fragments

at 252.07 m/z.
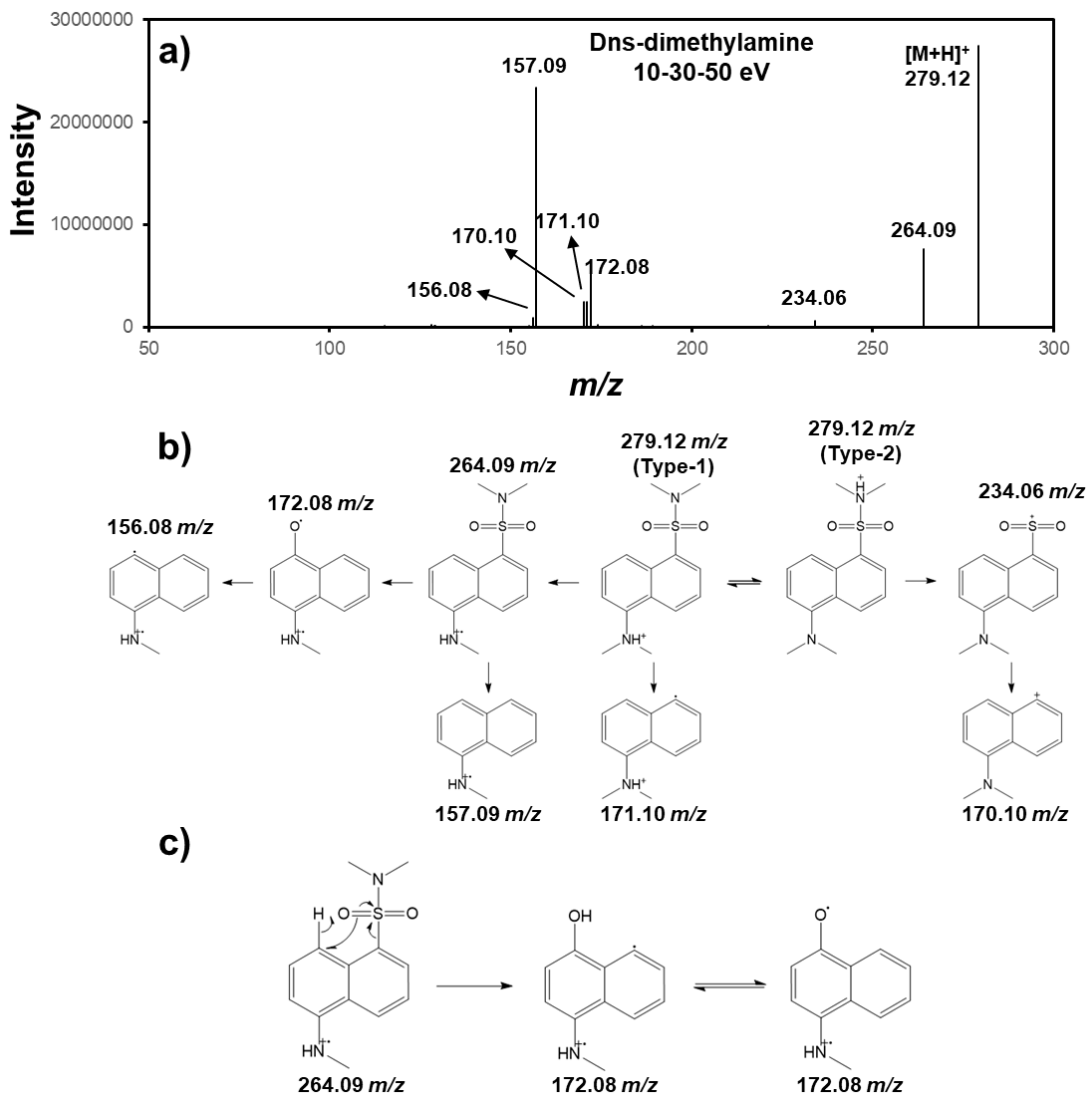
Figure 5.5 a) MS/MS spectrum of Dns-dimethylamine collected with a stepped collision

energy of 10-30-50 eV, b) fragmentation patterns of the major fragments present in the

MS/MS spectrum, and c) a proposed charge-remote fragmentation process for the

formation of fragment ions 172.08 m/z.
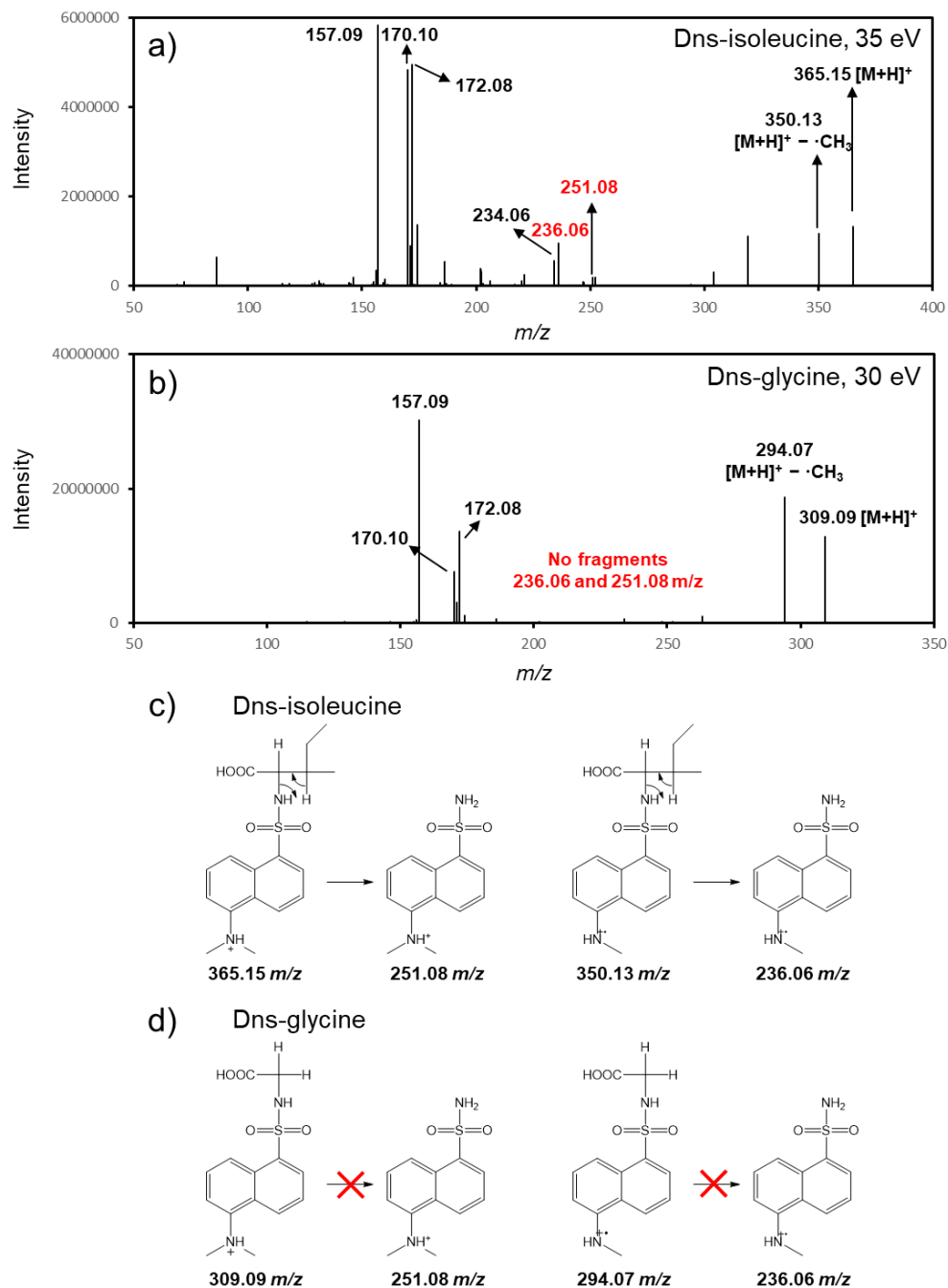
Figure 5.6 a) MS/MS spectrum of Dns-isoleucine collected with a collision energy of 35 eV, b) MS/MS spectrum of Dns-glycine collected with a collision energy of 30 eV. A proposed charge-remote hydrogen rearrangement for the formation of fragment ions 251.08/236.06 m/z for c) Dns-isoleucine and d) Dns-glycine, respectively.
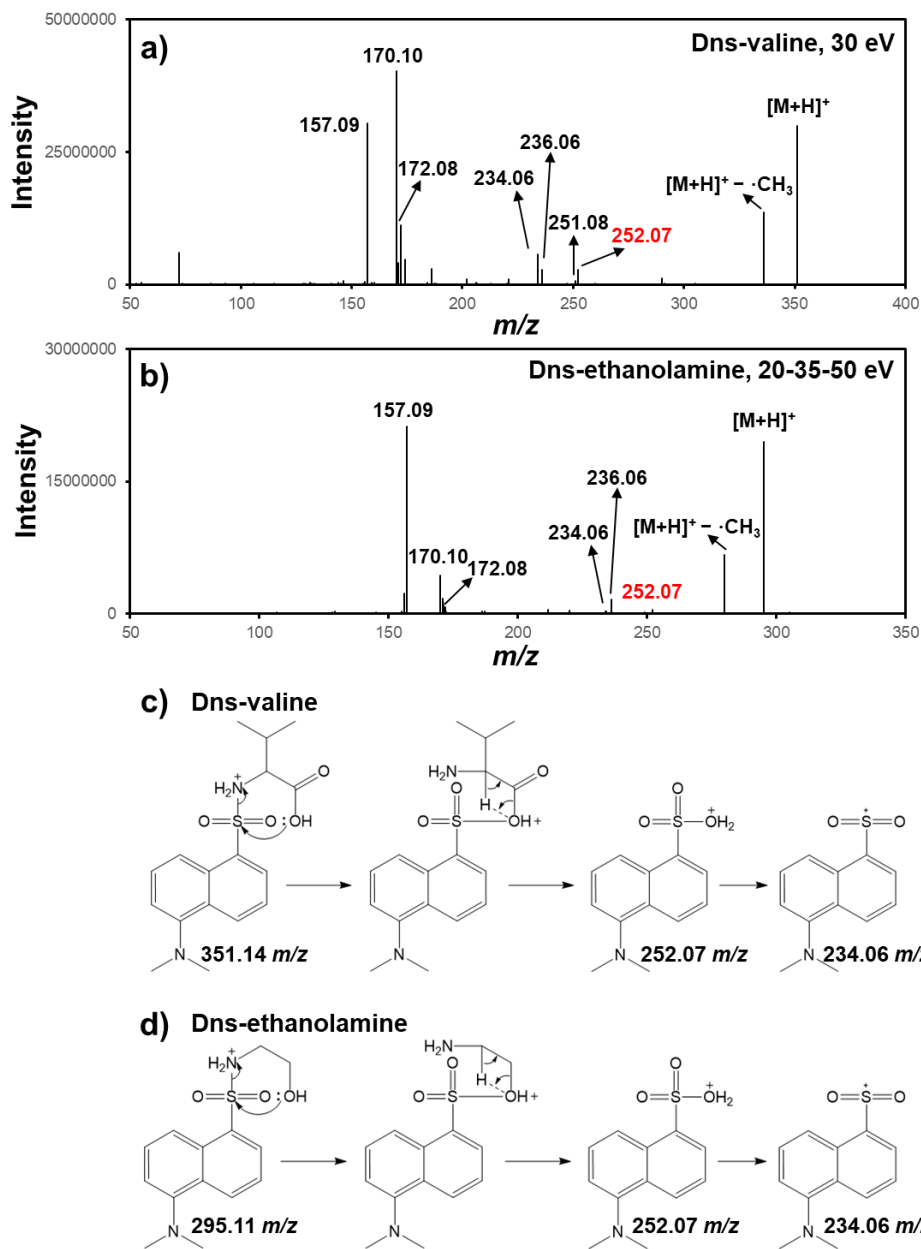
Figure 5.7 a) MS/MS spectrum of Dns-valine collected with a collision energy of 30 eV,

b) MS/MS spectrum of Dns-ethanolamine (a tier-1 metabolite) collected with a stepped

collision energy of 20-35-50 eV. A proposed charge-remote hydroxyl rearrangement for

the formation of fragment ions 252.07/234.06 m/z for c) Dns-valine and d) Dns-ethylamine,

respectively.
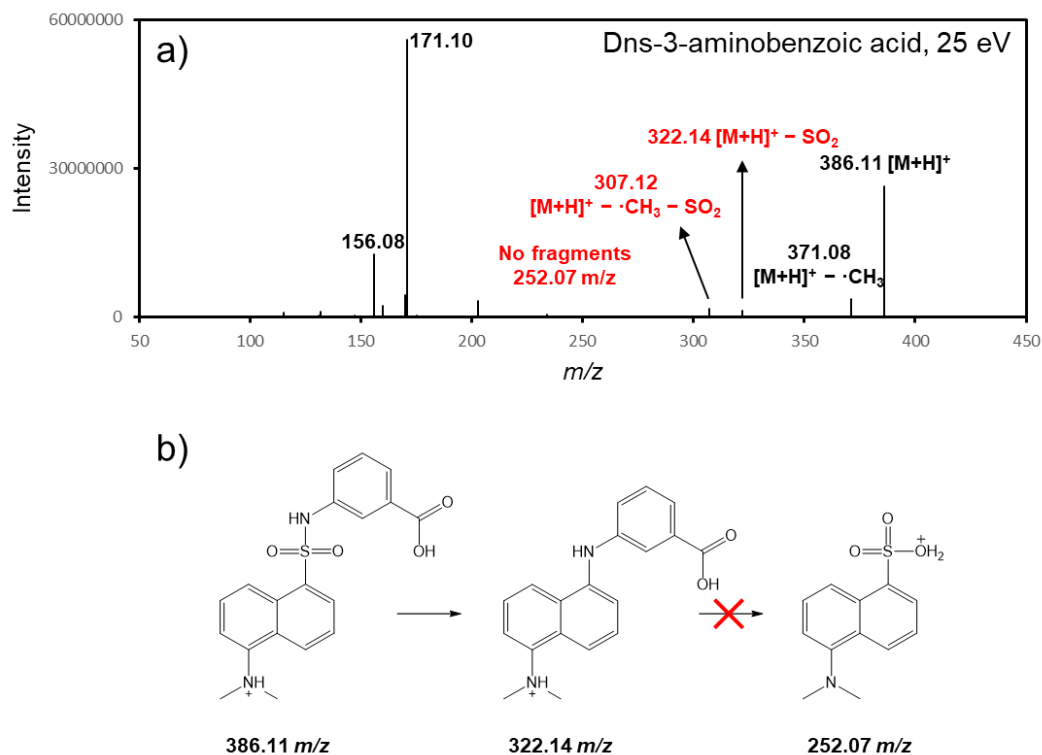
Figure 5.8 a) MS/MS spectrum of Dns-3-aminobenzoic acid collected with a collision energy of 25 eV and b) a proposed fragmentation showing the neutral loss of -SO$_2$ group and accounting for the absence of fragment ions 252.07 m/z.

5.3.1.2 Fragmentation in neutral loss region

Five typical neutral losses in terms of decarboxylation, dehydration, deamination, deacetylation, and desulfurization were found based on MS/MS elucidation of selected standards and tier-1 metabolites:

Decarboxylation The neutral loss of 46 m/z was commonly found in the MS/MS of dansyl-labeled alpha-amino acid, which was attributed to a decarboxylation process.[143,144] For example, as shown in the MS/MS of Dns-alanine (Figure 5.9a), a charge-

remote hydrogen rearrangement process was normally involved in the elimination of carboxyl groups (Figure 5.9b).

Dehydration Dehydration of dansyl-labeled metabolites can be attributed to both hydroxyl groups and carboxylic groups.[145,146] Hydroxyl groups (-OH) present in the structure of dansyl-labeled metabolites can undergo a similar hydrogen rearrangement process described in the decarboxylation section, resulting in the removal of a H_2O from the molecule. On the other hand, dehydration can also occur in dansyl-labeled gamma-or delta-amino acids. As shown in Figure 5.10a, the elimination of an H_2O molecule was found in dansyl-labeled gamma-aminobutyric acid, facilitated by a seven-membered cyclic transition state. Type-2 precursor ions were mainly involved during this process, which was supported by two main observations: i) No further radical loss of CH_3 was observed and ii) fingerprint fragments at 86.06 Da were detected.

Deamination Deamination process was only found in guanidino-containing metabolites in this work. The presence of the guanidino group offered multiple protonation sites during the ionization, as shown in Figure 5.11b. For type-3 protonated ions, a neutral loss of 17.03 Da was observed, the mechanism involved a simple inductive cleavage of the C-N bond (i.e., a simple charge migration fragmentation).

Deacetylation After protonation and subjected to HCD cell, N-acetyl derivatives underwent cleavage of the acetyl group attached to the nitrogen atom,[147] resulting in a

neutral loss of 42.01 Da. This reaction occurred with a similar hydrogen rearrangement as decarboxylation and dehydration (Figure 5.13).

Desulfurization There were two main types of desulfurization process during the fragmentation of dansyl-labeled metabolites, with structures of methyl sulfides and disulfides involved in, respectively. In most cases, the eliminations of sulfur-containing functional groups[142,148] were induced by a charge-remote rearrangement as demonstrated in Figure 5.12.
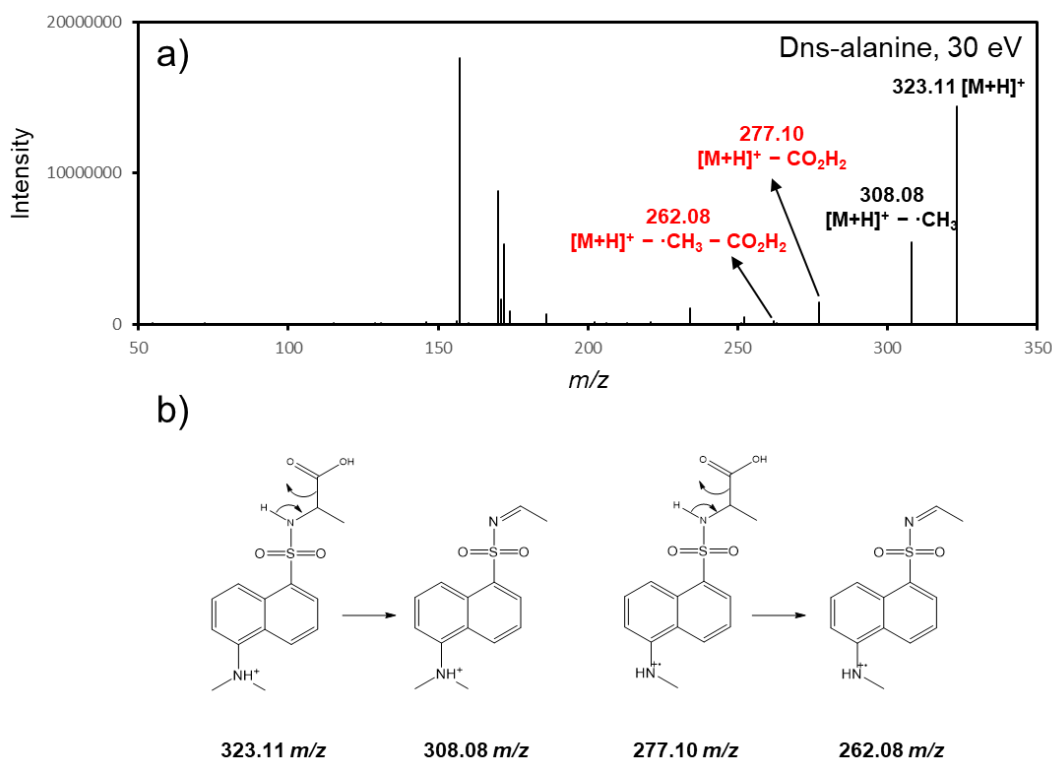
Figure 5.9 a) MS/MS spectrum of Dns-alanine collected with a collision energy of 30 eV and b) a proposed fragmentation showing the neutral loss of carboxylic group.
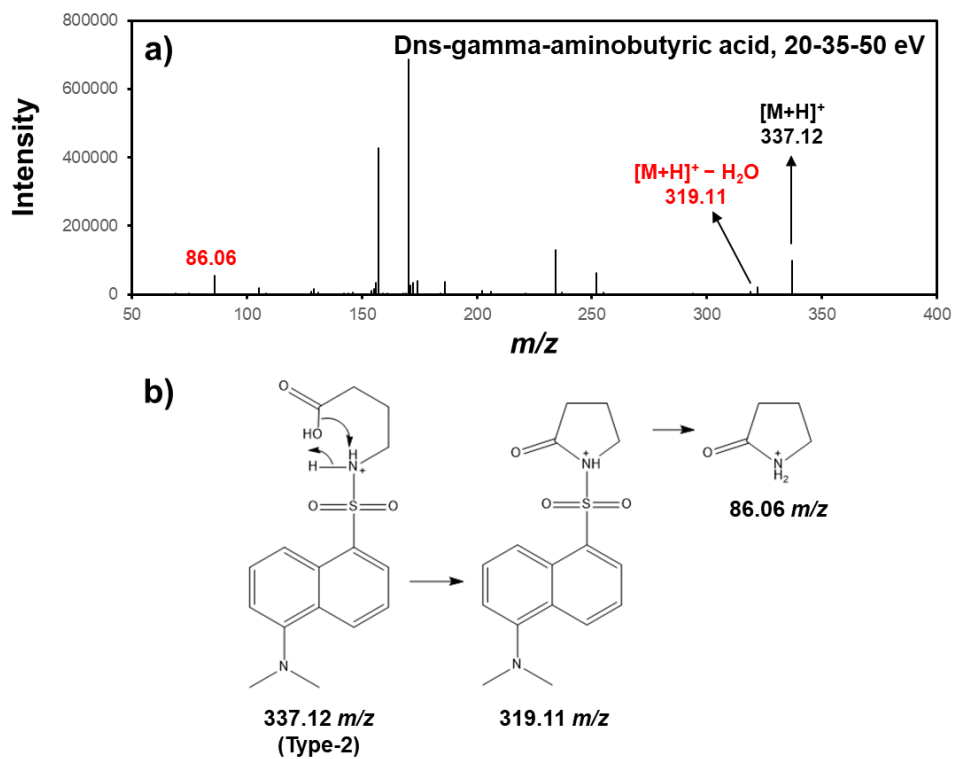
Figure 5.10 a) MS/MS spectrum of Dns-gamma-aminobutyric acid collected with a stepped collision energy of 20-35-50 eV and b) a proposed fragmentation showing the elimination of H_2O molecule.
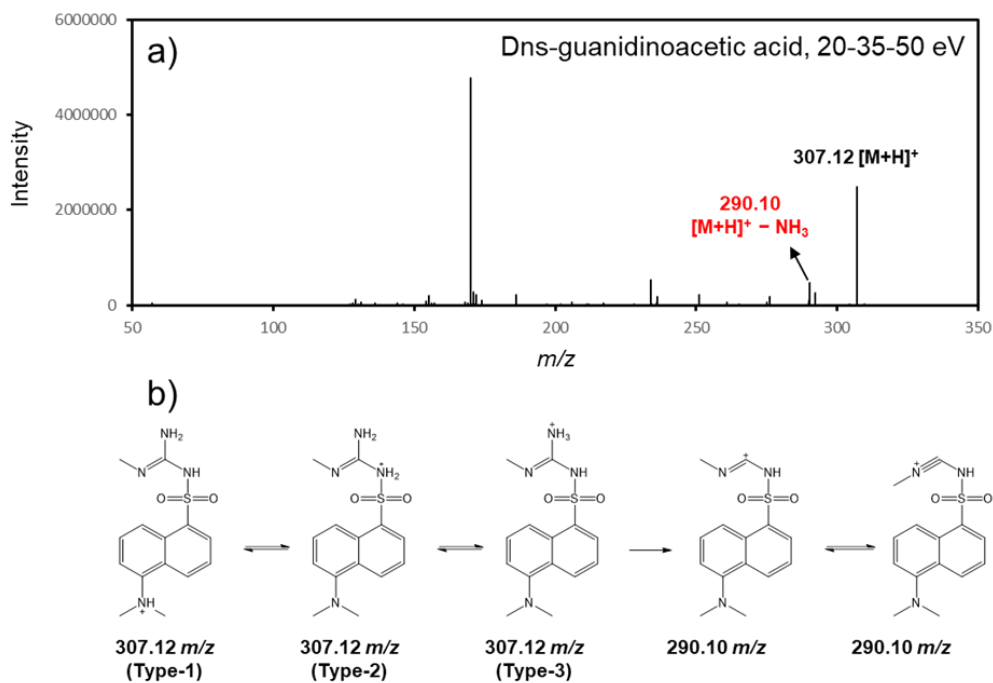
Figure 5.11 a) MS/MS spectrum of Dns-guanidinoacetic acid (tier-1 metabolite) collected with a stepped collision energy of 20-35-50 eV and b) a proposed fragmentation showing the deamination process.
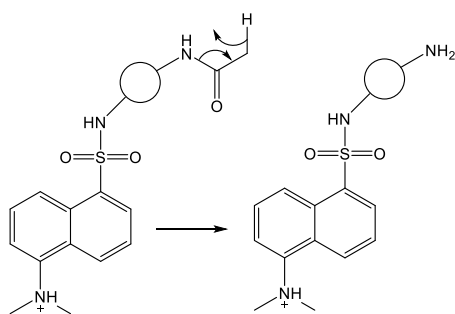
Figure 5.12 A proposed fragmentation showing the neutral loss of acetyl groups.
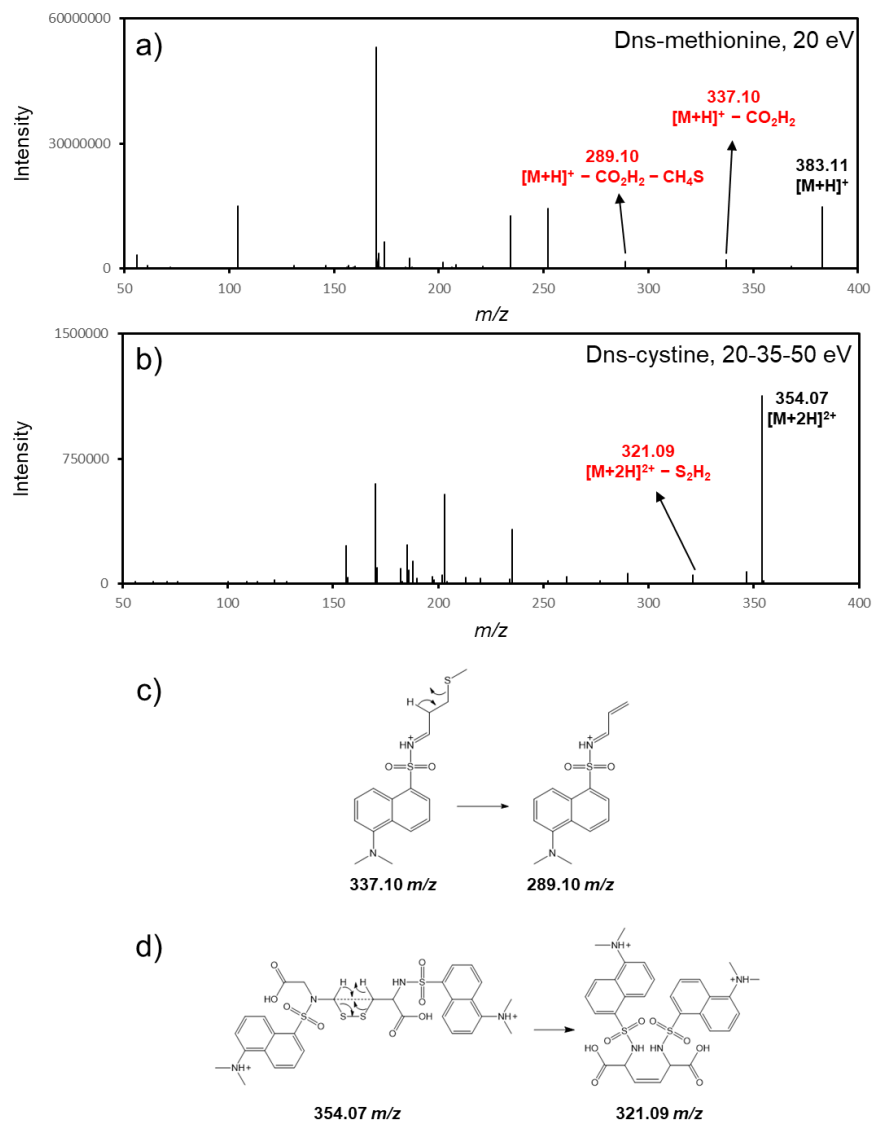
Figure 5.13 a) MS/MS spectrum of Dns-methionine collected with a collision energy of 20 eV, b) MS/MS spectrum of Dns-cystine collected with a stepped collision energy of 20-35-50 eV. A proposed fragmentation showing the elimination of c) methyl sulfide and d) disulfide.

5.3.1.3 Fragmentation in fingerprint region

Compared to neutral losses, which usually provide information about functional groups, the fragment ions within the fingerprint region[149,150] play a critical role in structure validation as they reveal essential information about the core molecular framework of the labeled metabolites. By identifying these backbone fragments, one can gain valuable insights into the connectivity of atoms and understand the overall arrangement of the chemical bonds in the metabolites. Table 5.1 lists all the fingerprint ions found after MS/MS interpretation of selected standards and tier-1 metabolites in this work, they can be further classified into three types:

Firstly, some metabolites contain stable structural units that can withstand the relatively moderate collision energy levels, typically not exceeding 50 eV, that are employed in this work. These stable units, like the imidazole unit found in histidine and the benzene ring present in phenylalanine, are highly robust thus are able to remain intact during the fragmentation process.

Secondly, as discussed earlier in the context of dehydration, some metabolites can undergo structure reorganization during the fragmentation process. When certain functional groups within the metabolite interact with each other, intramolecular bonds can be formed, leading to a stable cyclic structure. These stable intramolecular rings can also remain as fingerprint ions in MS/MS, serving as unique identifiers and aiding in the identification of metabolites.

Finally, a specific fingerprint was found during the fragmentation of dansyl-labeled alpha-amino acid, as shown in Figure 5.14. In principle, a neutral loss of alpha-carboxylic group occurred initially, followed by the formation of a new π-bond. Then fingerprint ions revealing the backbone structure of the alpha amino acid were generated via a charge-induced N-S bond cleavage.
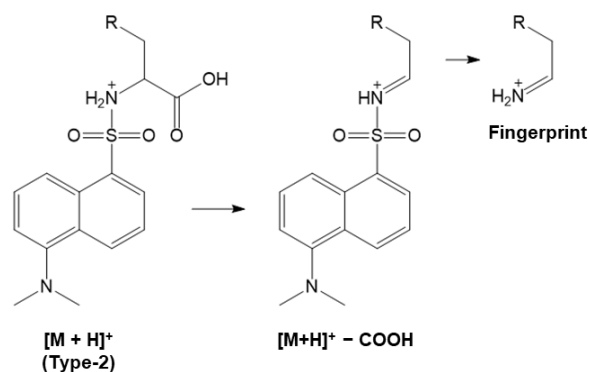
Figure 5.14 A proposed fragmentation showing formation of fingerprint ions specific to dansyl-labeled alpha-amino acids.

5.3.2 Tier-2 metabolites ID validation based on MS/MS interpretation

A total of 601 peak pairs were putatively identified as tier-2 metabolites (metabolites with a single tag are listed in Table 5.2). Those metabolites were validated by MS/MS interpretation based on the proposed fragmentation rules related to the dansyl-labeled metabolites. Peptides were excluded from the validation since a study focusing on the fragmentation of dipeptides and tripeptides is underway in our research group.[151]

The overall manual validation results are shown in Figure 5.15. Out of 601 tier-2 metabolites, 85 passed the validation. These metabolites displayed fragmentation patterns

in their MS/MS spectra that matched their identities. For reference, the MS/MS spectra and proposed fragmentation patterns of these tier-2 metabolites can be found in Appendix A. In metabolomics, isomers are commonly encountered. Notably, for most of the dansyl derivatives of stereoisomers and some structural isomers with closely related differences in bonding or functional group arrangements, they cannot be distinguished from each other based solely on the MS/MS data. For example, there were two peak pairs detected with the same m/z (i.e., 348.10/350.11) but different RT at 2.33 and 3.11 min, both of them were identified as 3-Cyano-L-alanine in tier-2 library matching. As shown in Figure 5.16, the MS/MS spectra collected for these two compounds presented similar patterns. The metabolites eluted at 2.33 and 3.11 were supposed to be a pair of structural isomers. In such cases, where constitutional isomers exhibiting small differences in structures were matched as the same identity due to the limited size of our tier-2 library and could not be confidently differentiated using MS/MS spectra, they were still considered as valid identifications but with a mark indicating them as isomers (included in Appendix A as well). In contrast, constitutional isomers with significant differences in their structures presented different fragmentation patterns in MS/MS. As shown in Figure 5.17, two peak pairs detected at RT 2.77 and 2.91 min were both identified as creatine, however, different MS/MS patterns were observed, indicating that the two metabolites contained different structures. The presence of fingerprint ions 113.05, 112.02, 98.03, and 69.04 m/z confirmed the identity of peak pair at 2.91 min as true creatine. However, the peak pair at

2.77 min, which also appeared to be creatine based on the identification results, requires further MS/MS interpretation to determine its real structure. Ten peak pairs in the same situation were filtered out and listed in Appendix B. In-depth fragmentation analysis and additional experiments are warranted to elucidate their exact structure in future work. To date, our tier-2 metabolite library contains approximately 7000 pathway-related metabolites, but it lacks representation of isomers. For future work, it is crucial to expand the library with a focus on increasing the coverage of isomers. With a larger and more diverse library, we can cross-reference experimental MS/MS spectra with multiple potential isomers and achieve more accurate structural validation. Additionally, incorporating isomers in the library will enhance the identification capabilities for challenging cases where MS/MS spectra of closely related structural isomers cannot be distinguished based on current data.
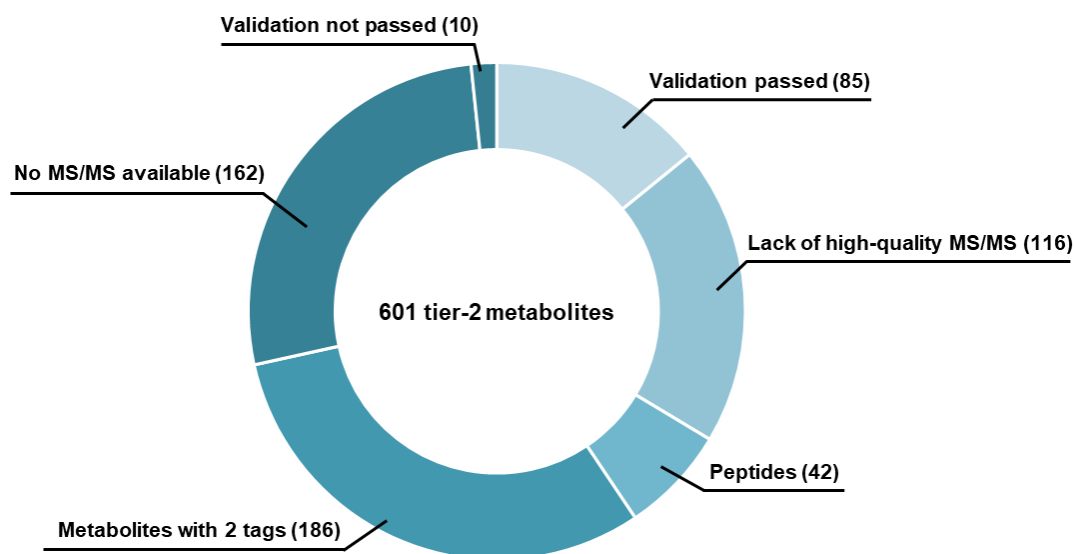
Figure 5.15 Distribution of 601 tier-2 metabolites after MS/MS spectra interpretation.

As shown in Figure 5.15, MS/MS spectra could not be comprehensively interpreted for 116 putative tier-2 metabolites because of the low quality, which was attributed to three reasons. Firstly, the intensities of those peak pairs were slightly higher than the threshold set to selected precursor ions for fragmentation, in this case, even if the MS/MS spectra were collected, their low intensities limited the spectra quality. Secondly, certain metabolites, such as *O*-glucuronide metabolites, are composed of several stable cyclic units, which could not generate multiple fragment ions even in HCD (e.g., neutral loss of a glucuronic acid unit was dominant in the MS/MS spectra collected from Dns-*O*-glucuronide metabolites). In order to improve the MS/MS quality in future research, a pseudo-MS3 strategy on Q-Orbitrap MS can be used for spectral acquisition (find details in Future Work). Lastly, for phenol-containing metabolites with simple structures, such as 2-methoxy-4-vinylphenol, 4-hydroxystyrene, and thymol, displayed uninformative MS/MS spectra when analyzed as dansyl-derivatives. As shown in Figure 5.18, only neutral losses of -SO_2 groups and fragmentations related to the dansyl group were found. However, given the limited number of constitutional isomers that are possible with the given chemical formula, additional structure validation can be accomplished by using reference standards.

MS/MS spectra of 162 tier-2 metabolites were not collected using either targeted or non-targeted acquisition method, because their intensities were too low to reach the threshold (1e5) set for precursor ion selections in MS/MS. Although modifications to the

current MS/MS acquisition method can be made, such as lowering the threshold, the quality

of the collected MS/MS in this case cannot be ensured. For those metabolites, together with

the metabolites only with low-quality MS/MS available due to the low abundance in

biological samples, a more sensitive LC-MS/MS platform will be required, or more efforts

should be made to develop a more sensitive targeted MS/MS acquisition method (find

details in Future Work).
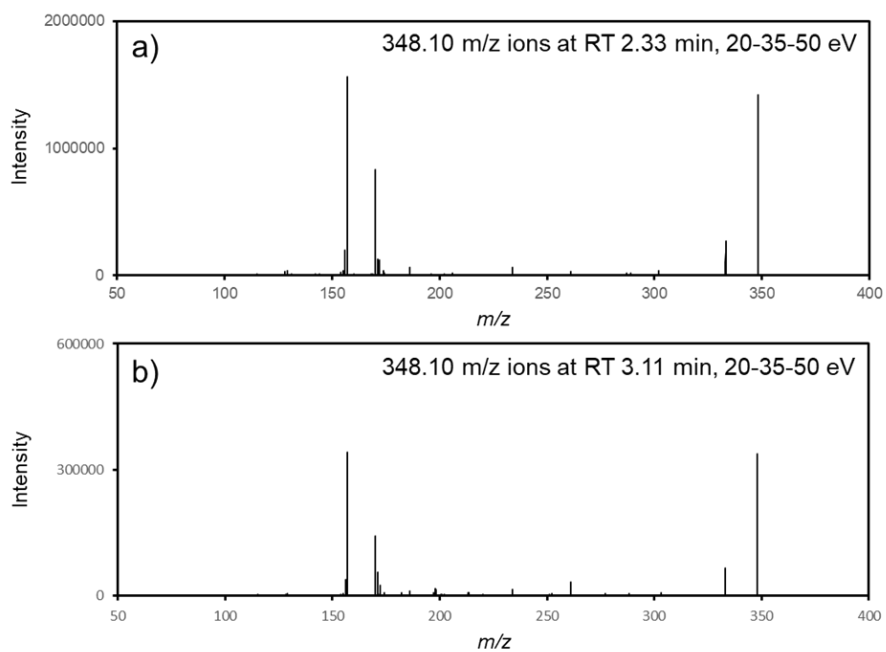
Figure 5.16 MS/MS spectra collected from a) 348.10 m/z ions at RT 2.33 min, and b)

348.10 m/z ions at RT 3.11 min.
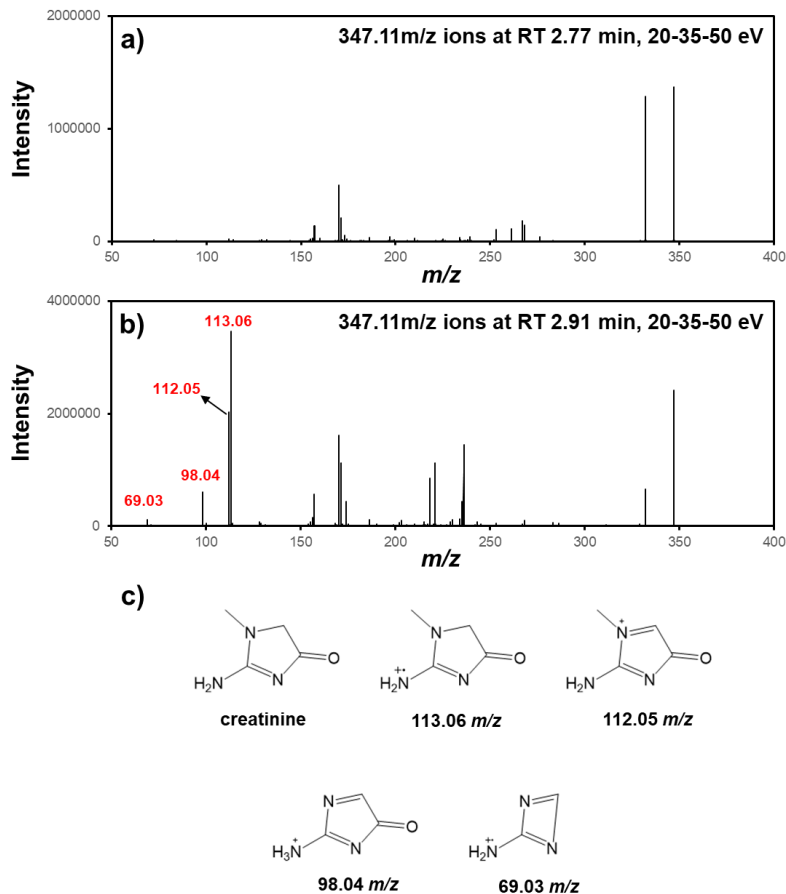
Figure 5.17 MS/MS spectra collected from a) 347.11 m/z ions at RT 2.77 min and b) 347.11 m/z ions at RT 2.91 min, and c) structures of creatine and four fingerprint ions detected in MS/MS spectra.

Finally, 186 metabolites with 2 tags were identified, accounting for approximately 31% of total tier-2 metabolites. However, the validation of these metabolites was skipped in this work as the main focus was on metabolites with a single tag. It is anticipated that metabolites with 2 tags may exhibit more complex fragmentation patterns, given their increased structural complexity. The presence of multiple tags may lead to more diverse fragmentation pathways, making their identification and interpretation more challenging.

