AN ABSTRACT OF THE THESIS OF

Ashish Vaswani for the degree of Doctor of Philosophy in Chemistry presented on July 30, 2021.

Title: Metabolomics in Conjunction with Computational Methods for Supporting Biomedical Research: to Improve Functional Resilience in Age-related Disorders.

Abstract approved: _____

Claudia S. Maier

Metabolomics and lipidomics lay the foundation of personalized medicine. The technological advancements in mass spectrometry techniques in combination with computational algorithms and methods have enabled the study of small molecules (metabolites and lipids) for understanding the disease state and biological pathways, the identification of biomarkers and the generation of predictive models for patient management, as well as the identification of natural products as new leads in drug discovery research. The computational methods utilize large, complex datasets to gather insights about underlying biological processes, trends, and non-random patterns.

This dissertation focuses on research studies in which the integration of metabolomics with computational methods enabled the discovery of active natural products as leads to combat Alzheimer's disease, utilization of metabolomics and lipidomics workflows for utilization of optimal cutting temperature compound stored heart tissues with mass spectrometry and assess the effect of doxycycline on biochemical pathways associated with breast cancer.

The methodical pipeline and associated workflows and technologies is described in Chapter 2. The optimal design of the preanalytical workflows as well as the integration with the appropriate measurement technologies are important for successful metabolomics and lipidomics studies. In this thesis, pre-analytical workflows were developed and applied for sample extraction procedures to separate metabolites and lipids from tissues, cells and botanical extracts, and subsequent chromatographic separation with ultra-performance liquid chromatography (UPLC). The metabolite and lipid profile were detected using high-resolution mass spectrometry in conjunction with tandem mass spectrometry. For characterization of isomers travelling wave ion mobility mass spectrometry was utilized. Metabolomics and lipidomics approaches were enhanced by computational methods for data processing, data visualization and interpretation.

In this thesis, we developed LC-MS metabolomics approaches for the characterization of botanical extracts and applied and evaluated innovative bio-chemometrics approaches to assign the bioactive principles. The natural products research in the thesis focused on Centella asiatica botanicals have gained popularity for their potential to enhance cognitive function and brain vitality in aging. An important contribution to improve effective clinical trials is the availability of standardized *Centella asiatica* extracts to facilitate reproducible use to account for substantial variability across natural products using LC-MS/MS workflow. A secondary goal in this thesis was method development aimed to reduce reliance on bioactivity guided fractionation by combining flow injection mass spectrometry with innovative computational methods that allow rapid dereplication of natural products and assigning of bioactive natural products. The methodological pipeline in conjunction with the applied computational approaches will lead to a decrease in time needed for moving bioactive natural products to preclinical testing.

In another study, methodology was evaluated to allow the use of bio-banked heart tissue samples for subsequent biomarker discovery research. The research on optimal cutting temperature (OCT) embedded heart tissue was designed to determine the compatibility of OCT storage with UPLC-MS/MS lipidomics studies. The results show that OCT stored heart tissue is compatible with LC-MS/MS lipidomics - facilitating the use of bio-banked

tissue samples for future studies. The critical evaluation of the developed workflow shows that LC-MS/MS lipidomics of OCT-banked tissues samples is reliable for the major lipid classes except for plasmalogens that would likely be underestimated with using the described protocol.

In the last research chapter in this thesis outlines studies designed to determine if a doxycycline (DOX)-dependent gene expression knockdown system is a viable strategy in the context of metabolomic studies of breast cancer cells for studying the biological effects of targeted gene silencing. This research utilized a workflow comprising of combination of NMR and mass spectrometry. NMR was utilized to identify polar metabolites. Hydrophilic interaction liquid chromatography was used in conjunction with MS/MS mass spectrometry to determine the effect of doxycycline on metabolites. Reversed phase ultra-performance liquid chromatography was utilized along with MS^E mass spectrometry to assess the impact of doxycycline on lipids. The research indicated DOX-based gene expression knockdown strategies unexpectedly affected metabolic pathways in the breast cancer cell lines. This serves as a cautionary tale for use of doxycycline in gene silencing in metabolomics and lipidomics experiments.

The conclusion of thesis provides a summary of the insights obtained by using computational methods in metabolomics. It provides perspectives on future of patient management, discovery of compounds with potential for treatment of diseases, obtaining in-depth understanding of disease state using mass spectrometry-based metabolomics and lipidomics.

Metabolomics in Conjunction with Computational Methods for Supporting Biomedical Research: to Improve Functional Resilience in Age-related Disorders.

by

Ashish Vaswani

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented July 30, 2021

Commencement June 2022

Doctor of Philosophy thesis of Ashish Vaswani presented on July 30, 2021.