Considering the high weight of metabolites with 2 tags, a follow-up study specifically focused on those metabolites is warranted, which will involve exploring and understanding the intricate fragmentation patterns, further validating their identities, and unraveling their unique structural characteristics.
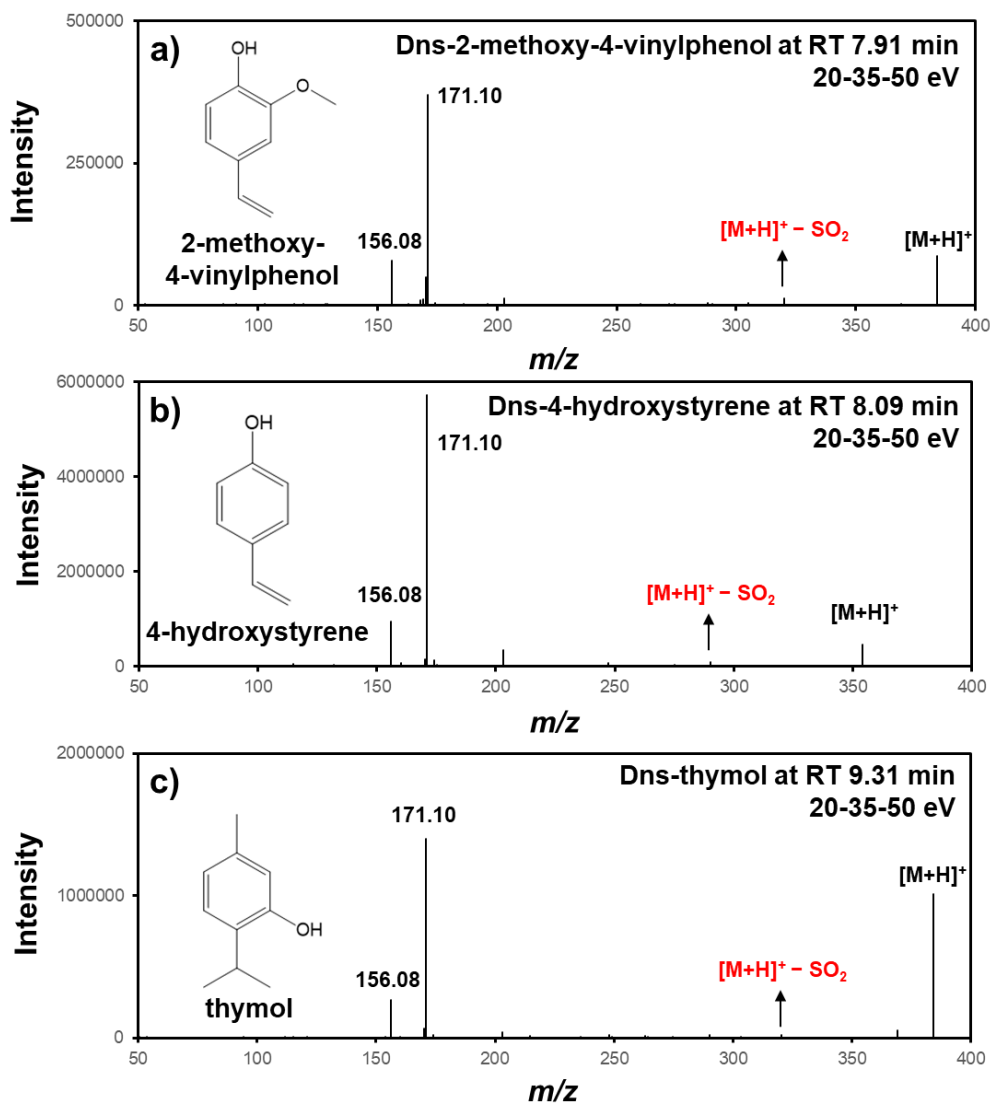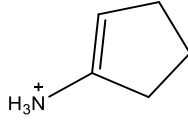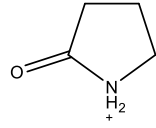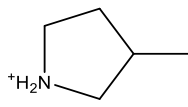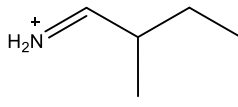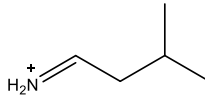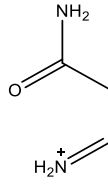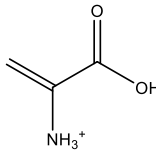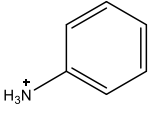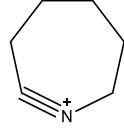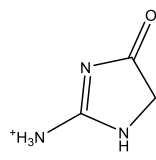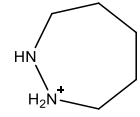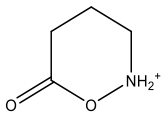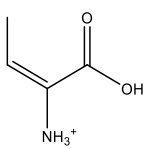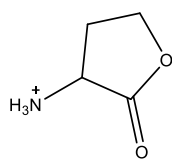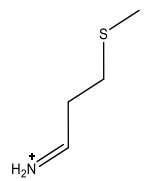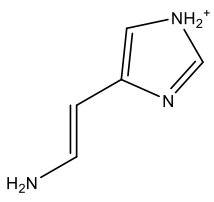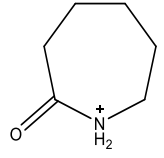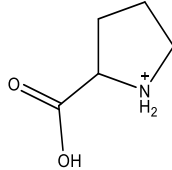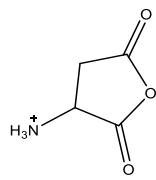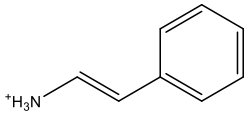
Figure 5.18 MS/MS spectra collected from a) Dns-2-methoxy-4-vinylphenol, b) Dns-4-hydroxystyrene, and c) Dns-thymol.

Table 5.1 Fingerprint ions observed in MS/MS interpretation of dansyl-labeled standards and tier-1 metabolites.

No.	m/z	Formula	Structure	Dns-precursor
1	56.0495	$C_3H_6N^+$		Methionine
2	60.0444	$C_2H_6NO^+$		Serine
3	61.0107	$C_2H_5S^+$		Methionine
4	70.0288	$C_3H_4NO^+$		Asparagine
5	70.0652	$C_4H_8N^+$		Proline
6	72.0808	$C_4H_{10}N^+$		Valine
7	74.0601	$C_3H_8NO^+$		Threonine
8	83.0604	$C_4H_7N_2^+$		Histidine
9	84.0444	$C_4H_6NO^+$		Glutamic acid
10	84.0444	$C_4H_6NO^+$		Glutamic acid
11	84.0808	$C_5H_{10}N^+$		D-Pipecolic acid

12	84.0808	$C_5H_{10}N^+$		N6-acetyl-L-lysine
13	86.0601	$C_4H_8NO^+$		Gamma-aminobutyric acid
14	86.0965	$C_5H_{12}N^+$		Isoleucine
15	86.0965	$C_5H_{12}N^+$		Isoleucine
16	86.0965	$C_5H_{12}N^+$		Leucine
17	87.0553	$C_3H_7N_2O^+$		Aspartic acid
18	88.0394	$C_3H_6NO_2^+$		Serine
19	94.0652	$C_6H_8N^+$		Kynurenine
20	96.0808	$C_6H_{10}N^+$		6-Aminocaproic acid
21	100.0506	$C_3H_6N_3O^+$		Guanidinoacetic acid
22	101.1074	$C_5H_{13}N_2^+$		N-alpha-Acetyl-L-lysine

23	102.0550	$C_4H_8NO_2^+$		Glutamic acid
24	102.0550	$C_4H_8NO_2^+$		3-Aminoisobutanoic acid
25	102.0550	$C_4H_8NO_2^+$		Glutamic acid
26	104.0529	$C_4H_{10}NS^+$		Methionine
27	110.0713	$C_5H_8N_3^+$		Histidine
28	114.0914	$C_6H_{12}NO^+$		6-Aminocaproic acid
29	116.0707	$C_5H_{10}NO_2^+$		Proline
30	116.0343	$C_4H_6NO_3^+$		Glutamic acid
31	120.0808	$C_8H_{10}N^+$		L-phenylalanine
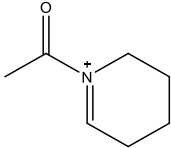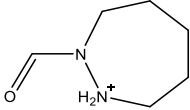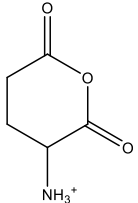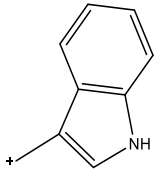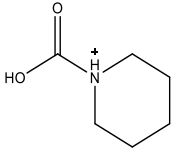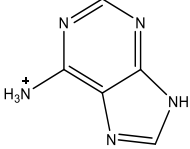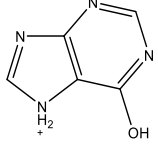
32	126.0914	$C_7H_{12}NO^+$		N6-acetyl-L-lysine
33	129.1023	$C_6H_{13}N_2O^+$		N-alpha-acetyl-L-lysine
34	130.0499	$C_5H_8NO_3^+$		Glutamic acid
35	130.0652	$C_9H_8N^+$		Tryptophan
36	130.0863	$C_6H_{12}NO_2^+$		N6-acetyl-L-lysine
37	136.0618	$C_5H_6N_5^+$		Adenosine
38	137.0458	$C_5H_5N_4O^+$		Hypoxanthine

Table 5.2 Peak pairs identified as tier-2 metabolites from a human urine sample, excluding metabolites with 2 tags.

No	RT (s)	m/z_light	NovaMT Library No.	Compound	Comments
3	76.2	413.138	APA00881000	Kanosamine	Appendix A
14	76.4	531.1664	AXA00237000	1-Methylguanosine	Appendix A
20	76.9	408.1704	AAA00424000	Isomer of L-Arginine/D-Arginine	Appendix A
21	77	403.1437	AYA00001000	Isomer of 1-Methylhistidine	MS/MS not available
22	77	518.1346	AAA00756A02	N2-Acetyl-5-Phosphooxy-L-lysine	Low-quality MS/MS
28	77.3	359.0728	AAA00263000	Isomer of taurine	Low-quality MS/MS
29	77.4	351.112	AAA00618000	Guanidinoacetic acid	Appendix A
35	78.3	454.128	AXA01060000	Aspartyl-Serine/Serylaspartic acid	Peptide
37	78.5	474.1809	AAA00772000	Homocarnosine	Low-quality MS/MS
38	78.5	502.1415	AXA01274000	Phenylalanyl-Cysteine	Peptide
39	78.7	422.1865	AXA01529000	L-Targinine	Appendix A
43	79.2	515.1711	AXA00408000	N6-Methyladenosine	Appendix A
45	79.4	436.2015	AXA00390000	Isomer of symmetric dimethylarginine	Appendix A
46	79.6	382.1065	AXA02944000	Hydroxylated lecithin	Low-quality MS/MS
54	80.9	425.1019	AXA03604000	gamma-Carboxyglutamic acid	Low-quality MS/MS
56	81.4	440.1484	AXA01322000	Serylthreonine	MS/MS not available
65	83.3	366.1127	AXA00651000	Glycyl-glycine	Peptide
66	83.8	597.1142	AXA00626000	8-Oxo-dGMP	MS/MS not available
73	84.9	365.1289	AXA00795000	Beta-Guanidinopropionic acid	Appendix A
80	85.4	482.123	AXA01049000	Aspartyl-Aspartic acid	Peptide
81	85.5	531.1658	AXA00237000	1-Methylguanosine	Appendix A
87	85.8	471.1431	AXA01802000	Deoxyeritadenine	Appendix A
91	86.3	563.111	AAA00503000	2',3'-Cyclic AMP	Low-quality MS/MS
93	86.8	556.1959	AAA01247000	2'-Deamino-2'-hydroxy-6'-dehydroparomamine	Low-quality MS/MS

98	87.4	438.1734	AAA00757A02	N2-Acetyl-L-Hydroxylysine	MS/MS not available
99	87.7	276.0803	AFA00001000	Cyanamide	Appendix A
116	91.6	732.1682	AAA01157000	Bis-gamma-glutamylcystine	MS/MS not available
120	91.9	436.2013	AXA00390000	Symmetric dimethylarginine	Appendix A
121	91.9	618.185	AAA00658000	S-Adenosyl-L-homocysteine	MS/MS not available
123	92	338.116	APA00963000	L-2,3-Diaminopropanoic acid	MS/MS not available
129	92.3	359.1539	AXA00246000	Isomer of 3-Methylhistamine	Low-quality MS/MS
133	92.5	380.1273	AAA00432000	Isomer of L-Glutamine/D-Glutamine	Appendix A
136	93	471.1434	AXA01802000	Deoxyeritadenine	Appendix A
137	93.1	515.1704	AXA00408000	N6-Methyladenosine	False identification
145	93.7	492.1411	APA00960000	L-gamma-Glutamyl-(3R)-L-beta-ethynylserine	Low-quality MS/MS
147	93.9	397.1428	AAA01231000	2-Deoxy-scyllo-inosamine	Low-quality MS/MS
151	94	426.1339	AXA01321000	Serylserine	MS/MS not available
155	94.3	417.0784	AXA00327000	Isomer of L-Homocysteic acid	MS/MS not available
168	99.1	521.1692	AXA00300000	N-Ribosylhistidine	Low-quality MS/MS
170	99.5	422.137	AFA00118000	Tabtoxinine-beta-lactam	Low-quality MS/MS
175	103.1	436.1536	AXA01320000	Serylproline	Peptide
178	104.1	617.1649	AXA00151000	Succinyladenosine	Appendix A
183	106.7	493.1745	AXA01148000	Hydroxyprolyl-Glutamine	Peptide
189	108.3	380.1274	AAA00842000	Isoglutamine	MS/MS not available
191	110.5	480.1443	AXA00576000	L-alpha-Aspartyl-L-hydroxyproline	Peptide
192	110.8	521.1688	AXA00300000	N-Ribosylhistidine	Low-quality MS/MS
202	113	408.1579	AAA00425A01	N5-Acetyl-L-Ornithine	MS/MS not available
211	114.5	399.1041	AAA00660000	L-Methionine S-oxide	MS/MS not available
214	114.8	422.1382	AFA00118000	Tabtoxinine-beta-lactam	Appendix A
215	115.3	471.0892	AXA01531000	S-Cysteinosuccinic acid	Low-quality MS/MS
220	115.7	498.1345	AAA00620A02	N2'-Acetyl-L-Cystathionine	Appendix A

221	116.2	308.1065	AXA03712000	Acetylhydrazine	Appendix A
223	116.3	353.1163	AXA04207000	Hydroxyethyl glycine	Appendix A
225	116.5	467.1964	AXA01242000	Lysyl-Serine	Peptide
226	117.1	347.1171	AAA00774000	Creatinine	False identification
232	117.7	466.1649	AXA01335000	Threoninyl-Hydroxyproline	Peptide
240	118.9	414.1217	APA00957000	4-Chloro-L-lysine	Low-quality MS/MS
249	121.5	480.1802	AXA01399000	Valyl-Glutamic acid	Peptide
250	121.5	408.1584	AAA00425A01	N5-Acetyl-L-Ornithine	Appendix A
254	122	452.1852	AXA01410000	Valyl-Threonine	Peptide
255	122.1	523.1861	AFA00121000	Isotabtoxin	Low-quality MS/MS
273	123.9	468.1438	APA00965000	2-[(2-Aminoethylcarbamoyl)methyl]-2-hydroxybutanedioic acid	Low-quality MS/MS
275	124.1	600.2014	AXA01830000	Tetrahydropentoxyline/1-(1,2,3,4,5-Pentahydroxypent-1-yl)-1,2,3,4-tetrahydro-beta-carboline-3-carboxylic acid	Low-quality MS/MS
276	124.1	480.1591	AYA00012000	Isomer of Gly-Gly-Gly-Gly	Peptide
285	125.4	516.0929	AAA00674A01	N-Acetyl-L-Cystine	Low-quality MS/MS
286	125.4	475.0667	AXA00505000	3-Mercaptolactate-cysteine disulfide	Low-quality MS/MS
292	126.3	438.1706	AAA00757A02	N2-Acetyl-L-Hydroxylysine	MS/MS not available
295	126.7	563.1331	AAA00999G01	3-Hydroxyanthranilate O-glucuronide	Low-quality MS/MS
301	127.8	450.1692	AAA00761000	gamma-Glutamyl-gamma-aminobutyraldehyde	Peptide
304	128.4	339.1011	AAA00627000	Isomer of L-Serine/D-Serine	Appendix A
305	129.3	597.1144	AXA00626000	8-Oxo-dGMP	MS/MS not available
310	129.8	523.1855	AFA00121000	Isotabtoxin	MS/MS not available
318	131.1	480.1441	AXA00576000	L-alpha-Aspartyl-L-hydroxyproline	Peptide
319	131.2	422.138	AXA00589000	L-glycyl-L-hydroxyproline	Peptide
322	131.7	451.0993	AAA00871S01	Tyramine O-sulfate	Appendix A
323	132	408.1588	AAA00425A01	N5-Acetyl-L-Ornithine	MS/MS not available
329	132.6	462.1695	AXA01156000	Hydroxyprolyl-Proline	Peptide

334	133.4	507.2274	AFA00136000	(2R,3R)-3-Methylglutamyl-5-semialdehyde-N6-lysine	Low-quality MS/MS
343	134.3	381.1113	AFA00141000	Isomer of Glutamic acid/D-Glutamic acid/L-Glutamic acid	Appendix A
344	134.4	365.1645	AAA00781000	N-Carbamoylputrescine	Appendix A
347	134.4	462.1696	AXA01156000	Hydroxyprolyl-Proline	Peptide
348	134.5	413.1199	AXA01549000	(R)C(R)S-S-Propylcysteine sulfoxide	Appendix A
353	135.1	515.1711	AXA00408000	N6-Methyladenosine	MS/MS not available
354	135.2	413.0845	AXA01528000	S-Carboxymethyl-L-cysteine	Low-quality MS/MS
356	135.5	379.132	AAA00733000	6-Amino-2-hydroxyl-3-hexenoic acid	Appendix A
357	135.6	363.1487	AAA00770000	4-Guanidinobutanal	False identification
376	140.8	348.1011	AFA00145000	3-Cyano-L-alanine	Appendix A
381	142.6	353.1163	AXA03778000	Aminoethoxyacetic acid	Appendix A
386	144.7	379.1316	AAA00733000	6-Amino-2-hydroxyl-3-hexenoic acid	Appendix A
404	149.4	454.1278	AXA01060000	Aspartyl-Serine/Serylaspartic acid	Peptide
406	150	463.2371	AXA03617000	N-[4-[[(E)-5-aminopent-2-enyl]amino]-2-hydroxybutyl]acetamide	Appendix A
417	151.5	308.1067	AXA03712000	Acetylhydrazine	Low-quality MS/MS
418	151.5	451.1007	AAA00871S01	Tyramine O-sulfate	MS/MS not available
420	151.7	409.1072	AAA01706000	7-Cyano-7-carbaguanine	MS/MS not available
425	152	480.1436	AXA00576000	L-alpha-Aspartyl-L-hydroxyproline	Peptide
434	154.1	482.0583	AAA00905S02	3,4-Dihydroxymandelaldehyde 4-O-sulfate	MS/MS not available
435	154.3	350.1533	AAA00740000	5-Aminopentanamide	False identification
439	154.8	351.1012	AAA00631000	L-Aspartate 4-semialdehyde	Low-quality MS/MS
440	155.3	353.117	AXA03778000	Aminoethoxyacetic acid	Appendix A
447	156.3	381.1494	AXA01555000	4-Amino-4-hydroxy-2,3-dimethylbutanoic acid	Appendix A
451	157.4	436.1904	AXA01185000	Isoleucyl-Alanine	Peptide
470	160	518.1709	AXA01418000	gamma-Glutamylhistidine	Peptide
472	160.7	578.133	AAA00905G02	3,4-Dihydroxymandelaldehyde 4-O-glucuronide	MS/MS not available
473	160.7	530.109	AXA01533000	N-Acetyldjenkolic acid	MS/MS not available

479	161.1	322.1589	AAA00758000	Putrescine	Appendix A
481	161.5	454.1431	AXA01085000	Cysteinyl-Valine	Peptide
484	161.7	381.1478	AXA01555000	(2R,3R,4R)-2-Amino-4-hydroxy-3-methylpentanoic acid	Low-quality MS/MS
496	163.6	395.1276	AFA00181000	4-Methyl-L-glutamic acid	Appendix A
499	163.7	473.1602	AXA00756000	O2'-4a-cyclic-tetrahydrobiopterin	Low-quality MS/MS
500	163.8	495.089	AAA00285S01	L-Tyrosine O-sulfate	MS/MS not available
507	165.1	467.0578	AAA00999S01	3-Hydroxyanthranilate O-sulfate	Low-quality MS/MS
510	166.1	348.1376	AXA00006000	3-Amino-2-piperidone	Appendix A
514	166.5	494.1953	AXA01191000	Isoleucyl-Glutamic acid	Peptide
517	166.7	379.1322	AAA00733000	6-Amino-2-oxohexanoic acid	MS/MS not available
519	166.9	347.1174	AAA00774000	Creatinine	False identification
520	167.1	492.1437	APA00960000	L-gamma-Glutamyl-(3R)-L-beta-ethynylserine	Peptide
534	168.7	295.1112	AAA01327000	Ethanolamine	Appendix A
545	170.1	441.1142	AXA02079000	(2R,2'S)-Isobuteine	MS/MS not available
550	170.5	496.0741	AXA03643000	3-hydroxy-3-(3-hydroxyphenyl)propanoic acid-O-sulphate	MS/MS not available
551	170.5	468.0416	AAA00868S02	2,5-Dihydroxybenzoate 5-O-sulfate	MS/MS not available
554	171.3	540.0996	AXA03651000	4-Hydroxy-5-(dihydroxyphenyl)-valeric acid-O-sulphate/4-Hydroxy-5-(dihydroxyphenyl)-valeric acid-O-sulphate III	Low-quality MS/MS
558	171.6	564.1173	AAA00868G01	2,5-Dihydroxybenzoate 2-O-glucuronide	Low-quality MS/MS
563	171.8	531.1659	AXA00476000	3'-O-Methylguanosine	MS/MS not available
564	172.2	372.1123	AXA00672000	8-Hydroxypurine	Low-quality MS/MS
565	172.4	367.1324	AXA04549000	O-Methyl-L-threonine	Appendix A
571	173.2	409.1416	AXA00325000	N-Carboxyethyl-g-aminobutyric acid	Low-quality MS/MS
574	173.4	446.1743	AXA00596000	L-prolyl-L-proline	Peptide
578	173.8	347.1171	AAA00774000	Creatinine	Appendix A
579	174	379.1327	AAA00733000	6-Amino-2-oxohexanoic acid	MS/MS not available
582	174.3	502.1757	AXA01151000	Hydroxyprolyl-Histidine	Peptide

583	174.4	481.0736	AXA04242000	2-Hydroxyacetaminophen sulfate	Low-quality MS/MS
586	174.8	366.0642	AAA01479000	2-Hydroxyethylenedicarboxylic acid	Low-quality MS/MS
600	176	436.1917	AXA01185000	Isoleucyl-Alanine	Peptide
604	176.3	362.1167	AFA00143000	gamma-Amino-gamma-cyanobutanoic acid	Low-quality MS/MS
605	176.4	512.1827	AXA00592000	L-phenylalanyl-L-hydroxyproline	Peptide
623	179.2	592.1487	AXA03556000	Dihydrocaffeic acid 3-O-glucuronide	Low-quality MS/MS
626	179.5	441.1149	AXA02079000	(2R,2'S)-Isobuteine	Appendix A
628	179.9	481.0737	AXA04242000	2-Hydroxyacetaminophen sulfate	Low-quality MS/MS
640	180.7	618.1641	AXA03661000	5-(3',4'-Dihydroxyphenyl)-gamma-valerolactone-4'-O-glucuronide	MS/MS not available
647	181	409.1058	AAA01706000	7-Cyano-7-carbaguanine	Low-quality MS/MS
648	181	351.1371	AAA00735000	Isomer of 5-Aminopentanoic acid	False identification
654	181.5	295.1117	AAA01327000	Ethanolamine	Appendix A
657	181.7	478.2005	AXA01193000	Isoleucyl-Hydroxyproline	Peptide
659	182.5	494.0565	AAA00872S01	2-Hydroxy-3-(4-hydroxyphenyl)propenoate O-sulfate	MS/MS not available
670	183.4	441.1153	AXA02079000	(2R,2'S)-Isobuteine	MS/MS not available
672	184	482.0578	AAA00905S02	3,4-Dihydroxymandelaldehyde 4-O-sulfate	Low-quality MS/MS
678	185	393.1482	APA00768000	Calystegin A3	MS/MS not available
683	186.6	607.1597	AAA00891G01	3,4-Dihydroxy-L-phenylalanine 3-O-glucuronide	Low-quality MS/MS
687	187.4	348.1015	AFA00145000	3-Cyano-L-alanine	Appendix A
689	187.7	378.1847	AXA00663000	1-(3-Aminopropyl)-4-aminobutanal	Appendix A
690	187.7	429.1154	AXA00280000	4-Carboxyphenylglycine	MS/MS not available
691	187.8	409.1072	AAA01706000	7-Cyano-7-carbaguanine	Low-quality MS/MS
692	188.2	441.1157	AXA02079000	(2R,2'S)-Isobuteine	MS/MS not available
700	190	478.2013	AXA01193000	Isoleucyl-Hydroxyproline	Peptide
703	190.4	562.1324	AAA00895G02	3,4-Dihydroxyphenylacetaldehyde 4-O-glucuronide	Low-quality MS/MS
705	190.6	385.1255	AXA00329000	Isomer1 of 2-Phenylglycine	MS/MS not available
708	191.6	482.0582	AAA00905S01	3,4-Dihydroxymandelaldehyde 3-O-sulfate	MS/MS not available

716	192.3	309.127	AAA00639000	(R)-1-Aminopropan-2-ol	Appendix A
722	192.6	512.1852	AYA00023000	Leucyl-Methionine sulfoxide	Peptide
725	193.4	540.0998	AXA03651000	4-Hydroxy-5-(dihydroxyphenyl)-valeric acid-O-sulphate/4-Hydroxy-5-(dihydroxyphenyl)-valeric acid-O-sulphate III	MS/MS not available
729	194.6	618.1641	AXA03661000	5-(3',4'-Dihydroxyphenyl)-gamma-valerolactone-4'-O-glucuronide	Appendix A
731	194.7	520.1272	AXA01580000	3'-Methoxyfukiic acid	MS/MS not available
733	195.1	432.1334	AAA00533000	5-Acetylamino-6-amino-3-methyluracil	Low-quality MS/MS
736	195.5	378.1863	AXA00663000	4-(3-Aminobutylamino)butan-2-one	Appendix A
738	196.1	461.1491	AAA00573000	Deoxycytidine	MS/MS not available
740	196.5	365.1169	AAA01774000	(S)-4-Amino-5-oxopentanoic acid	MS/MS not available
743	197.1	497.1508	AXA01037000	Asparaginyl-Methionine	Peptide
745	197.6	488.185	AXA01511000	L-Furosine	Low-quality MS/MS
748	197.7	618.1641	AXA03661000	5-(3',4'-Dihydroxyphenyl)-gamma-valerolactone-4'-O-glucuronide	Appendix A
749	197.7	359.1171	AXA00602000	3-Methylcytosine	Low-quality MS/MS
752	197.9	543.2275	AXA01383000	Tyrosyl-Lysine	Peptide
761	199	349.1216	AXA02077000	4-Amino-2-methylenebutanoic acid	Appendix A
764	199.2	496.0733	AXA04182000	3-(3,5-dihydroxyphenyl)-1-propanoic acid sulphate	Appendix A
788	201.9	410.0909	AAA00132000	Ascorbic acid	Low-quality MS/MS
792	202.3	385.1229	AXA00329000	Isomer2 of 2-Phenylglycine	MS/MS not available
793	202.3	393.1483	APA00768000	Calystegin A3	Low-quality MS/MS
794	202.3	379.1326	AAA00744000	(S)-5-Amino-3-oxohexanoic acid	MS/MS not available
799	202.9	349.1218	AXA02077000	4-Amino-2-methylenebutanoic acid	Low-quality MS/MS
803	203.1	580.1495	AAA00906G01	3,4-Dihydroxyphenylethyleneglycol 3-O-glucuronide	MS/MS not available
804	203.2	468.0422	AAA00868S01	2,5-Dihydroxybenzoate 2-O-sulfate	MS/MS not available
807	203.5	563.1344	AAA00999G01	3-Hydroxyanthranilate O-glucuronide	MS/MS not available
816	205	531.1652	AXA00476000	3'-O-Methylguanosine	MS/MS not available
826	206.1	528.1806	AXA00066000	gamma-Glutamylphenylalanine	MS/MS not available

835	207.2	618.1646	AXA03661000	5-(3',4'-Dihydroxyphenyl)-gamma-valerolactone-4'-O-glucuronide	Low-quality MS/MS
844	208.7	520.1277	AXA01580000	3'-Methoxyfukiic acid	Low-quality MS/MS
846	208.9	307.1115	AAA00636000	Aminoacetone	Appendix A
847	209	393.148	APA00768000	Calystegin A3	Low-quality MS/MS
850	209.4	415.1547	AXA00337000	6-Methyltetrahydropterin	MS/MS not available
874	212	553.1676	AXA03788000	Salbutamol 4-O-sulfate	MS/MS not available
877	212.6	321.0909	AAA00593000	Aminoacrylic acid	Low-quality MS/MS
881	213	365.1534	ABA00614000	6-Aminohexanoic acid	Appendix A
884	213.8	448.1063	AAA00940000	2-Hydroxy-6-oxonona-2,4-diene-1,9-dioic acid	MS/MS not available
894	215	578.1332	AAA00283G02	Homogentisic acid 5-O-glucuronide	MS/MS not available
895	215.1	571.1383	AXA00620000	2,8-Dihydroxyquinoline-beta-D-glucuronide	MS/MS not available
898	215.6	536.1596	AYA00025000	Histidinyl-Methionine sulfoxide	Peptide
900	216	369.1125	AAA00455000	Adenine	Appendix A
904	216.5	520.1275	AXA01580000	3'-Methoxyfukiic acid	Low-quality MS/MS
909	217.1	460.1535	AAA01757000	Porphobilinogen	Low-quality MS/MS
910	217.2	482.0577	AAA00283S01	Homogentisic acid 2-O-sulfate	MS/MS not available
912	217.5	494.0577	AAA00872S01	2-Hydroxy-3-(4-hydroxyphenyl)propenoate O-sulfate	Low-quality MS/MS
913	217.5	414.1229	APA00957000	4-Chloro-L-lysine	MS/MS not available
916	217.6	580.1484	AAA00906G01	3,4-Dihydroxyphenylethyleneglycol 3-O-glucuronide	Low-quality MS/MS
917	217.6	470.1382	AAA00988000	L-Formylkynurenine/N'-Formylkynurenine	MS/MS not available
923	218.2	321.1271	AAA00759000	4-Aminobutyraldehyde	Appendix A
928	218.5	463.1168	APA00808000	4,5-seco-Dopa/4-(L-Alanin-3-yl)-2-hydroxy-cis,cis-muconate 6-semialdehyde	MS/MS not available
934	219.1	571.1383	AXA00620000	2,8-Dihydroxyquinoline-beta-D-glucuronide	MS/MS not available
938	219.3	410.0909	AAA00132000	Ascorbic acid	MS/MS not available
955	222.4	601.1491	AXA00776000	Xanthurenate-8-O-beta-D-glucoside	MS/MS not available
961	223	444.1593	AAA01598A01	N-Acetyl-Pyridoxamine	MS/MS not available

967	223.3	592.1843	AXA01646000	Phlorisobutyrophenone 2-glucoside	MS/MS not available
968	223.3	496.0734	AXA03557000	Dihydrocaffeic acid 3-sulfate	MS/MS not available
974	223.7	393.1482	APA00768000	Calystegin A3	Low-quality MS/MS
977	224.5	321.127	AAA00759000	4-Aminobutyraldehyde	Appendix A
979	225.1	578.1331	AAA00283G01	Homogentisic acid 2-O-glucuronide	MS/MS not available
984	226.1	421.1797	AAA01669000	8-Amino-7-oxononanoic acid	MS/MS not available
986	226.2	594.196	AXA00480000	Dityrosine	MS/MS not available
988	226.8	544.1757	AXA01124000	Glutamyltyrosine	Peptide
994	227.6	468.0422	AAA00868S01	2,5-Dihydroxybenzoate 2-O-sulfate	MS/MS not available
997	228.3	393.1121	AFA00180000	4-Methylene-L-glutamic acid	Appendix A
1001	228.4	592.1488	AXA03556000	Dihydrocaffeic acid 3-O-glucuronide	MS/MS not available
1004	229.3	369.1126	AAA00455000	Adenine	Appendix A
1008	229.9	450.2062	AXA01413000	Valyl-Valine	Peptide
1026	232.2	463.0631	AAA00889S01	5,6-Dihydroxyindole 5-O-sulfate/5,6-Dihydroxyindole 6-O-sulfate	Appendix A
1027	232.5	488.185	AXA01550000	L-Pyridosine	MS/MS not available
1035	234.3	592.1486	AXA03556000	Dihydrocaffeic acid 3-O-glucuronide	MS/MS not available
1036	234.4	436.0687	ABA00129000	(1E,3E)-4-Hydroxybuta-1,3-diene-1,2,4-tricarboxylic acid	MS/MS not available
1042	235.2	424.0524	ABA00149000	3-Sulfocatechol	Low-quality MS/MS
1044	235.6	460.1545	AAA01757000	Porphobilinogen	MS/MS not available
1047	236.8	563.1337	AAA00999G01	3-Hydroxyanthranilate O-glucuronide	Appendix A
1057	237.9	415.1187	AAA01719000	7,8-Dihydroxanthopterin	Appendix A
1060	238.1	484.1905	AXA01777000	Paucine	MS/MS not available
1065	238.7	548.1226	AAA00869G02	Gentisate aldehyde 5-O-glucuronide	MS/MS not available
1066	238.7	321.0905	AAA00593000	Aminoacrylic acid	MS/MS not available
1067	238.7	514.1648	AXA04461000	Isomer of L-Aspartyl-L-phenylalanine	Peptide
1072	239.3	578.1328	AAA00283G01	Homogentisic acid 2-O-glucuronide	MS/MS not available
1074	239.5	571.1386	AXA00620000	2,8-Dihydroxyquinoline-beta-D-glucuronide	MS/MS not available

1076	239.6	494.0577	AXA03550000	Caffeic acid 4-sulfate	Low-quality MS/MS
1079	239.7	396.0939	AAA00646000	1,2-Dihydroxy-5-(methylthio)pent-1-en-3-one	Low-quality MS/MS
1080	239.9	563.1334	AAA00999G01	3-Hydroxyanthranilate O-glucuronide	Appendix A
1092	241.8	448.1903	AXA01303000	Prolyl-Valine	Peptide
1094	241.9	337.0854	AAA00589000	(Z)-3-Peroxyaminoacrylic acid	Low-quality MS/MS
1098	242.2	522.089	AXA01429000	5'-(3',4'-Dihydroxyphenyl)-gamma-valerolactone sulfate	Low-quality MS/MS
1103	242.6	592.1488	AXA03556000	Dihydrocaffeic acid 3-O-glucuronide	MS/MS not available
1117	244.1	335.1425	APA00767000	5-Aminopentanal	Appendix A
1118	244.2	445.1067	AAA00980000	5-(2'-Formylethyl)-4,6-dihydroxypicolinic acid	MS/MS not available
1125	245.2	365.1165	AAA01774000	(S)-4-Amino-5-oxopentanoic acid	MS/MS not available
1136	246.5	563.134	AAA00999G01	3-Hydroxyanthranilate O-glucuronide	Low-quality MS/MS
1138	246.6	415.1327	AXA00320000	L-Threo-3-Phenylserine	Peptide
1141	247.3	462.1215	ABA00389000	2-Hydroxy-3-carboxy-6-oxo-7-methylocta-2,4-dienoic acid	MS/MS not available
1148	247.8	321.1271	AXA02555000	Morpholine	Appendix A
1149	248.1	578.1333	AAA00283G01	Homogentisic acid 2-O-glucuronide	Low-quality MS/MS
1157	248.5	424.1065	AAA00882000	2,4-Dihydroxyhept-2-enedioic acid	Low-quality MS/MS
1165	249.4	397.1254	APA00823000	Homomethionine	Low-quality MS/MS
1166	249.5	466.0622	AXA03544000	2,4-Dihydroxyacetophenone 5-sulfate	MS/MS not available
1167	249.5	335.1425	APA00767000	5-Aminopentanal	Appendix A
1179	250.5	668.18	APA00456000	8-C-Glucosylnaringenin	MS/MS not available
1183	250.9	452.1625	AAA01013000	Isomer of N-Acetylserotonin	MS/MS not available
1186	251.2	363.1371	AXA01551000	2-Pyrrolidineacetic acid	Appendix A
1194	252.5	484.0736	AXA03670000	Hydroxymethoxyphenylcarboxylic acid-O-sulphate	MS/MS not available
1203	253.1	473.1377	AAA00891A01	N-Acetyl-3,4-Dihydroxy-L-phenylalanine	MS/MS not available
1212	254.6	417.1118	AAA01610000	4-Pyridoxic acid	MS/MS not available
1221	255.7	668.1803	APA00456000	8-C-Glucosylnaringenin	MS/MS not available
1233	256.7	486.1697	AXA00999000	Isomer of Alanyl-Tyrosine	Peptide

1239	258.1	430.0975	AAA00892000	3-(3,4-Dihydroxyphenyl)pyruvic acid	MS/MS not available
1248	259	443.1271	AAA00861000	p-Hydroxyphenylacetylglycine	MS/MS not available
1272	260.7	648.2117	AXA04113000	N-desalkyl delavirdine	MS/MS not available
1285	262.3	571.1385	AXA00620000	2,8-Dihydroxyquinoline-beta-D-glucuronide	Low-quality MS/MS
1289	262.8	395.1064	AXA01547000	trans-S-(1-Propenyl)-L-cysteine	Low-quality MS/MS
1297	263.8	279.1165	AXA00797000	Ethylamine	Appendix A
1314	265.8	462.1691	AXA00537000	Prolylhydroxyproline/Prolyl-Hydroxyproline	Peptide
1328	267.2	418.1322	AAA00907000	3-Methoxy-4-hydroxyphenylethyleneglycol	Low-quality MS/MS
1331	267.6	571.1389	AXA00620000	2,8-Dihydroxyquinoline-beta-D-glucuronide	Low-quality MS/MS
1335	268.4	335.1426	AYA00051000	4-Hydroxypiperidine	Low-quality MS/MS
1337	268.5	494.1382	AXA02834000	1,2,3,4-Tetrahydro-b-carboline-1,3-dicarboxylic acid	MS/MS not available
1346	269.4	522.0893	AXA01429000	5'-(3',4'-Dihydroxyphenyl)-gamma-valerolactone sulfate	MS/MS not available
1353	270.1	351.0832	AXA00344000	Homocysteine thiolactone	Appendix A
1355	270.2	321.127	AXA02555000	Morpholine	MS/MS not available
1379	271.9	446.1272	AXA00152000	Vanillactic acid	MS/MS not available
1382	272	424.0524	ABA00149000	3-Sulfocatechol	Appendix A
1385	272.1	446.0906	ABA00561000	5-Carboxyvanillic acid	Low-quality MS/MS
1390	272.8	604.1484	AXA03564000	Isoferuloyl C1-glucuronide	MS/MS not available
1393	273.1	592.1487	AXA03556000	Dihydrocaffeic acid 3-O-glucuronide	Low-quality MS/MS
1395	273.4	335.1424	AYA00051000	4-Hydroxypiperidine	Low-quality MS/MS
1399	273.7	668.18	APA00456000	8-C-Glucosylnaringenin	Low-quality MS/MS
1409	274.8	344.1066	AXA03891000	3-Hydroxy-4-aminopyridine	Low-quality MS/MS
1412	275.4	429.1463	AAA00893A01	N-Acetyl-Dopamine	MS/MS not available
1416	276.2	279.1165	AXA00797000	Ethylamine	Appendix A
1421	276.7	429.1113	AXA00485000	3-Hydroxyhippuric acid	Appendix A
1426	277.3	321.0906	AAA00593000	Aminoacrylic acid	Low-quality MS/MS
1432	277.9	473.1384	AAA00891A01	N-Acetyl-3,4-Dihydroxy-L-phenylalanine	MS/MS not available

1439	279.1	377.1529	AXA02078000	L-2-Amino-3-methylenehexanoic acid	MS/MS not available
1440	279.1	459.122	AAA01682000	4-Amino-4-deoxychorismic acid	Low-quality MS/MS
1441	279.1	379.1319	AXA01536000	L-trans-5-Hydroxy-2-piperidinecarboxylic acid	MS/MS not available
1468	283.1	446.127	AXA00152000	Vanillactic acid	Low-quality MS/MS
1487	284.5	496.1549	AAA00994A01	N-Acetyl-5-Hydroxy-L-tryptophan	Low-quality MS/MS
1488	284.6	666.1621	AXA01668000	1,8-Dihydroxy-3-hydroxymethylanthraquinone 8-O-b-D-glucoside	Low-quality MS/MS
1501	285.4	416.0835	ABA00434000	4-Hydroxyisophthalic acid	MS/MS not available
1502	285.4	638.1328	AXA01432000	Urolithin A-3-O-glucuronide/Urolithin A-8-O-glucuronide	Low-quality MS/MS
1504	285.6	424.0517	ABA00149000	3-Sulfocatechol	Appendix A
1505	285.6	416.1162	AXA00365000	3-(3-Hydroxyphenyl)-3-hydroxypropanoic acid	Appendix A
1517	287.2	371.106	AXA04432000	3-hydroxybenzamide	Appendix A
1536	289.3	399.1377	AAA00935000	Isomer of L-Phenylalanine/D-Phenylalanine	Appendix A
1537	289.7	373.0862	AXA00784000	3-Hydroxypicolinic acid	Low-quality MS/MS
1541	290	448.189	AXA01303000	Prolyl-Valine	Peptide
1554	291.4	446.1268	AXA00152000	Vanillactic acid	Low-quality MS/MS
1566	292.9	357.1239	AYA00036000	2-Aminobenzyl alcohol	Appendix A
1572	294.5	471.1588	AXA00769000	Salsoline-1-carboxylic acid	MS/MS not available
1580	295.4	576.1904	AXA01873000	3-Hydroxy-3-(3,4-dihydroxy-4-methylpentanoyl)-5-(3-methylbutyl)-1,2,4-cyclopentanetrione	MS/MS not available
1585	295.6	460.1427	AXA03887000	3-(4-Hydroxy-3-methoxyphenyl)-2-methyllactic acid	MS/MS not available
1586	295.6	391.0956	AAA00995000	2-Aminomuconic acid	Low-quality MS/MS
1596	296.2	321.1263	AXA02555000	Morpholine	Low-quality MS/MS
1613	297.9	429.1114	AAA00999A01	N-Acetyl-3-Hydroxyanthranilic acid	MS/MS not available
1627	299.4	446.1264	AXA00152000	Vanillactic acid	MS/MS not available
1630	299.8	496.1538	AAA00994A01	N-Acetyl-5-Hydroxy-L-tryptophan	Low-quality MS/MS
1641	300.9	402.1003	AAA00905000	3,4-Dihydroxyphenylacetic acid	Appendix A
1661	302.6	432.1951	APA00781000	Slaframine	MS/MS not available

1671	304.7	335.1423	AYA00051000	4-Hydroxypiperidine	Appendix A
1680	305.3	425.1168	AAA01010000	Isomer of 4-Hydroxyindoleacetic acid	Appendix A
1682	305.5	444.1473	AXA01435000	3,4-Dihydroxyphenylvaleric acid	Low-quality MS/MS
1696	308	279.1161	AFA00212000	Dimethylamine	Appendix A
1703	310.3	399.1376	AXA00479000	Norsalsolinol	Low-quality MS/MS
1719	312.9	410.153	AAA01008000	Isomer of Serotonin	MS/MS not available
1762	318.1	588.1533	AXA02170000	1-O-Caffeoylquinic acid	MS/MS not available
1767	318.8	392.0827	ABA00132000	3-Hydroxy-cis,cis-muconic acid	MS/MS not available
1781	319.8	444.1473	AXA01435000	3,4-Dihydroxyphenylvaleric acid	MS/MS not available
1808	322.3	396.0938	AAA00646000	1,2-Dihydroxy-5-(methylthio)pent-1-en-3-one	Low-quality MS/MS
1810	322.6	509.2111	APA00753000	Elwesine	MS/MS not available
1815	323.4	406.0956	AAA00859000	2-Hydroxyhepta-2,4-dienedioic acid	MS/MS not available
1821	324.3	480.1611	AXA03944000	cyclic 6-Hydroxymelatonin	MS/MS not available
1845	328.1	425.1563	APA00830000	Trihomomethionine	Appendix A
1860	330.3	416.1164	AXA00009000	Isohomovanillic acid	Low-quality MS/MS
1864	330.9	376.0851	ABA00058000	cis,cis-4-Hydroxymuconic semialdehyde	Low-quality MS/MS
1883	333	305.132	AXA02590000	Pyrrolidine	Appendix A
1891	333.6	399.1012	AAA01596000	4-Pyridoxolactone	Appendix A
1899	335.1	432.1106	ABA00559000	Syringic acid	Low-quality MS/MS
1901	335.5	444.1098	APA00407000	5-Hydroxyferulic acid methyl ester	MS/MS not available
1905	335.6	460.1423	AXA03887000	3-(4-Hydroxy-3-methoxyphenyl)-2-methyllactic acid	Low-quality MS/MS
1956	342.1	430.1319	AXA03886000	3-(3,4-Dihydroxyphenyl)-2-methylpropionic acid	Low-quality MS/MS
1958	342.2	480.1591	AXA03944000	cyclic 6-Hydroxymelatonin	Low-quality MS/MS
1972	343.8	372.0897	ABA00593000	2-Hydroxy-5-methylquinone	Low-quality MS/MS
1973	343.9	507.1953	APA00701000	4'-O-Methylnorbelladine	MS/MS not available
1983	344.7	307.1475	AXA03600000	Diethylamine	Appendix A
1984	344.7	396.0934	AXA04524000	Umbelliferone	Low-quality MS/MS

2001	347.2	416.1161	AXA01641000	2',6'-Dihydroxy-4'-methoxyacetophenone	MS/MS not available
2007	347.8	348.0901	AAA00919000	2-Hydroxy-2,4-pentadienoic acid	MS/MS not available
2009	348.2	521.2144	AXA03747000	Dihydromorphine/6alpha-Hydroxy-hydromorphone/6beta-Hydroxy-hydromorphone/Dihydroisomorphine/Nordihydroisomorphine	MS/MS not available
2016	348.9	430.1322	AXA03886000	3-(3,4-Dihydroxyphenyl)-2-methylpropionic acid	Low-quality MS/MS
2037	350.9	423.1006	AXA04543000	alpha-Cyano-4-hydroxycinnamic acid	Low-quality MS/MS
2058	352.8	458.1267	APA00419000	Sinapic acid	Low-quality MS/MS
2064	353.6	476.164	AXA01314000	Serylhistidine	Low-quality MS/MS
2070	354.1	507.1951	APA00701000	4'-O-Methylnorbelladine	MS/MS not available
2073	354.2	460.1422	AXA03887000	3-(4-Hydroxy-3-methoxyphenyl)-2-methyllactic acid	Low-quality MS/MS
2092	356.7	480.159	AXA03944000	cyclic 6-Hydroxymelatonin	MS/MS not available
2102	357.6	462.1332	AXA01748000	Benzyl salicylic acid	Low-quality MS/MS
2107	358.3	523.1936	AXA03621000	Norcocaine	MS/MS not available
2111	358.7	423.1012	AXA04543000	alpha-Cyano-4-hydroxycinnamic acid	False identification
2113	358.7	430.1329	APA00414000	5-Hydroxyconiferyl alcohol	MS/MS not available
2124	359.8	307.1475	AXA03600000	Diethylamine	MS/MS not available
2158	363.5	519.1951	AXA04149000	norhydrocodone	Low-quality MS/MS
2164	364.3	307.1478	AXA03600000	Diethylamine	MS/MS not available
2195	366.7	456.1109	ABA00498000	2-Hydroxy-4-hydroxymethylbenzalpyruvic acid	MS/MS not available
2239	371.7	451.1715	AXA04065000	N-desisopropylpropranolol	Low-quality MS/MS
2261	374.7	416.1164	AXA01641000	2',6'-Dihydroxy-4'-methoxyacetophenone	MS/MS not available
2266	375.3	432.1112	ABA00559000	Syringic acid	MS/MS not available
2314	382.5	442.132	AXA01427000	5-(3',4'-Dihydroxyphenyl)-gamma-valerolactone	MS/MS not available
2325	383.4	523.23	AXA00912000	Donepezil metabolite M4	MS/MS not available
2338	385.2	412.1212	APA00440000	Coniferyl aldehyde	MS/MS not available
2351	386.9	456.1473	APA00455000	Coniferyl acetic acid	MS/MS not available

2382	390.1	414.1369	AXA03389000	4-Hydroxy-2,6,6-trimethyl-3-oxo-1,4-cyclohexadiene-1-carboxaldehyde	MS/MS not available
2388	390.6	408.1227	ABA00504000	1,2-Dihydroxy-8-methylnaphthalene	False identification
2400	391.5	523.23	AXA00912000	Donepezil metabolite M4	MS/MS not available
2413	393.1	523.194	AXA03621000	Norcocaine	MS/MS not available
2425	394.4	426.1373	APA00447000	Coumaryl acetic acid	Low-quality MS/MS
2455	398.3	523.23	AXA00912000	Donepezil metabolite M4	MS/MS not available
2464	399.1	436.1692	AXA01758000	Tetrahydroharmol	MS/MS not available
2483	401.8	333.1633	AXA02988000	2-Methylpiperidine	Appendix A
2488	402.6	453.1881	AXA03629000	Ritalinic acid	False identification
2510	406.1	519.1957	APA00643000	(S)-Coclaurine	MS/MS not available
2516	407.1	389.1284	AAA00832000	Isomer of L-Histidine	MS/MS not available
2534	408.6	403.144	AXA00786000	4-Aminobiphenyl	False identification
2538	408.9	386.1057	ABA00433000	Isomer of 4-Hydroxy-3-methylbenzoic acid	Low-quality MS/MS
2575	413	400.1217	AXA01658000	Ethyl vanillin	Low-quality MS/MS
2589	414.5	389.1281	AAA00832000	L-Histidine	MS/MS not available
2592	414.7	533.2104	APA00644000	(S)-N-Methylcoclaurine/(R)-N-Methylcoclaurine	MS/MS not available
2605	416.2	370.1106	APA00413000	3,4-Dihydroxystyrene	Low-quality MS/MS
2679	423.2	525.2457	AXA03299000	Norcapsaicin	MS/MS not available
2742	431.5	517.143	AXA01445000	Avenanthramide 1p	Low-quality MS/MS
2926	459.3	566.2573	APA00234000	11,20-Dihydroxysugiol	MS/MS not available
2960	463.3	498.1946	AXA00260000	gamma-CEHC	Appendix A
2981	465.5	505.2192	AXA03907000	4-Hydroxyatomoxetine	MS/MS not available
2986	467.1	497.1533	AXA02036000	Dehydroanonaine	MS/MS not available
2997	469.5	462.177	AXA02849000	Bisphenol A	MS/MS not available
3043	475.9	384.1265	AXA00845000	2-Methoxy-4-vinylphenol	Low-quality MS/MS
3070	481.2	577.2407	APA00665000	Laudanine	MS/MS not available

3077	482	354.1157	APA00409000	4-Hydroxystyrene	Low-quality MS/MS
3098	486.2	505.1834	APA00702000	N-Demethylnarwedine	MS/MS not available
3102	486.5	354.1157	APA00409000	4-Hydroxystyrene	Low-quality MS/MS
3103	486.5	512.2108	AXA00230000	alpha-CEHC	Appendix A
3170	498.9	505.1829	APA00702000	N-Demethylnarwedine	MS/MS not available
3179	500.8	354.1156	APA00409000	4-Hydroxystyrene	MS/MS not available
3208	505.1	524.2108	AXA01573000	[6]-Dehydrogingerdione	MS/MS not available
3228	508	378.1158	AXA00699000	2-Naphthol	Appendix A
3237	509.5	496.2159	AAA00307000	3"-Hydroxy-geranylhydroquinone	MS/MS not available
3289	518.2	527.2581	AXA03300000	Nordihydrocapsaicin	MS/MS not available
3306	520.1	404.131	AXA03073000	2-Biphenylol	MS/MS not available
3419	535.3	540.2421	AXA00723000	5'-Carboxy-gamma-chromanol	MS/MS not available
3421	535.7	497.2471	AXA03732000	O-Desmethylvenlafaxine	MS/MS not available
3514	556.5	384.1626	AXA04525000	Thymol	Low-quality MS/MS
3525	558.7	428.1887	AXA00936000	4-Hydroxypropofol	MS/MS not available
3573	575.1	398.1785	AXA02814000	2-Butyl-4-methylphenol	MS/MS not available
3606	582.7	398.1783	AXA02814000	2-Butyl-4-methylphenol	MS/MS not available
3687	605.6	898.2046	AXA04195000	Clofazimine glucuronide	MS/MS not available
3720	616	496.2153	AAA00307000	3"-Hydroxy-geranylhydroquinone	MS/MS not available
3734	630.4	440.2256	AXA00859000	4-(1,1,3,3-Tetramethylbutyl)-phenol	Low-quality MS/MS
3744	636.5	440.2254	AXA00859000	4-(1,1,3,3-Tetramethylbutyl)-phenol	MS/MS not available

The number of identified metabolites in the first column refers to the count of corresponding peak pairs out of the 3772 detected peak pairs.

5.4 Conclusions

We investigated the fragmentation patterns of dansyl-labeled amine-/phenol-containing metabolites using HCD on Q-Orbitrap MS. Informative MS/MS spectra that encompassed structural information related to both the dansyl group and the metabolite moiety were collected from selected standards and tier-1 metabolites. MS/MS elucidation revealed distinct dansyl label cleavage for phenol-containing and amine-containing metabolites. For the fragmentation related to the metabolite moiety, five types of neural losses including decarboxylation, dehydration, deamination, desulfurization, and deacetylation were found in dansyl-labeled metabolites. Additionally, we identified crucial fingerprint ions within the MS/MS spectra that revealed the core molecular framework of the labeled metabolites. Based on those observations, identities of 85 tier-2 metabolites were validated, showing the promising potential of MS/MS-based metabolite ID validation using HCD in CIL LC-MS metabolomics. Ten metabolites exhibited false tier-2 identities, indicating the need for further expanding the size of current tier-2 library size and improving the RT matching accuracy. For some tier-2 metabolites, only low-quality or even no MS/MS was available, making it challenging to analyze their structures accurately. To address these issues, future work with a focus on the development of a targeted and high-quality MS/MS spectral collection method specifically for low-intensity peaks is needed. Furthermore, we should also direct our efforts towards studying metabolites with

two CIL tags. The presence of two tags adds complexity to the spectra, complicating the

interpretation and identification of these metabolites.

Chapter 6

Conclusion and Future Work

6.1 Thesis summary

CIL LC-MS is a powerful technique in metabolomics with high coverage of metabolome profiling. In CIL LC-MS, metabolites are divided into different sub-groups based on the functional groups they contain, followed by a labeling reaction specific to the functional group. With a rational design of the labeling reagents, significant improvement in LC separation and MS detection can be achieved. Moreover, the usage of differential isotope labeling allows for accurate relative quantification. To date, a four-channel labeling protocol has been developed in our lab for CIL LC-MS metabolome analysis, i.e., ^{12}C-/^{13}C-dansylation for amine-/phenol-containing sub-metabolome, ^{12}C-/^{13}C-dimethylaminophenacyl (DmPA) bromide for carboxyl-containing sub-metabolome, base-activated ^{12}C-/^{13}C-dansylation for hydroxyl-containing sub-metabolome, and ^{12}C-/^{13}C-dansylhydrazine (DnsHz) for carbonyl-containing metabolome. The combination of these four sub-metabolomes can achieve a total 86%-95% non-lipid metabolite coverage.

In our previous studies, quadrupole time-of-flight (QTOF) MS has been extensively used for CIL LC-MS. My research aims to implement CIL LC-MS analysis on Q-Orbitrap MS, with a focus on maximizing the instrumental performance for metabolomics.

In Chapter 2, we examined the effects of instrument type on the detectability of true metabolites with a focus on the comparison of TOF MS vs. Orbitrap MS, regarding on four

main factors: a) detection sensitivity, b) detection dynamic range, c) ionization efficiency, and d) ion transportation efficiency. The results from this study clearly indicate that instrument type can have a profound effect on metabolite detection in CIL LC-MS. Therefore, comparison of metabolome data generated using different instruments needs to be carefully done, particular attention should be paid to the peak pairs that were only detected by a specific LC-MS platform. If necessary, we suggest combining datasets obtained from QTOF and Orbitrap to provide a higher coverage of the metabolome.

Building on the conclusions from Chapter 2, Chapter 3 introduced a segmented spectrum scan method using Orbitrap MS in CIL LC-MS to enhance metabolite detection efficiency. This method yielded a significant increase in detectable metabolites, while maintaining accurate relative quantification and precise peak ratio measurements. Application of the method to dansyl-labeled human urine samples resulted in the detection of 5867 peak pairs or metabolites, a 56% gain compared to 3765 peak pairs detectable with full scan. Out of these, 5575 peak pairs (95.0%) could be identified or mass-matched.

In Chapter 4, we evaluated the performance of the segment-scan-assisted CIL LC-MS across different sample types. The segment scan method outperformed conventional full scan in samples with higher metabolic complexity, such as feces and urine, detecting up to 94% more peak pairs. For less complex samples like saliva, the gain in peak pairs was approximately 25%. The validated 120-m/z segment scan method proved to be a

reliable approach for CIL LC-Q-Orbitrap MS, enhancing metabolite detection efficiency in samples with high or medium metabolic complexity.

Chapter 5 delved into the fragmentation patterns of dansyl-labeled amine-/phenol-containing metabolites using HCD on Q-Orbitrap MS. Informative MS/MS spectra were collected, providing structural information for both the dansyl group and the metabolite moiety. The fragmentation analysis revealed distinct patterns for phenol-containing and amine-containing metabolites, including five types of neutral losses (find details on neutral losses in Table 6.1). Crucial fingerprint ions within the MS/MS spectra unveiled the core molecular framework of the labeled metabolites. Using selected standards and tier-1 metabolites, we validated the identities of 85 tier-2 metabolites, demonstrating the potential of MS/MS-based metabolite ID validation using CID technique.

Table 6.1 Neutral losses found in dansyl-labeled amine-/phenol-containing metabolites.

Compound class	Neutral loss	Mass (Da)
Hydroxyls	H_2O	18.0153
Carboxylic acids	H_2O	18.0153
	CO_2H_2	46.0254
Guanidino-containing compounds	NH_3	17.0305
N-acetyl derivatives	CH_2CO	42.0100
Methyl sulfides	CH_4S	48.1075
Disulfides	H_2S_2	66.1459

6.2 Future work

Based on the results presented in Chapter 2, future research should be focused on hardware modification to minimize the differences in order to generate more reproducible metabolome data from different types of instruments. We can always optimize MS conditions to reduce the differences in dynamic ranges among different instruments, so the hardware modification will be focused on using the same configuration of ESI and ionization housing for different instruments. However, it will be challenging to design a universal ESI configuration that can be compatible with different designs of mass spectrometers.

In Chapter 3-4, an optimal segment scan method has been developed and implemented for routine CIL LC−MS metabolome analysis of samples with different metabolomic complexity. However, data processing still involves some manual steps, such as subset extraction and sub-lists merging. Future work needs to focus on developing a dedicated function to be integrated into IsoMS as a routine data processing procedure for segment-scan-assisted CIL LC-Orbitrap MS.

In Chapter 5, MS/MS interpretation has been proven to be promising to improve the confidence level for putatively identified tier-2 metabolites. Future work in terms of high-quality MS/MS spectral acquisition should be focused on three aspects as follows:

Firstly, acquiring high-quality MS/MS spectra for low-abundance metabolites remains challenging. This is not only due to the inherent dependence of MS/MS spectral

quality on the intensity of precursor ions but also because the MS/MS acquisition of extremely low-intensity precursor ions of interest tends to face contamination from other chemical backgrounds present in the isolation window at the same time. To address this issue, we proposed a targeted MS/MS method focused on low-abundance metabolites for future work. The intensities of low-abundance metabolites can be boosted by increasing the accumulation time during the precursor ion selection process. By prolonging the accumulation time, a higher number of precursor ions can be accumulated, leading to increased ion population and improved sensitivity during MS/MS analysis. Then fragmentations specific to the ions of interest amid other chemical noises can be identified via a "pace-screening" strategy. As shown in Figure 7.1, MS/MS spectra of the precursor of interest are collected consecutively within a period of time. During this process, fragment ions related to the targeted metabolites are expected to display the same pace in intensity changes over time as the protonated metabolite in the MS1 scan. This unique pattern allows for the clear distinction of fragment ions of interest from other chemical noises, enhancing the accuracy and reliability of the MS/MS data analysis.

Secondly, certain metabolites are composed of several stable units, leading to insufficient fragmentation even in HCD. In this case, a pseudo-MS3 strategy on Q-Orbitrap MS can be used for spectral acquisition. The pseudo-MS3 approach involves two steps. First, the protonated metabolites containing stable units undergo an in-source CID in the skimmer region of the mass spectrometer, generating fragment ions corresponding to the

stable units present in the metabolites. These stable unit ions are then introduced into the

HCD cell for further fragmentation. In the second step, the stable unit ions undergo HCD,

leading to additional fragmentation. The HCD-induced fragments can offer insights into

the bonding arrangements and chemical groups within the stable units.
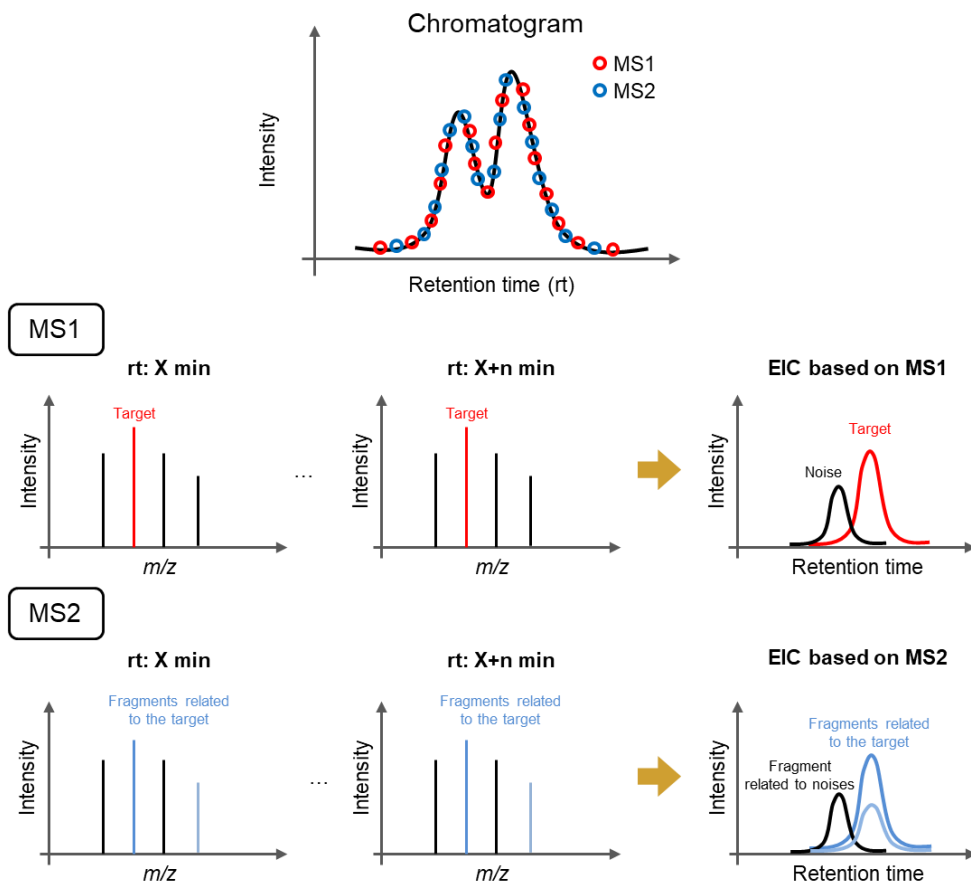
Figure 6.1 Schematic of the "pace-screening" strategy for selective fragment filtering in

MS/MS Analysis.

Finally, a study specifically focused on metabolites with 2 tags should be carried

out in the future as we anticipated that those metabolites may exhibit more complex

fragmentation patterns, given their increased structural complexity.

156

Reference

(1) Idle, J. R.; Gonzalez, F. J. Metabolomics. *Cell Metab* **2007**, *6* (5), 348–351.

(2) Roessner, U.; Bowne, J. What Is Metabolomics All About? *Biotechniques* **2009**, *46* (5), 363–365.

(3) Liu, X.; Locasale, J. W. Metabolomics: A Primer. *Trends Biochem Sci* **2017**, *42* (4), 274–284.

(4) Newsholme, E. A.; Start, C. *Regulation in Metabolism.*; 1973.

(5) Kennedy, E. P. Metabolism of Lipides. *Annu Rev Biochem* **1957**, *26* (1), 119–148.

(6) Patti, G. J.; Yanes, O.; Siuzdak, G. Metabolomics: The Apogee of the Omics Trilogy. *Nat Rev Mol Cell Biol* **2012**, *13* (4), 263–269.

(7) Bustamante, C. D.; De La Vega, F. M.; Burchard, E. G. Genomics for the World. *Nature* **2011**, *475* (7355), 163–165.

(8) Ulitsky, I.; Bartel, D. P. LincRNAs: Genomics, Evolution, and Mechanisms. *Cell* **2013**, *154* (1), 26–46.

(9) Fields, S. Proteomics in Genomeland. *Science (1979)* **2001**, *291* (5507), 1221–1224.

(10) Tyers, M.; Mann, M. From Genomics to Proteomics. *Nature* **2003**, *422* (6928), 193–197.

(11) Beecher, C. W. W. The Human Metabolome. In *Metabolic profiling: Its role in biomarker discovery and gene function analysis*; Springer, 2003; pp 311–319.

(12) Psychogios, N.; Hau, D. D.; Peng, J.; Guo, A. C.; Mandal, R.; Bouatra, S.; Sinelnikov, I.; Krishnamurthy, R.; Eisner, R.; Gautam, B. The Human Serum Metabolome. *PLoS One* **2011**, *6* (2), e16957.

(13) Ilyin, S. E.; Belkowski, S. M.; Plata-Salamán, C. R. Biomarker Discovery and Validation: Technologies and Integrative Approaches. *Trends Biotechnol* **2004**, *22* (8), 411–416.

(14) Goossens, N.; Nakagawa, S.; Sun, X.; Hoshida, Y. Cancer Biomarker Discovery and Validation. *Transl Cancer Res* **2015**, *4* (3), 256.

(15) Buchholz, A.; Hurlebaus, J.; Wandrey, C.; Takors, R. Metabolomics: Quantification of Intracellular Metabolite Dynamics. *Biomol Eng* **2002**, *19* (1), 5–15.

(16) Aretz, I.; Meierhofer, D. Advantages and Pitfalls of Mass Spectrometry Based Metabolome Profiling in Systems Biology. *Int J Mol Sci* **2016**, *17* (5), 632.

(17) Segers, K.; Declerck, S.; Mangelings, D.; Heyden, Y. Vander; Eeckhaut, A. Van. Analytical Techniques for Metabolomic Studies: A Review. *Bioanalysis* **2019**, *11* (24), 2297–2318.

(18) Zhou, R.; Li, L. Effects of Sample Injection Amount and Time-of-Flight Mass Spectrometric Detection Dynamic Range on Metabolome Analysis by High-Performance Chemical Isotope Labeling LC–MS. *J Proteomics* **2015**, *118*, 130–139.

(19) Worley, B.; Powers, R. Multivariate Analysis in Metabolomics. *Curr Metabolomics* **2013**, *1* (1), 92–107.

(20) Liland, K. H. Multivariate Methods in Metabolomics–from Pre-Processing to

Dimension Reduction and Statistical Analysis. *TrAC Trends in Analytical Chemistry* **2011**, *30* (6), 827–841.

(21) Wishart, D. S. Quantitative Metabolomics Using NMR. *TrAC trends in analytical chemistry* **2008**, *27* (3), 228–237.

(22) Markley, J. L.; Brüschweiler, R.; Edison, A. S.; Eghbalnia, H. R.; Powers, R.; Raftery, D.; Wishart, D. S. The Future of NMR-Based Metabolomics. *Curr Opin Biotechnol* **2017**, *43*, 34–40.

(23) Emwas, A.-H.; Roy, R.; McKay, R. T.; Tenori, L.; Saccenti, E.; Gowda, G. A. N.; Raftery, D.; Alahmari, F.; Jaremko, L.; Jaremko, M. NMR Spectroscopy for Metabolomics Research. *Metabolites* **2019**, *9* (7), 123.

(24) Wu, H.; Southam, A. D.; Hines, A.; Viant, M. R. High-Throughput Tissue Extraction Protocol for NMR-and MS-Based Metabolomics. *Anal Biochem* **2008**, *372* (2), 204–212.

(25) Wehrens, R.; Hageman, J. A.; van Eeuwijk, F.; Kooke, R.; Flood, P. J.; Wijnker, E.; Keurentjes, J. J. B.; Lommen, A.; van Eekelen, H. D. L. M.; Hall, R. D. Improved Batch Correction in Untargeted MS-Based Metabolomics. *Metabolomics* **2016**, *12*, 1–12.

(26) Annesley, T. M. Ion Suppression in Mass Spectrometry. *Clin Chem* **2003**, *49* (7), 1041–1044.

(27) Nagana Gowda, G. A.; Abell, L.; Lee, C. F.; Tian, R.; Raftery, D. Simultaneous Analysis of Major Coenzymes of Cellular Redox Reactions and Energy Using Ex Vivo 1H NMR Spectroscopy. *Anal Chem* **2016**, *88* (9), 4817–4824.

(28) Zhou, B.; Xiao, J. F.; Tuli, L.; Ressom, H. W. LC-MS-Based Metabolomics. *Mol Biosyst* **2012**, *8* (2), 470–481.

(29) Becker, S.; Kortz, L.; Helmschrodt, C.; Thiery, J.; Ceglarek, U. LC–MS-Based Metabolomics in the Clinical Laboratory. *Journal of Chromatography B* **2012**, *883*, 68–75.

(30) Stettin, D.; Poulin, R. X.; Pohnert, G. Metabolomics Benefits from Orbitrap GC–MS—Comparison of Low-and High-Resolution GC–MS. *Metabolites* **2020**, *10* (4), 143.

(31) Beale, D. J.; Pinu, F. R.; Kouremenos, K. A.; Poojary, M. M.; Narayana, V. K.; Boughton, B. A.; Kanojia, K.; Dayalan, S.; Jones, O. A. H.; Dias, D. A. Review of Recent Developments in GC–MS Approaches to Metabolomics-Based Research. *Metabolomics* **2018**, *14*, 1–31.

(32) Ramautar, R.; Somsen, G. W.; de Jong, G. J. CE-MS in Metabolomics. *Electrophoresis* **2009**, *30* (1), 276–291.

(33) Ramautar, R.; Demirci, A.; de Jong, G. J. Capillary Electrophoresis in Metabolomics. *TrAC Trends in Analytical Chemistry* **2006**, *25* (5), 455–466.

(34) Faria, M.; Halquist, M. S.; Yuan, M.; Mylott Jr, W.; Jenkins, R. G.; Karnes, H. T. Comparison of a Stable Isotope Labeled (SIL) Peptide and an Extended SIL Peptide as Internal Standards to Track Digestion Variability of an Unstable Signature Peptide

during Quantification of a Cancer Biomarker, Human Osteopontin, from Plasma Using Capill. *Journal of Chromatography B* **2015**, *1001*, 156–168.

(35) Zheng, J. J.; Shields, E. E.; Snow, K. J.; Nelson, D. M.; Olah, T. V; Reily, M. D.; Robertson, D. G.; Shipkova, P. A.; Stryker, S. A.; Xin, B. The Utility of Stable Isotope Labeled (SIL) Analogues in the Bioanalysis of Endogenous Compounds by LC-MS Applied to the Study of Bile Acids in a Metabolomics Assay. *Anal Biochem* **2016**, *503*, 71–78.

(36) Bruheim, P.; Kvitvang, H. F. N.; Villas-Boas, S. G. Stable Isotope Coded Derivatizing Reagents as Internal Standards in Metabolite Profiling. *J Chromatogr A* **2013**, *1296*, 196–203.

(37) Higashi, T.; Ogawa, S. Isotope-Coded ESI-Enhancing Derivatization Reagents for Differential Analysis, Quantification and Profiling of Metabolites in Biological Samples by LC/MS: A Review. *J Pharm Biomed Anal* **2016**, *130*, 181–193.

(38) Guo, K.; Li, L. Differential 12C-/13C-Isotope Dansylation Labeling and Fast Liquid Chromatography/Mass Spectrometry for Absolute and Relative Quantification of the Metabolome. *Anal Chem* **2009**, *81* (10), 3919–3932.

(39) Zhao, S.; Li, H.; Han, W.; Chan, W.; Li, L. Metabolomic Coverage of Chemical-Group-Submetabolome Analysis: Group Classification and Four-Channel Chemical Isotope Labeling LC-MS. *Anal Chem* **2019**, *91* (18), 12108–12115.

(40) Zhao, S.; Li, L. Dansylhydrazine Isotope Labeling LC-MS for Comprehensive Carboxylic Acid Submetabolome Profiling. *Anal Chem* **2018**, *90* (22), 13514–13522.

(41) Zhao, S.; Luo, X.; Li, L. Chemical Isotope Labeling LC-MS for High Coverage and Quantitative Profiling of the Hydroxyl Submetabolome in Metabolomics. *Anal Chem* **2016**, *88* (21), 10617–10623.

(42) Zhao, S.; Dawe, M.; Guo, K.; Li, L. Development of High-Performance Chemical Isotope Labeling LC–MS for Profiling the Carbonyl Submetabolome. *Anal Chem* **2017**, *89* (12), 6758–6765.

(43) Glish, G. L.; Vachet, R. W. The Basics of Mass Spectrometry in the Twenty-First Century. *Nat Rev Drug Discov* **2003**, *2* (2), 140–150.

(44) Wilm, M. Principles of Electrospray Ionization. *Molecular & cellular proteomics* **2011**, *10* (7).

(45) Dawson, P. H. *Quadrupole Mass Spectrometry and Its Applications*; Elsevier, 2013.

(46) Mamyrin, B. A. Time-of-Flight Mass Spectrometry (Concepts, Achievements, and Prospects). *Int J Mass Spectrom* **2001**, *206* (3), 251–266.

(47) Scigelova, M.; Makarov, A. Orbitrap Mass Analyzer–Overview and Applications in Proteomics. *Proteomics* **2006**, *6* (S2), 16–21.

(48) Brown, S. C.; Kruppa, G.; Dasseux, J. Metabolomics Applications of FT-ICR Mass Spectrometry. *Mass Spectrom Rev* **2005**, *24* (2), 223–231.

(49) de Hoffmann, E. Tandem Mass Spectrometry: A Primer. *Journal of mass spectrometry* **1996**, *31* (2), 129–137.

(50) Llorach, R.; Favari, C.; Alonso, D.; Garcia-Aloy, M.; Andres-Lacueva, C.; Urpi-

Sarda, M. Comparative Metabolite Fingerprinting of Legumes Using LC-MS-Based Untargeted Metabolomics. *Food Research International* **2019**, *126*, 108666.

(51) Jandrić, Z.; Roberts, D.; Rathor, M. N.; Abrahim, A.; Islam, M.; Cannavan, A. Assessment of Fruit Juice Authenticity Using UPLC–QToF MS: A Metabolomics Approach. *Food Chem* **2014**, *148*, 7–17.

(52) Nikou, T.; Witt, M.; Stathopoulos, P.; Barsch, A.; Halabalaki, M. Olive Oil Quality and Authenticity Assessment Aspects Employing FIA-MRMS and LC-Orbitrap MS Metabolomic Approaches. *Front Public Health* **2020**, *8*, 558226.

(53) Godzien, J.; Ciborowski, M.; Martínez-Alcázar, M. P.; Samczuk, P.; Kretowski, A.; Barbas, C. Rapid and Reliable Identification of Phospholipids for Untargeted Metabolomics with LC–ESI–QTOF–MS/MS. *J Proteome Res* **2015**, *14* (8), 3204–3216.

(54) Garcia, C. J.; García-Villalba, R.; Garrido, Y.; Gil, M. I.; Tomás-Barberán, F. A. Untargeted Metabolomics Approach Using UPLC-ESI-QTOF-MS to Explore the Metabolome of Fresh-Cut Iceberg Lettuce. *Metabolomics* **2016**, *12*, 1–13.

(55) Lemonakis, N.; Poudyal, H.; Halabalaki, M.; Brown, L.; Tsarbopoulos, A.; Skaltsounis, A.-L.; Gikas, E. The LC–MS-Based Metabolomics of Hydroxytyrosol Administration in Rats Reveals Amelioration of the Metabolic Syndrome. *Journal of Chromatography B* **2017**, *1041*, 45–59.

(56) Han, L.; Wang, P.; Wang, Y.; Zhao, Q.; Zheng, F.; Dou, Z.; Yang, W.; Hu, L.; Liu, C. Rapid Discovery of the Potential Toxic Compounds in Polygonum Multiflorum by UHPLC/Q-Orbitrap-MS-Based Metabolomics and Correlation Analysis. *Front Pharmacol* **2019**, *10*, 329.

(57) Lee, J. W.; Ji, S.-H.; Choi, B.-R.; Choi, D. J.; Lee, Y.-G.; Kim, H.-G.; Kim, G.-S.; Kim, K.; Lee, Y.-H.; Baek, N.-I. UPLC-QTOF/MS-Based Metabolomics Applied for the Quality Evaluation of Four Processed Panax Ginseng Products. *Molecules* **2018**, *23* (8), 2062.

(58) Li, Y.; Li, L. Retention Time Shift Analysis and Correction in Chemical Isotope Labeling Liquid Chromatography/Mass Spectrometry for Metabolome Analysis. *Rapid Communications in Mass Spectrometry* **2020**, *34* (S1), e8643. https://doi.org/https://doi.org/10.1002/rcm.8643.

(59) Zhou, R.; Tseng, C.-L.; Huan, T.; Li, L. IsoMS: Automated Processing of LC-MS Data Generated by a Chemical Isotope Labeling Metabolomics Platform. *Anal Chem* **2014**, *86* (10), 4675–4679.

(60) Lange, H. C.; Eman, M.; Van Zuijlen, G.; Visser, D.; Van Dam, J. C.; Frank, J.; de Mattos, M. J. T.; Heijnen, J. J. Improved Rapid Sampling for in Vivo Kinetics of Intracellular Metabolites in Saccharomyces Cerevisiae. *Biotechnol Bioeng* **2001**, *75* (4), 406–415.

(61) Sáez, M. J.; Lagunas, R. Determination of Intermediary Metabolites in Yeast. Critical Examination of the Effect of Sampling Conditions and Recommendations for Obtaining True Levels. *Mol Cell Biochem* **1976**, *13*, 73–78.

(62) Wiza, J. L. Microchannel Plate Detectors. *Nucl. Instrum. Methods* **1979**, *162* (1–3), 587–601.

(63) Wells, J. M.; McLuckey, S. A. Collision-induced Dissociation (CID) of Peptides and Proteins. *Methods Enzymol* **2005**, *402*, 148–185.

(64) Wang, Y.; Kora, G.; Bowen, B. P.; Pan, C. MIDAS: A Database-Searching Algorithm for Metabolite Identification in Metabolomics. *Anal Chem* **2014**, *86* (19), 9496–9503.

(65) Shen, Y.; Tolić, N.; Xie, F.; Zhao, R.; Purvine, S. O.; Schepmoes, A. A.; Moore J, R.; Anderson, G. A.; Smith, R. D. Effectiveness of CID, HCD, and ETD with FT MS/MS for Degradomic-Peptidomic Analysis: Comparison of Peptide Identification Methods. *J Proteome Res* **2011**, *10* (9), 3929–3943.

(66) Medzihradszky, K. F.; Burlingame, A. L. The Advantages and Versatility of a High-Energy Collision-Induced Dissociation-Based Strategy for the Sequence and Structural Determination of Proteins. *Methods* **1994**, *6* (3), 284–303.

(67) Demarque, D. P.; Crotti, A. E. M.; Vessecchi, R.; Lopes, J. L. C.; Lopes, N. P. Fragmentation Reactions Using Electrospray Ionization Mass Spectrometry: An Important Tool for the Structural Elucidation and Characterization of Synthetic and Natural Products. *Nat Prod Rep* **2016**, *33* (3), 432–455.

(68) Pezzatti, J.; Boccard, J.; Codesido, S.; Gagnebin, Y.; Joshi, A.; Picard, D.; González-Ruiz, V.; Rudaz, S. Implementation of Liquid Chromatography–High Resolution Mass Spectrometry Methods for Untargeted Metabolomic Analyses of Biological Samples: A Tutorial. *Anal Chim Acta* **2020**, *1105*, 28–44.

(69) Vuckovic, D. Improving Metabolome Coverage and Data Quality: Advancing Metabolomics and Lipidomics for Biomarker Discovery. *Chemical Communications* **2018**, *54* (50), 6728–6749.

(70) Cao, G.; Song, Z.; Hong, Y.; Yang, Z.; Song, Y.; Chen, Z.; Chen, Z.; Cai, Z. Large-Scale Targeted Metabolomics Method for Metabolite Profiling of Human Samples. *Anal Chim Acta* **2020**, *1125*, 144–151.

(71) Zhao, S.; Li, L. Chemical Derivatization in LC-MS-Based Metabolomics Study. *TrAC Trends in Analytical Chemistry* **2020**, *131*, 115988.

(72) Jia, S.; Xu, T.; Huan, T.; Chong, M.; Liu, M.; Fang, W.; Fang, M. Chemical Isotope Labeling Exposome (CIL-EXPOSOME): One High-Throughput Platform for Human Urinary Global Exposome Characterization. *Environ Sci Technol* **2019**, *53* (9), 5445–5453.

(73) Chen, G.; Zhang, Q. Simultaneous Quantification of Free Fatty Acids and Acylcarnitines in Plasma Samples Using Dansylhydrazine Labeling and Liquid Chromatography–Triple Quadrupole Mass Spectrometry. *Anal Bioanal Chem* **2020**, *412*, 2841–2849.

(74) Han, J.; Lin, K.; Sequeira, C.; Borchers, C. H. An Isotope-Labeled Chemical Derivatization Method for the Quantitation of Short-Chain Fatty Acids in Human Feces by Liquid Chromatography–Tandem Mass Spectrometry. *Anal Chim Acta*

2015, *854*, 86–94.

(75) Arashida, N.; Nishimoto, R.; Harada, M.; Shimbo, K.; Yamada, N. Highly Sensitive Quantification for Human Plasma-Targeted Metabolomics Using an Amine Derivatization Reagent. *Anal Chim Acta* **2017**, *954*, 77–87.

(76) Huang, T.; Toro, M.; Lee, R.; Hui, D. S.; Edwards, J. L. Multi-Functional Derivatization of Amine, Hydroxyl, and Carboxylate Groups for Metabolomic Investigations of Human Tissue by Electrospray Ionization Mass Spectrometry. *Analyst* **2018**, *143* (14), 3408–3414.

(77) Takayama, T.; Mizuno, H.; Toyo'oka, T.; Akatsu, H.; Inoue, K.; Todoroki, K. Isotope Corrected Chiral and Achiral Nontargeted Metabolomics: An Approach for High Accuracy and Precision Metabolomics Based on Derivatization and Its Application to Cerebrospinal Fluid of Patients with Alzheimer's Disease. *Anal Chem* **2019**, *91* (7), 4396–4404.

(78) Yu, Y.; Li, G.; Wu, D.; Liu, J.; Chen, J.; Hu, N.; Wang, H.; Wang, P.; Wu, Y. Thiol Radical-Based Chemical Isotope Labelling for Sterols Quantitation through High Performance Liquid Chromatography-Tandem Mass Spectrometry Analysis. *Anal Chim Acta* **2020**, *1097*, 110–119.

(79) Martin, J.-C.; Maillot, M.; Mazerolles, G.; Verdu, A.; Lyan, B.; Migné, C.; Defoort, C.; Canlet, C.; Junot, C.; Guillou, C. Can We Trust Untargeted Metabolomics? Results of the Metabo-Ring Initiative, a Large-Scale, Multi-Instrument Inter-Laboratory Study. *Metabolomics* **2015**, *11*, 807–821.

(80) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J. A Cross-Platform Toolkit for Mass Spectrometry and Proteomics. *Nat Biotechnol* **2012**, *30* (10), 918–920.

(81) Huan, T.; Li, L. Counting Missing Values in a Metabolite-Intensity Data Set for Measuring the Analytical Performance of a Metabolomics Platform. *Anal Chem* **2015**, *87* (2), 1306–1313.

(82) Huan, T.; Wu, Y.; Tang, C.; Lin, G.; Li, L. DnsID in MyCompoundID for Rapid Identification of Dansylated Amine-and Phenol-Containing Metabolites in LC–MS-Based Metabolomics. *Anal Chem* **2015**, *87* (19), 9838–9845.

(83) Li, L.; Li, R.; Zhou, J.; Zuniga, A.; Stanislaus, A. E.; Wu, Y.; Huan, T.; Zheng, J.; Shi, Y.; Wishart, D. S. MyCompoundID: Using an Evidence-Based Metabolome Library for Metabolite Identification. *Anal Chem* **2013**, *85* (6), 3401–3408.

(84) Maxwell, E. J.; Chen, D. D. Y. Twenty Years of Interface Development for Capillary Electrophoresis–Electrospray Ionization–Mass Spectrometry. *Anal Chim Acta* **2008**, *627* (1), 25–33.

(85) Kang, Y.; Schneider, B. B.; Bedford, L.; Covey, T. R. Design Characteristics to Eliminate the Need for Parameter Optimization in Nanoflow ESI-MS. *J Am Soc Mass Spectrom* **2019**, *30* (11), 2347–2357.

(86) Wang, W.; Bajic, S.; John, B.; Emerson, D. R. Numerical Simulation of Ion Transport in a Nano-Electrospray Ion Source at Atmospheric Pressure. *J Am Soc*

Mass Spectrom **2018**, *29* (3), 600–612.

(87) Leito, I.; Herodes, K.; Huopolainen, M.; Virro, K.; Künnapas, A.; Kruve, A.; Tanner, R. Towards the Electrospray Ionization Mass Spectrometry Ionization Efficiency Scale of Organic Compounds. *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry* **2008**, *22* (3), 379–384.

(88) Shaffer, S. A.; Prior, D. C.; Anderson, G. A.; Udseth, H. R.; Smith, R. D. An Ion Funnel Interface for Improved Ion Focusing and Sensitivity Using Electrospray Ionization Mass Spectrometry. *Anal Chem* **1998**, *70* (19), 4111–4119.

(89) Kalli, A.; Smith, G. T.; Sweredoski, M. J.; Hess, S. Evaluation and Optimization of Mass Spectrometric Settings during Data-Dependent Acquisition Mode: Focus on LTQ-Orbitrap Mass Analyzers. *J Proteome Res* **2013**, *12* (7), 3071–3086.

(90) Kilpatrick, L. E.; Kilpatrick, E. L. Optimizing High-Resolution Mass Spectrometry for the Identification of Low-Abundance Post-Translational Modifications of Intact Proteins. *J Proteome Res* **2017**, *16* (9), 3255–3265.

(91) Giera, M.; Yanes, O.; Siuzdak, G. Metabolite Discovery: Biochemistry's Scientific Driver. *Cell Metab* **2022**, *34* (1), 21–34.

(92) Hu, S. Cancer Metabolomics. Springer 2022.

(93) Veenstra, T. D. Omics in Systems Biology: Current Progress and Future Outlook. *Proteomics* **2021**, *21* (3–4), 2000235.

(94) Nagy, P.; Thorgeirsson, S. S.; Grisham, J. W. Organizational Principles of the Liver. *The Liver: Biology and Pathobiology* **2020**, 1–13.

(95) Rustam, Y. H.; Reid, G. E. Analytical Challenges and Recent Advances in Mass Spectrometry Based Lipidomics. *Anal Chem* **2018**, *90* (1), 374–397.

(96) Li, Y.; Li, L. Improving Accuracy of Peak-Pair Intensity Ratio Measurement in Differential Chemical Isotope Labeling LC–MS for Quantitative Metabolomics. *Int J Mass Spectrom* **2018**, *434*, 202–208.

(97) Krajewski, L. C.; Rodgers, R. P.; Marshall, A. G. 126 264 Assigned Chemical Formulas from an Atmospheric Pressure Photoionization 9.4 T Fourier Transform Positive Ion Cyclotron Resonance Mass Spectrum. *Anal Chem* **2017**, *89* (21), 11318–11324.

(98) Southam, A. D.; Payne, T. G.; Cooper, H. J.; Arvanitis, T. N.; Viant, M. R. Dynamic Range and Mass Accuracy of Wide-Scan Direct Infusion Nanoelectrospray Fourier Transform Ion Cyclotron Resonance Mass Spectrometry-Based Metabolomics Increased by the Spectral Stitching Method. *Anal Chem* **2007**, *79* (12), 4595–4602.

(99) Gavard, R.; Palacio Lozano, D. C.; Guzman, A.; Rossell, D.; Spencer, S. E. F.; Barrow, M. P. Rhapso: Automatic Stitching of Mass Segments from Fourier Transform Ion Cyclotron Resonance Mass Spectra. *Anal Chem* **2019**, *91* (23), 15130–15137.

(100) Hohenester, U. M. Barbier Saint-Hilaire P., Fenaille F., Cole RB Investigation of Space Charge Effects and Ion Trapping Capacity on Direct Introduction Ultra-High-

Resolution Mass Spectrometry Workflows for Metabolomics. *J. Mass Spectrom* **2020**, *55*, e4613.

(101) Baira, E.; Dagla, I.; Siapi, E.; Zoumpoulakis, P.; Tsarbopoulos, A.; Simitzis, P.; Goliomytis, M.; Deligeorgis, S. G.; Skaltsounis, A.-L.; Gikas, E. Development of a Validated UHPLC-ESI (-)-HRMS Methodology for the Simultaneous Quantitative Determination of Hesperidin, Hesperetin, Naringin, and Naringenin in Chicken Plasma. *Food Anal Methods* **2019**, *12*, 1187–1196.

(102) Han, L.; Kaesler, J.; Peng, C.; Reemtsma, T.; Lechtenfeld, O. J. Online Counter Gradient LC-FT-ICR-MS Enables Detection of Highly Polar Natural Organic Matter Fractions. *Anal Chem* **2020**, *93* (3), 1740–1748.

(103) Su, X.; Lu, W.; Rabinowitz, J. D. Metabolite Spectral Accuracy on Orbitraps. *Anal Chem* **2017**, *89* (11), 5940–5948.

(104) Barrientos, R. C.; Zhang, Q. Differential Isotope Labeling by Permethylation and Reversed-Phase Liquid Chromatography–Mass Spectrometry for Relative Quantification of Intact Neutral Glycolipids in Mammalian Cells. *Anal Chem* **2019**, *91* (15), 9673–9681.

(105) Hao, L.; Zhu, Y.; Wei, P.; Johnson, J.; Buchberger, A.; Frost, D.; Kao, W. J.; Li, L. Metandem: An Online Software Tool for Mass Spectrometry-Based Isobaric Labeling Metabolomics. *Anal Chim Acta* **2019**, *1088*, 99–106.

(106) Li, S.; Liu, F.-L.; Zhang, Z.; Yin, X.-M.; Ye, T.-T.; Yuan, B.-F.; Feng, Y.-Q. Ultrasensitive Determination of Sugar Phosphates in Trace Samples by Stable Isotope Chemical Labeling Combined with RPLC–MS. *Anal Chem* **2022**, *94* (11), 4866–4873.

(107) Khamis, M. M.; Adamko, D. J.; El-Aneed, A. Mass Spectrometric Based Approaches in Urine Metabolomics and Biomarker Discovery. *Mass Spectrom Rev* **2017**, *36* (2), 115–134.

(108) Guo, K.; Li, L. Differential 12C-/13C-Isotope Dansylation Labeling and Fast Liquid Chromatography/Mass Spectrometry for Absolute and Relative Quantification of the Metabolome. *Anal Chem* **2009**, *81* (10), 3919–3932. https://doi.org/10.1021/ac900166a.