APPROVED:

Major Professor, representing Chemistry

Head of the Department of Chemistry

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Ashish Vaswani, Author

ACKNOWLEDGEMENTS

I am sincerely grateful to my advisor Professor Claudia Maier, for providing me opportunities to discover and pursue new ideas, presenting results on multiple occasions, supporting me and being flexible with schedules and mentoring me to develop scientific attitude and managing projects. I would like to thank her for her contribution with novel ideas, involving me in multiple projects aligned with my interest in improvement of human health and her patience through my PhD journey. I would like to express gratitude towards my committee members Dr. Siva Kolluri, Dr. Jennifer Field, Dr. Gerd Bobe, Dr. Neal Sleszynski, Dr Yuan Jiang and Dr. Kim Anderson. They have provided me helpful feedback, motivated and guided me and have been flexible with their schedules. I would like to appreciate the learning I have acquired in the field of biostatistics from Dr. Bobe and understanding of machine learning techniques obtained from Dr. Jiang.

I would like to express gratitude to key-note leaders, cardio-thoracic surgeons Dr. Eric Zimmermann and Dr. Jaishankar Raman, and doctors in cardiovascular medicine Dr. Nabil Alkayed and Dr. Sanjiv Kaul, who played key roles in guiding me with clinical outcomes of my research and how it can be translated to help doctors and patients in the long run. I was grateful for opportunity to work with Dr. Amala Soumyanath and contribute to her amazing research of using Ayurvedic plant, *Centella asiatica* (Bramhi) for improving cognition and memory which also resulted in NIH grant funding. I would like to acknowledge the contribution of my mentor postdoctoral fellow Dr. Armando Alcazar for explaining the trickiest concepts with simple examples, guiding me with experimental design, troubleshooting instruments, and mentoring me for conceptualizing the project to presenting it in conferences or publication, motivating me at each step. I would like to express my gratitude to Professor Fred Stevens, Dr. Jaewoo Choi, Dr. Wenbin Wu, Linus Pauling Institute, and Professor Brown, College of Pharmacy, for the opportunity to explore computational methods for the discovery of bioactive natural product and dietary markers of grain intake, for their guidance, support, and motivation. I would like to thank Jeffrey More and Dr. Liping Yang for all the help with mass spectrometry experiments, training and resources provided by them. I would like to acknowledge guidance and training in proteomics from my fellow lab member Stanislau Stanisheuski. I am grateful

for the support and valuable feedback from present and past lab members Dr. Manuel Garcia-Jaramillo, Dr. Fereshteh Zandkarimi, Dr. Zifeng Song, Md Nure Alam, Mona Khorani, Suji Park, Arpa Ebrahimi, Rony Koluda and Keenan Tenoyo. I would also like to thank my parents and family for their support and love throughout my PhD journey.

CONTRIBUTION OF AUTHORS

Dr. Amala Soumyanath contributed to the experimental design used for research described in chapter 3. Dr. Armando Alcazar Magana contributed to experimental design, method development of studies described in chapters 3, 4 and 5 as well as writing to sections thereof. Dr. Jaishankar Raman and Dr. Eric Zimmermann contributed tissue samples, experimental design and writing for chapter 4. Dr. Sarah Courtneidge and Dr. Ronn Leon contributed breast cancer cells, Dr. Patrick Reardon conducted the NMR analyses and Stanislau Stanisheuski contributed protein quantification of samples pertinent for the studies described in chapter 5. Dr. Claudia Maier was the principal investigator and advisor on all chapters.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

LIST OF FIGURES

LIST OF FIGURES (Continued)

Figure Page

LIST OF FIGURES (Continued)

Figure Page

LIST OF FIGURES (Continued)

Figure Page

LIST OF FIGURES (Continued)

Figure Page

LIST OF FIGURES (Continued)

Figure Page

LIST OF FIGURES (Continued)

Figure Page

LIST OF FIGURES (Continued)

Figure Page

LIST OF FIGURES (Continued)

Figure Page

LIST OF FIGURES (Continued)

Figure Page

LIST OF TABLES

Table Page

LIST OF APPENDIX FIGURES

LIST OF APPENDIX FIGURES (Continued)

Figure Page

LIST OF APPENDIX FIGURES (Continued)

Figure Page

LIST OF APPENDIX TABLES

LIST OF APPENDIX TABLES (Continued)

Table Page

LIST OF APPENDIX TABLES (Continued)

Table Page

LIST OF APPENDIX TABLES (Continued)

Table Page

1. Introduction

Advances in "omics' technologies combined with innovative use of computational methods pave the way for personalized medicine in peruse of optimal health in hitherto unprecedented ways. Previously treatments provided to all patients were uniform. Analysis of omics data will help in transforming the uniform treatment into personalized medicine to provide better care to patients and improve the reliability in terms of safety and efficacy (Alyass, Turcotte, and Meyre 2015; Vogenberg, Barash, and Pursel 2010). The omics profile which comprises of genomic (DNA), transcriptomic (mRNA), proteomic (protein), metabolomic and lipidomic data will help in understanding the complex patient specific interactions occurring in diseases, such as cancer and cardiovascular diseases. Omics provide information about changes across patients due to changes in gene pool, environmental factors, alterations in gut microbiota and modified enzyme levels (Beger 2013; Nicholson and Wilson 2003). Statistical analyses and data integration will help in establishing links between different omics regimes. Data enrichment tools may aid in the biological contextualization of "omics" data. Multi-omics strategies provide integrated system-level understanding of host environment interactions, and when applied in medicine may have implication for patient subtyping and treatment stratification.