(109) Zhou, R.; Tseng, C. L.; Huan, T.; Li, L. IsoMS: Automated Processing of LC-MS Data Generated by a Chemical Isotope Labeling Metabolomics Platform. *Anal Chem* **2014**, *86* (10), 4675–4679. https://doi.org/10.1021/ac5009089.

(110) Najdekr, L.; Friedecky, D.; Tautenhahn, R.; Pluskal, T.; Wang, J.; Huang, Y.; Adam, T. Influence of Mass Resolving Power in Orbital Ion-Trap Mass Spectrometry-Based Metabolomics. *Anal Chem* **2016**, *88* (23), 11429–11435.

(111) Huan, T.; Li, L. Quantitative Metabolome Analysis Based on Chromatographic Peak Reconstruction in Chemical Isotope Labeling Liquid Chromatography Mass Spectrometry. *Anal Chem* **2015**, *87* (14), 7011–7016.

(112) Southam, A. D.; Payne, T. G.; Cooper, H. J.; Arvanitis, T. N.; Viant, M. R. Dynamic Range and Mass Accuracy of Wide-Scan Direct Infusion Nanoelectrospray Fourier

Transform Ion Cyclotron Resonance Mass Spectrometry-Based Metabolomics Increased by the Spectral Stitching Method. *Anal Chem* **2007**, *79* (12), 4595–4602. https://doi.org/10.1021/ac062446p.

(113) Gavard, R.; Palacio Lozano, D. C.; Guzman, A.; Rossell, D.; Spencer, S. E. F.; Barrow, M. P. Rhapso: Automatic Stitching of Mass Segments from Fourier Transform Ion Cyclotron Resonance Mass Spectra. *Anal Chem* **2019**, *91* (23), 15130–15137. https://doi.org/10.1021/ACS.ANALCHEM.9B03846/SUPPL_FILE/AC9B03846_SI_001.PDF.

(114) Zhao, S.; Li, H.; Han, W.; Chan, W.; Li, L. Metabolomic Coverage of Chemical-Group-Submetabolome Analysis: Group Classification and Four-Channel Chemical Isotope Labeling LC-MS. *Anal Chem* **2019**, *91* (18), 12108–12115. https://doi.org/10.1021/acs.analchem.9b03431.

(115) Hollywood, K.; Brison, D. R.; Goodacre, R. Metabolomics: Current Technologies and Future Trends. *Proteomics* **2006**, *6* (17), 4716–4723.

(116) Shulaev, V. Metabolomics Technology and Bioinformatics. *Brief Bioinform* **2006**, *7* (2), 128–139.

(117) Gibney, M. J.; Walsh, M.; Brennan, L.; Roche, H. M.; German, B.; Van Ommen, B. Metabolomics in Human Nutrition: Opportunities and Challenges–. *Am J Clin Nutr* **2005**, *82* (3), 497–503.

(118) Johnson, C. H.; Gonzalez, F. J. Challenges and Opportunities of Metabolomics. *J Cell Physiol* **2012**, *227* (8), 2975–2981.

(119) Lozano, D. C. P.; Gavard, R.; Arenas-Diaz, J. P.; Thomas, M. J.; Stranz, D. D.; Mejía-Ospino, E.; Guzman, A.; Spencer, S. E. F.; Rossell, D.; Barrow, M. P. Pushing the Analytical Limits: New Insights into Complex Mixtures Using Mass Spectra Segments of Constant Ultrahigh Resolving Power. *Chem Sci* **2019**, *10* (29), 6966–6978.

(120) Southam, A. D.; Weber, R. J. M.; Engel, J.; Jones, M. R.; Viant, M. R. A Complete Workflow for High-Resolution Spectral-Stitching Nanoelectrospray Direct-Infusion Mass-Spectrometry-Based Metabolomics and Lipidomics. *Nat Protoc* **2017**, *12* (2), 310–328.

(121) Wang, C.-F.; Li, L. Segment Scan Mass Spectral Acquisition for Increasing the Metabolite Detectability in Chemical Isotope Labeling Liquid Chromatography–Mass Spectrometry Metabolome Analysis. *Anal Chem* **2022**, *94* (33), 11650–11658.

(122) Zheng, J.; Dixon, R. A.; Li, L. Development of Isotope Labeling LC–MS for Human Salivary Metabolomics and Application to Profiling Metabolome Changes Associated with Mild Cognitive Impairment. *Anal Chem* **2012**, *84* (24), 10802–10811.

(123) Wang, C.-F.; Li, L. Instrument-Type Effects on Chemical Isotope Labeling LC-MS Metabolome Analysis: Quadrupole Time-of-Flight MS vs. Orbitrap MS. *Anal Chim Acta* **2022**, *1226*, 340255.

(124) Luo, X.; Gu, X.; Li, L. Development of a Simple and Efficient Method of Harvesting and Lysing Adherent Mammalian Cells for Chemical Isotope Labeling LC-MS-Based Cellular Metabolomics. *Anal Chim Acta* **2018**, *1037*, 97–106.

(125) Su, X.; Wang, N.; Chen, D.; Li, Y.; Lu, Y.; Huan, T.; Xu, W.; Li, L.; Li, L. Dansylation Isotope Labeling Liquid Chromatography Mass Spectrometry for Parallel Profiling of Human Urinary and Fecal Submetabolomes. *Anal Chim Acta* **2016**, *903*, 100–109.

(126) Mandel, I. D. The Functions of Saliva. *J Dent Res* **1987**, *66* (2_suppl), 623–627.

(127) Kaufman, E.; Lamster, I. B. The Diagnostic Applications of Saliva—a Review. *Critical Reviews in oral biology & medicine* **2002**, *13* (2), 197–212.

(128) Bouatra, S.; Aziat, F.; Mandal, R.; Guo, A. C.; Wilson, M. R.; Knox, C.; Bjorndahl, T. C.; Krishnamurthy, R.; Saleem, F.; Liu, P. The Human Urine Metabolome. *PLoS One* **2013**, *8* (9), e73076.

(129) Zierer, J.; Jackson, M. A.; Kastenmüller, G.; Mangino, M.; Long, T.; Telenti, A.; Mohney, R. P.; Small, K. S.; Bell, J. T.; Steves, C. J. The Fecal Metabolome as a Functional Readout of the Gut Microbiome. *Nat Genet* **2018**, *50* (6), 790–795.

(130) Karu, N.; Deng, L.; Slae, M.; Guo, A. C.; Sajed, T.; Huynh, H.; Wine, E.; Wishart, D. S. A Review on Human Fecal Metabolomics: Methods, Applications and the Human Fecal Metabolome Database. *Anal Chim Acta* **2018**, *1030*, 1–24.

(131) Guo, K.; Li, L. High-Performance Isotope Labeling for Profiling Carboxylic Acid-Containing Metabolites in Biofluids by Mass Spectrometry. *Anal Chem* **2010**, *82* (21), 8789–8793.

(132) Luo, X.; Li, L. Metabolomics of Small Numbers of Cells Using Chemical Isotope Labeling Combined with Nanoflow LC-MS. *Metabolomics* **2021**, 49–60.

(133) Lemmel, C.; Weik, S.; Eberle, U.; Dengjel, J.; Kratt, T.; Becker, H.-D.; Rammensee, H.-G.; Stevanović, S. Differential Quantitative Analysis of MHC Ligands by Mass Spectrometry Using Stable Isotope Labeling. *Nat Biotechnol* **2004**, *22* (4), 450–454.

(134) Johnson, A. R.; Carlson, E. E. Collision-Induced Dissociation Mass Spectrometry: A Powerful Tool for Natural Product Structure Elucidation. ACS Publications 2015.

(135) Sartori, L. R.; Vessecchi, R.; Humpf, H.; Da Costa, F. B.; Lopes, N. P. A Systematic Investigation of the Fragmentation Pattern of Two Furanoheliangolide C-8 Stereoisomers Using Electrospray Ionization Mass Spectrometry. *Rapid Communications in Mass Spectrometry* **2014**, *28* (7), 723–730.

(136) Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jørgensen, T. J. D. Highly Selective Enrichment of Phosphorylated Peptides from Peptide Mixtures Using Titanium Dioxide Microcolumns. *Molecular & cellular proteomics* **2005**, *4* (7), 873–886.

(137) Davies, V.; Wandy, J.; Weidt, S.; Van Der Hooft, J. J. J.; Miller, A.; Daly, R.; Rogers, S. Rapid Development of Improved Data-Dependent Acquisition Strategies. *Anal Chem* **2021**, *93* (14), 5676–5683.

(138) Goodwin, K. J.; Gangl, E.; Sarkar, U.; Pop-Damkov, P.; Jones, N.; Borodovsky, A.;

Woessner, R.; Fretland, A. J. Development of a Quantification Method for Adenosine in Tumors by LC-MS/MS with Dansyl Chloride Derivatization. *Anal Biochem* **2019**, *568*, 78–88.

(139) Jansons, M.; Pugajeva, I.; Bartkevics, V.; Karkee, H. B. LC-MS/MS Characterisation and Determination of Dansyl Chloride Derivatised Glyphosate, Aminomethylphosphonic Acid (AMPA), and Glufosinate in Foods of Plant and Animal Origin. *Journal of Chromatography B* **2021**, *1177*, 122779.

(140) Beaudry, F.; Guénette, S. A.; Winterborn, A.; Marier, J.-F.; Vachon, P. Development of a Rapid and Sensitive LC–ESI/MS/MS Assay for the Quantification of Propofol Using a Simple off-Line Dansyl Chloride Derivatization Reaction to Enhance Signal Intensity. *J Pharm Biomed Anal* **2005**, *39* (3–4), 411–417.

(141) Zheng, J.; Li, L. Fragmentation of Protonated Dansyl-Labeled Amines for Structural Analysis of Amine-Containing Metabolites. *Int J Mass Spectrom* **2012**, *316*, 292–299.

(142) Wang, Z.; Hop, C. E. C. A.; Kim, M.; Huskey, S.; Baillie, T. A.; Guan, Z. The Unanticipated Loss of SO2 from Sulfonamides in Collision-induced Dissociation. *Rapid communications in mass spectrometry* **2003**, *17* (1), 81–86.

(143) Hao, G.; Wang, D.; Gu, J.; Shen, Q.; Gross, S. S.; Wang, Y. Neutral Loss of Isocyanic Acid in Peptide CID Spectra: A Novel Diagnostic Marker for Mass Spectrometric Identification of Protein Citrullination. *J Am Soc Mass Spectrom* **2008**, *20* (4), 723–727.

(144) Hao, G.; Wang, D.; Gu, J.; Shen, Q.; Gross, S. S.; Wang, Y. Neutral Loss of Isocyanic Acid in Peptide CID Spectra: A Novel Diagnostic Marker for Mass Spectrometric Identification of Protein Citrullination. *J Am Soc Mass Spectrom* **2008**, *20* (4), 723–727.

(145) Winter, D.; Lehmann, W. D. Individual b 2 Ion Fragmentation Profiles Combined with AspN Digestion Improve N-Terminal Peptide Sequencing. *Anal Bioanal Chem* **2009**, *393*, 1587–1591.

(146) Rousu, T.; Herttuainen, J.; Tolonen, A. Comparison of Triple Quadrupole, Hybrid Linear Ion Trap Triple Quadrupole, Time-of-flight and LTQ-Orbitrap Mass Spectrometers in Drug Discovery Phase Metabolite Screening and Identification in Vitro–Amitriptyline and Verapamil as Model Compounds. *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry* **2010**, *24* (7), 939–957.

(147) Triebl, A.; Weissengruber, S.; Trötzmüller, M.; Lankmayr, E.; Köfeler, H. Quantitative Analysis of N-acylphosphatidylethanolamine Molecular Species in Rat Brain Using Solid-phase Extraction Combined with Reversed-phase Chromatography and Tandem Mass Spectrometry. *J Sep Sci* **2016**, *39* (13), 2474–2480.

(148) Guvvala, V.; Subramanian, V. C.; Anireddy, J. S. A Study on Structural

Characterization of Degradation Products of Cangrelor Using LC/QTOF/MS/MS and NMR. *J Pharm Biomed Anal* **2019**, *170*, 327–334.

(149) Ding, S.; Dudley, E.; Plummer, S.; Tang, J.; Newton, R. P.; Brenton, A. G. Fingerprint Profile of Ginkgo Biloba Nutritional Supplements by LC/ESI-MS/MS. *Phytochemistry* **2008**, *69* (7), 1555–1564.

(150) Nicolas, E. C.; Scholz, T. H. Active Drug Substance Impurity Profiling: Part II. LC/MS/MS Fingerprinting. *J Pharm Biomed Anal* **1998**, *16* (5), 825–836.

(151) Cheng, Z.; Li, L. Development of Chemical Isotope Labeling Liquid Chromatography Orbitrap Mass Spectrometry for Comprehensive Analysis of Dipeptides. *Anal Chem* **2023**, *95* (16), 6629–6636.

Appendix A. MS/MS spectra and proposed fragmentation patterns of 85 tier-2

metabolites that passed MS/MS-based identification validation.

No.3 Kanosamine

No. 14 1-Methylguanosine (isomer)

531.1657 m/z   166.0724 m/z   149.0458 m/z
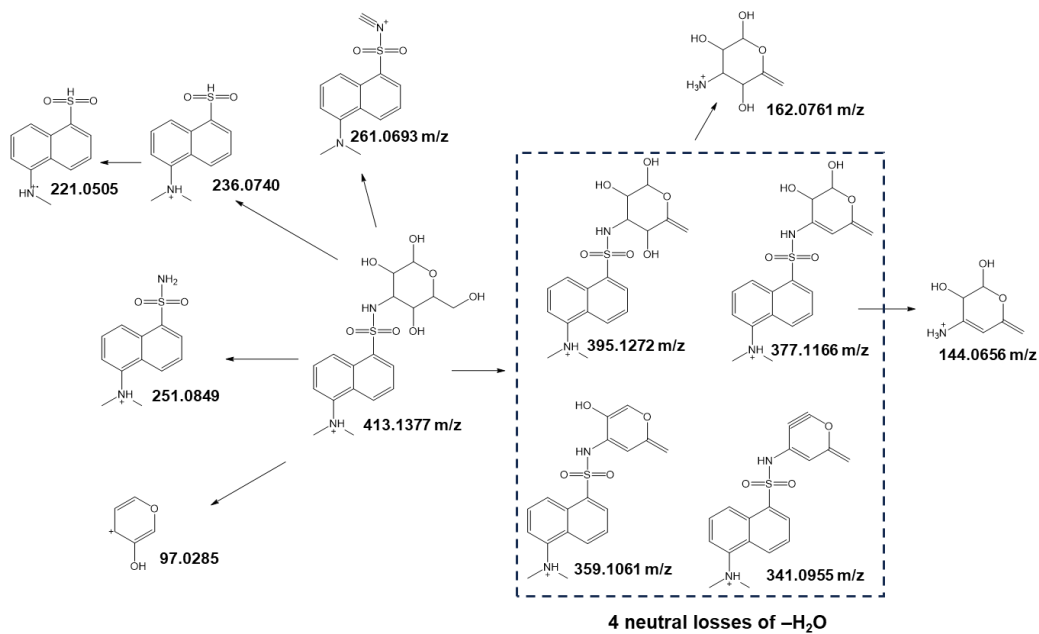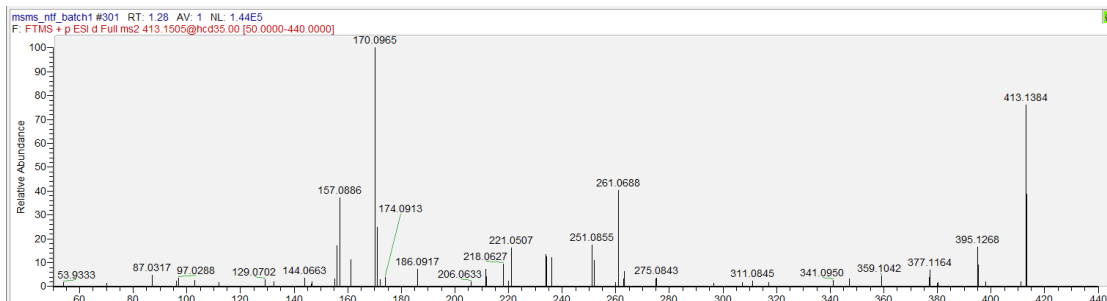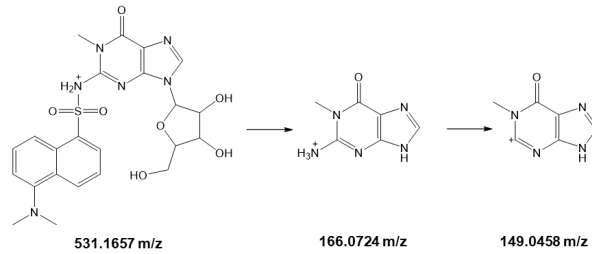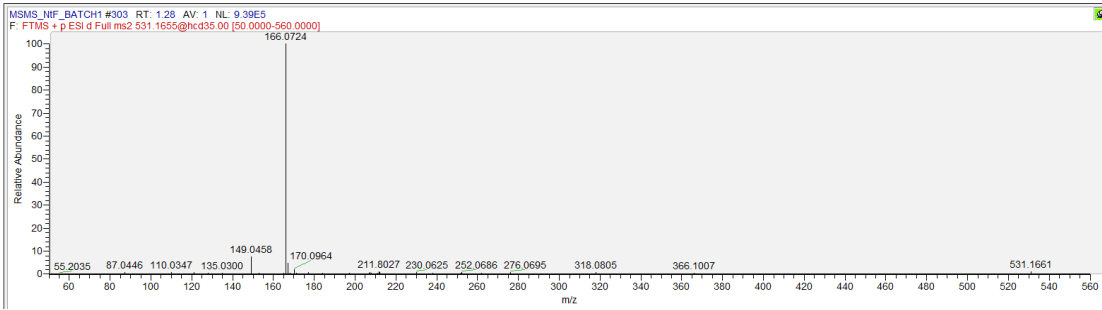
No.20 Isomer of L-Arginine/D-Arginine
(MS2 was identical with L-Arginine/D-Arginine)

70.0652 m/z

112.0870 m/z

408.1701 m/z   391.1435 m/z   366.1483 m/z   349.1217 m/z   303.1163 m/z
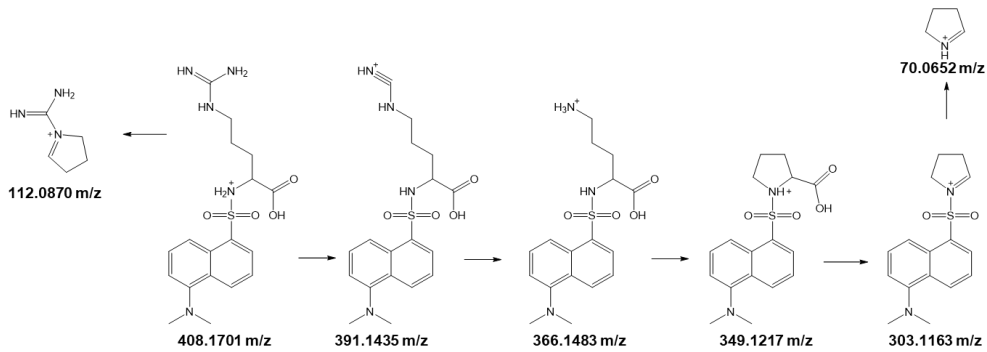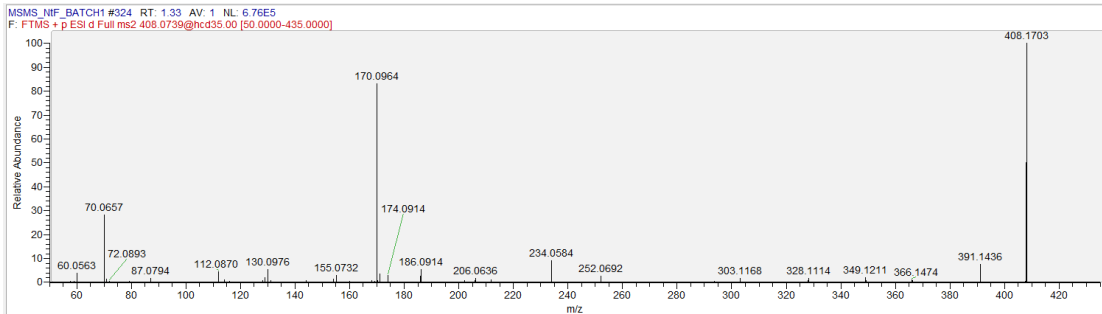
No. 29 Guandinoacetic acid
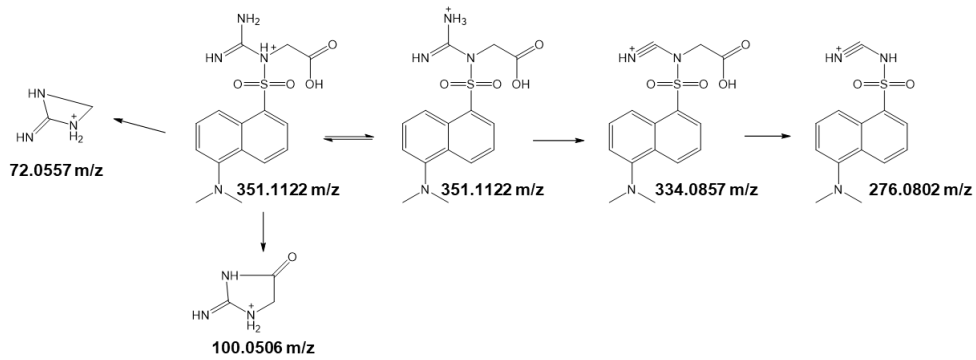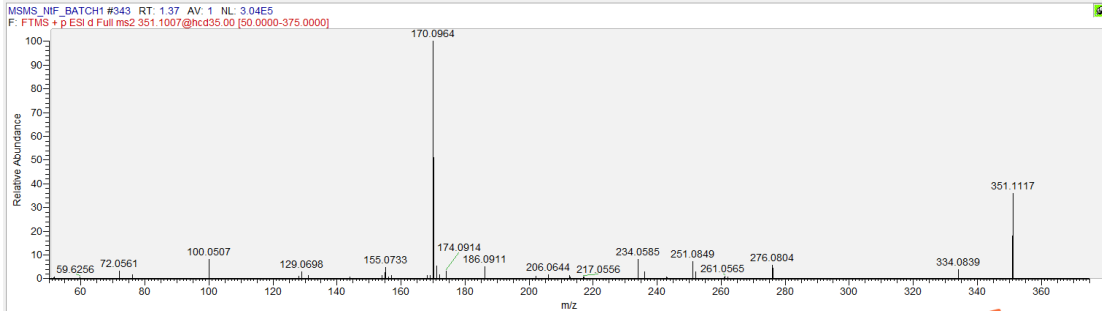
No.39 L-Targinine
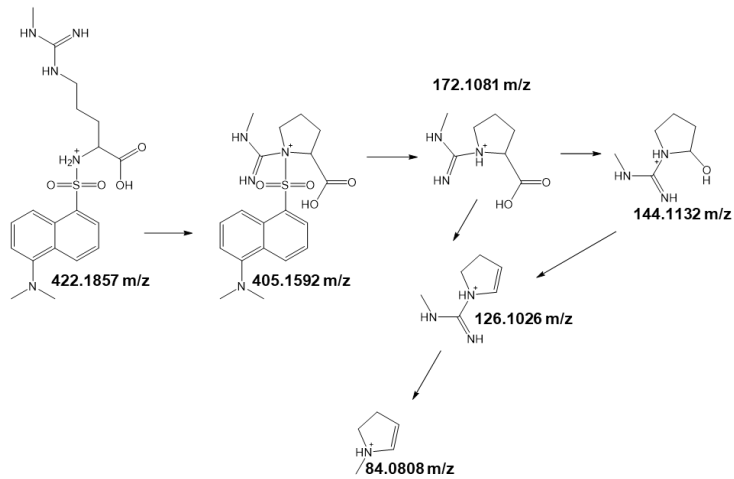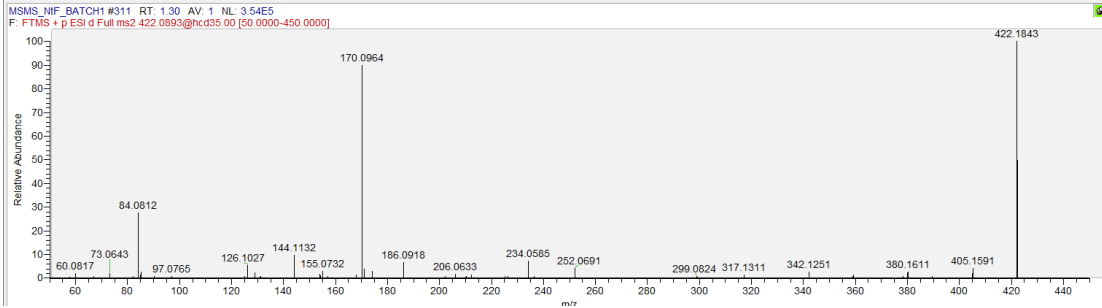
No. 43 N6-Methyladenosine
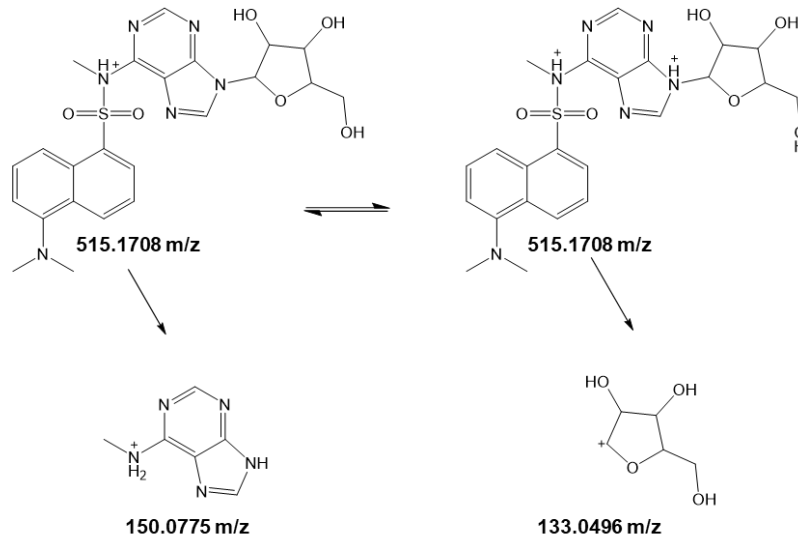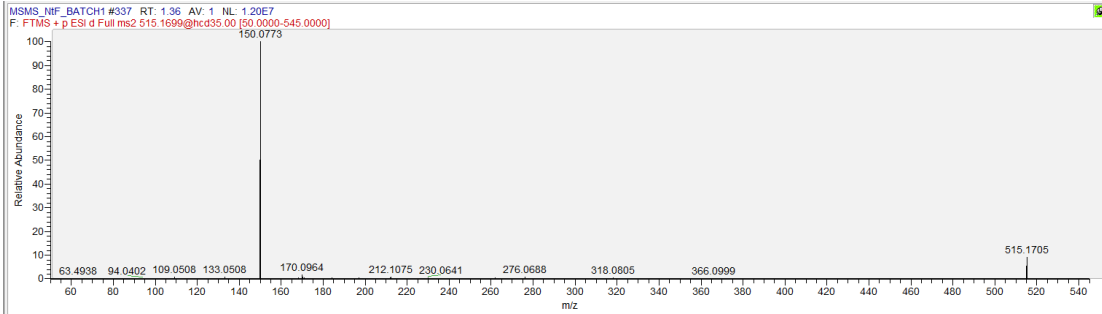
515.1708 m/z ⇌ **515.1708 m/z**

150.0775 m/z   **133.0496 m/z**

172

No. 45 Isomer of dimethylarginine
(MS2 was identical with dimethylarginine)

436.2014 m/z   436.2014 m/z   405.1592 m/z

88.0757 m/z

366.1483 m/z   349.1217 m/z   349.1217 m/z   303.1162 m/z   70.0652 m/z

No. 73 Beta-Guanidinopropionic acid

90.0550 m/z

365.1279 m/z   **348.1013 m/z**   **114.0662 m/z**

276.0802 m/z   **251.0849 m/z**   **234.0584 m/z**
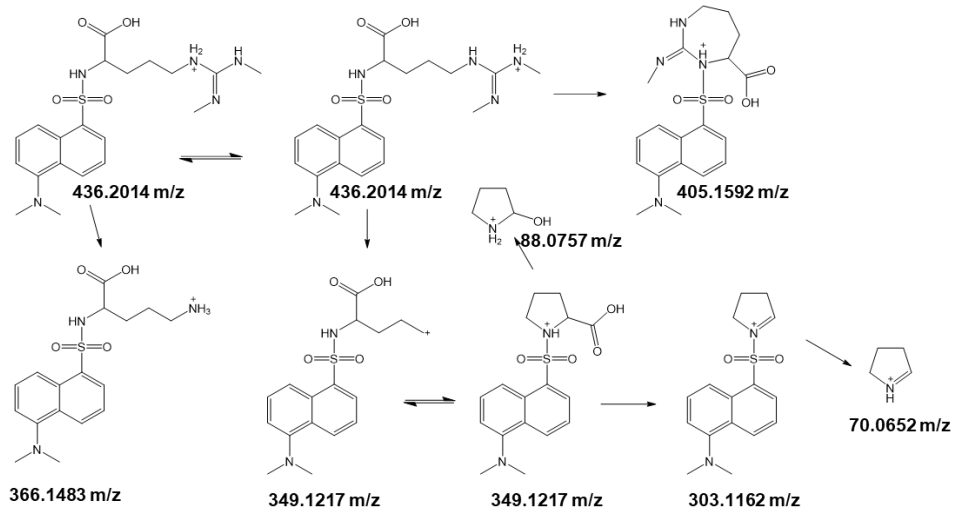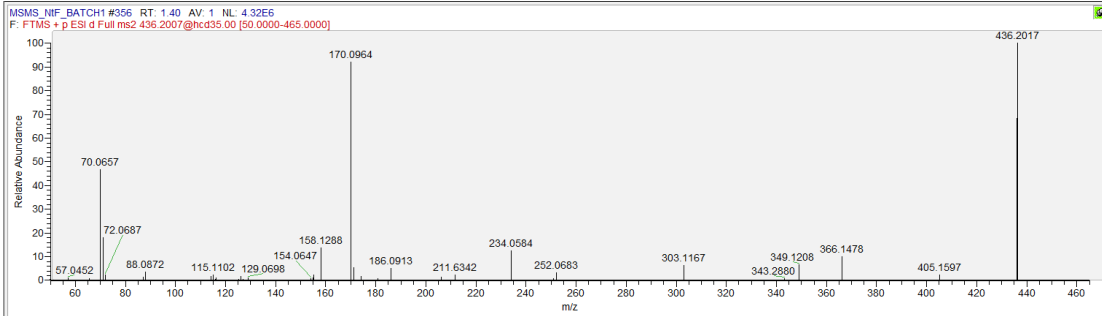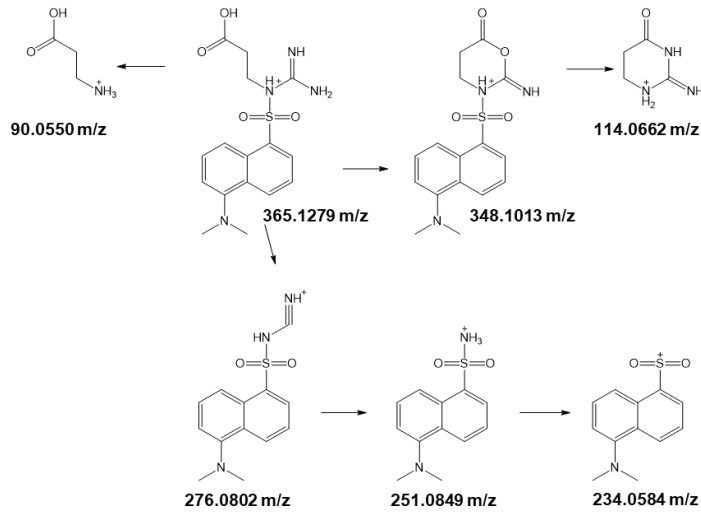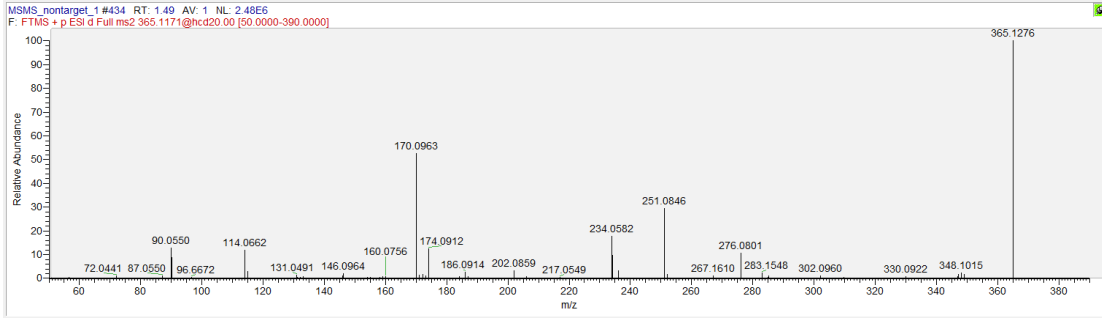
No. 81 1-Methylguanosine (isomer)

531.1657 m/z   **166.0724 m/z**   **149.0458 m/z**
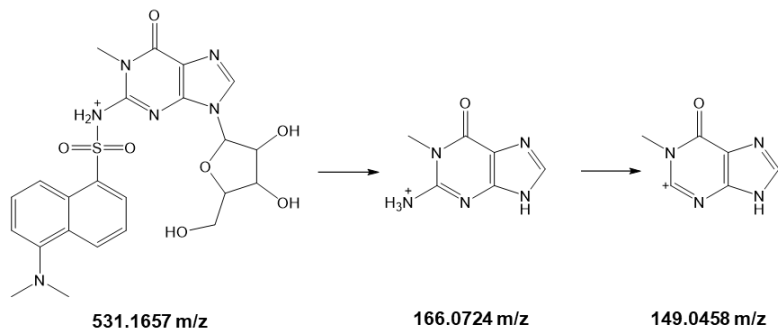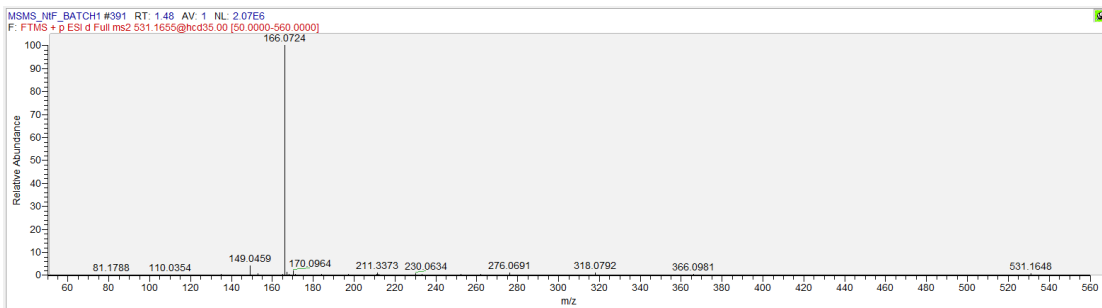
No. 87 Deoxyeritadenine (isomer)
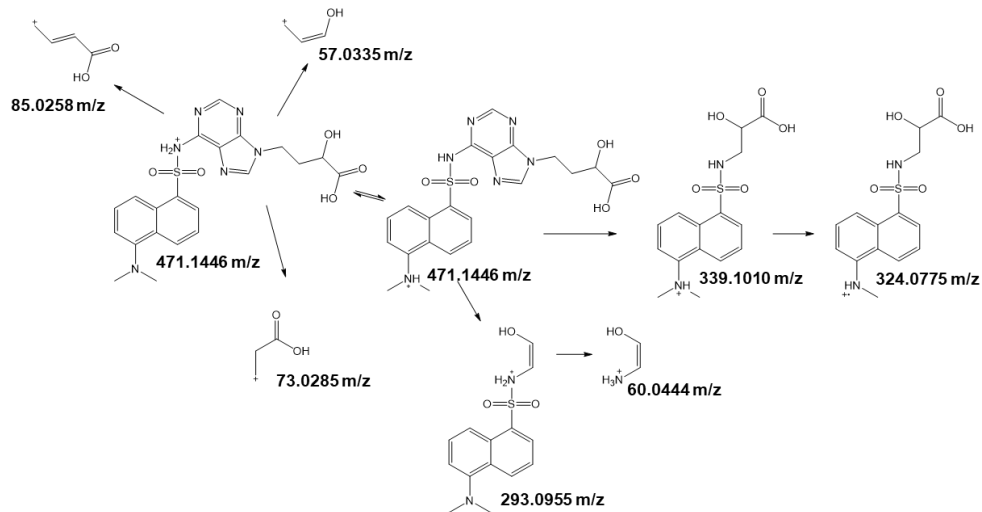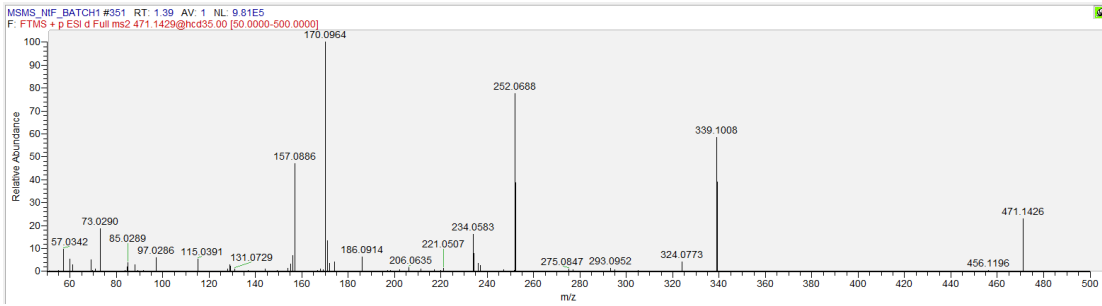
No. 99 Cyanamide (two precursor available, 276.0802 and 276.0821 m/z)
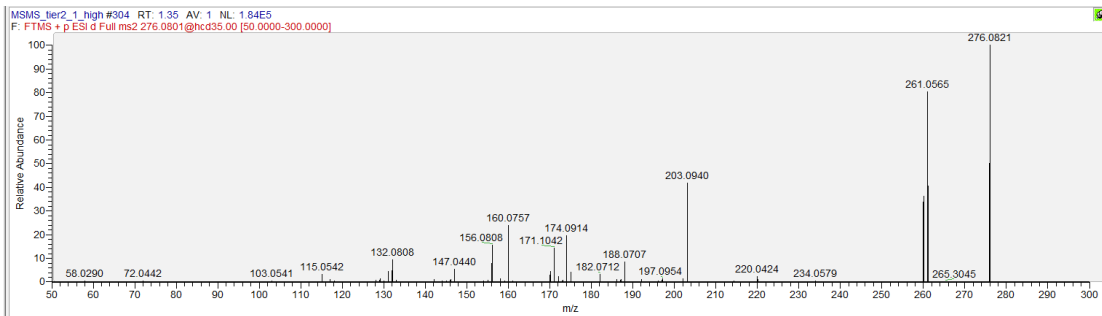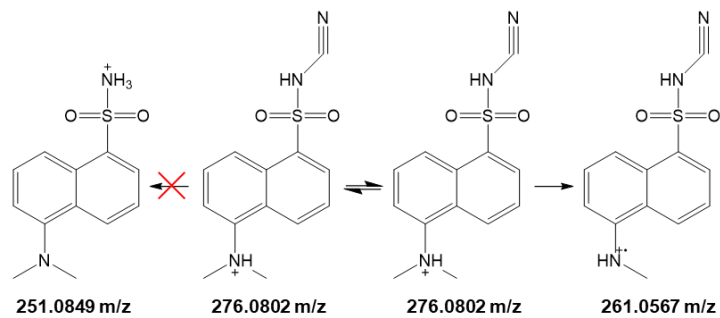
No. 120 Isomer of dimethylarginine
(MS2 was identical with dimethylarginine)
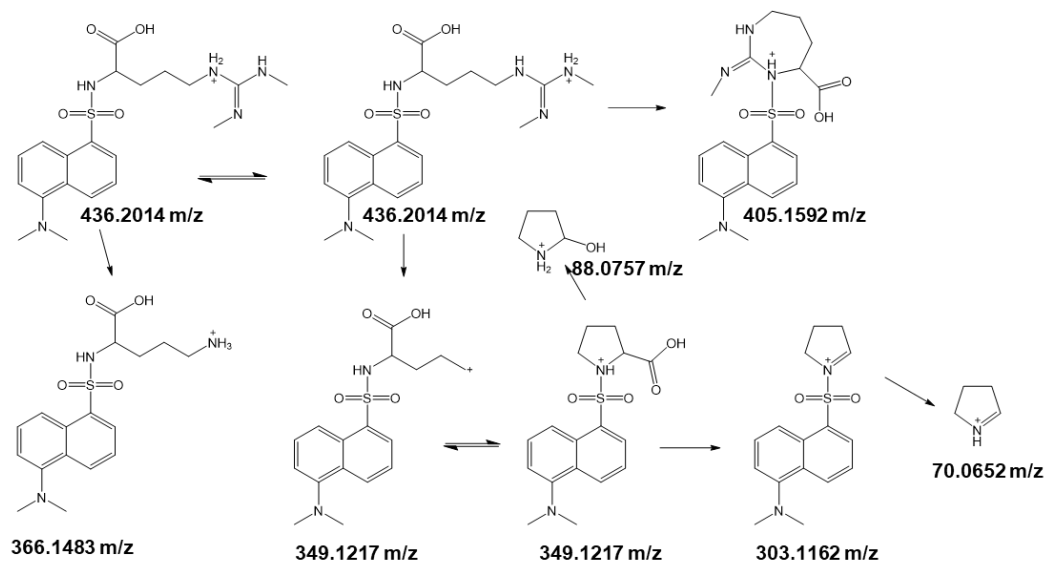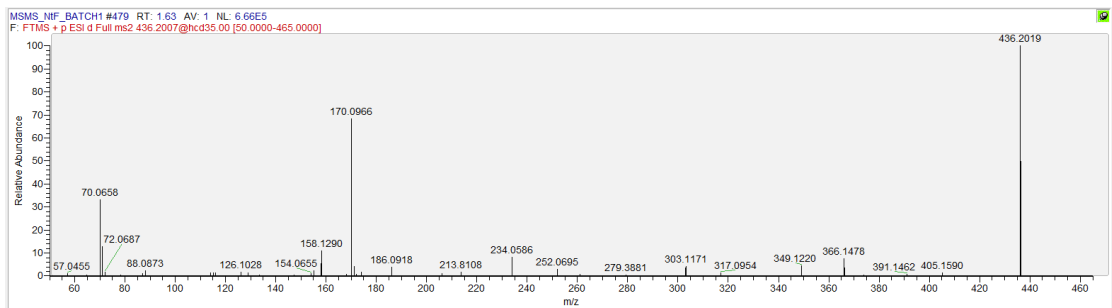
No. 133 Isomer of L-Glutamine/D-Glutamine
(MS2 was identical with L-Glutamine/D-Glutamine)
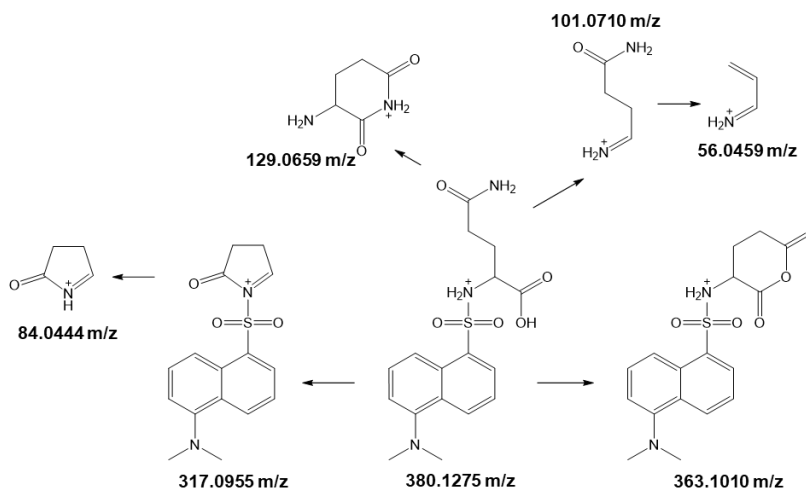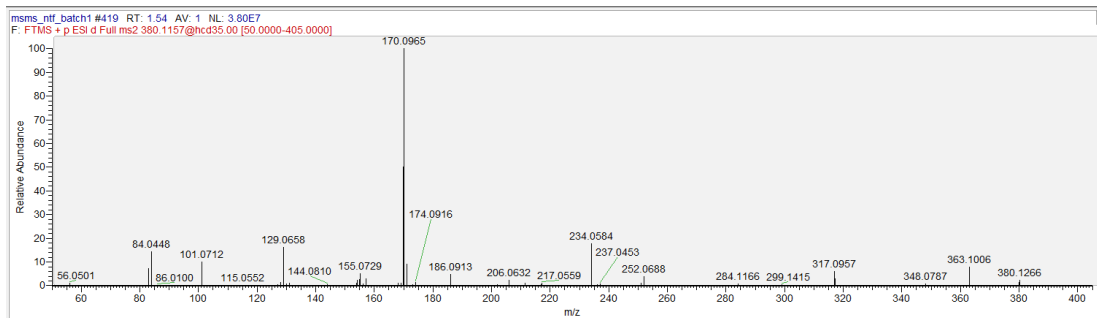
No. 136 Deoxyeritadenine (isomer)
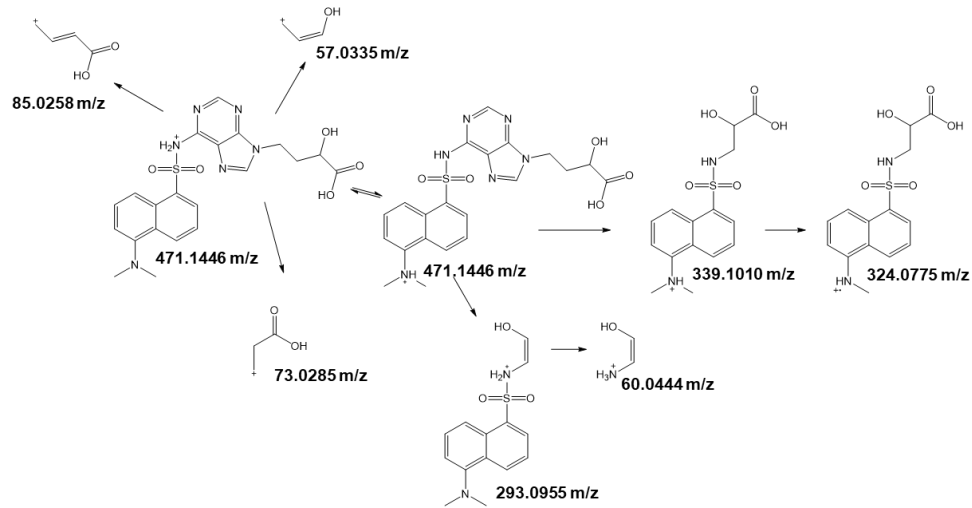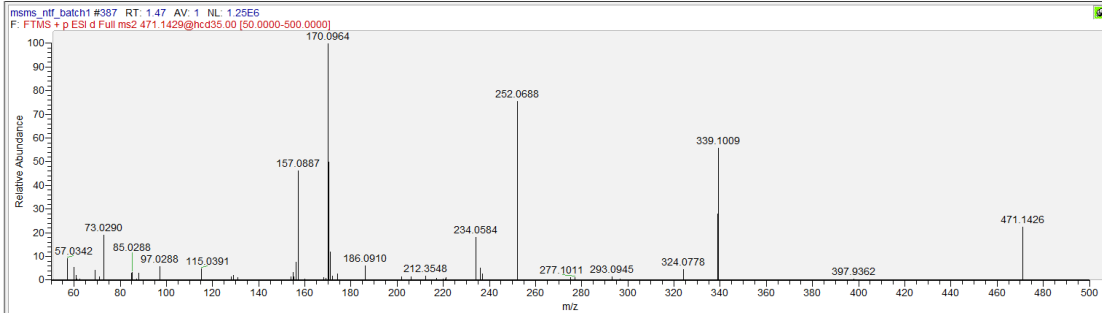
No. 178 Succinyladenosine