Metabolomics strives to obtain a complete snapshot of the metabolites involved in a biological system. As such, metabolomics provides chemical fingerprints of biological systems that can be further interrogated. Due the vast chemical diversity of metabolites chemical analysis strategies is needed that allow determining the chemical structures and provide access to quantitative information at a given time point. As such, the comprehensive analysis of the metabolome of cells, tissues or plant materials necessitates use of robust analytical platforms, such as liquid chromatography – mass spectrometry (LC-MS) or nuclear magnetic resonance (NMR) spectroscopy, for the detection, identification, and quantification of metabolites.

1.1 Metabolomics for the discovery of aberrant biochemical pathways associated with disease

Metabolomics is defined as a method to analyze metabolites in a biological entity (Beger 2013). The metabolites will include phytochemicals, organic acids, lipids, small molecules of metabolic pathways. These metabolites could serve as intermediates in metabolic pathways that impact biological function. Their effect on biological function also makes them an ideal candidate for biomarkers. Quantification of levels of metabolites across treatment and control samples allows identifying the effect of treatment on metabolic pathways (Johnson et al. 2015). Disparate metabolites point to deregulation of metabolic pathways that are associated with diseases. Metabolomics has been extensively applied to study metabolic processes associated in cancer cells.

Energy production takes place in cancer cells mainly by aerobic glycolysis instead of oxidative phosphorylation (Hirayama et al. 2009). Oxidative phosphorylation occurs in mitochondria using the TCA (tricarboxylic acid cycle) (Acin-Perez et al. 2009). Study of intermediates in glycolysis and TCA cycle would reveal mechanism cancer cells use to synthesize amino acids, lipids required for rapid cell division and migration (Beger 2013; Lawrence, Willoughby, and Gilroy 2002; Van Meer, Voelker, and Feigenson 2008; Menendez and Lupu 2007).This would help in developing novel targets for diagnostics and therapeutics (Beger 2013).

1.2 Metabolomics in conjunction with molecular networks in the context of natural products research

Metabolomics is useful for identifying unknown natural products in natural product mixtures. Botanical extracts are analyzed using fractionation followed flow injection and mass spectrometry. The mass spectral data is used to construct the molecular networks by aligning the MS/MS spectra of the parent ions. Global natural products social networking (GNPS) helps us visualize metabolites in natural products and their associations with other metabolites using spectral similarity. GNPS utilizes MS/MS data and creates network based on the similarity of their MS^2 fragmentation. The spectral similarity is computed using cosine score between two spectra. The cosine score ranges between 0 and 1. 0 represents no similarity and 1 represents the spectra are identical with each other (Mingxun

Wang et al. 2016). Each precursor ion is represented as a node, and they are connected using an edge based on the cosine score. GNPS was used to identify metabolites in natural products that could be used as potential remedies to treat illnesses along with statistical predictive modeling techniques. GNPS constructs the molecular networks by aligning MS/MS spectra of the parent ions.

1.3 Lipidomics, a sub-discipline of metabolomics dedicated to the analysis of lipids

The lipidome is part of metabolome composed of lipid species. Lipidomics deals with identifying the lipid components involved and understanding the interactions of lipids with other lipids and other chemical entities to provide structure and maintain function in biological systems (Wenk 2005). Along with metabolites lipids play in important role in regulating biological pathways, maintaining structure and functions in cells and tissues (Beger 2013). Lipids help maintain structure of cell membrane, impart membrane fluidity play an important role in cell signalling and apoptosis (X. Han and Gross 2003; Wenk 2005). The lipids act as transporters for release of energy and also act as energy reservoirs. As a result deregulation of lipids is linked to cancer. Targeting lipids will help in controlling cell division and migration of cancer cells (Yan Lim and Yee Kwan 2020). Lipid species are classified on the basis of their functional backbone. Lipids can be categorized into eight categories, polyketides, fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, prenols, sterol lipids and saccharolipids (Fahy et al. 2005).

Lipids have a polar head group charachteristic of the lipid species and an aliphatic tails with different number sof carbon atoms and levels of unsaturations. The lipids are represented as XX (NC:UB) with XX representing two letter code for the lipid species, NC representing number of carbons and UB representing number of unsaturated bonds.

1.4 Technologies

Metabolome analyis is performed either by targeted or untargeted methods. Targeted metabolomics is used for detection and quantitation of a selected set of metabolites or lipids, whereas the untargeted approach aims to detect and measure signals for all

metabolites and then annotate them based on spectral matching with metabolomics databases (Bingol 2018). Metabolomics is performed by acquiring data using high throughput platforms, such as liquid chromatography (LC), mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR). NMR and MS are used together for performing metabolomics to obtain higher level of confidence. Mass spectrometry is a highly sensitive and selective techniques that allows identification and quantification that can be used across a wide range of metabolites and lipids. Mass spectrometry can be coupled with ion mobility spectrometry (IMS) that increases the peak capacity and resolves the peaks for coeluting isomers that cannot be separated by liquid chromatography. Hydrophilic interaction liquid chromatography (HILIC) and reversed phase liquid chromatography techniques are used to separate polar metabolites and lipids respectively and are coupled with mass spectrometry for their detection and quantitation.