178
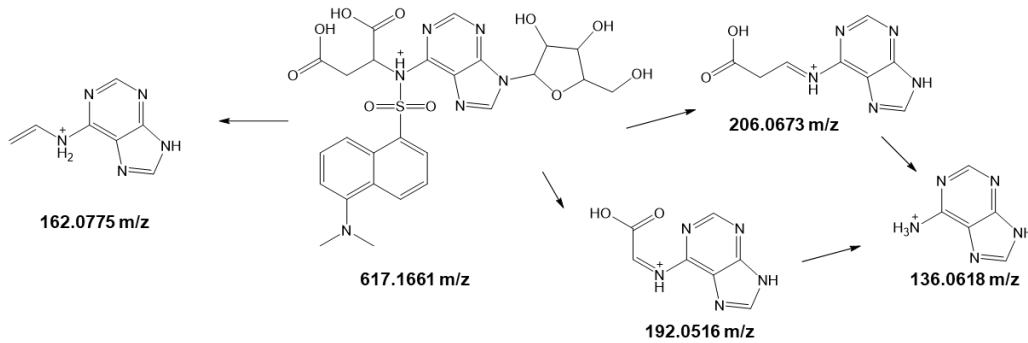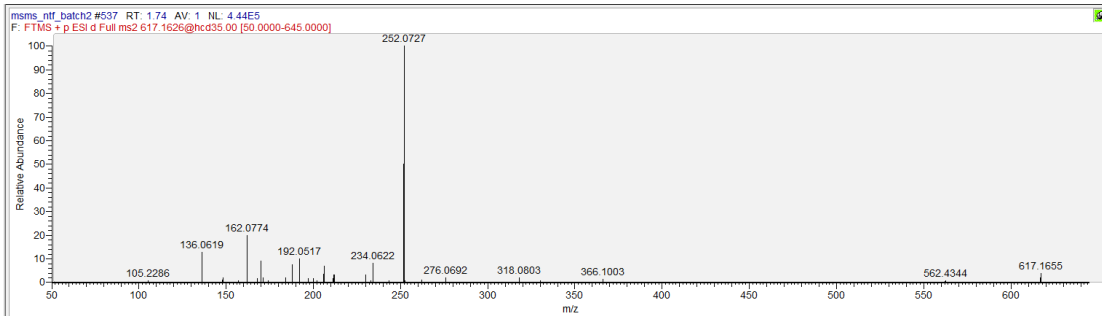
No. 214 Tabtoxinine-beta-lactam

No. 220 N2'-Acetyl-L-Cystathionine

90.0372 m/z

136.0427 m/z

112.0394 m/z

498.1364 m/z

480.1258 m/z

456.1258 m/z

452.1309 m/z

160.0427 m/z

393.0938 m/z

367.0781 m/z

410.1203 m/z

347.0883 m/z

56.0495 m/z

289.1006 m/z
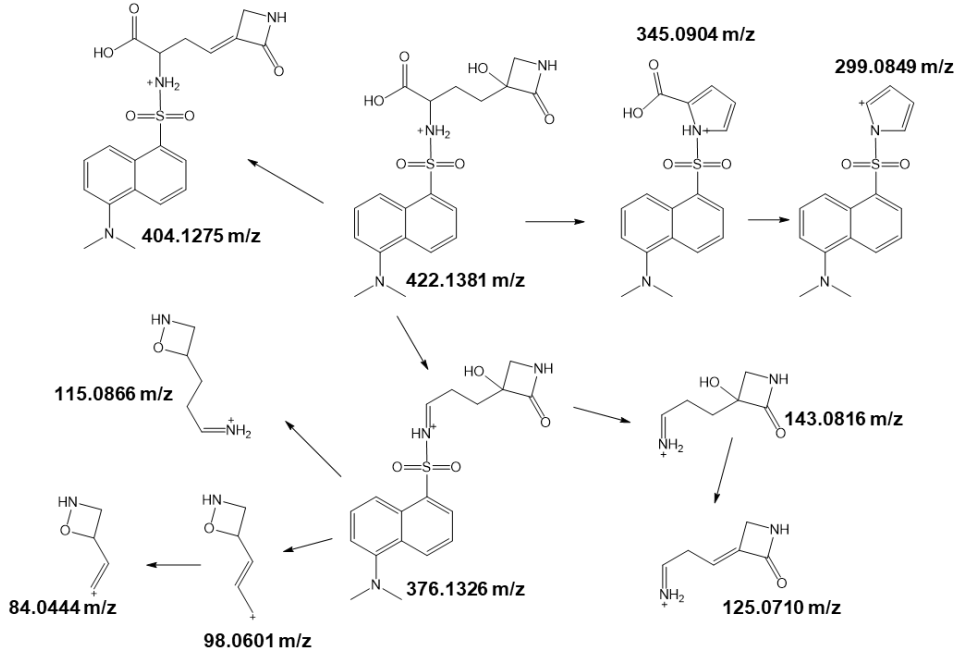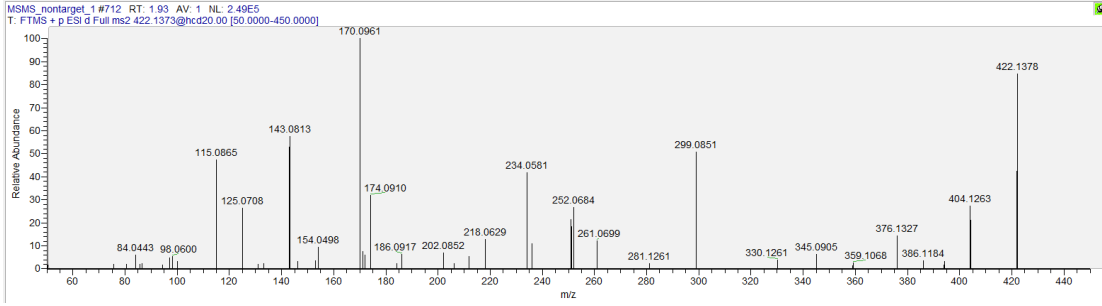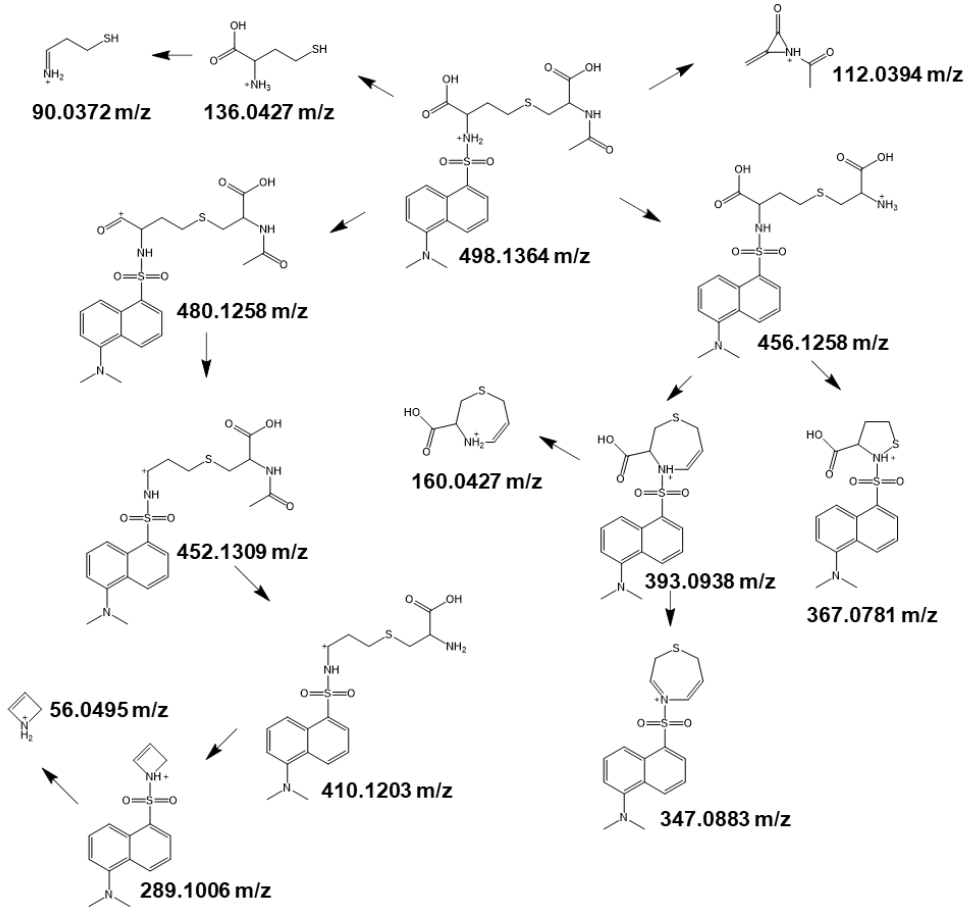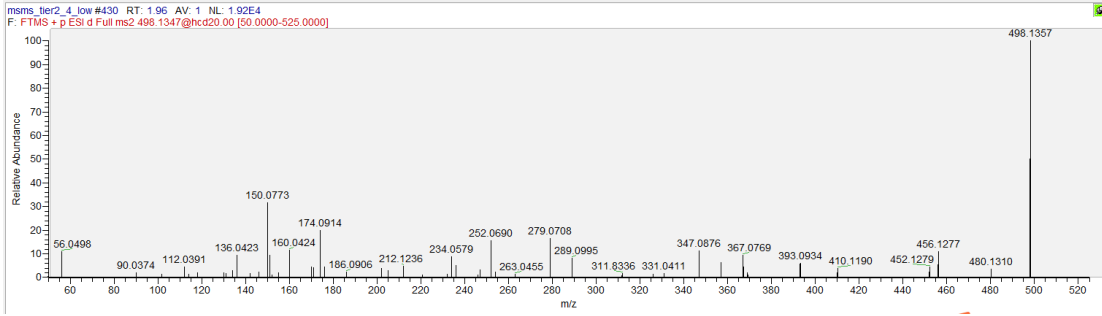
No. 221 Acetylhydrazine (two precursors were available, 308.1855 and 308.1064 m/z)
(Fragments at 249.11 and 85.02 m/z were generated from precursors at 308.1855 m/z)
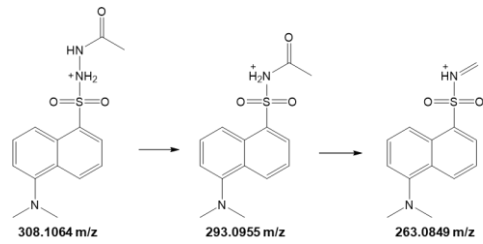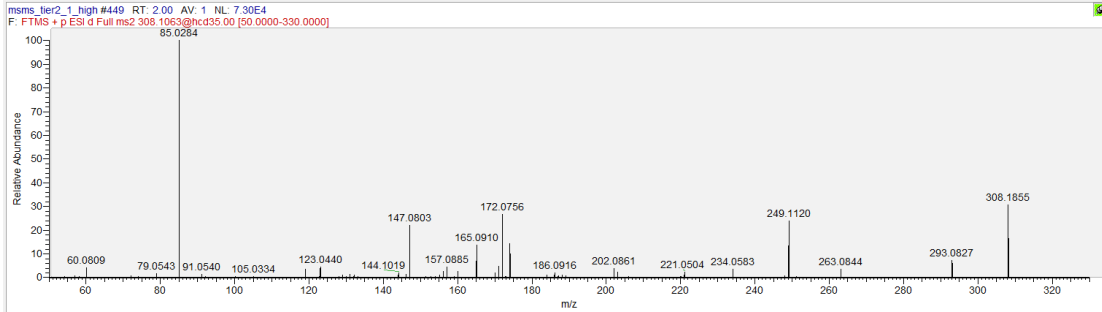
No. 223 Hydroxyethyl glycine
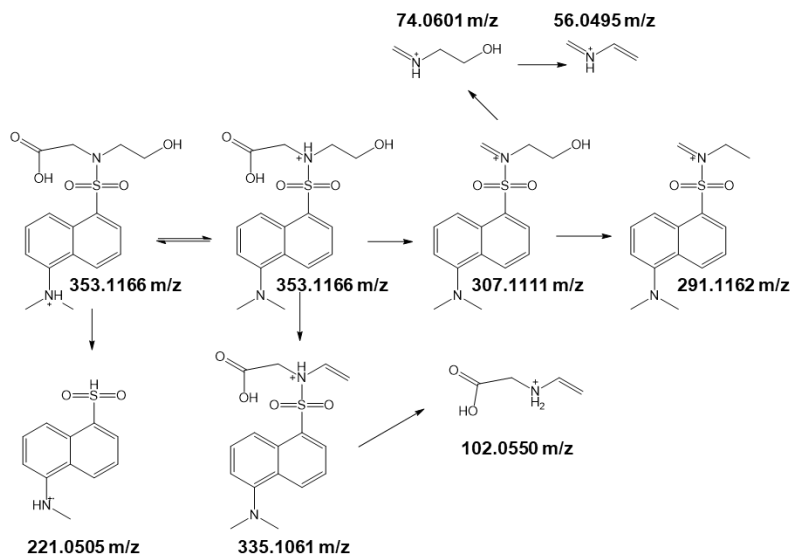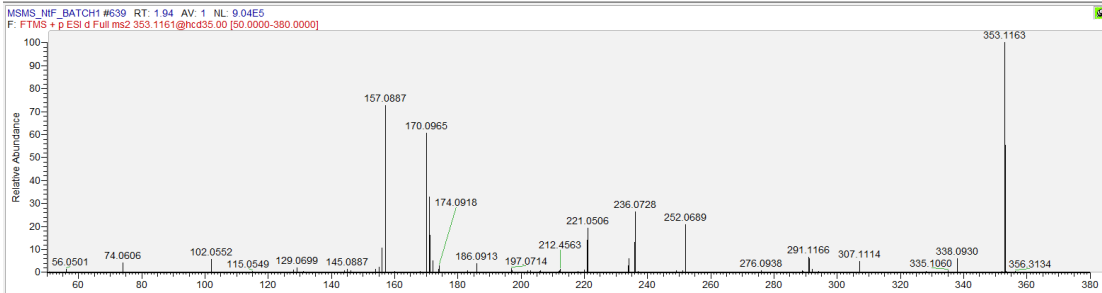
No. 250 N5-Acetyl-L-Ornithine
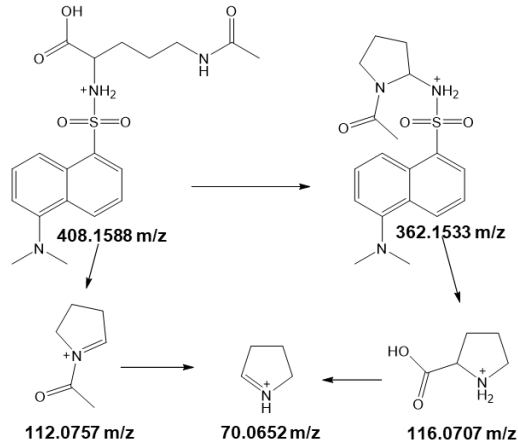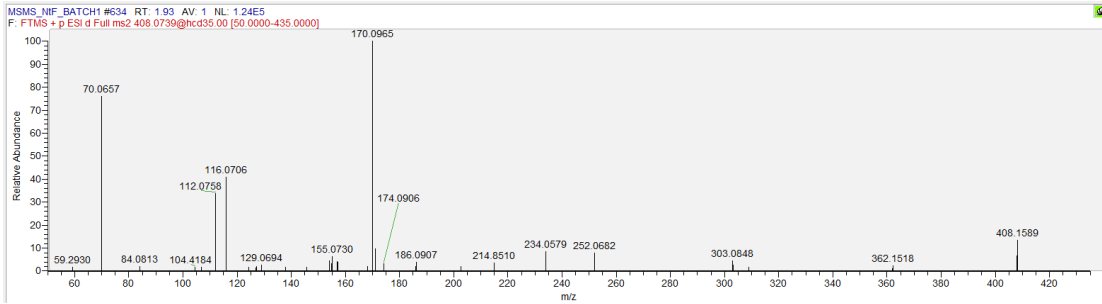
No. 304 Isomer of L-Serine/D-Serine
(MS2 was identical with L-Serine/D-Serine)
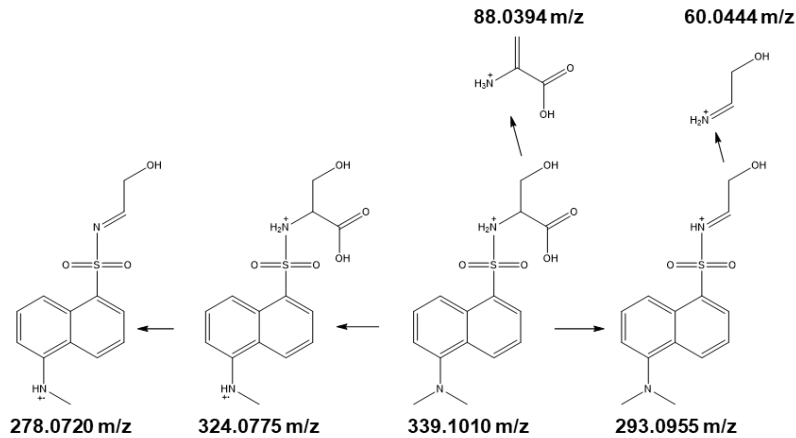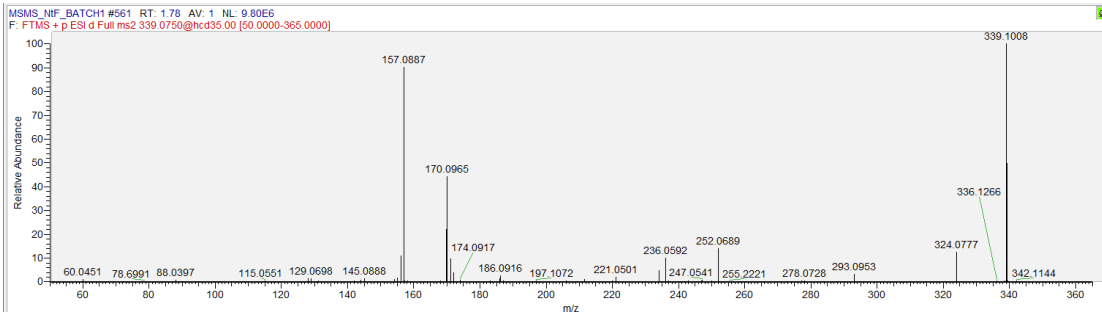
No. 322 Tyramine O-sulfate
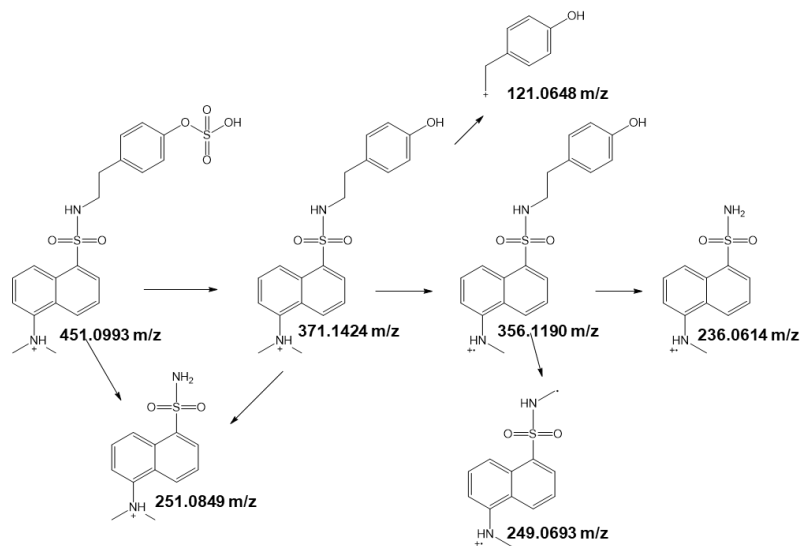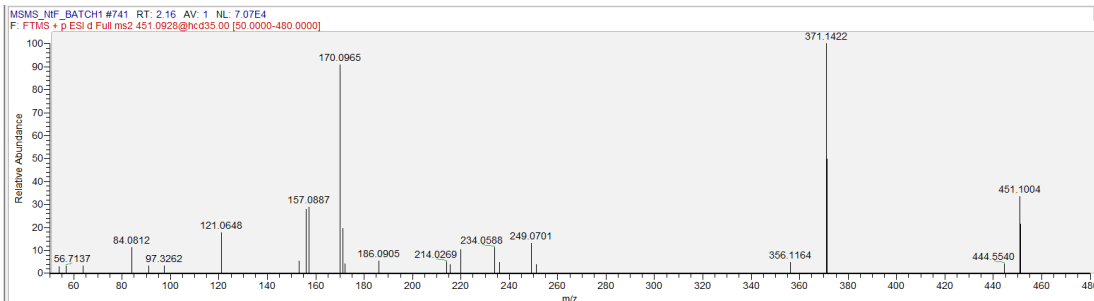
No. 343 Isomer of Glutamic acid/D-Glutamic acid/L-Glutamic acid
(MS2 was identical with Glutamic acid/D-Glutamic acid/L-Glutamic acid)
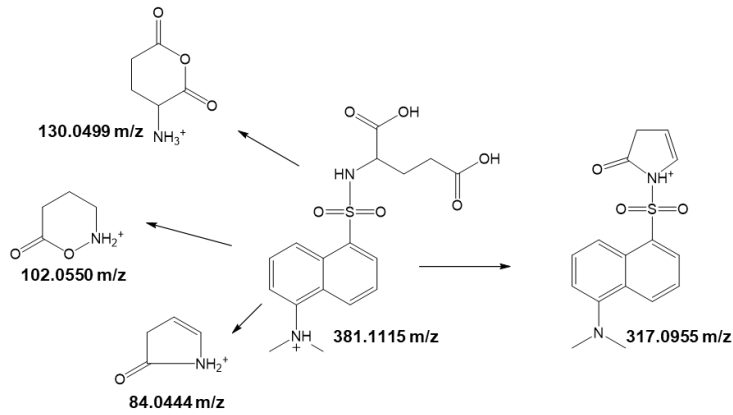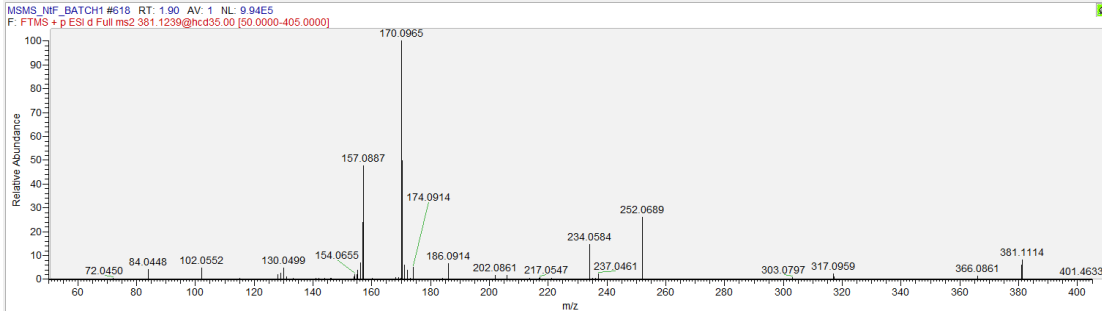
No. 344 N-Carbamoylputrescine
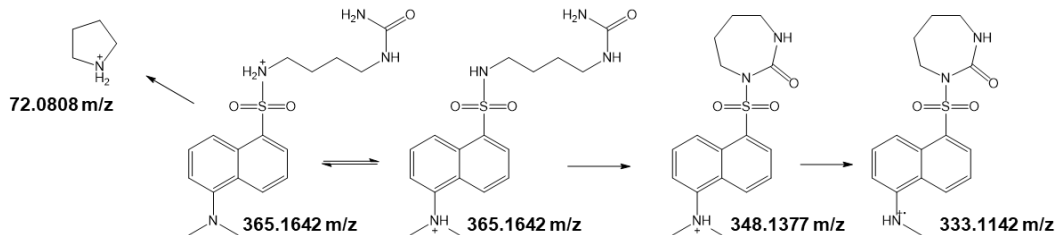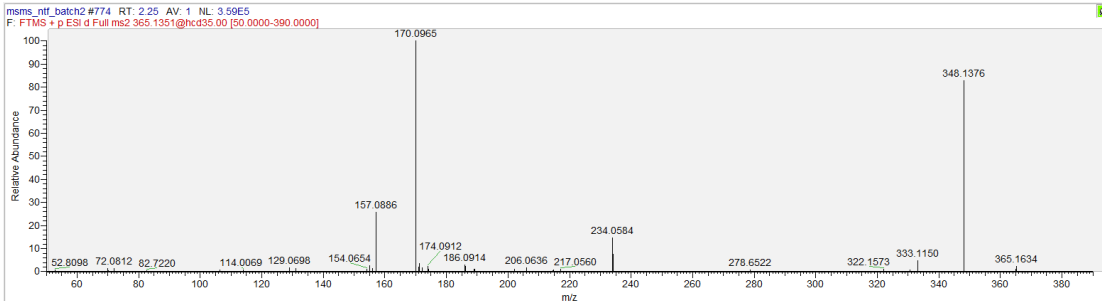
No.348 Propylcysteine sulfoxide (two precursors available, 413.1239 and 413.1200) (fragments at 76.02 m/z were generated from precursors at 413.1239 m/z)
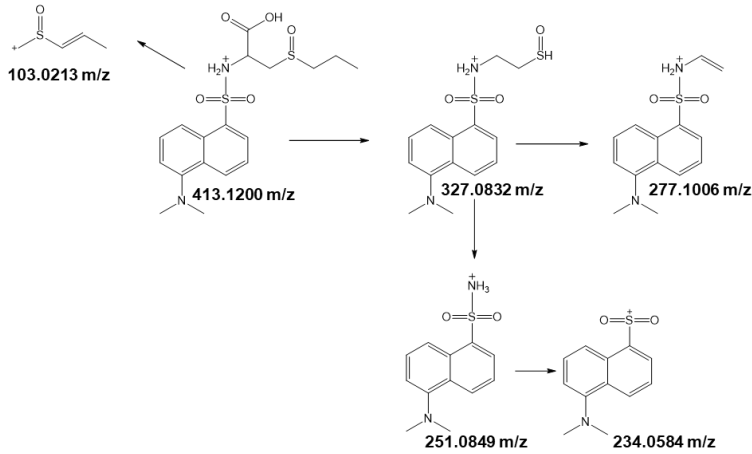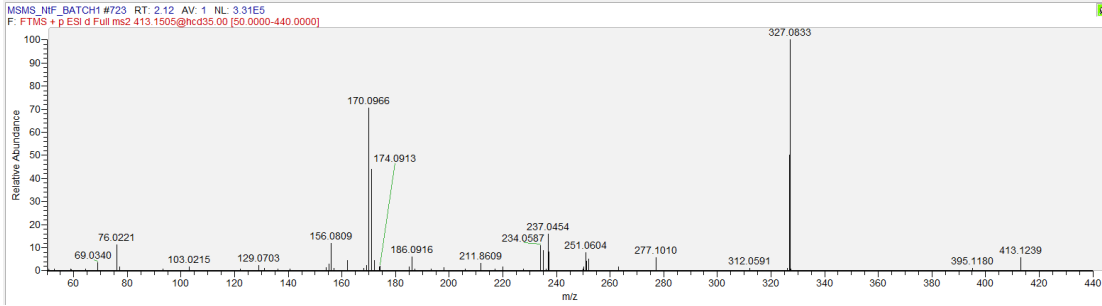
No. 356 6-Amino-2-hydroxyhex-5-enoic acid (Isomer)
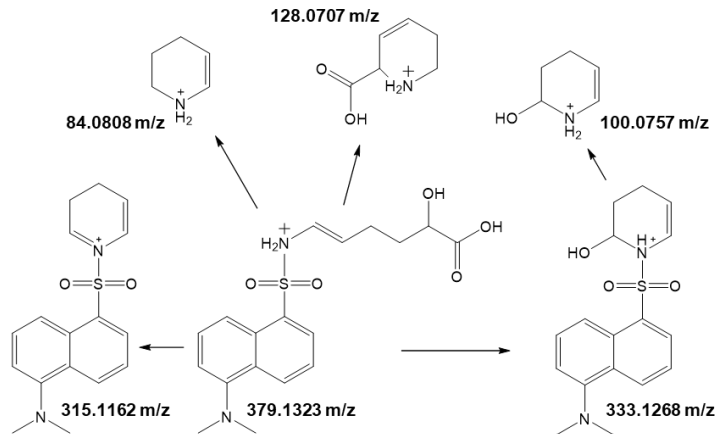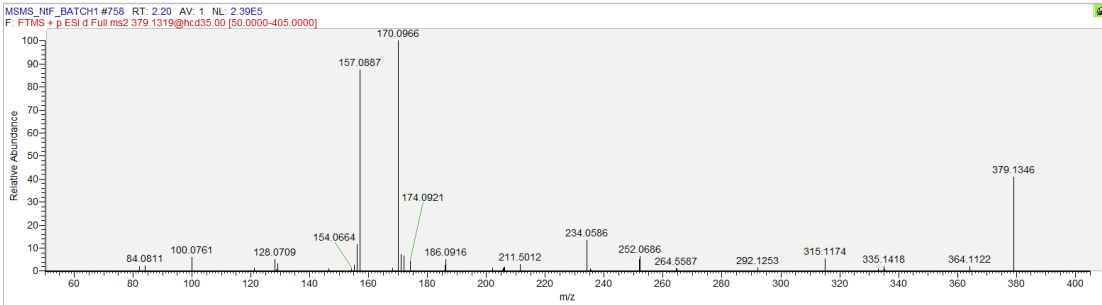
No. 376 3-Cyano-L-alanine (isomer)
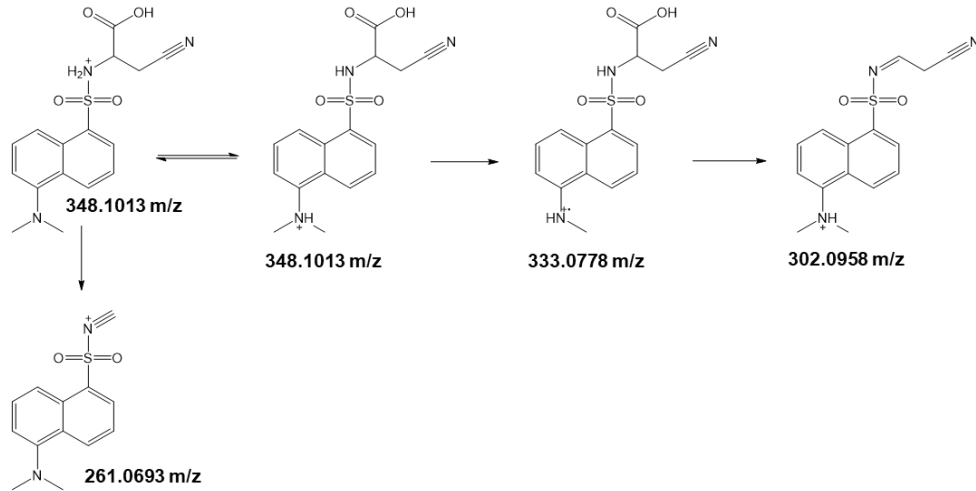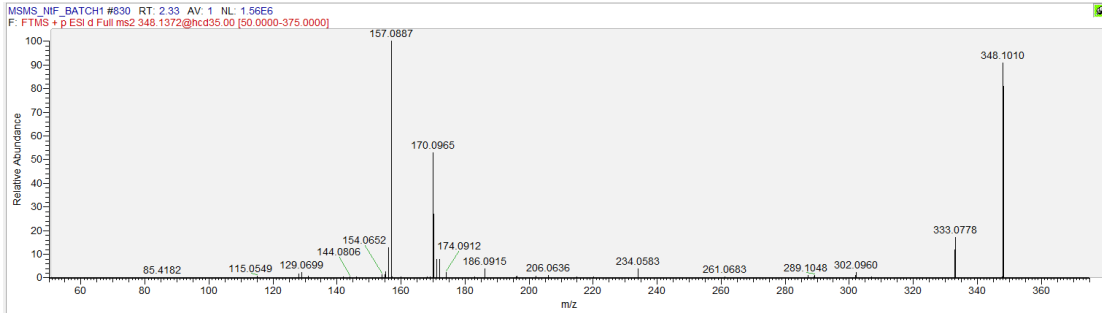
NO. 381 Aminoethoxyacetic acid
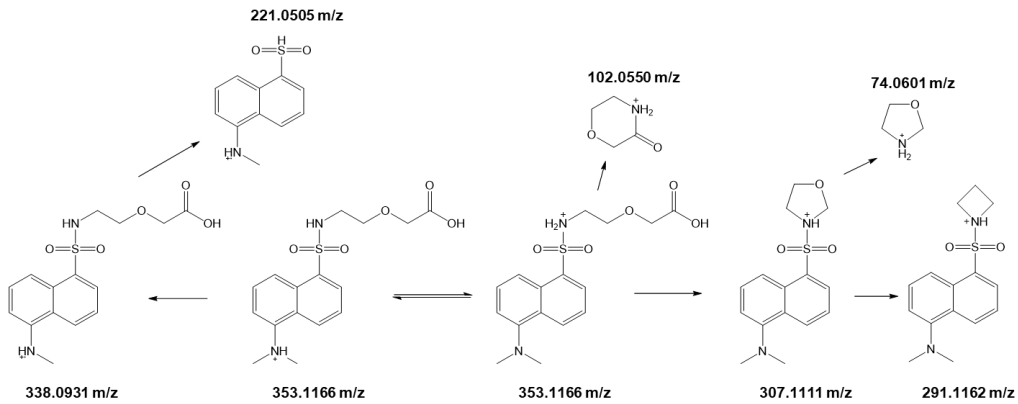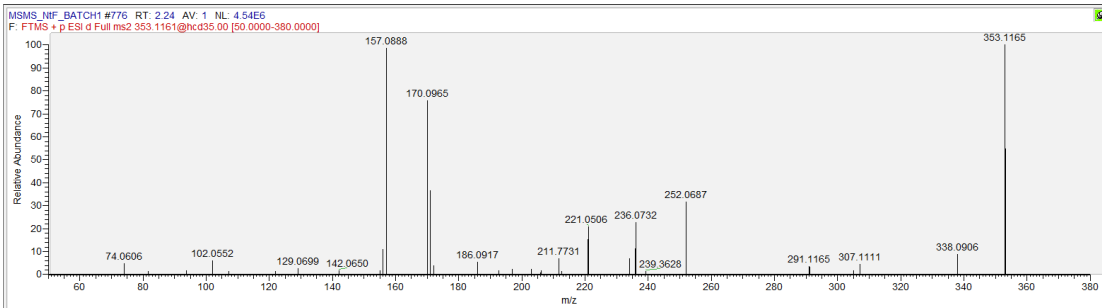
No. 386 6-Amino-2-hydroxyhex-5-enoic acid (isomer)
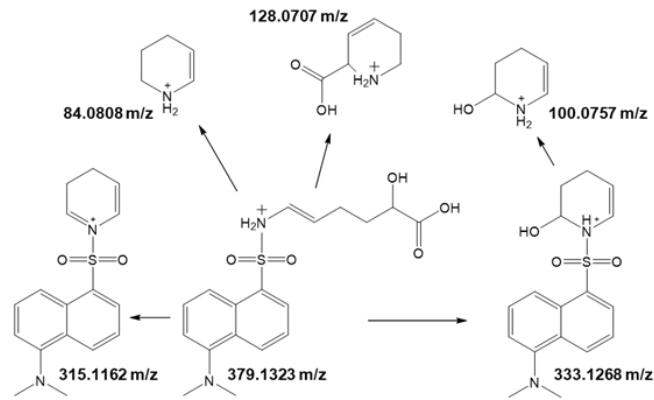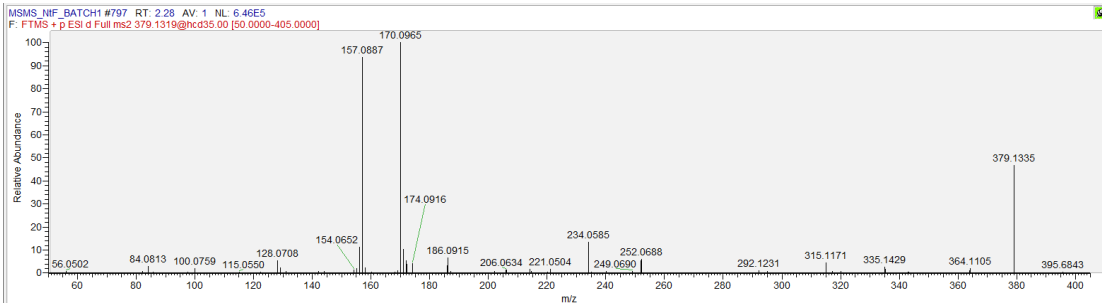
No. 406 N-[4-[[(E)-5-aminopent-2-enyl]amino]-2-hydroxybutyl]acetamide
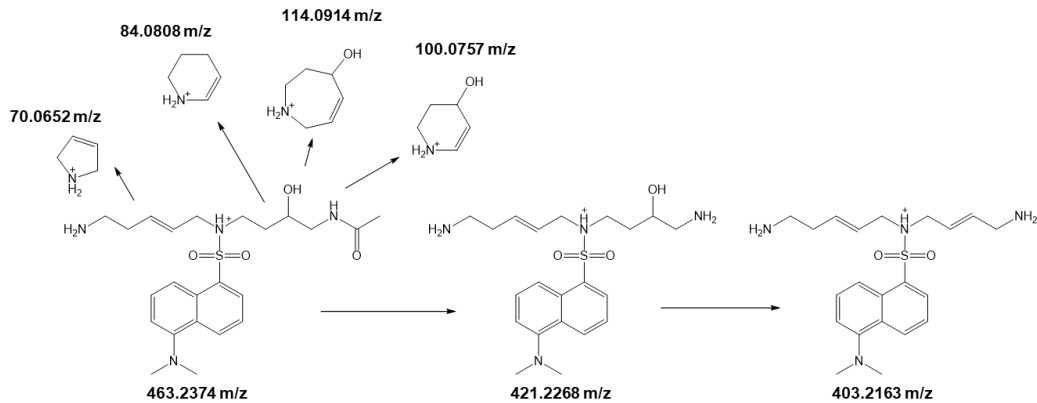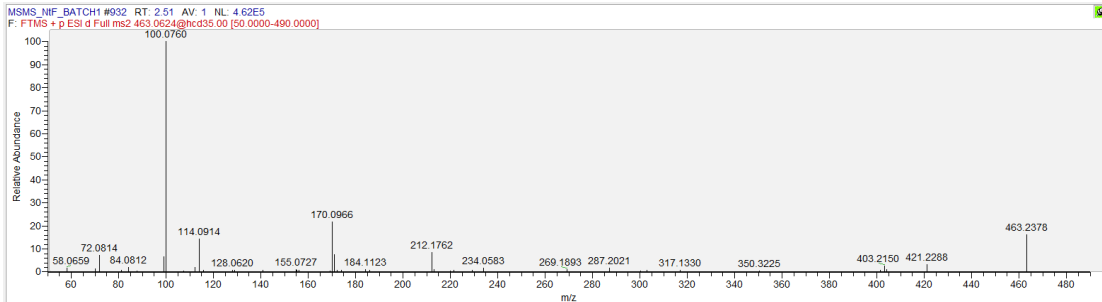
No. 440 Aminoethoxyacetic acid (isomer, it might be 4-amino-4-hydroxybutanoic acid)
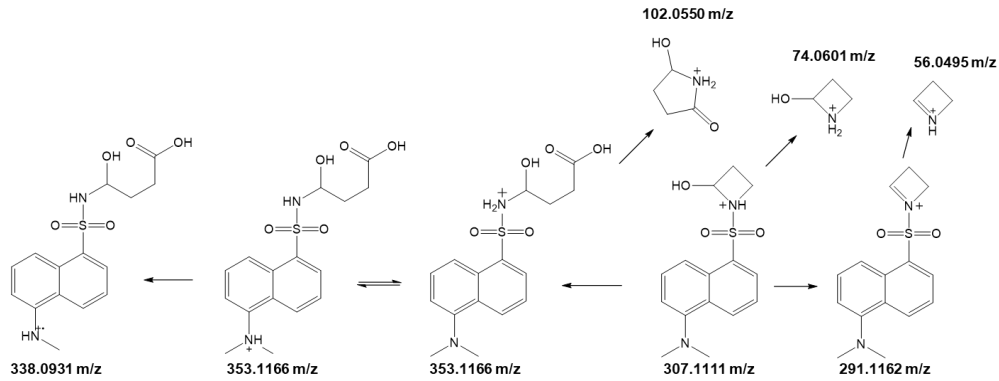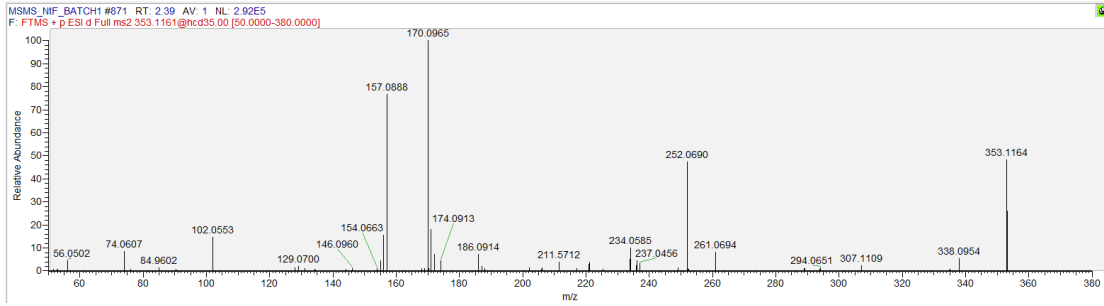
No. 447 4-Amino-4-hydroxy-2,3-dimethylbutanoic acid
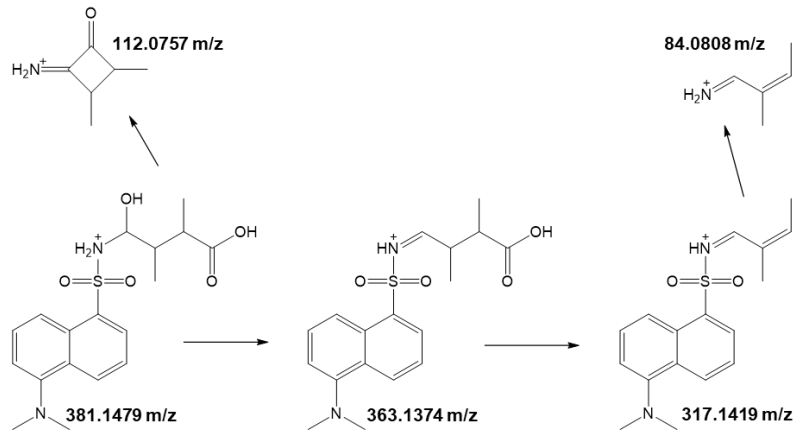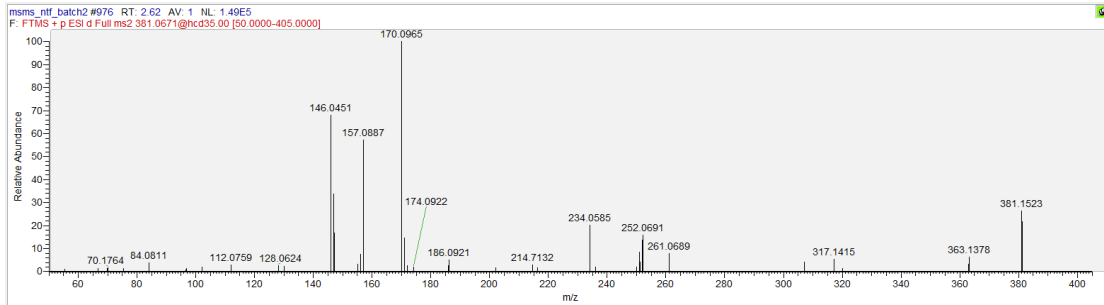
Two precursors were available in the isolation window for MS/MS, fragments at 146.0451 were generated from precursors 381.1523.

No. 479 Putrescine
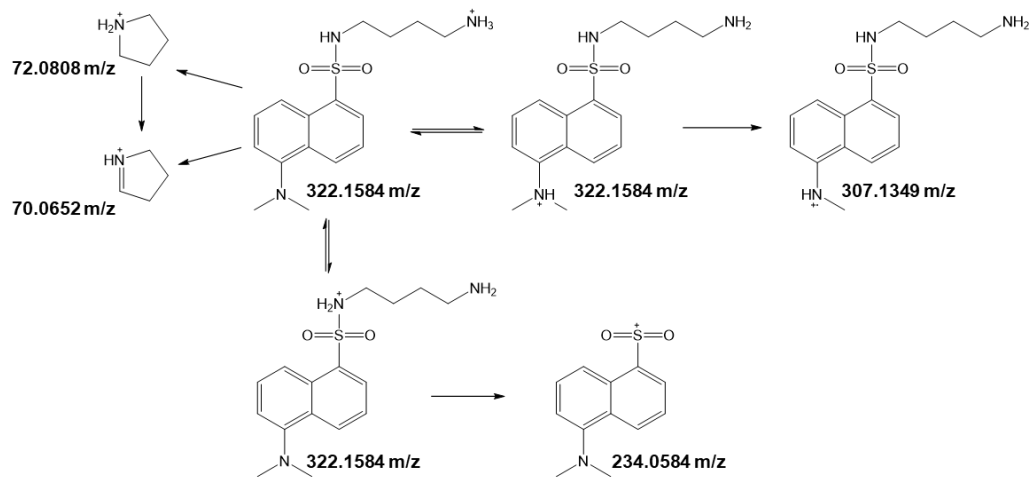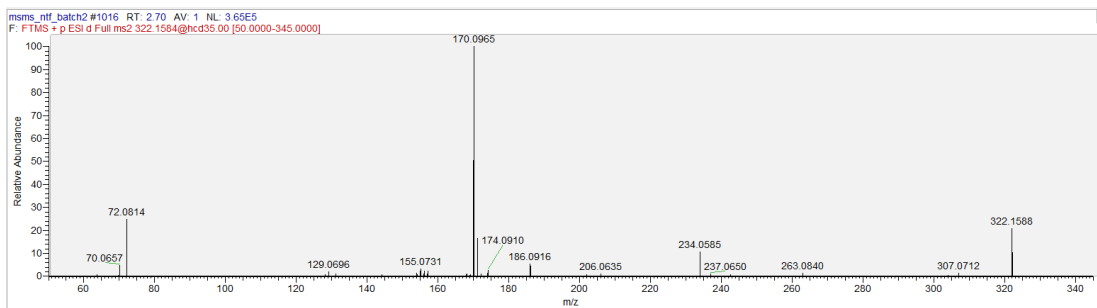
No. 496 4-Methyl-L-glutamic acid

144.0656 m/z

116.0707 m/z

70.0652 m/z

395.1272 m/z

98.0601 m/z

331.1111 m/z

No. 510 3-Amino-2-piperidone

113.0710 m/z

85.0761 m/z

70.0562 m/z

221.0505 m/z

348.1377 m/z

348.1377 m/z

No. 534 Ethanolamine (isomer)

252.0689 m/z	295.1111 m/z	295.1111 m/z	280.0877 m/z	236.0614 m/z
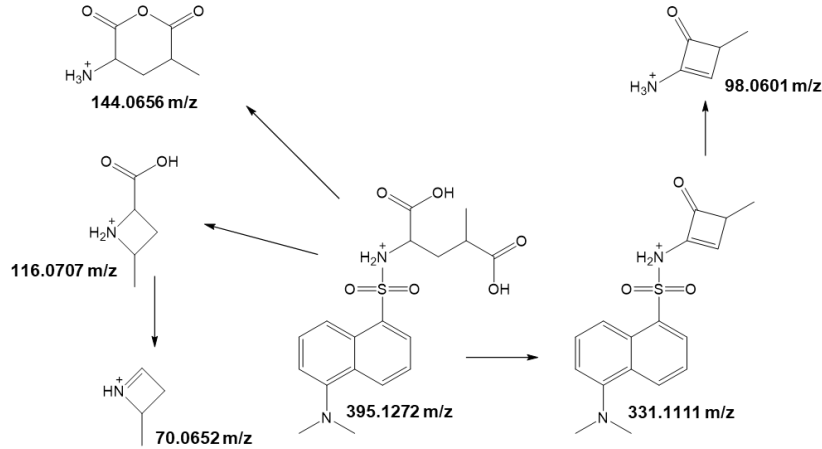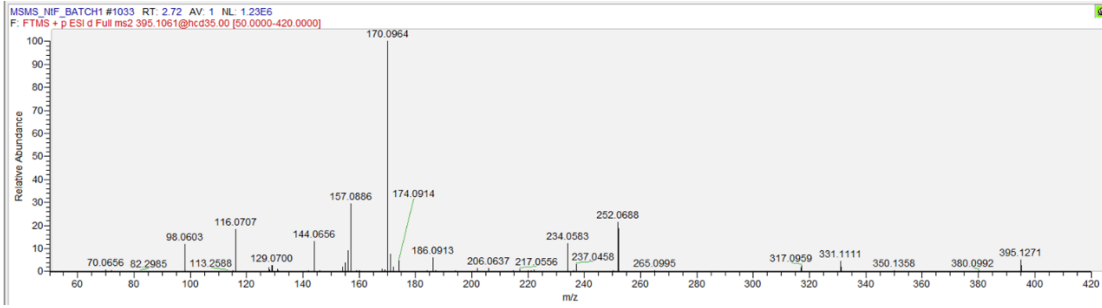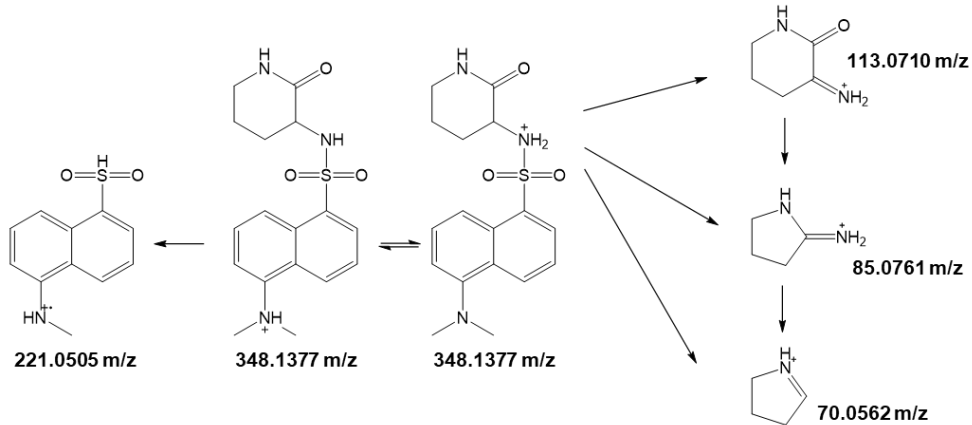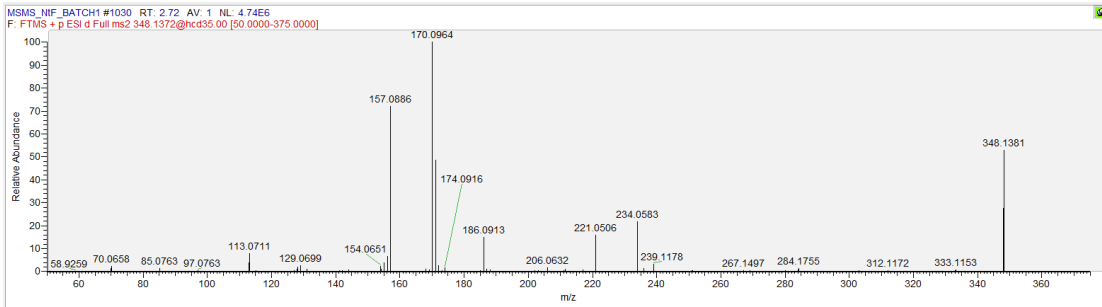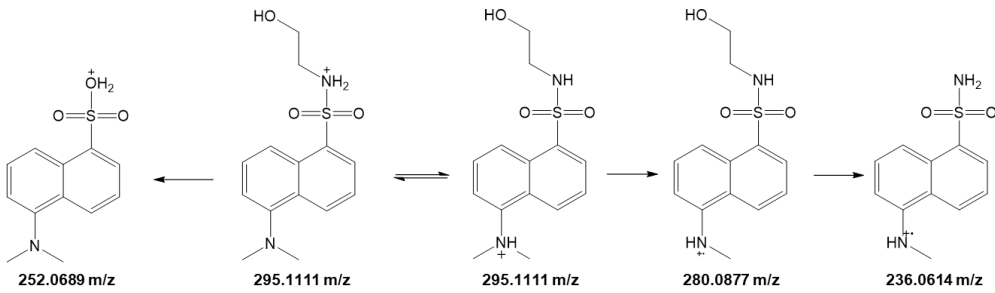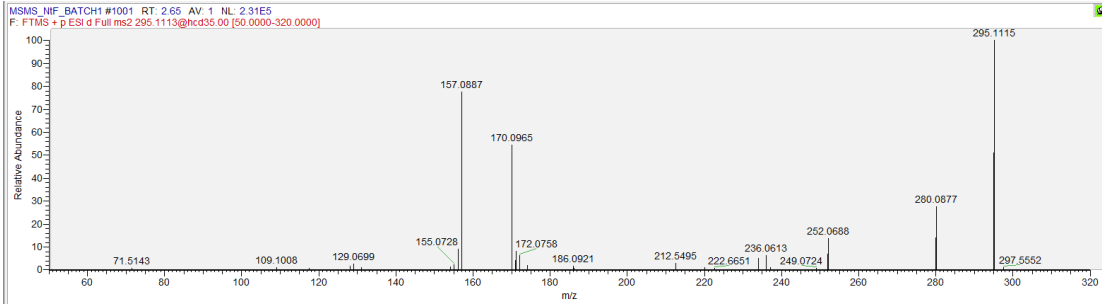
No. 565 O-Methyl-L-threonine (two precursors available, 367.1323 and 367.0688)
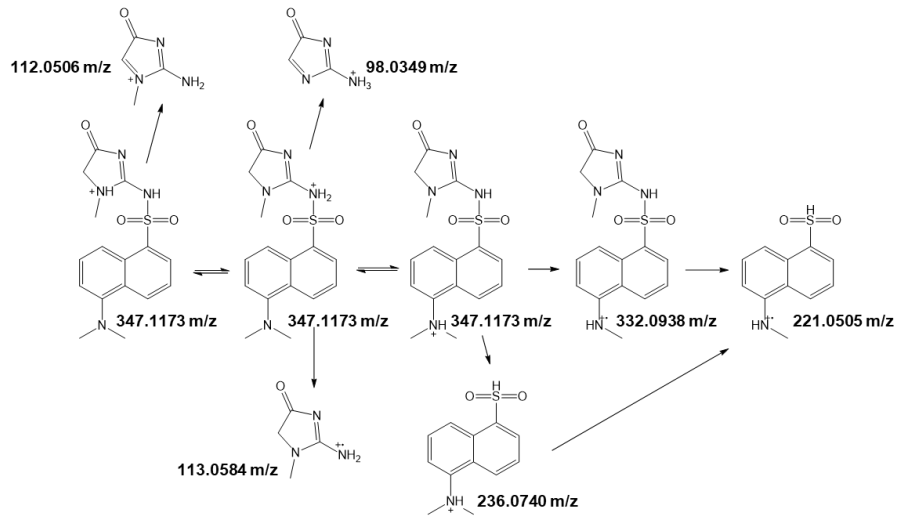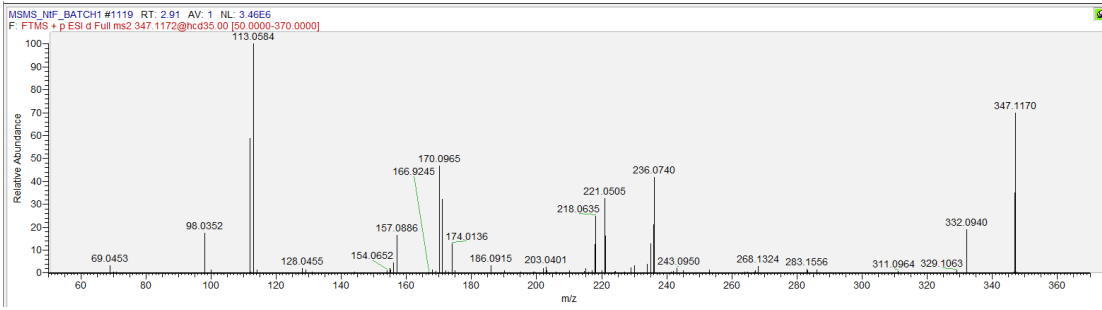
116.0707 m/z

88.0757 m/z

367.1323 m/z	321.1268 m/z	275.0849 m/z

No. 578 Creatine

No. 626 (2R,2'S)-Isobuteine
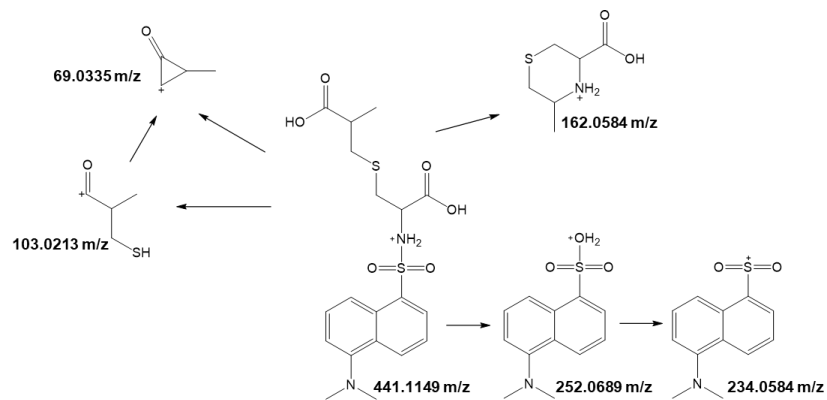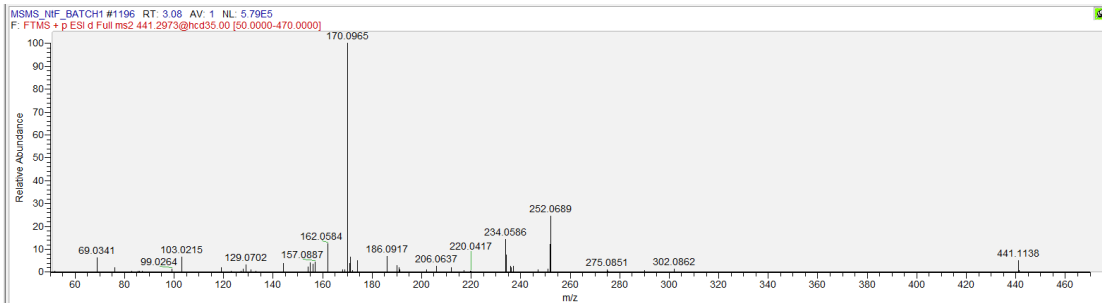
192

No. 654 Ethanolamine (isomer)
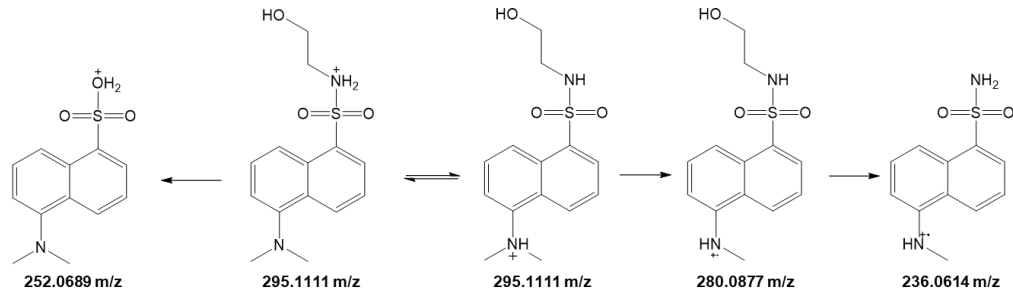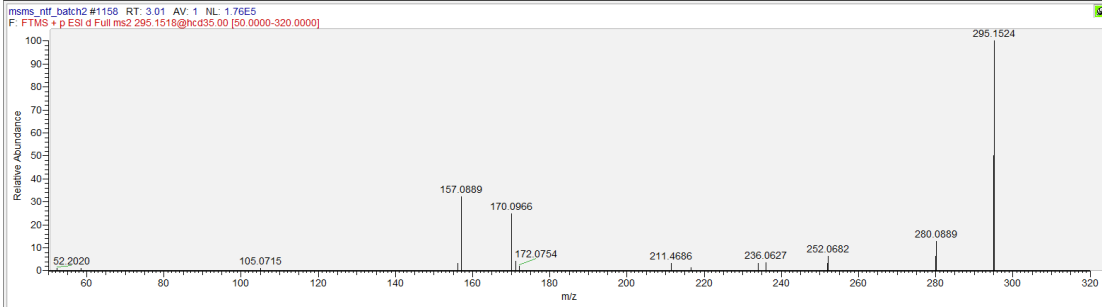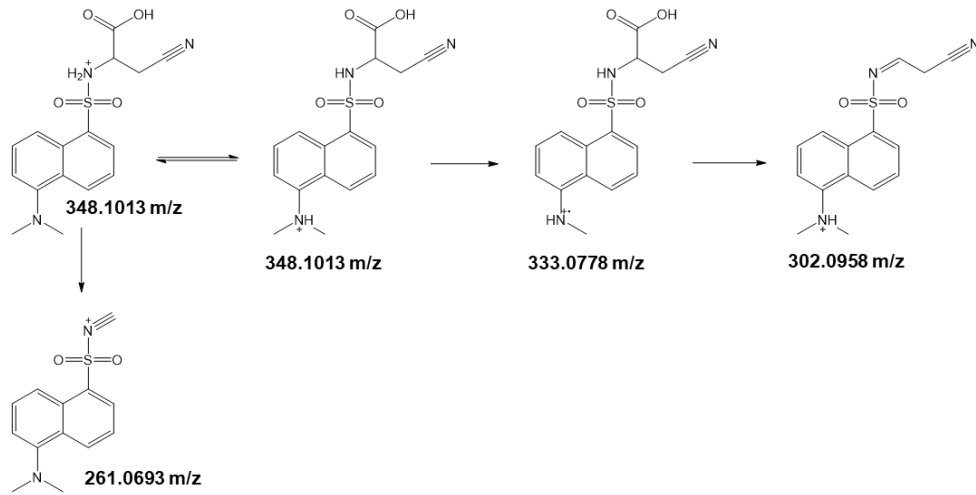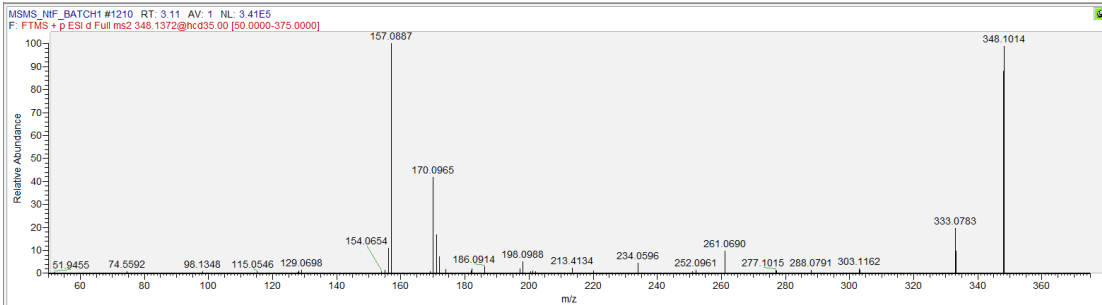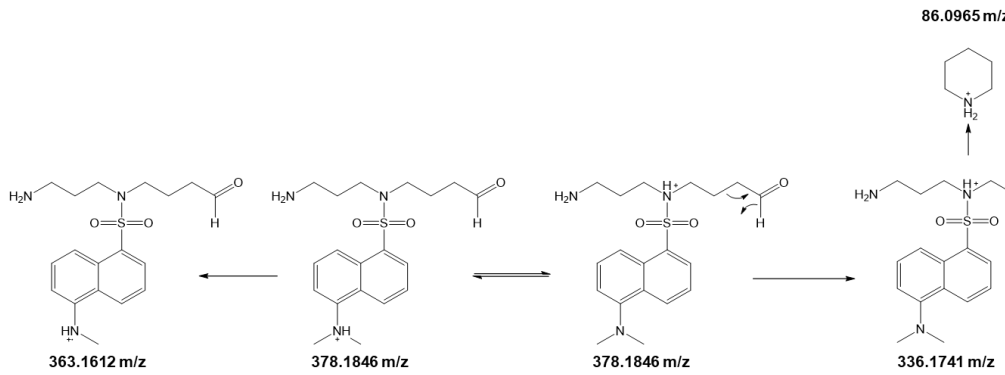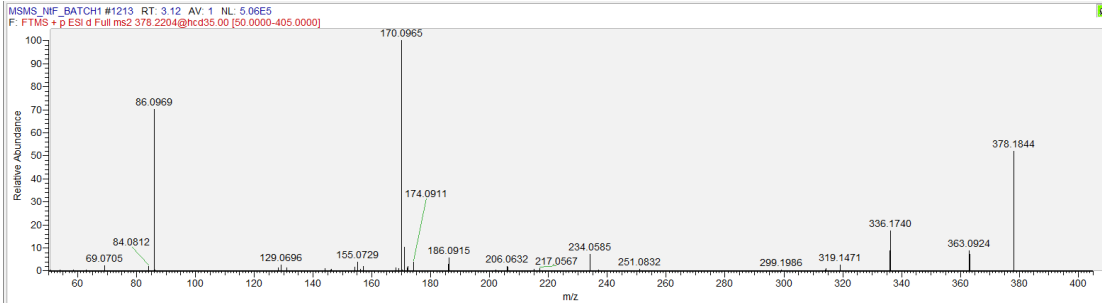
252.0689 m/z	295.1111 m/z	295.1111 m/z	280.0877 m/z	236.0614 m/z

No. 687 3-Cyano-L-alanine (isomer)

348.1013 m/z

348.1013 m/z

333.0778 m/z

302.0958 m/z

261.0693 m/z

No. 689 1-(3-Aminopropyl)-4-aminobutanal

86.0965 m/z

363.1612 m/z   **378.1846 m/z**   **378.1846 m/z**   **336.1741 m/z**

No. 716 (R)-1-Aminopropan-2-ol

58.0652 m/z

294.1033 m/z   **309.1268 m/z**   **309.1268 m/z**   **263.0849 m/z**

194

No. 729 5-(3',4'-Dihydroxyphenyl)-gamma-valerolactone-4'-O-glucuronide

618.1640 m/z   442.1319 m/z   378.1700 m/z   360.1595 m/z

No. 736 4-(3-Aminobutylamino)butan-2-one

72.0808 m/z

378.1846 m/z   322.1584 m/z
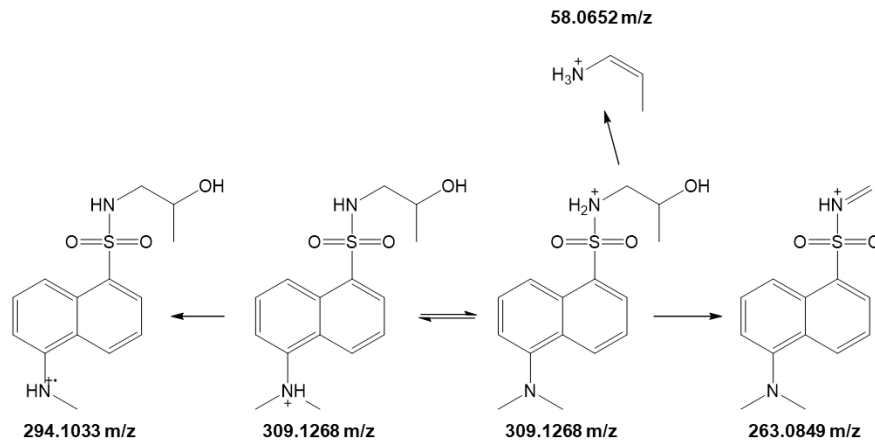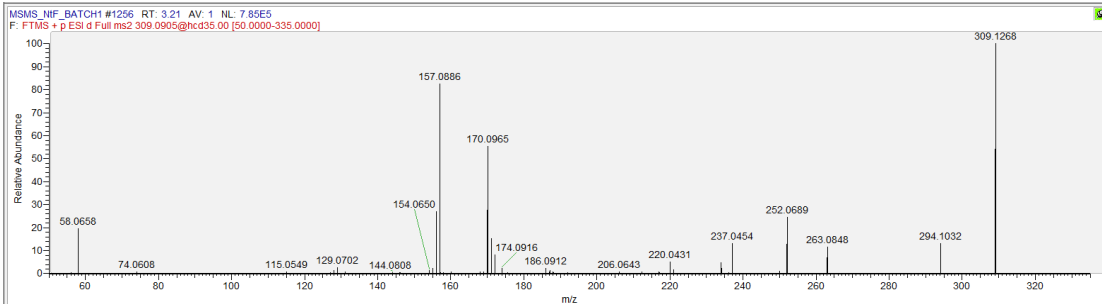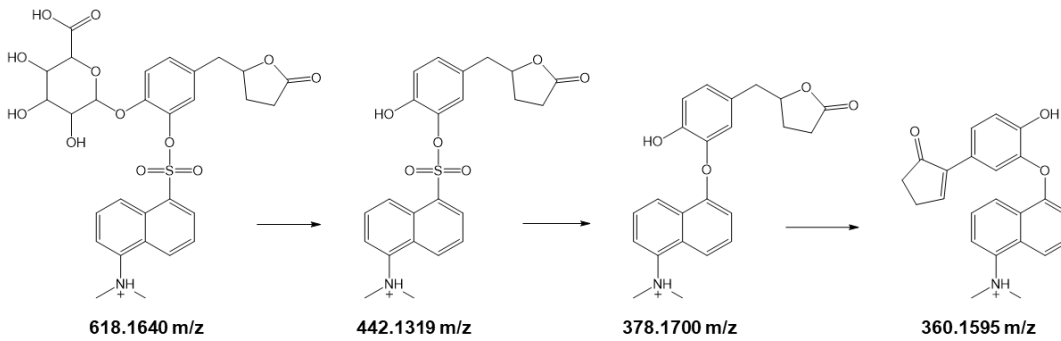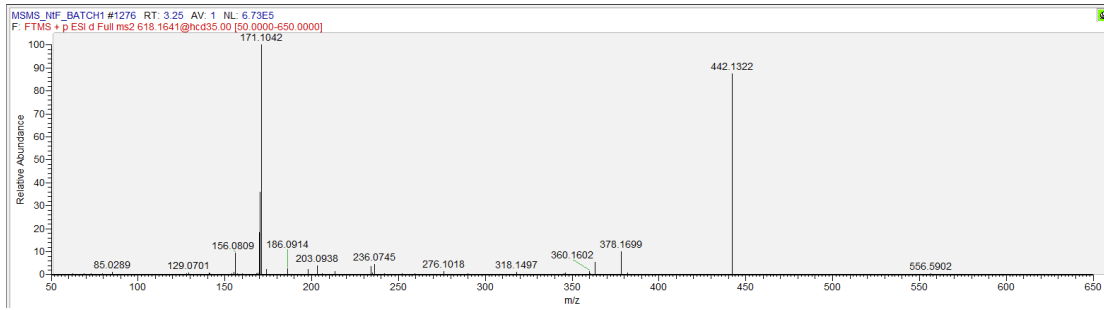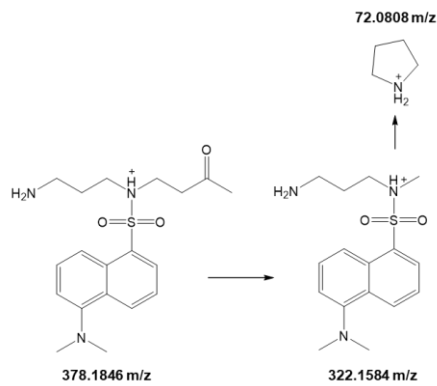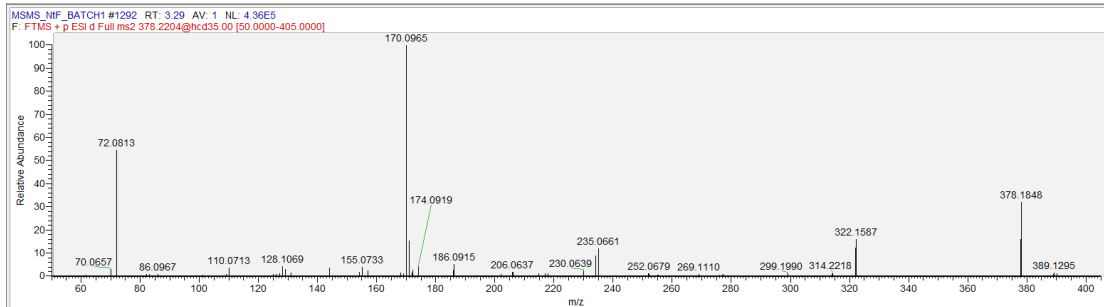
No. 748 5-(3',4'-Dihydroxyphenyl)-gamma-valerolactone-4'-O-glucuronide (isomer)
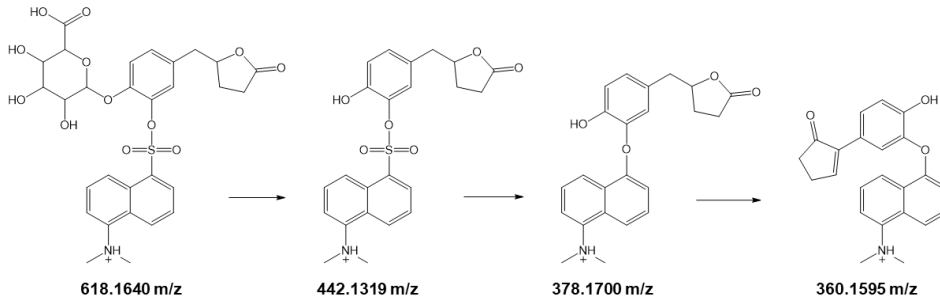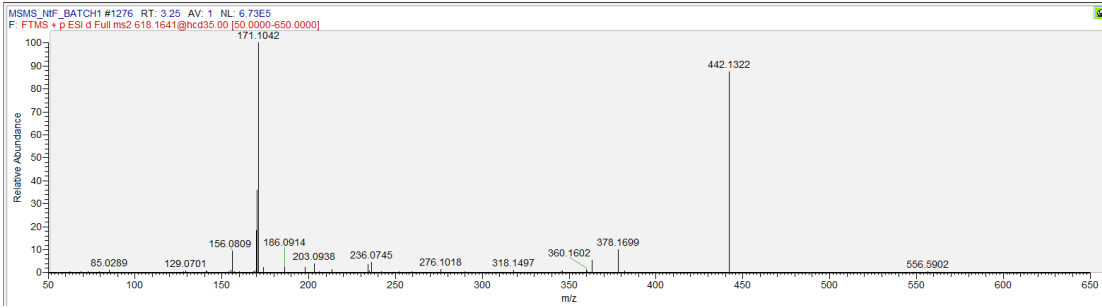
618.1640 m/z   442.1319 m/z   378.1700 m/z   360.1595 m/z

No. 761 4-Amino-2-methylenebutanoic acid

98.0601 m/z

334.0982 m/z   349.1217 m/z   349.1217 m/z   234.0584 m/z

196

No. 764 3-(3,5-dihydroxyphenyl)-1-propanoic acid sulphate

416.1163 m/z

496.0731 m/z

398.1057 m/z

352.1544 m/z

337.1309 m/z
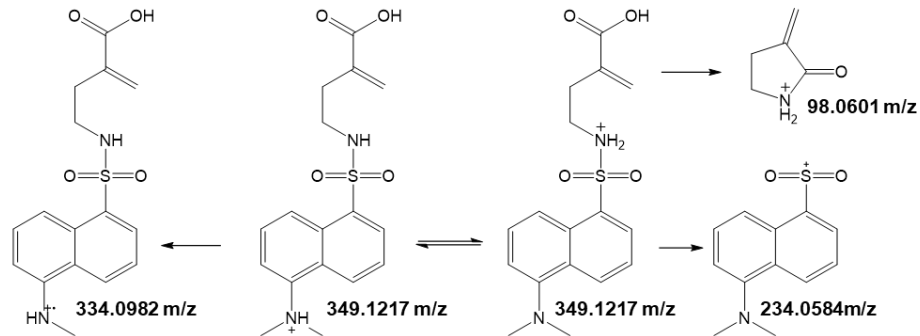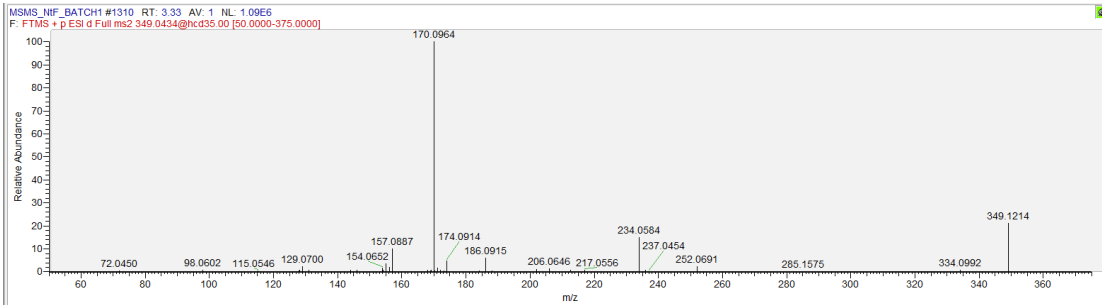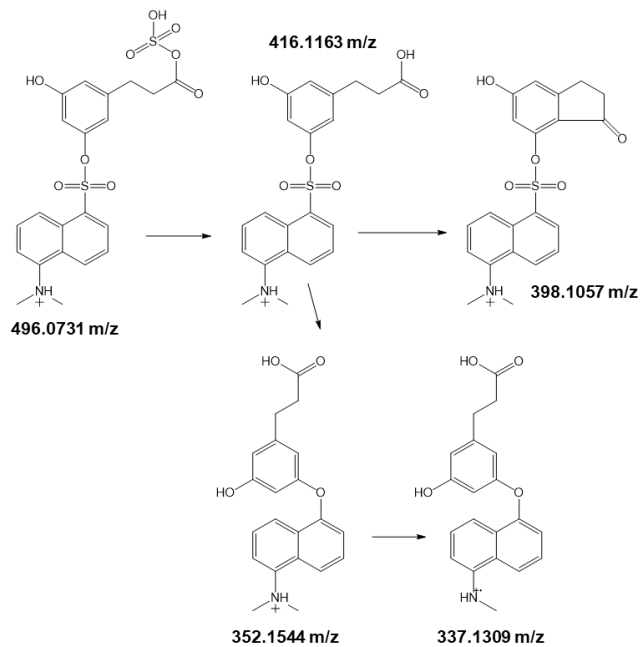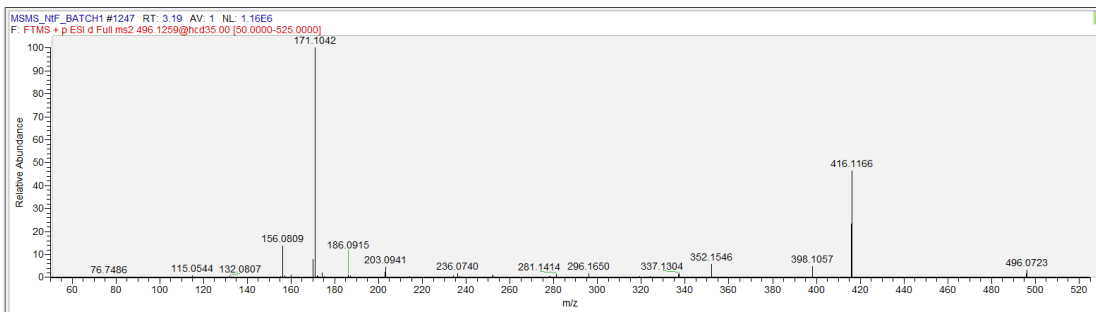
No. 799 4-Amino-2-methylenebutanoic acid
(Isomer, it might be (E)-4-amino-3-methylbut-2-enoic acid)
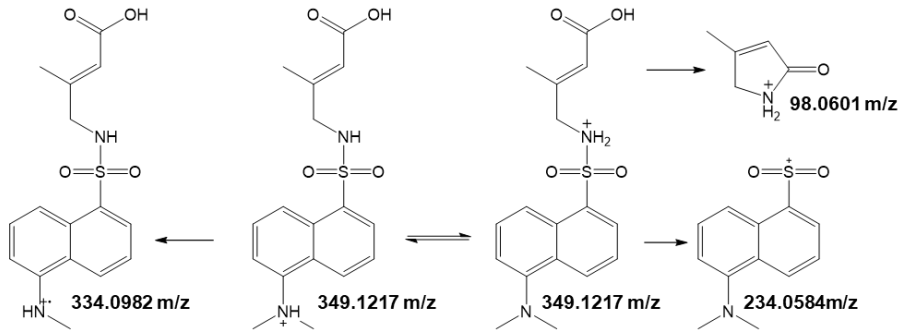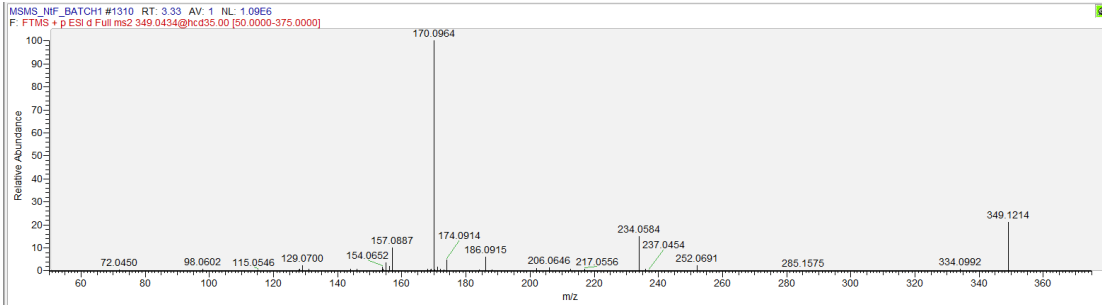
No. 846 Aminoacetone