1.5 Aims of thesis

The focus of my thesis research was to utilize mass spectrometry techniques, chemometrics and bioinformatics tools (1) to obtain an understanding of mechanisms associated with disease and associated physiological perturbation, and (2) for characterization of botanicals and leverage computational chemistry tools, utilize statistical analyses and machine learning algorithms and open-access databases for natural product drug discovery.

The second chapter provides an overview of the methods, technologies and computational tools that are commonly utilized in metabolomics/lipidomics for various biological matrices such as tissue and plasma with focus on workflows relevant to this thesis. Specifically, methodological details are discussed concerning liquid chromatography techniques coupled with mass spectrometry for chemical fingerprinting of botanical extracts. and workflows, technologies and data processing strategies followed to ensure robustness and repeatibility of the analytical workflows used for obtaining metabolite and lipid profiles for mammalian cells and tissues.

The second chapter also reviews statistical analyses and machine learning algorithms for testing hypothesis, assessing the effect of compounds on disease mechanisms, visualization

of similarities and differences of metabolites developing prediction models based on spectral data that can be used for drug discovery, and performing biomarker discovery.

The third chapter deals with the characterizing of botanical extracts (focus on *Centella asiatica*), how chemical fingerprinting can be utilized for evaluating extract reproducibility This will help in producing botanical supplements with reliable efficacy and in reducing time involved in testing them in clinical trials. In this chapter the combination of mass spectrometry-based metabolomics was explored for facilitating dereplication of botanical extracts and reducing the reliance on bioactivity-guided fractionation using computational tools for discovery of bioactive principles.

The fourth chapter explores untargeted lipidomics for studying integrity of bio-banked tissue for subsequent lipid profiling. This study developed analytical workflow to utilize optimal cutting temperature compound (OCT) embedded heart tissue typically used to store bio-banked tissues as it imparts stability to the tissue for long term storage. This study performed comparative lipidomics on OCT stored heart tissues and previously used liquid nitrogen stored heart tissues with the help of reversed phase liquid chromatography coupled with mass spectrometry.

The fifth chapter reports on the critical evaluation of the compatibility of a widely used doxycycline-based inducible gene-silencing strategy with metabolomics utilizing hydrophilic liquid chromatography coupled with MS/MS mass spectrometry and lipidomics using reversed phase liquid chromatography coupled with MS^E mass spectrometry for studying breast cancer metabolism. The findings provide evidence that doxycycline knockout strategies cause disturbances in biochemical pathways, which should be considered when the study design includes metabolomics assays.

The sixth chapter provides a summary about the application of mass spectrometry in detection and identification of biomolecular fingerprints in biomedical research through metabolomics and lipidomics. It highlights the importance of mass spectrometry technology in conjunction with computational methods such as global natural product social molecular networking (GNPS), statistical and machine learning techniques for patient management, biomarker discovery, identification of bioactive molecules and understanding of disease states. It provides a future perspective of how metabolomics can

be translated into accurate point of care diagnostic technique with advances in mass spectrometry technologies.

2. Methodical Pipeline

This chapter focuses on methodologies involved in metabolomics and lipidomics. This chapter will provide information about sample preparation, liquid chromatography and mass spectrometry, high throughput technologies, bioinformatics, and computational tools to visualize and interpret the underlying biology. The conceptual methodological workflow is described in the Figure 2.1.

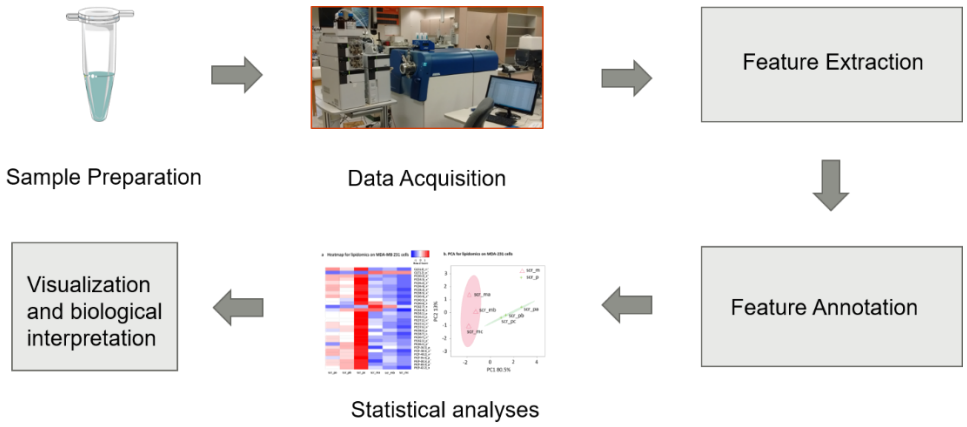


Figure 2.1 Workflow for LC-MS based metabolomics.