No. 900 Adenine

136.0618 m/z   369.1129 m/z   369.1129 m/z   305.1150 m/z

No. 923 4-Aminobutyraldehyde

70.0652 m/z

306.1033 m/z   321.1268 m/z   321.1268 m/z   278.1084 m/z
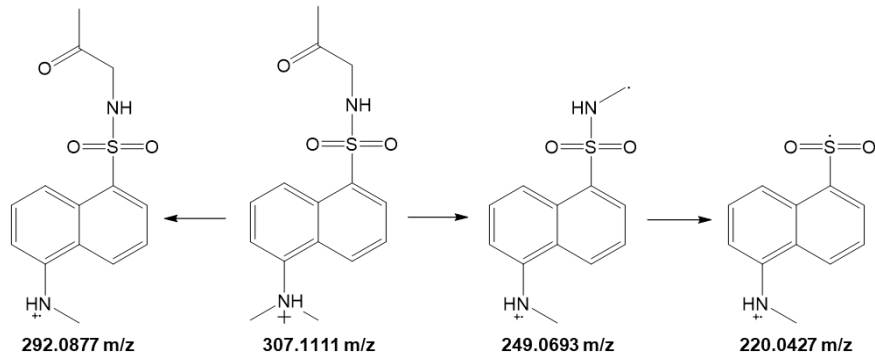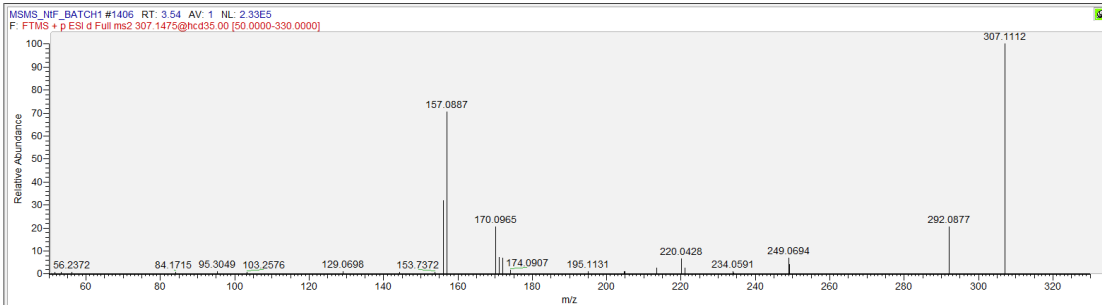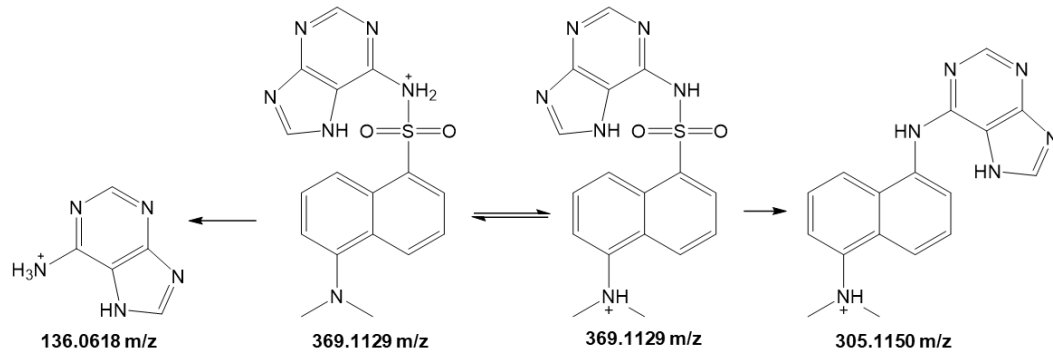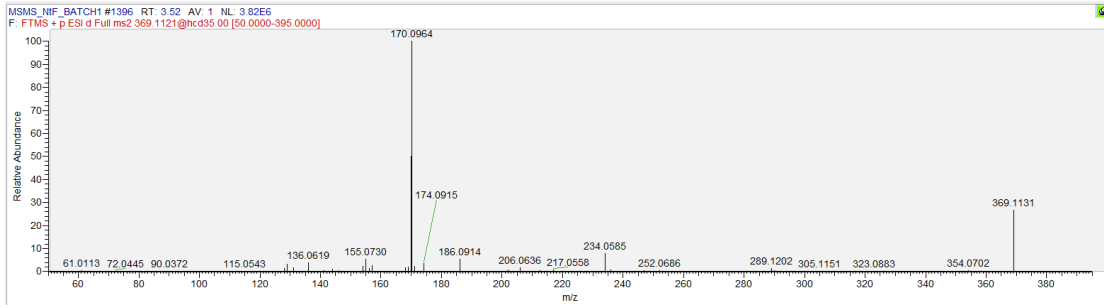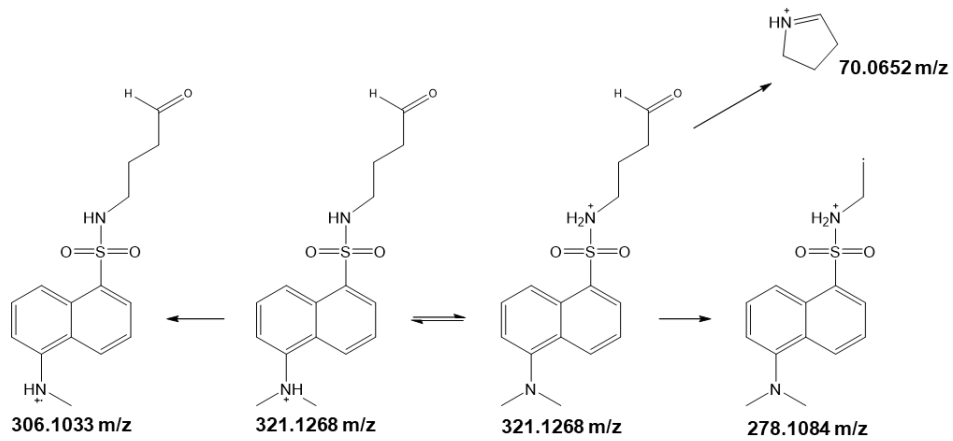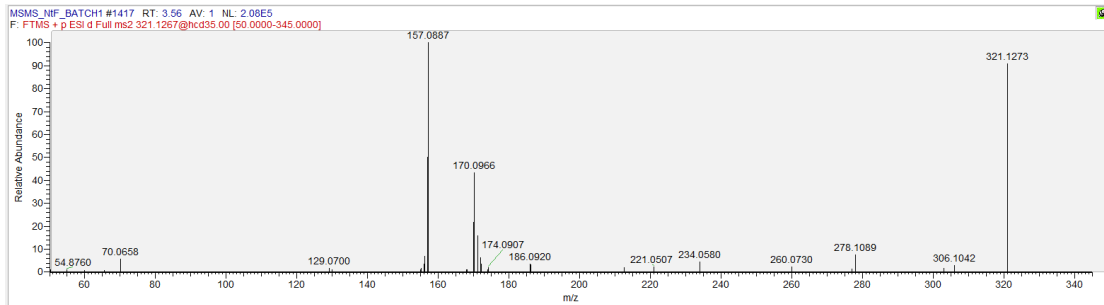
No. 977 4-Aminobutyraldehyde (isomer, it might be 3-aminobutanal)
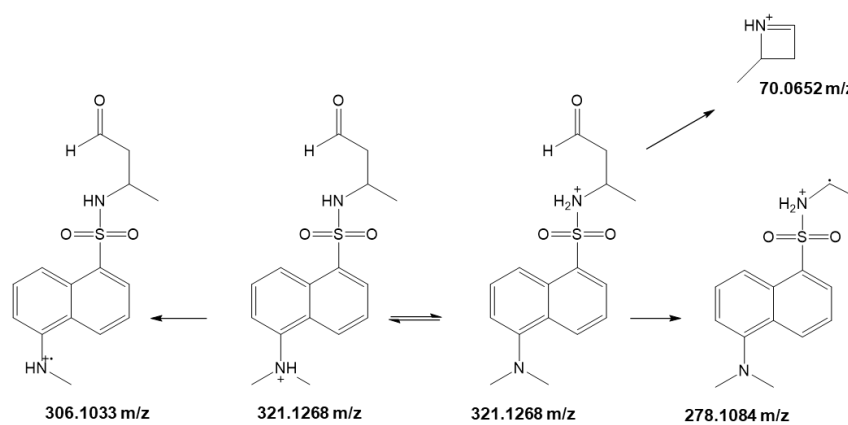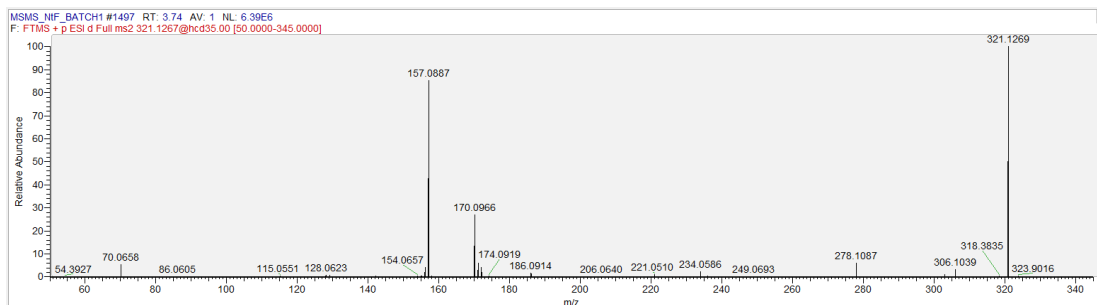
No. 997 4-Methylene-L-glutamic acid
(two precursors available, 393.0944 and 393.1115)
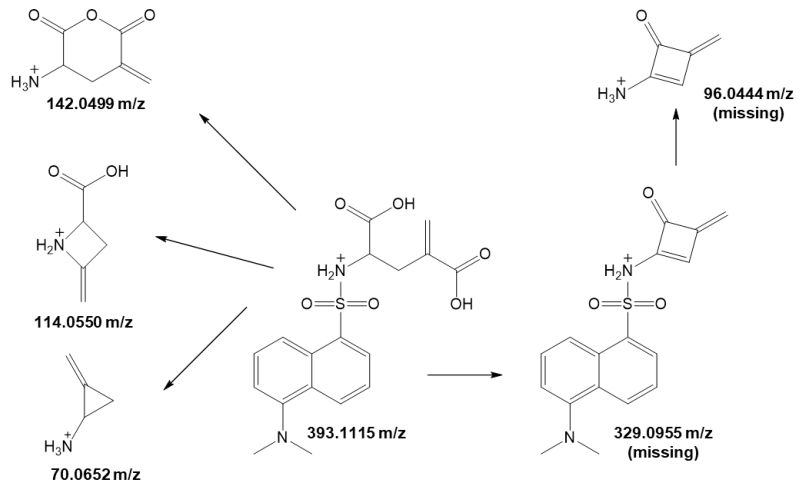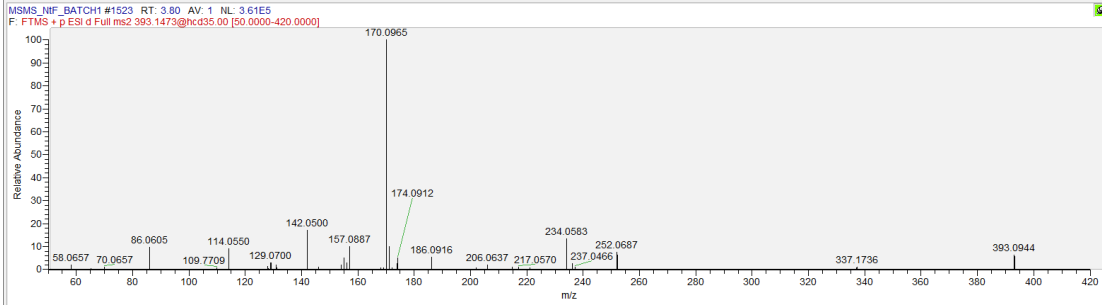
Fragments at 86.0605 were generated from precursors 393.0944 m/z.

No. 1004 Adenine (different labeling position)
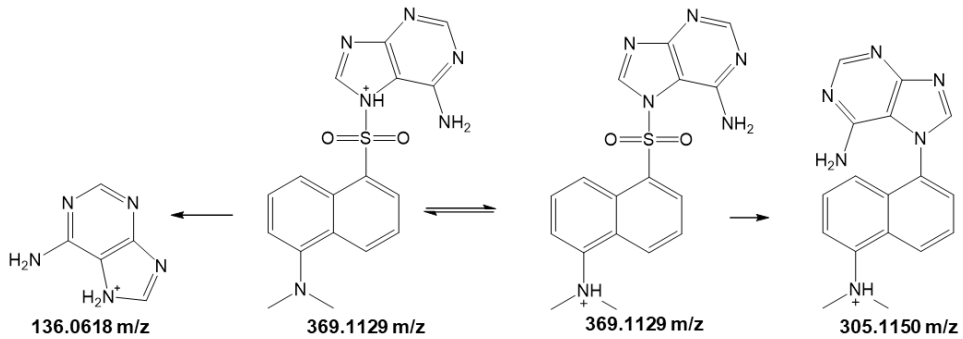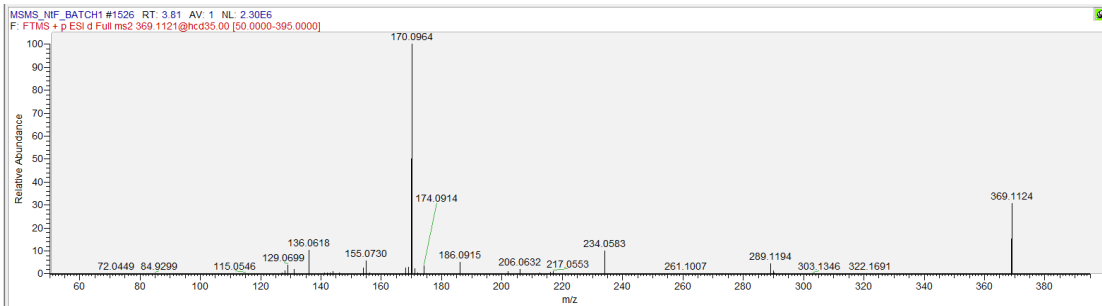
No. 1026 5,6-Dihydroxyindole 5-O-sulfate/5,6-Dihydroxyindole 6-O-sulfate

436.0629 m/z → **383.1061 m/z** → **319.1442 m/z**
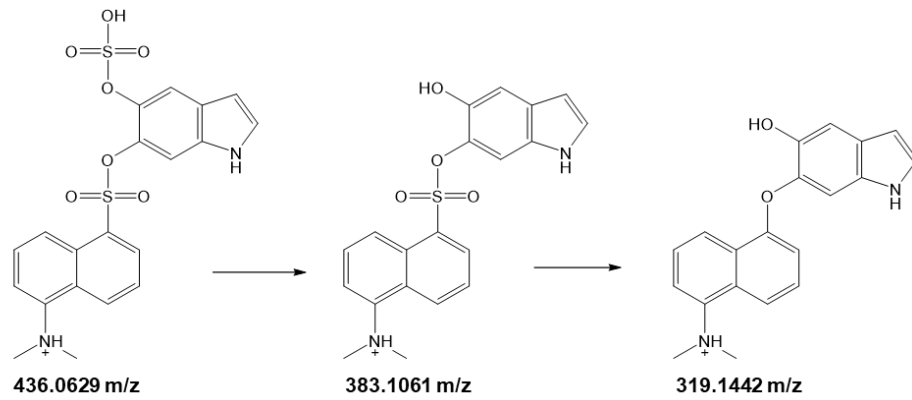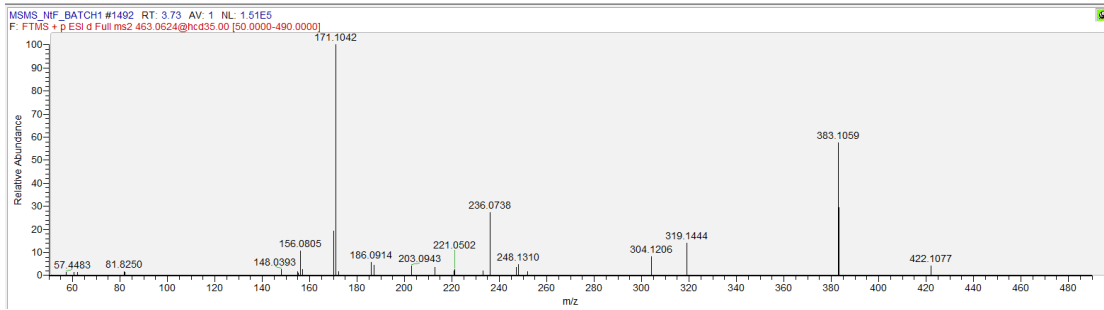
No. 1047 3-Hydroxyanthranilate O-glucuronide

203
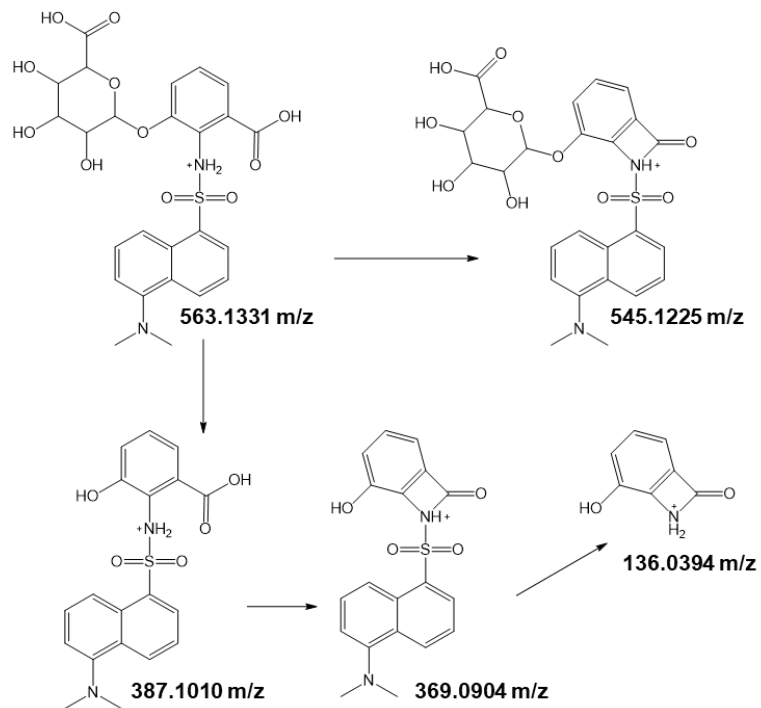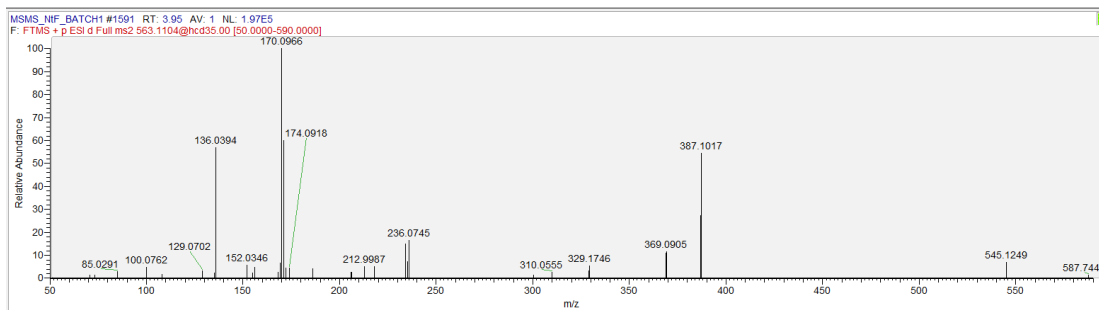
No. 1057 7,8-Dihydroxanthopterin

415.1183 m/z

181.0595 m/z
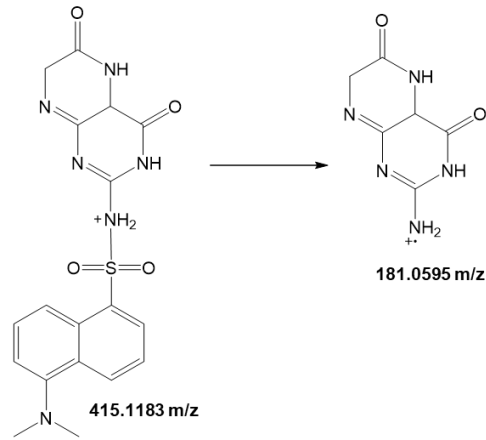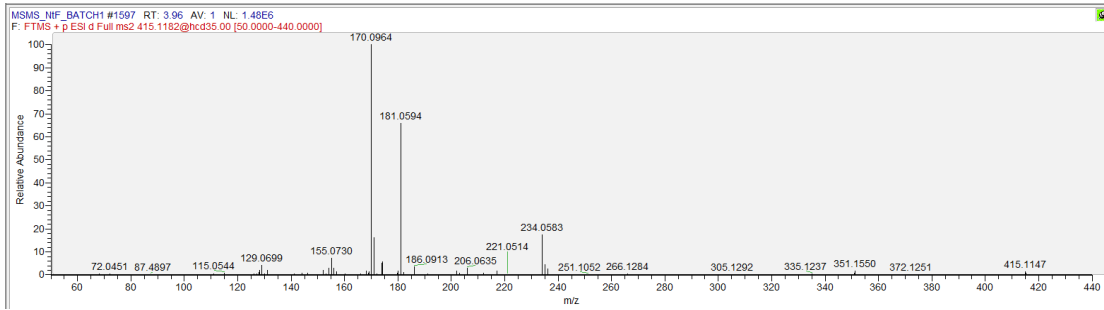
No. 1080 3-Hydroxyanthranilate O-glucuronide (isomer)
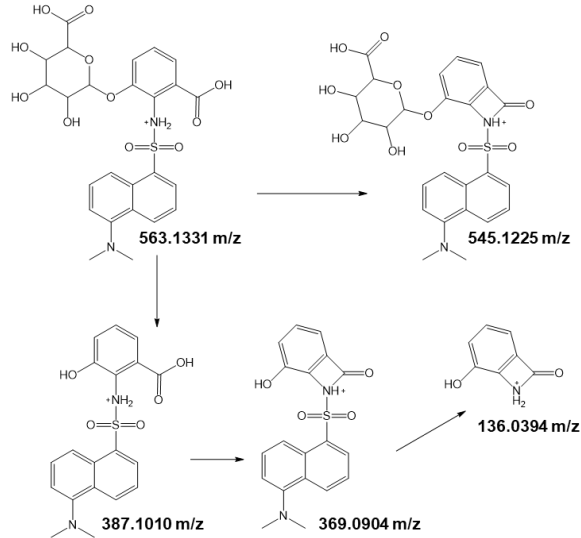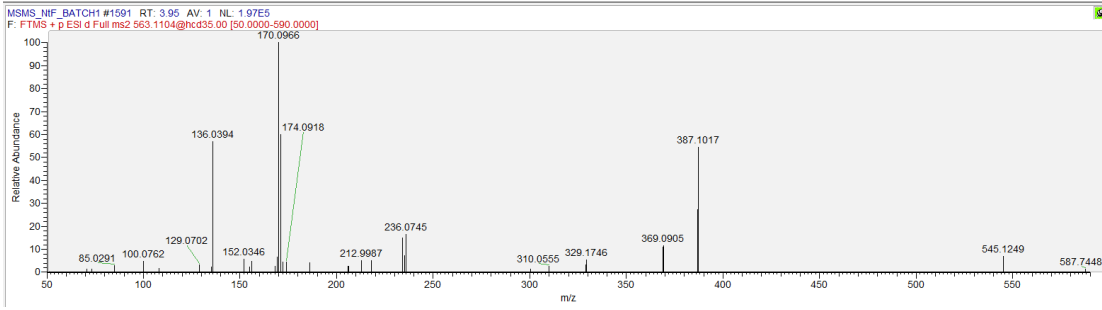
No. 1117 5-Aminopentanal
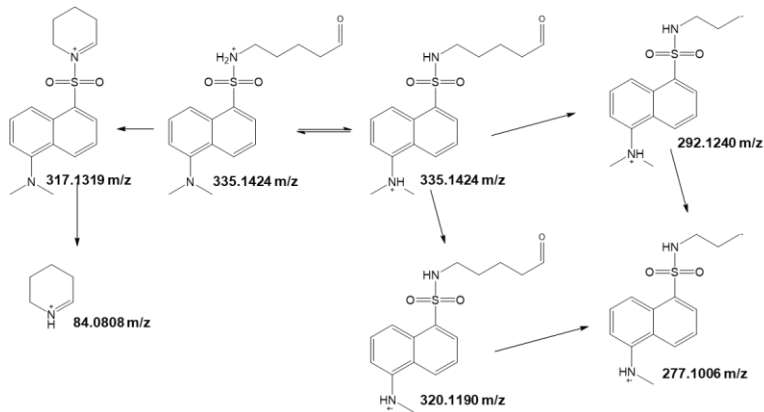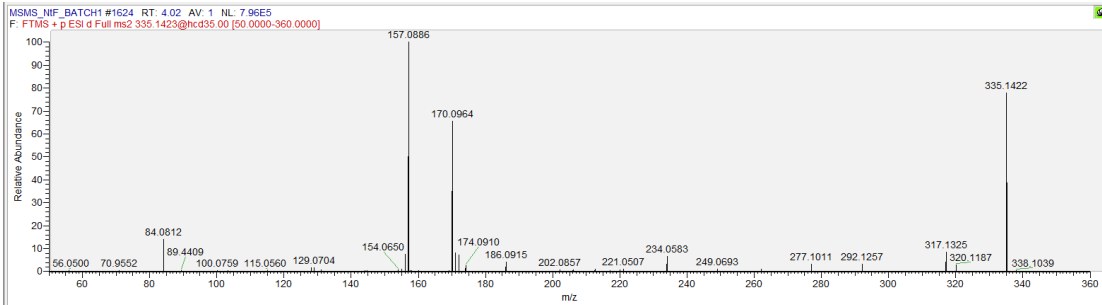
No. 1148 Morpholine
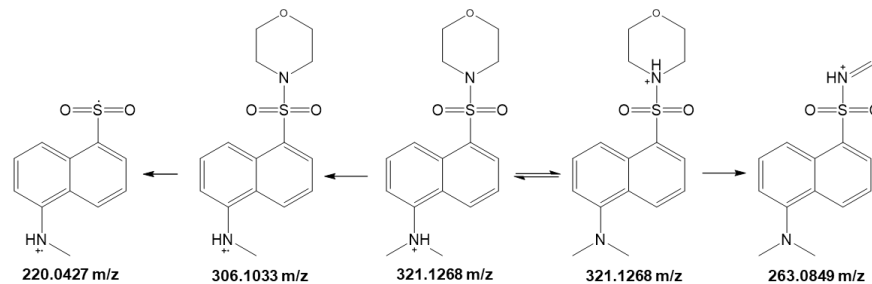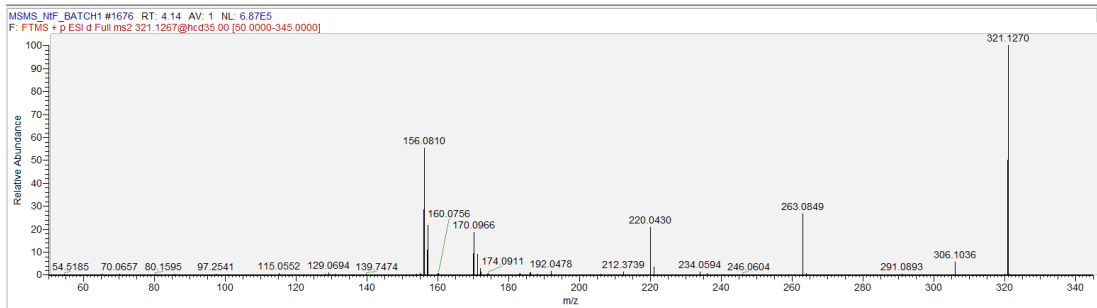
A ring-opening fragmentation was involved in the formation of ions at 263.0849 m/z.

No. 1167 5-Aminopentanal (isomer, it might be 4-aminopentanal)
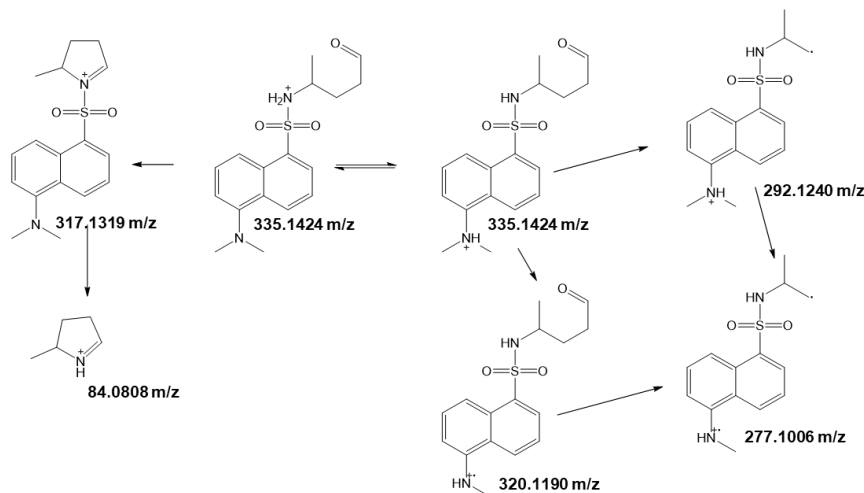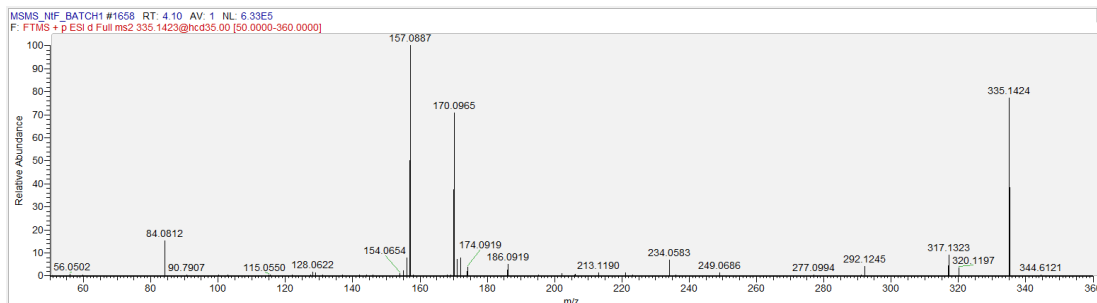
No. 1186 2-Pyrrolidineacetic acid
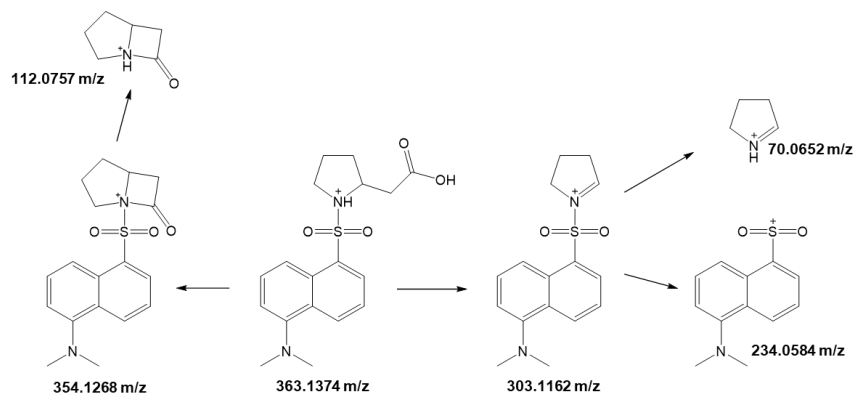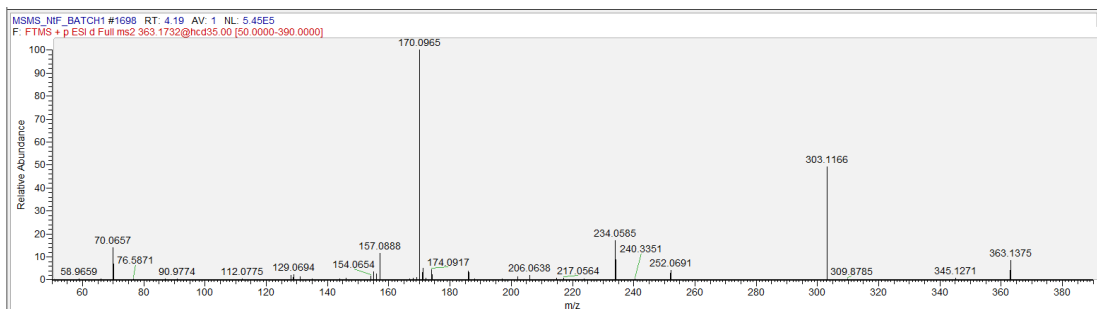
No. 1297 Ethylamine
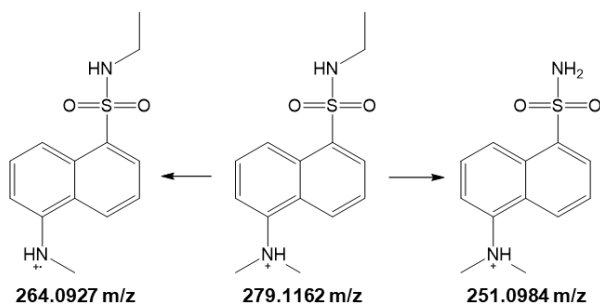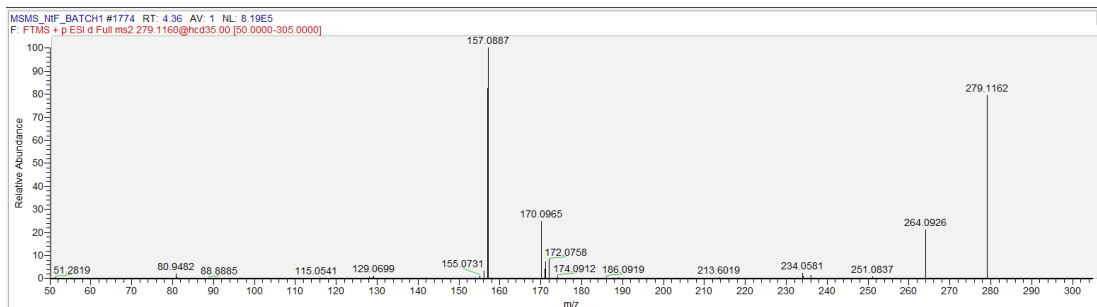
No. 1353 Homocysteine thiolactone
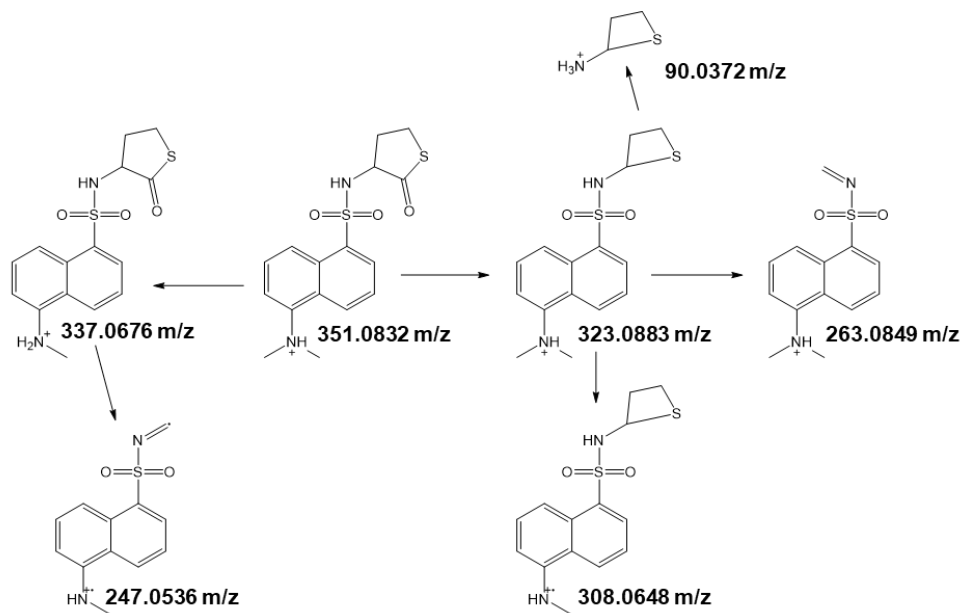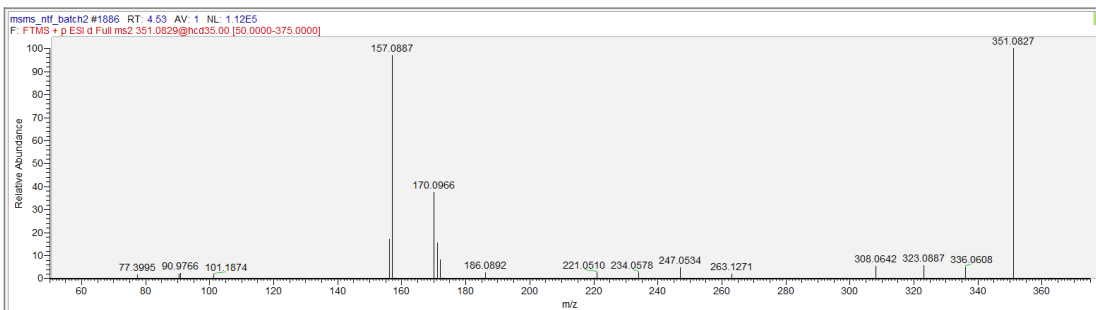
No. 1382 3-Sulfocatechol
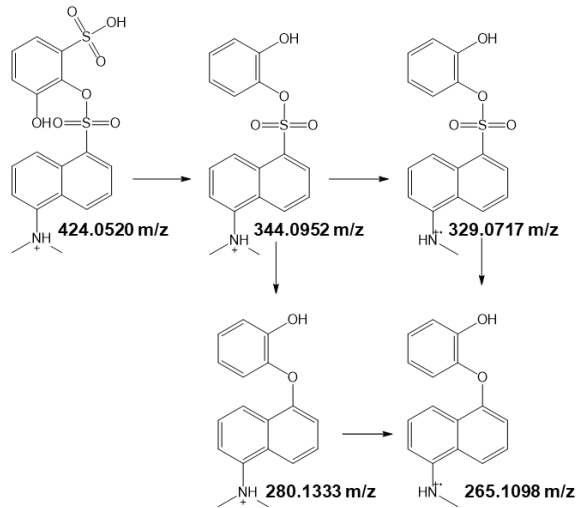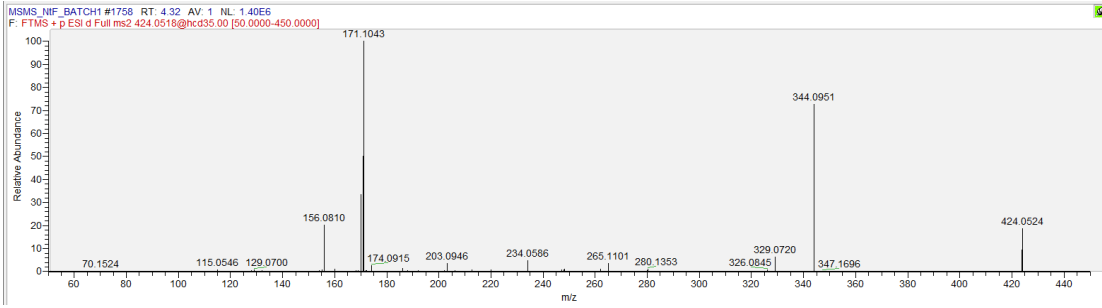
No. 1416 Ethylamine
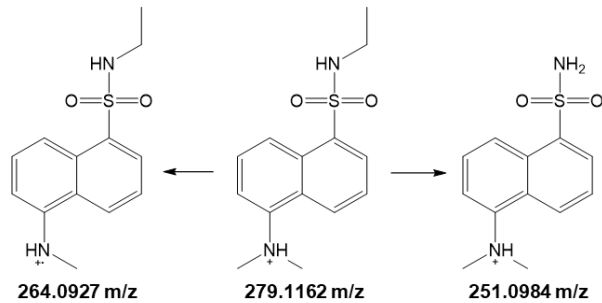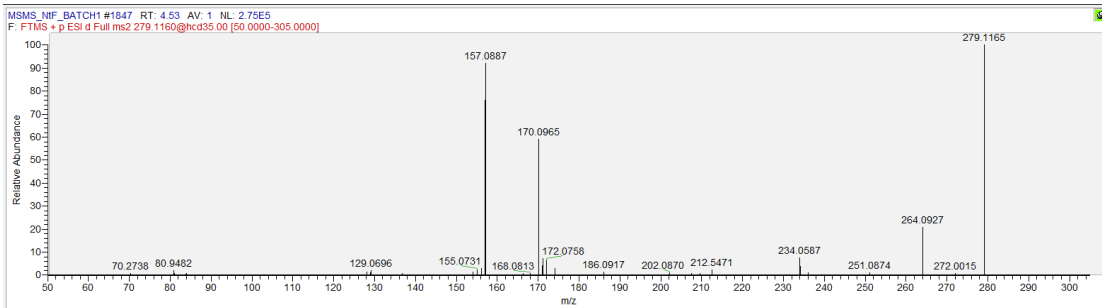
No. 1421 3-Hydroxyhippuric acid
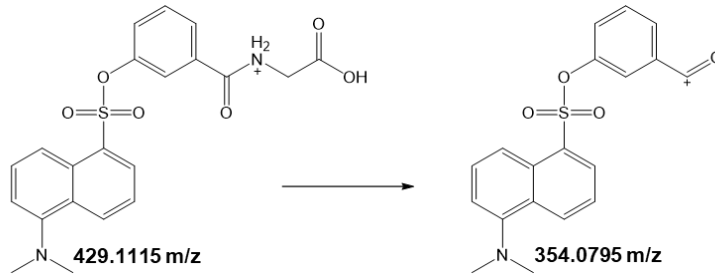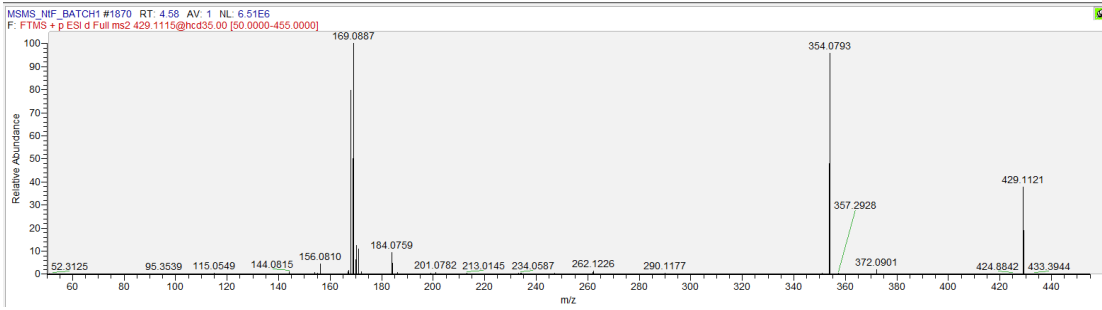
No. 1504 3-Sulfocatechol (different labeling position)
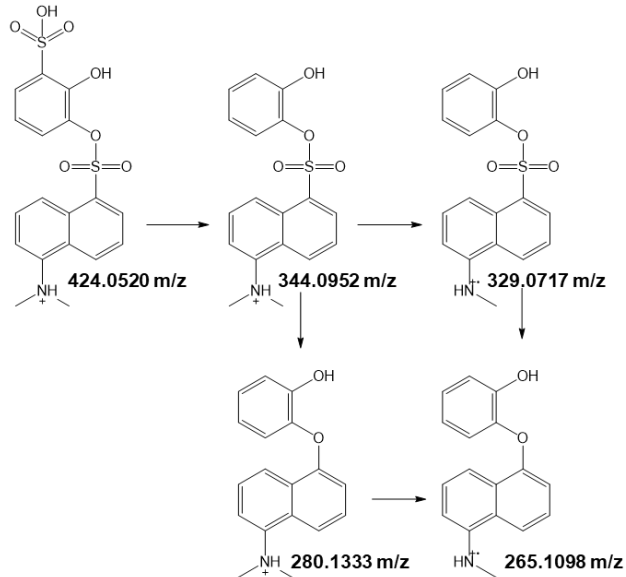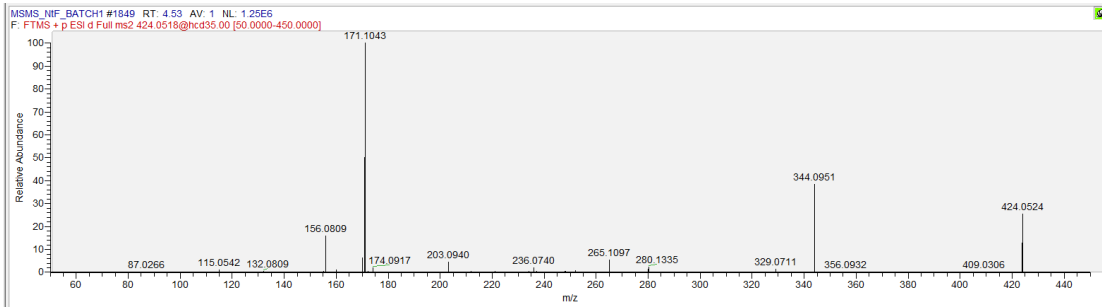
No. 1505 3-(3-Hydroxyphenyl)-3-hydroxypropanoic acid

339.0924 m/z   354.1159 m/z   416.1163 m/z   398.1057 m/z   380.0952 m/z
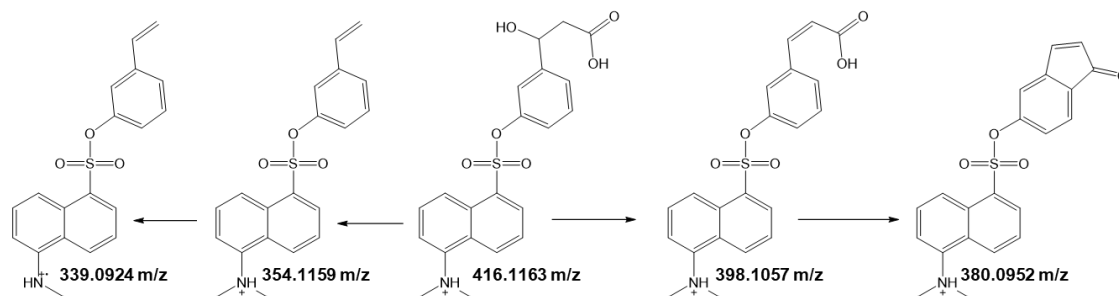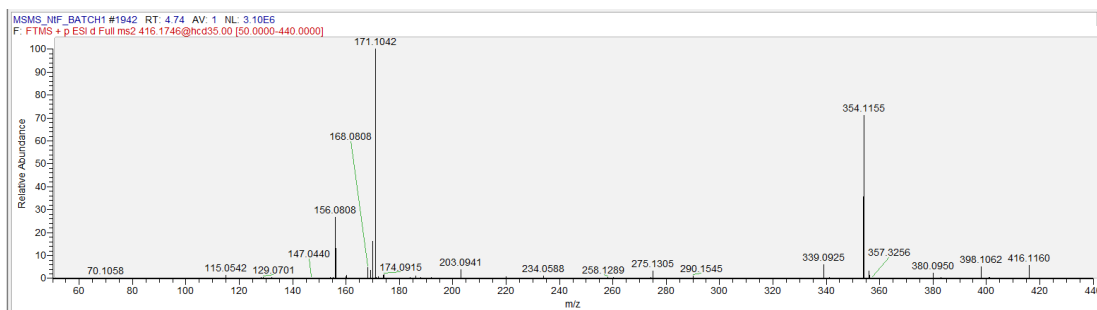
No. 1517 3-hydroxybenzamide