Analysis of metabolites and lipids can be divided into two categories of targeted and un-targeted approaches (Checa, Bedia, and Jaumot 2015). Targeted analysis is performed by measuring only a specific set of metabolites and lipids whereas untargeted analysis involves measuring all the detectable metabolites and lipids from the sample (Patti 2011; Vinayavekhin and Saghatelian 2010). This thesis focuses on untargeted analysis which provides comprehensive chemical fingerprints and allows us to look for novel compounds.

2.1 Sample preparation in metabolomics and lipidomics

Sample preparation procedures were developed to perform high throughput metabolomics using LC-MS analytical workflows to detect phytochemicals in botanical extracts and metabolites and lipids from mammalian cells and tissues. These analytical procedures have been described in this section.

2.1.1 Homogenization

Homogenization lyses the tissues or cells to release metabolites, lipids and proteins into the solvent. The energy involved in homogenization needs to be controlled otherwise compounds of interest will be potentially destroyed. The tissues or cells are suspended in a solvent along with 0.2 mg/mL of butylated hydroxytoluene to prevent lipid oxidation in 2ml reinforced tube with ceramic beads of 0.1-0.5 mm diameter. These tubes are placed in a spinning agitator. The agitator is set at 5000 or 6000 rpm depending on the strength of tissue or cells that need to be lysed. The samples should be prechilled, agitator should only be spun for short durations with idle time between each run and the samples should be rechilled to prevent the sample in the tubes from getting hot. Heart tissues are hard to be homogenized (Goldberg 2008) so they had to be homogenized at 6000 rpm with 2.8 mm beads for six repeated cycles of 20 seconds with breaks of 30 seconds in between them. The homogenate is subjected to extraction to separate the compounds into multiple phases.

2.1.2 Metabolite and lipid extraction

In liquid-liquid extraction (LLE), the compounds in the homogenate are partitioned using solvent systems. Polar and non-polar metabolites for metabolomics and lipidomics, respectively, are extracted with a combination of methanol, chloroform and water. The metabolite fractions in this method are collected in the top portion comprising of methanol and water. The chloroform portion below contains the lipid portion. Sometimes, solid phase extraction (SPE) is performed before to collect lipid classes with similar structure (Wolf and Quinn 2008). In a collaborative project, the solid phase extraction was performed to extract oxylipins from plasma to perform biomarker analysis in plasma samples obtained from coronary artery disease patients (Le et al. 2021).

The single step liquid-liquid extraction procedure was used for the studies described in chapters 4, 5 and 6. This extraction protocol is also compatible for metabolomic analysis with NMR. Methanol/chloroform/water extraction exhibited lesser variability than methanol/water extraction (X. Han 2016).

Specifically, the tissues or cells were washed with PBS and suspended in 300 µl water:methanol 1:1. They were homogenized to release metabolites, lipids and proteins

from the biological specimen. The homogenate was transferred to vials with 500 µl chilled chloroform and 300 µl chilled water followed by centrifugation at 10000 rcf at 4°C for 10 min. This resulted in a formation of three phase mixture. The top metabolite portion was divided into 2 parts for NMR and LC-MS analysis. The sample for metabolic profiling with NMR was suspended in 10mM PBS and NMR buffer comprising of DSS and D_2O whereas the sample for metabolic profiling with HILIC-MS/MS was transferred to 95% acetonitrile and 5% water. Acetonitrile and water are fully miscible with each other and can be used with mobile phase additives and buffer. The mobile phase comprising of acetonitrile and water have lower viscosity and creates lower back pressure. This allows the LC system to operate at higher flow rate and decreases the run times. Acetonitrile is also used in reversed phase chromatography as the elution strength of acetonitrile is higher than methanol which allows for reduced analyte retention which decreases the total run time. The bottom lipid portion was resuspended in 200 µl 50% acetonitrile:isopropanol. Isopropanol dissolves lipids and provides higher elution strength needed for eluting of lipids in reversed-phase LC systems. Under these conditions all major lipid classes can be chromatographically separated and runs are between 12 and 20 minutes (X. Han 2016).

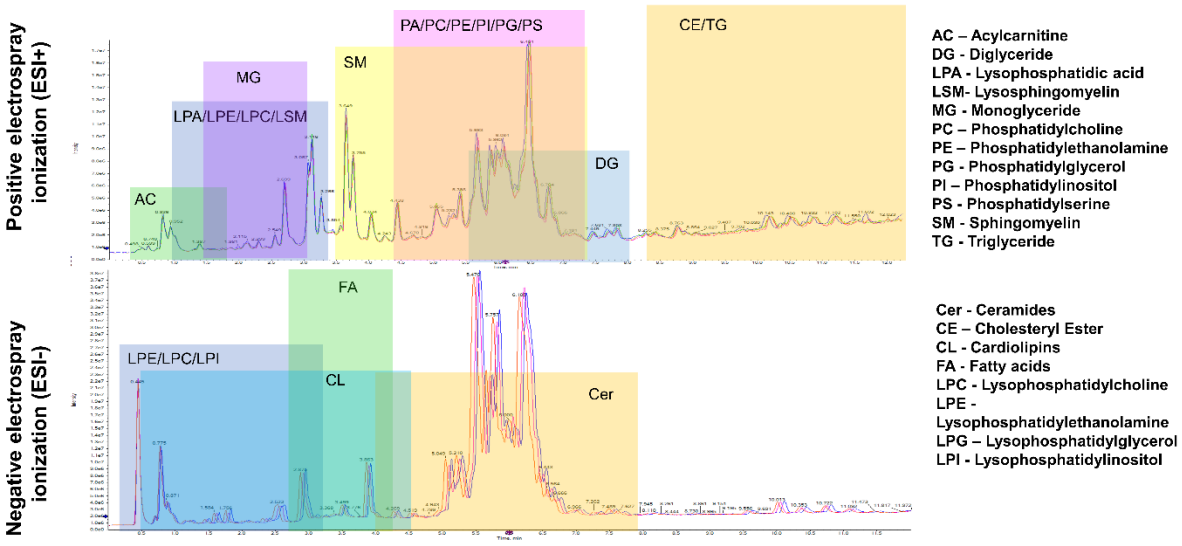


Figure 2.2 Lipid elution pattern for QCs for ESI+ and ESI- mode with elution windows for lipid classes in chapter 4.

2.1.3 Internal standards

Internal standards are used in untargeted and targeted metabolomics and lipidomics. Internal standards (IS) are isotopically-coded version of commonly observed metabolites and lipids. IS are typically used for marking retention windows, platform performance monitoring, determining recovery and quantification. Multiple standards are used in case of untargeted lipid profiling wherein each standard is a representative of the lipid class because of high price of labelled lipid standards. This approach helps in alleviating the problem of normalization because of different ionization efficiency across different lipid classes (Bowden et al. 2017; Miao Wang, Wang, and Han 2017). SPLASH lipidomics standard mixture comprises of lipids with odd acyl carbon number for normalization (Abdullah et al. 1950). SPLASH Lipidomics mixture containing internal standard for each lipid class of interest was spiked to QCs (matrix) at different dilutions for calibration, as the ionization efficiency for different lipid species is different (Abdullah et al. 1950; Miao Wang, Wang, and Han 2017). SRM 1950 NIST sample (NIST1950 Metabolites in human

plasma n.d.) was injected at different dilutions to test for LC mass spectrometer platform performance.

However, use of internal standards is difficult in metabolomics profiling because of the large differences in the physicochemical properties of metabolites will not be able to provide coverage across all the metabolite classes. This issue was resolved to some extent using in house authentic compound reference libraries.

2.1.4 In house authentic compound reference libraries

In house library for different chromatographic methods were created using IROA Mass Spectrometry Metabolite Library of Standards comprising of 650 metabolites, associated with primary metabolic pathways. This library was used for detecting and metabolites of interest in metabolic pathways by comparing the retention time, accurate mass, and MS/MS fragments. In addition to the metabolic databases that provided putative assignments, standards helped in providing additional confidence in their annotation.

2.1.5 Handling and storage of metabolites and lipids

There is a risk of degradation, oxidation, or aggregation associated with metabolites and lipids on exposure to light or air, or higher temperatures. Incorrect sampling or sample pretreatment procedures could potentially lead to changes in the metabolites or lipids of interest resulting in biased results (Álvarez-Sánchez, Priego-Capote, and Luque de Castro 2010). Along with internal standards, careful sample handling and storage helps preserve the reliability and reproducibility of metabolomics experiments. The best practice of handling metabolites and lipids is to keep the samples chilled. When there is a delay in using the biological samples, they need to be stored at -80^{o}C or snap frozen in liquid nitrogen until needed.

In chapter 4, the impact of different storage conditions on the lipidome of heart tissue samples is studied. Specifically, the heart tissue samples had been stored by embedding them in optimum cutting temperature (OCT) compound or were flash frozen in liquid nitrogen. Sample preparation strategies involve homogenization and vortexing procedures which might heat up the sample leading to lipid peroxidation. Lipid peroxidation is prevented by keeping samples chilled in ice and use of inhibitors to the enzymes

responsible for lipid oxidation such as butylated hydroxytoluene (BHT) to decrease oxidation of unsaturated lipids (Wolf and Quinn 2008).

2.2 Analytical platforms for lipidomics and metabolomics

Analysis of metabolites is performed using NMR and mass spectrometry and lipids are analyzed using mass spectrometry. High throughput analysis allows analysis of tens to thousands of metabolites and lipids in a single run. Mass spectrometry had a higher sensitivity than NMR but NMR has the advantage of quantitative analysis to determine the concentration of metabolites in the pico- and femtomolar range (L W Sumner, Lei, and Huhman 2011).

2.3 NMR

NMR is ideal technique for metabolomics on small polar metabolites. NMR provides information about the structure, concentration for metabolites in the micro-molar range with higher reliablity without the loss of sample (Naz et al. 2014). The drawbacks with NMR technique are lower sensitivity when compared to mass spectrometry and inability to identify non-polar lipids (L W Sumner, Lei, and Huhman 2011).