307.1442 m/z   371.1061 m/z   371.1061 m/z   354.0795 m/z
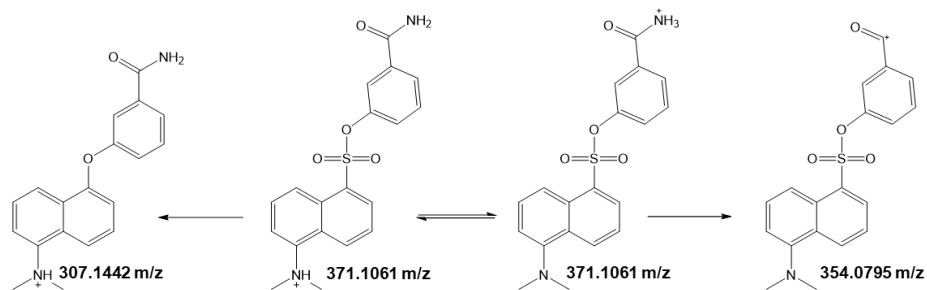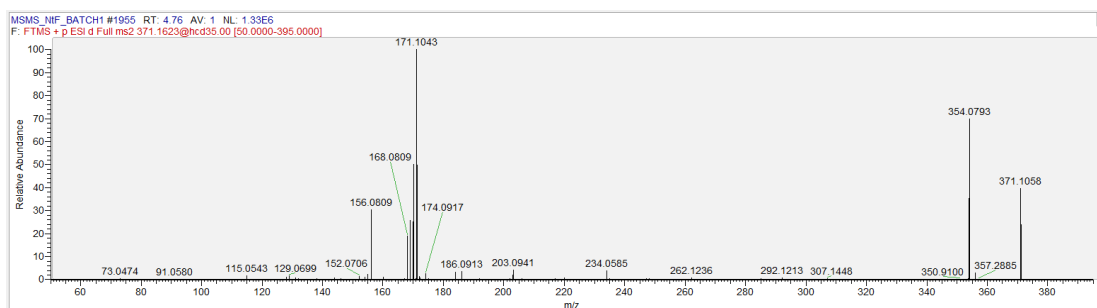
No. 1536 Isomer of L-Phenylalanine/D-Phenylalanine
(MS2 was identical with L-Phenylalanine/D-Phenylalanine)
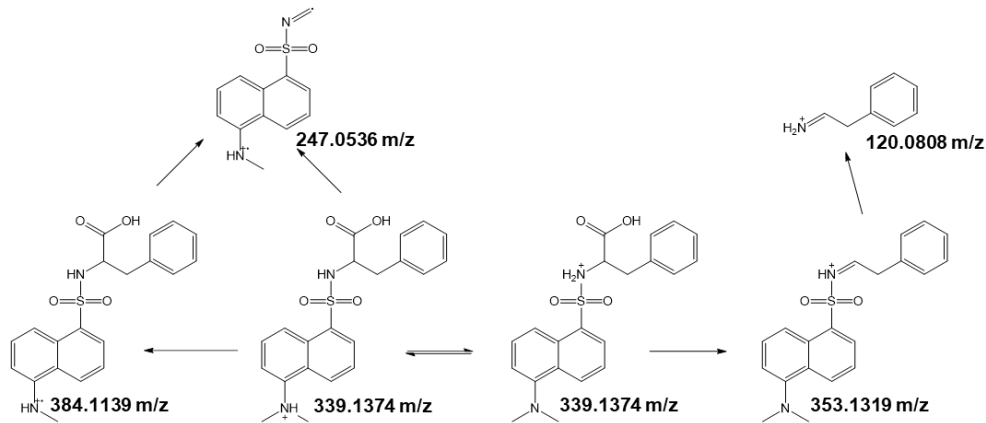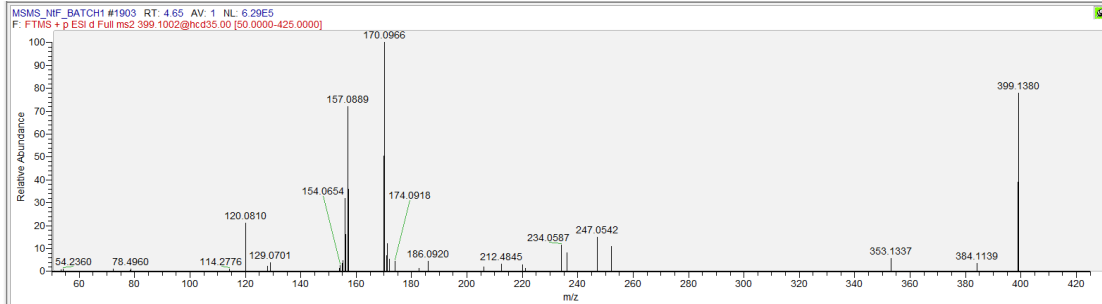
No. 1566 2-Aminobenzyl alcohol
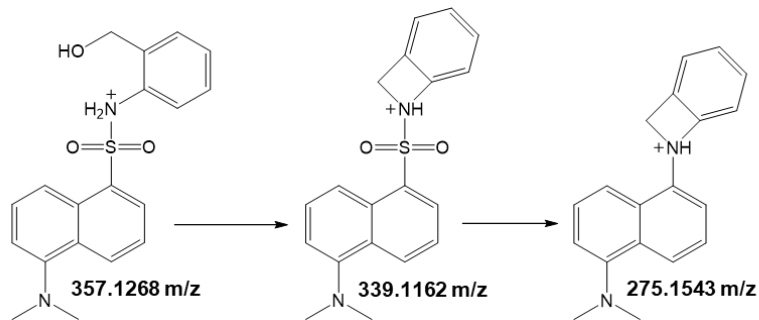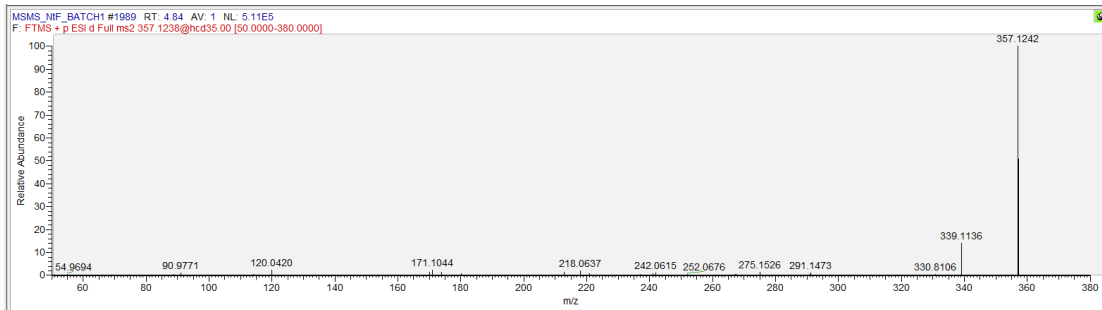
No. 1641 3,4-Dihydroxyphenylacetic acid
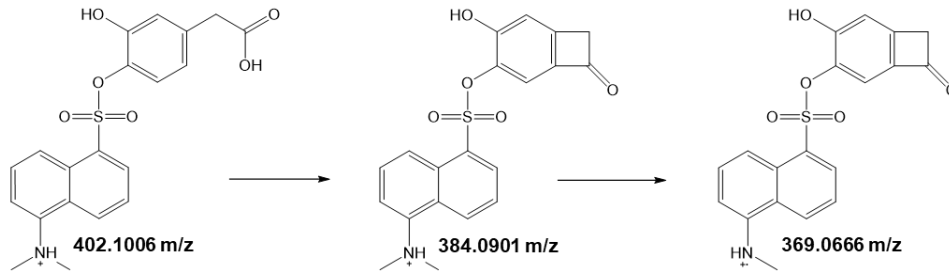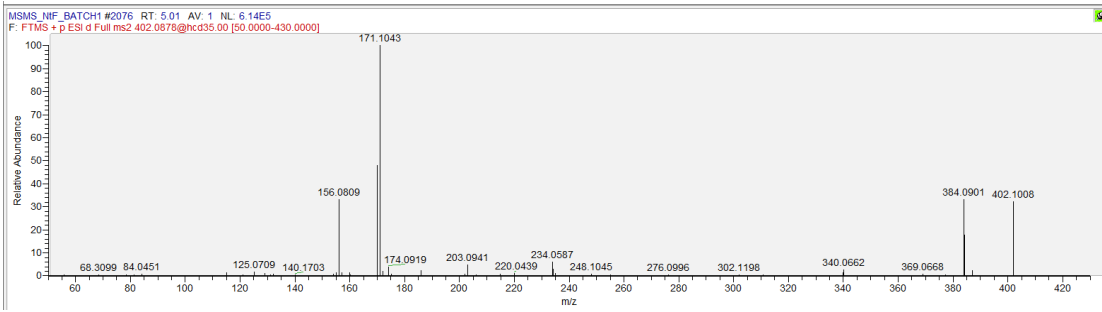
No. 1671 4-Hydroxypiperidine
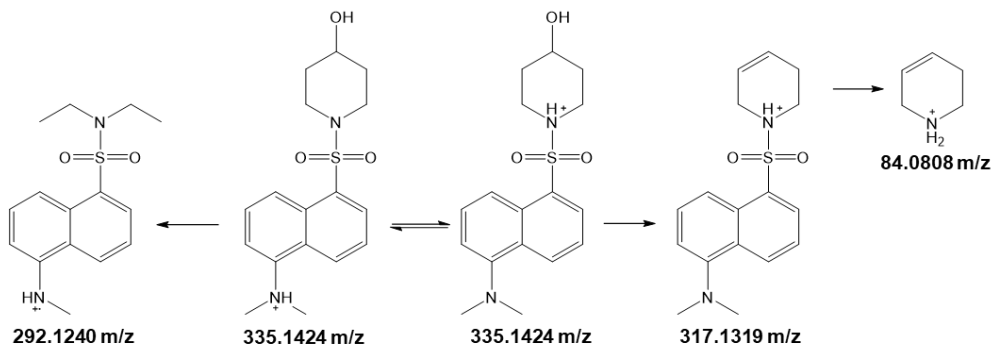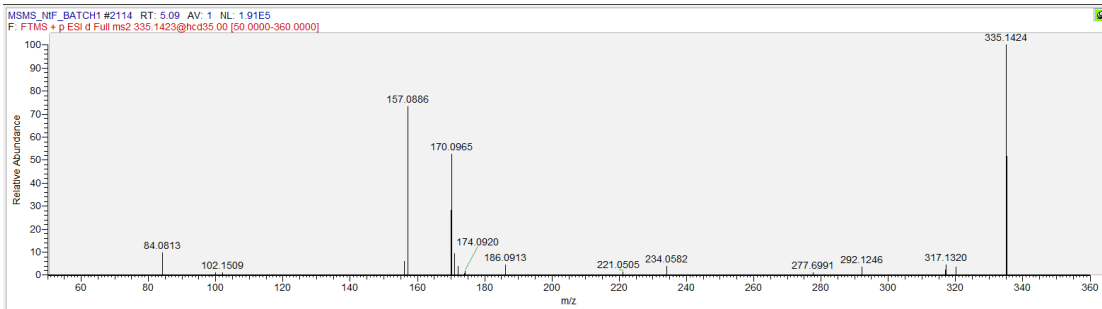
No. 1680 4-Hydroxyindoleacetic acid

381.1268 m/z

425.1166 m/z

379.1111 m/z

146.0601 m/z

315.1492 m/z
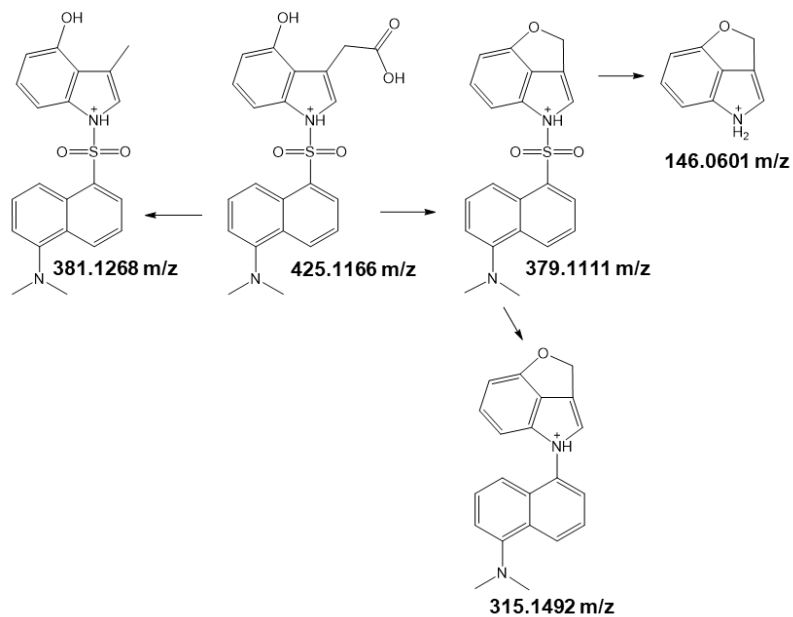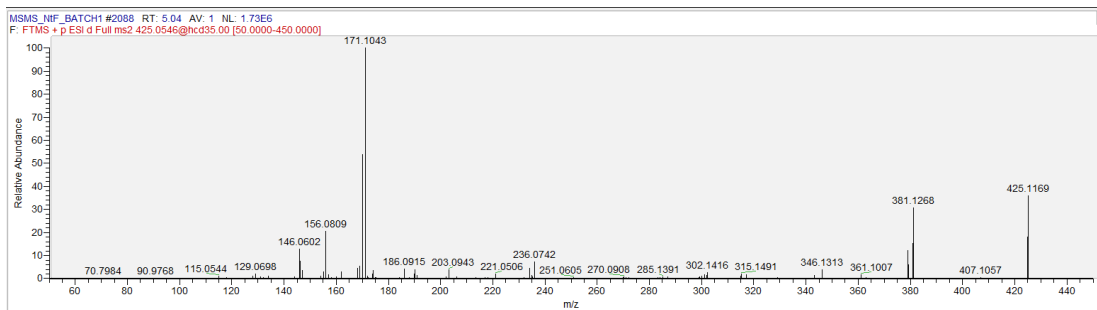
No. 1845 Trihomomethionine

175.0788 m/z

115.0754 m/z

425.1564 m/z

377.1530 m/z

97.0648 m/z

61.0107 m/z

69.0699 m/z

No. 1883 Pyrrolidine

70.0652 m/z

290.1084 m/z

305.1319 m/z

305.1319 m/z
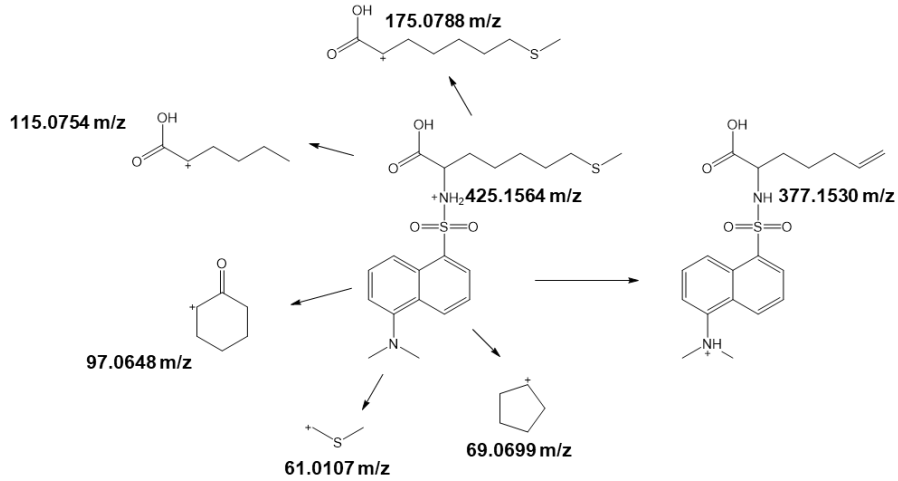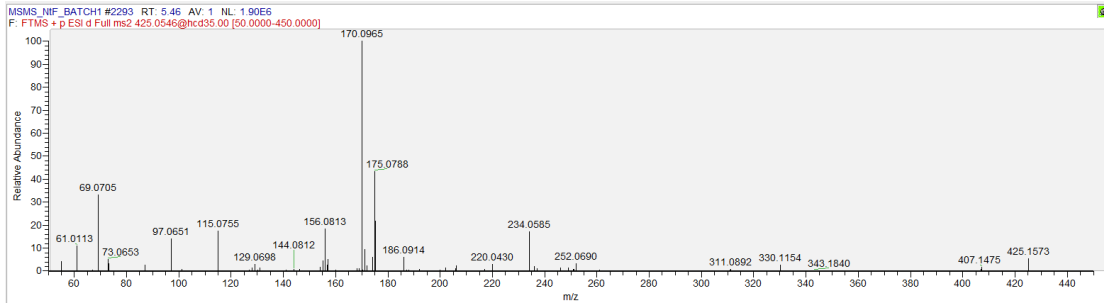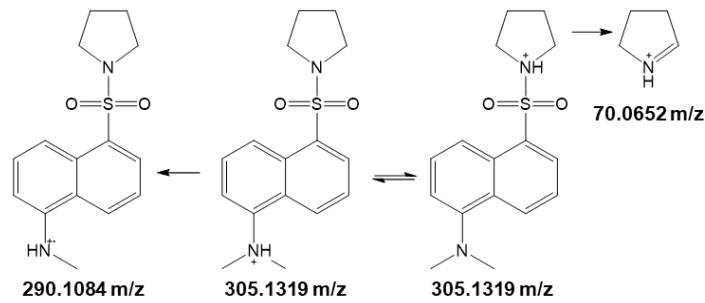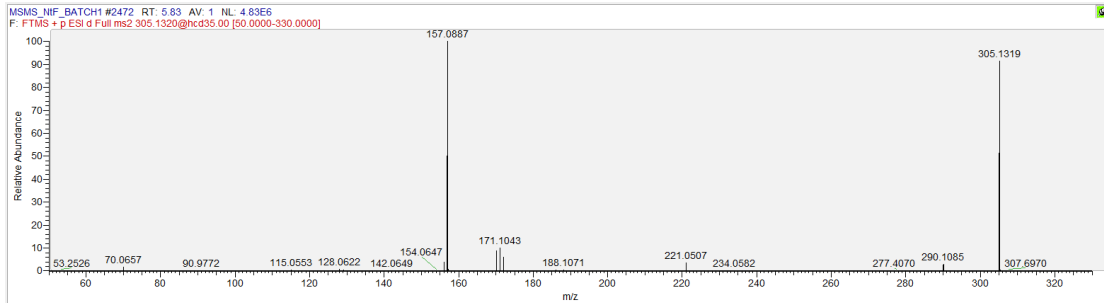
No. 1891 4-Pyridoxolactone
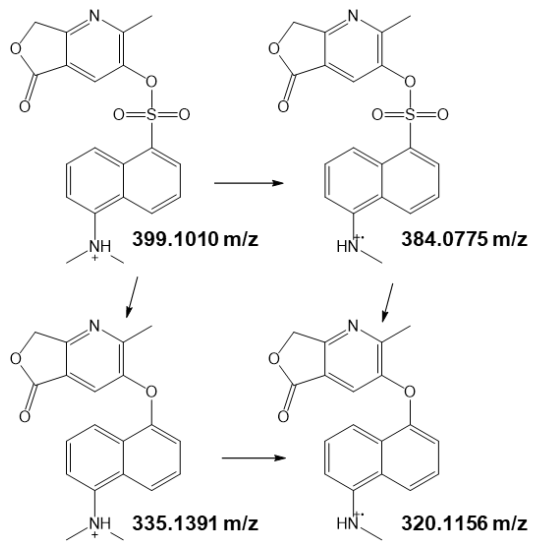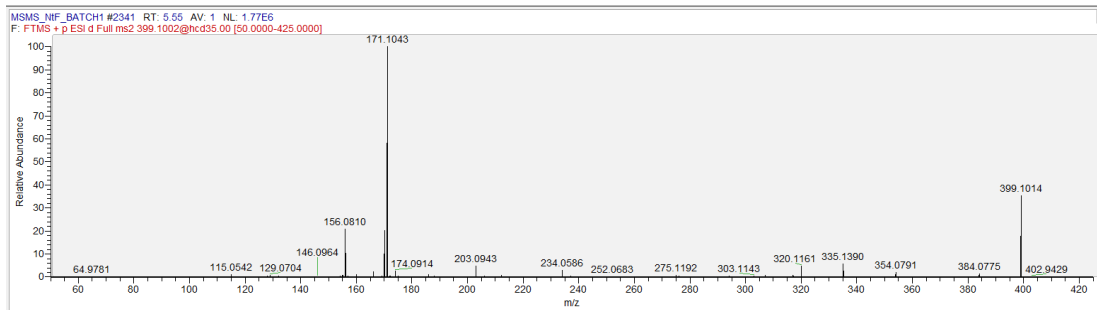
No. 1983 Diethylamine

72.0808 m/z

307.1475 m/z   307.1475 m/z   292.1240 m/z

No. 2483 2-Methylpiperidine

98.0965 m/z

263.0614 m/z   318.1397 m/z   333.1632 m/z   333.1632 m/z

No. 2483 2-Methylpiperidine

98.0965 m/z

263.0614 m/z   318.1397 m/z   333.1632 m/z   333.1632 m/z
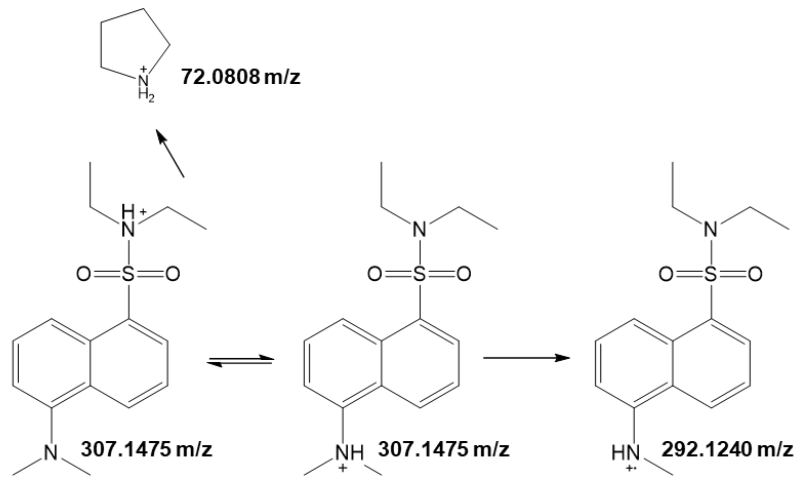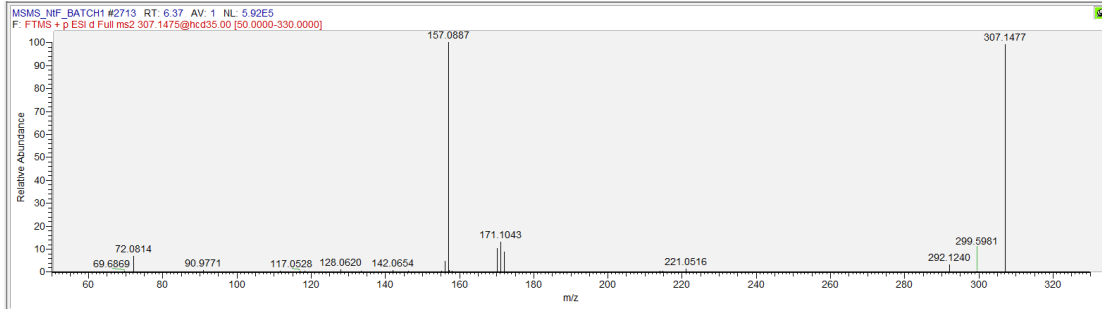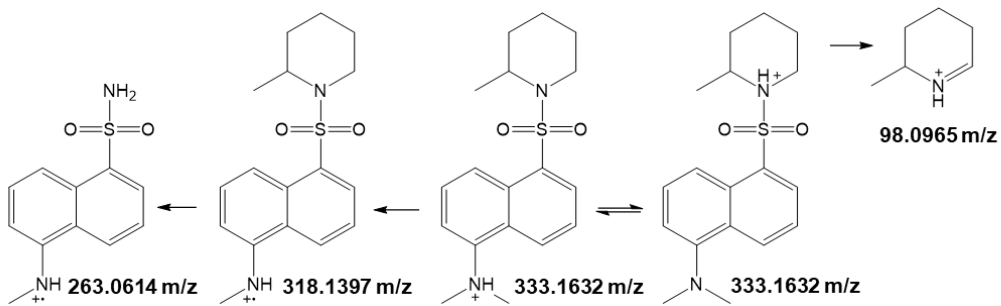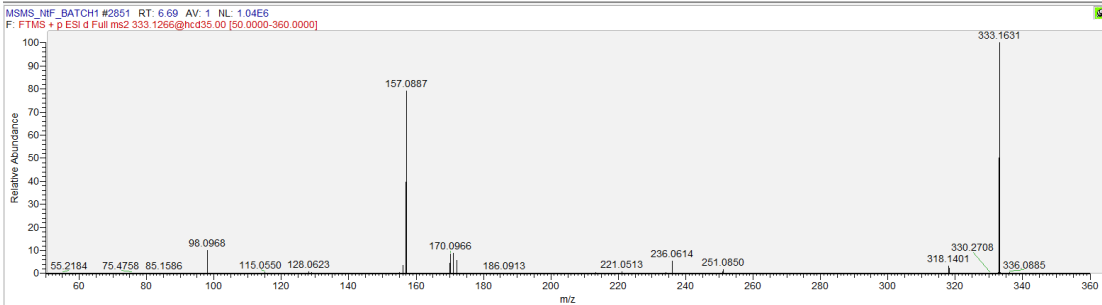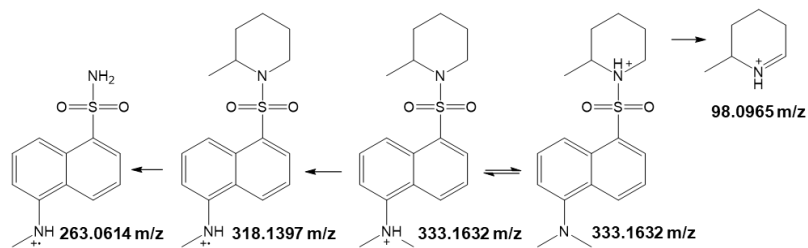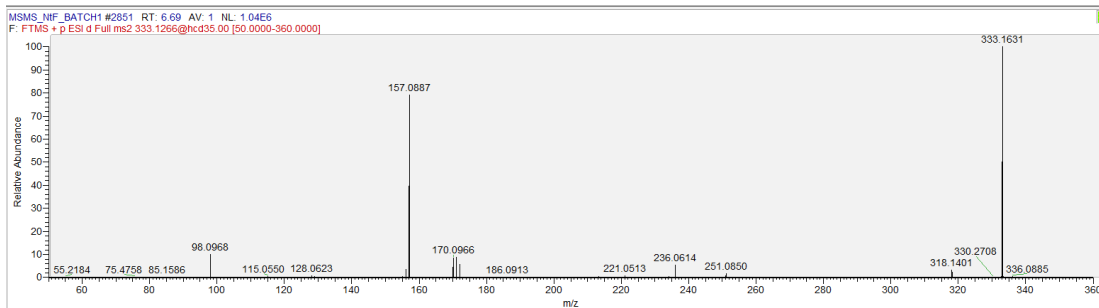
No. 2960 gamma-CEHC
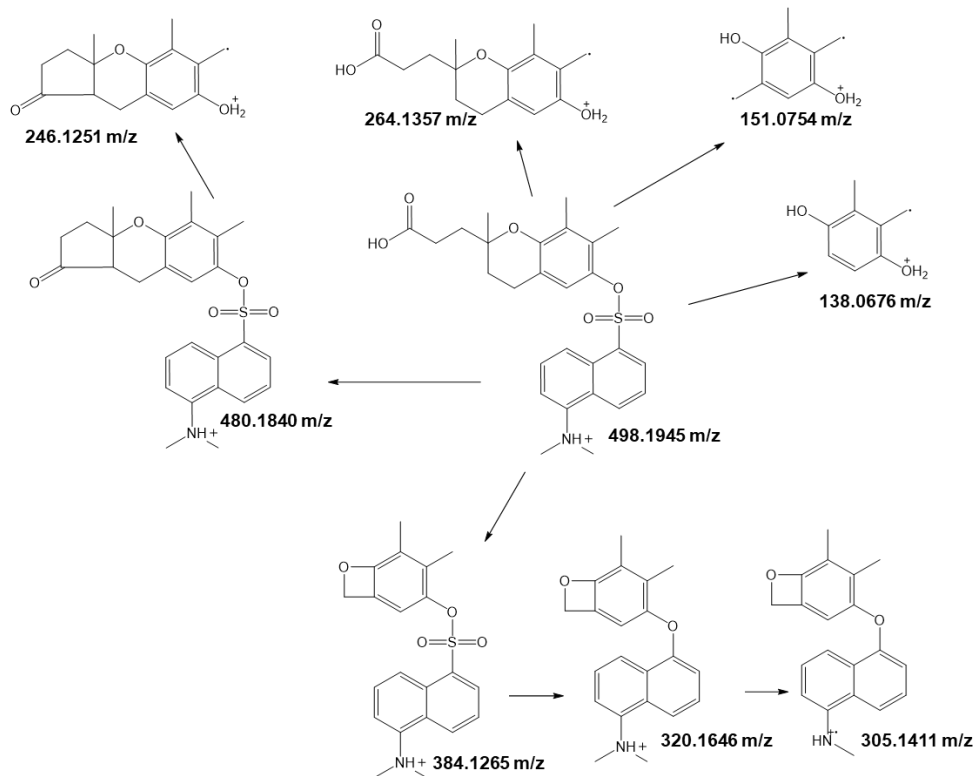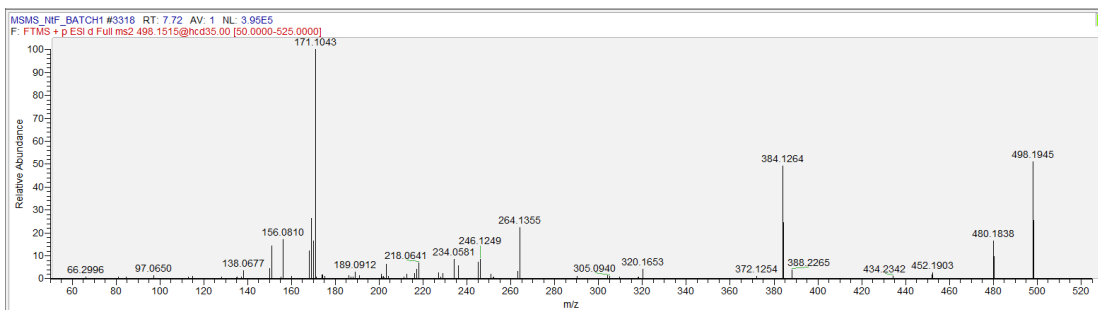
No. 3103 alpha-CEHC

No. 3228 2-Naphthol

378.1159 m/z   314.1540 m/z   229.1305 m/z
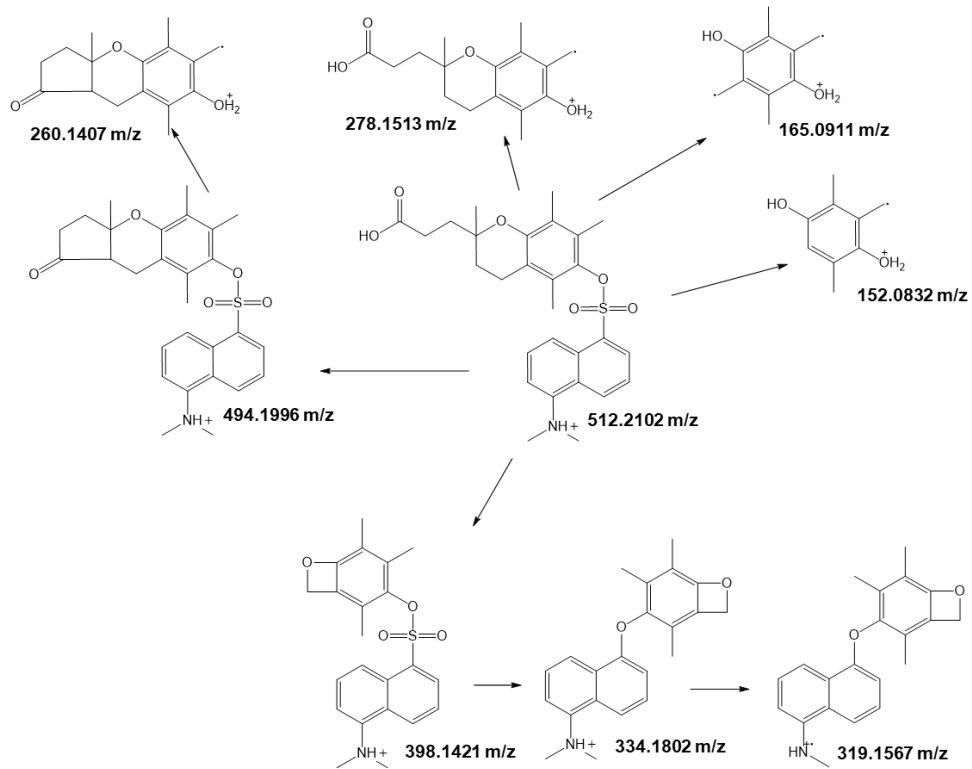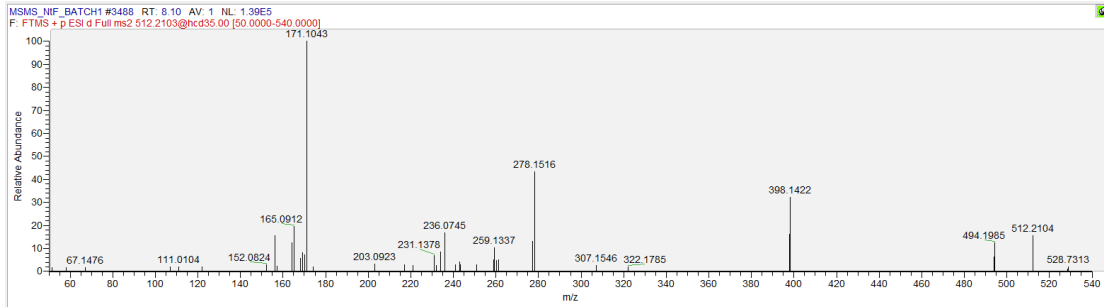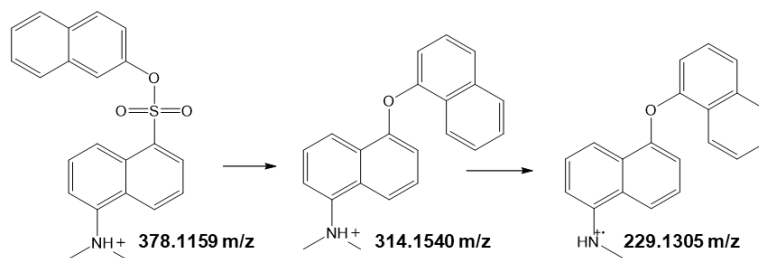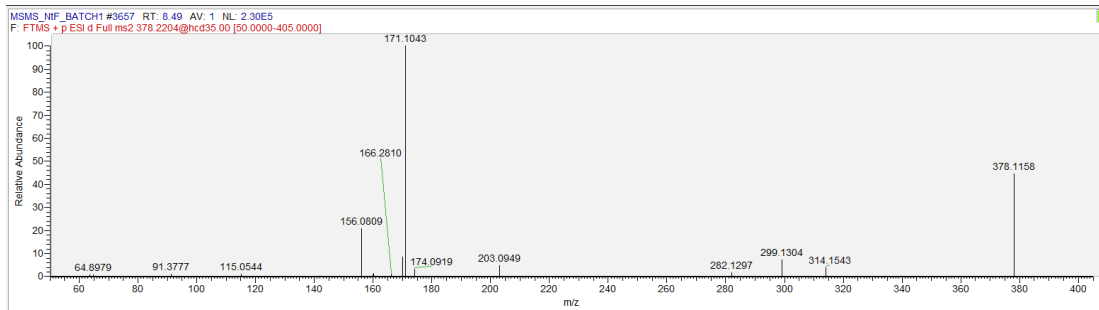
Appendix B. MS/MS spectra of 10 tier-2 metabolites that did not pass

MS/MS-based identification validation.

No. 137 N6-Methyladenosine (False identification, fragmentation patterns displayed in the collected MS/MS were different from Dns-labeled N6-Methyladenosine)
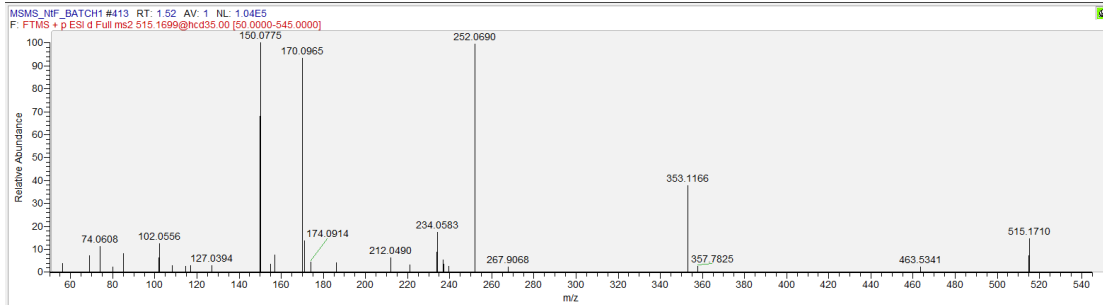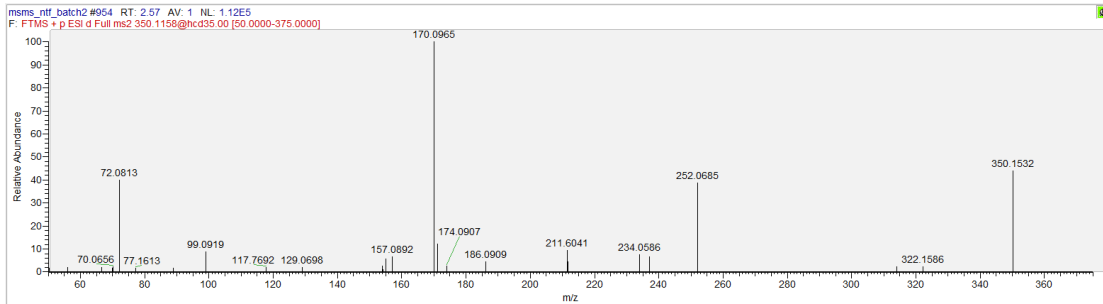

No. 226 Creatinine (False identification, no fingerprint specific to creatinine was found) (Two precursor ions available, 347.1170 and 347.2216 m/z)
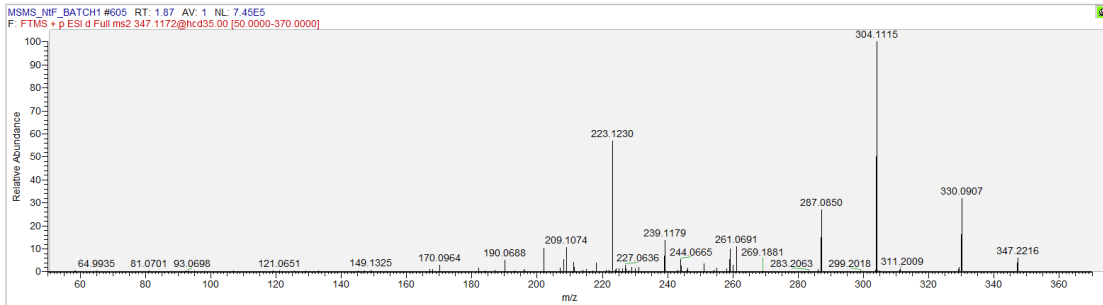
No. 357 4-Guanidinobutanal
(False identification, fragmentation patterns displayed in the collected MS/MS were different from Dns-labeled 4-Guanidinobutanal, it was not supposed to see a neutral loss of H_2O molecule)
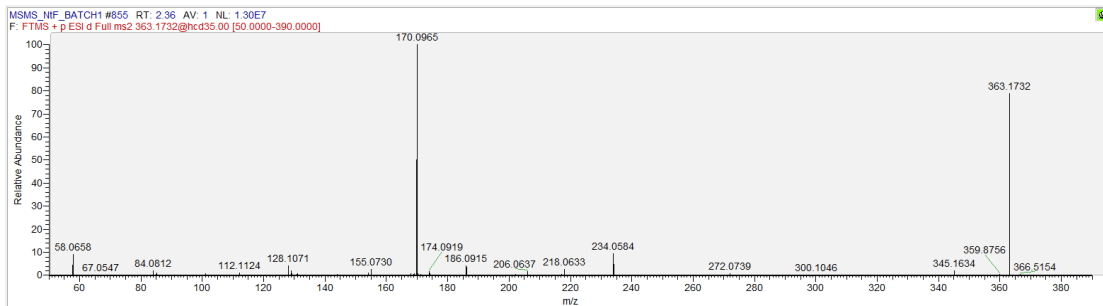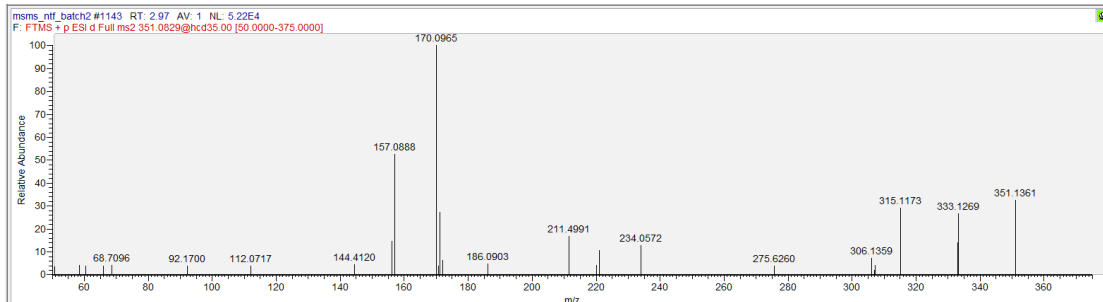
No. 435 5-Aminopentanamide (False identification, ions 252.0685 were attributed to a hydroxyl/carboxyl-containing metabolite)
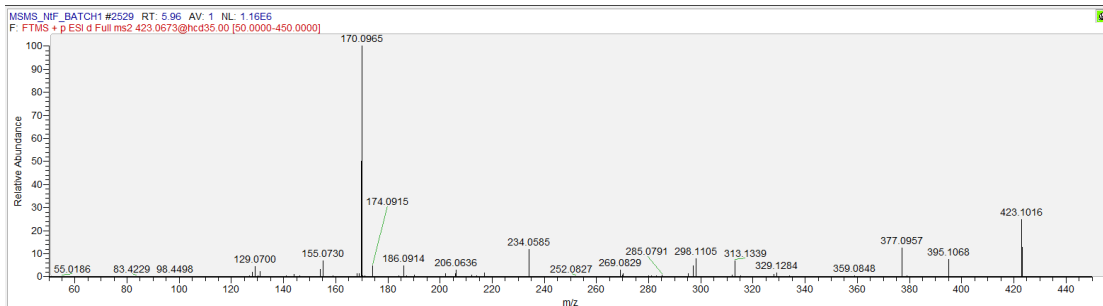

No. 519 Creatine (False identification, no fingerprint specific to creatinine was found)
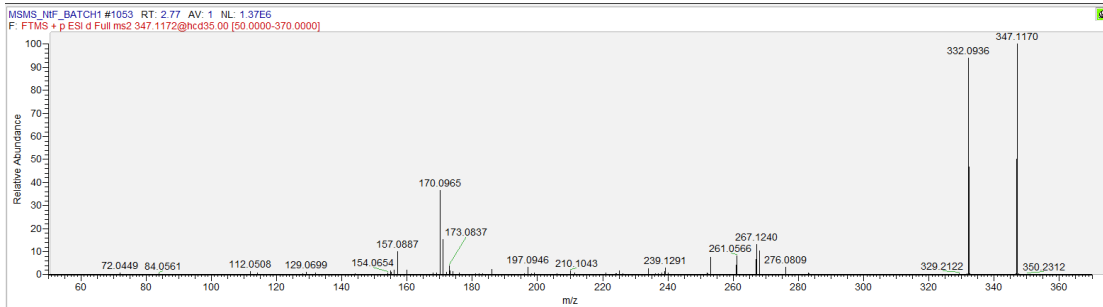
No. 648 Isomer of 5-Aminopentanoic acid (No ions 252 m/z were found, two neutral losses of -OH group were present in the collected spectra)
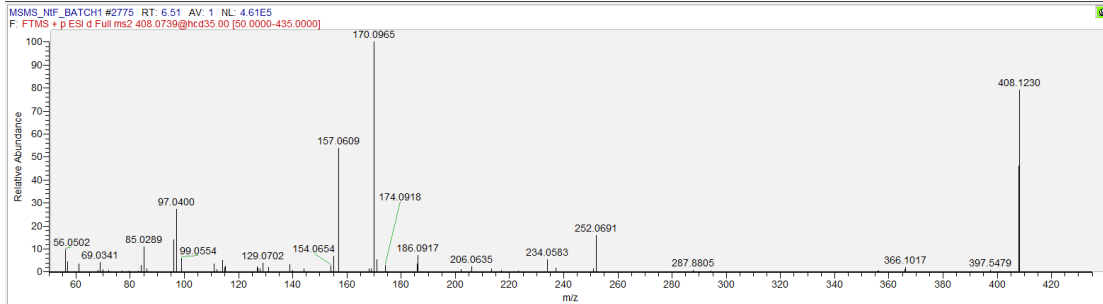

No. 2111 alpha-Cyano-4-hydroxycinnamic acid
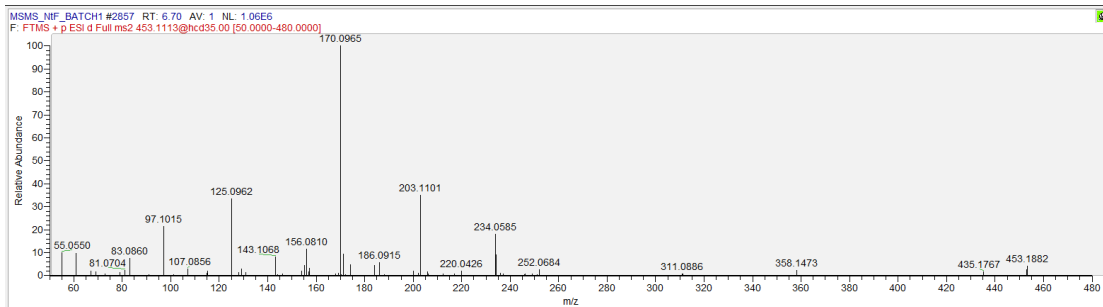(False identification, ions 423.1016 were attributed to an amine-containing metabolite)

No. 2388 1,2-Dihydroxy-8-methylnaphthalene

(False identification, ions 408.1230 were attributed to an amine-containing metabolite)

No. 2488 Ritalinic acid

(False identification, ions 453.1882 were attributed to a phenol-containing metabolite)

No. 2534 4-Aminobiphenyl

(False identification due to false mass match, theoretical m/z of Dns-4-aminobiphenyl should be 403.1402)