2.4 Liquid Chromatography coupled with mass spectrometry

The drawbacks of lower sensitivity and inability to identify non-polar lipids by NMR is addressed by liquid chromatography coupled with mass spectrometry (LC-MS). LC-MS approaches enable us to obtain comprehensive chemical fingerprints of complex systems. LC-MS approaches provide access to accurate mass information, structural information when combined with gas phase fragmentation techniques, abundance. High resolution mass spectrometry platforms offer the needed resolving power, dynamic range, and sensitivity to obtained comprehensive coverage of chemical entities present in complex

mixtures. In this thesis, untargeted analysis were performed with ultra-high-performance chromatography for both polar and non-polar compounds (Dettmer, Aronov, and Hammock 2007; Griffiths and Wang 2009; Patti, Yanes, and Siuzdak 2012).

There are two major types of liquid chromatography used in metabolomics and lipidomics, hydrophilic interaction liquid chromatography (HILIC) and reversed-phase liquid chromatography (RPLC) (Gika et al. 2014).

2.4.1 HILIC

Hydrophilic interaction liquid chromatography (HILIC) uses hydrophilic stationary phase. HILIC column separates the metabolites based on primarily their partitioning in mobile and stationary phases. The additional mechanism that helps in retention of polar metabolites is adsorption. Zwitterionic stationary phases are used because of their ion exchange interactions. The stationary phase adsorb water using hydrogen bonding that controls retention of metabolites. These columns are referred to as ZIC-HILIC columns. The mobile phase used in HILIC consists of polar organic solvents miscible in water like acetonitrile (Buszewski and Noga 2012). Isocratic or gradient modes can be used for HILIC. Isocratic modes use higher concentration of organic solvent and gradient mode use a gradient from high percentage of organic solvent to high percentage of water (Alpert 1990). We used a gradient method as were performing untargeted metabolomics and were trying to detect as many metabolites as possible. Buffers like ammonium acetate and ammonium formate are added to the mobile phase. This helps in maintaining the pH and ion strength. pH was maintained at neutral for both ESI positive and ESI negative modes as that would allow us to detect wide range of metabolites. This also helped in preventing asymmetric shapes of peaks and tailing of chromatographic peaks (Buszewski and Noga 2012).

HILIC optimization and HILIC library creation

HILIC method is usually utilized for separation of polar metabolites. Metabolites are commonly represented by a few isomers in human cells. This elevates the importance of chromatographic separation for identification and quantification of these metabolites.

Optimization of chromatographic separation would also help in minimizing ion suppression. The mechanism of separation in HILIC chromatography involves hydrophilic partitioning of polar metabolites in stationary polar phase. The analytes are retained using dipole-dipole interactions, hydrogen bonding and ion exchange with the stationary phase in the column. Therefore, the mobile phase gradient was optimized to prevent non-polar metabolite elution in dead volume and peak tailing. Using authentic IROA standards, 3 different gradients, 80%, 90% and 95% of acetonitrile were tested in order to optimize the analysis (avoiding non-polar compounds at dead volume, accumulation of compounds and/or equilibration time after each run), peak suppression (at high concentration of acetonitrile) and peak shape (Figure 2.4). At least 5% water was maintained in the mixture during the process of gradient selection to improve the life of the column (Snyder et al. 2013). The chromatograms for the conditions tested are shown in the following Figure 2.4.

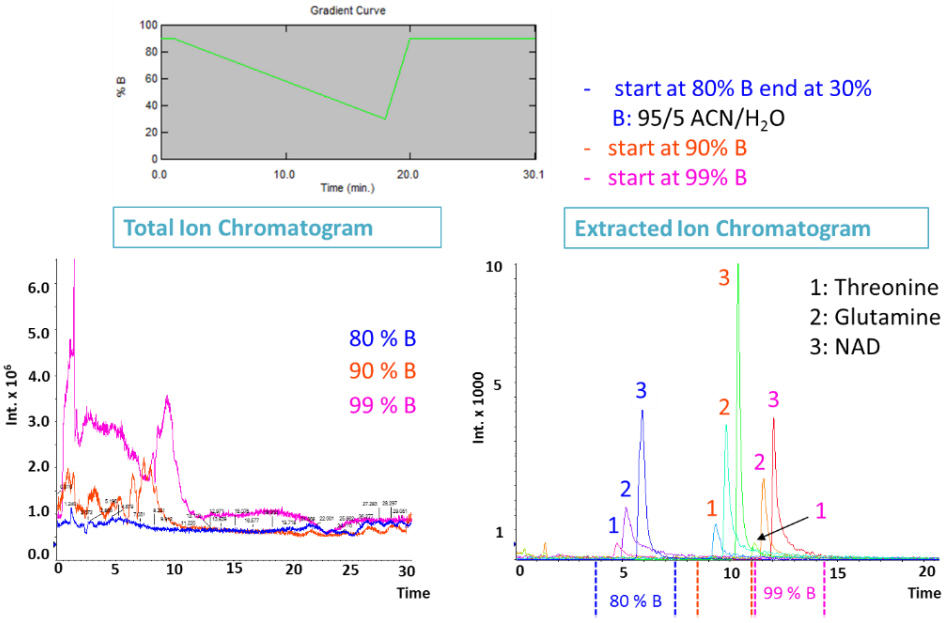


Figure 2.3 HILIC LC-MS/MS method optimization using authentic standards. Final gradient is shown in the graph.

The gradient with 90% B was selected to obtain sharp peaks of metabolites without losing any metabolite information. Once the HILIC method was optimized, a library was created from authentic standards for the HILIC-MS method that helped in metabolite assignments

based on matching, retention time and mass spectral data with the curated spectral information.

2.4.2 RPLC

The other chromatographic method used in this thesis is reversed phase chromatography (RPLC) that uses a C8 or C18 column where the stationary phase consists of silica particles with C8 or C18 aliphatic carbon tails (Chester 2013). This makes the stationary phase hydrophobic, and the hydrophobic interactions help in retaining the non-polar metabolites. Analysis of non-polar metabolites was conducted using a gradient of isopropanol and other solvents such as acetonitrile and methanol with mixture of organic solvents (methanol and acetonitrile) and water (Cajka and Fiehn 2014, 2016b; Sarafian et al. 2014). The mobile phase also consisted of buffer 0.1% formic acid and ammonium formate in negative ion mode of ESI or ammonium acetate in the positive ion mode of ESI (Cajka and Fiehn 2016a; Hyötyläinen and Orešič 2015; Triebl et al. 2017). Charged surface hybrid (CSH) C18 column was used because of its higher selectivity with mobile phase containing formic acid (Nováková, Vlčková, and Petr 2012). Chromatographic peaks obtained with CSH C18 were narrow and did not show tailing on account of reduced silanol activity (Škrášková et al. 2013). The non-polar metabolite species with shorter aliphatic chains elute before longer chains and the metabolites with higher level of unsaturation elute before the analogues without unsaturation (Hyötyläinen and Orešič 2015).

2.4.3 Mass spectrometry

Liquid chromatography is used to separate different chemical moieties and mass spectrometry is used for identifying chemical compound by comparing their spectra, retention time and exact mass with that of the compounds in mass spectral libraries. The chemical compounds present in natural products, cells and tissues are ionized using electrospray ionization before transferring them to the mass analyzer.

ESI is a soft ionization technique, and it performs analysis of metabolites without disturbing chemical nature of the biomolecule before mass analysis. The ESI works on the principle of applying high voltage (2-5 kV) on the injection needle. Liquid is sprayed from the needle at high voltage in a cone shape. The exit of the needle and orifice inside the chamber. The temperature of the chamber is increased, and the environment of the chamber

is kept dry with the flow of nitrogen gas. Due to the rise in temperature and dry environment the droplets injected from the orifice evaporate which undergo evaporation and explode because of the repulsion forces from the same charged ions resulting in "Coulomb fission". This process repetitively leads to formation of single charged molecule/chemical (ion) with no solvent. This process of obtaining charged ions from intact metabolites was first demonstrated by John Fenn et al. (Banerjee and Mazumdar 2012; Bedair and Sumner 2008; El-Aneed, Cohen, and Banoub 2009; Gross 2011; Kind and Fiehn 2013).

The resulting ions were single charged of the form ($[M+H]^+$ and $[M-H]^-$). However, there might be adduct formation with a common ion Na^+ forming a sodium adduct $[M+Na]^+$ or a loss of water resulting in $[M-H2O-H]^-$ ions. Although these ions have same retention time, they have different *m/z*. The adducts can interfere detection, quantitation, and identification of biomolecules. The ions will be later separated by mass spectrometer according to the mass-to-charge-ratio. The mass analyzers popularly used are quadrupole analyzer, the ion trap analyzer, and time of flight analyzer (Gross 2011). In the thesis, an ABSciex QTOF 5600 and a SYNAPT G2 High definition mass spectrometer (Waters Corporation), both hybrid quadruple orthogonal acceleration Time-of-flight (ToF) MS instruments, were used for untargeted analysis. The instruments are a combination of quadrupole and ToF analyzers.

In this instrument configuration the quadrupole (q) mass analyzer acts as a "mass filter". The quadrupole provides low resolution wherein the separation of ions occurs using four charged hyperbolic rods. The ions of a particular *m/z* range move through these rods using superimposed radio frequency (RF) and constant direct current (DC) potentials between four parallel rods.

A Time-of-Flight (ToF) analyzer is a mass analyzer wherein separation of ions is carried out by applying voltage to a flight tube. The principle of separation is the time taken by the ions to reach the detector. As the voltage applied is the same for all the ions the kinetic energy is the same, so the time taken by the ions is proportional to the mass (de Hoffmann and Stroobant 2007). In modern ToF instrument the analyzer is enhanced with a reflectron ("ion mirror") that reflects and focus ions of a particular *m/z*. The ion mirror ensures that the ions with same *m/z* but with different kinetic energies reach detector at same time. This

arrangement allows ToF to achieve resolution range greater than 10000 (FWHM) as compared to resolution of 1000 (FWHM) for linear ToF (Chernushevich, Loboda, and Thomson 2001).

The schematic of AB Sciex QToF 5600 time-of-flight mass analyzer with a reflectron is shown in Figure 2.5.

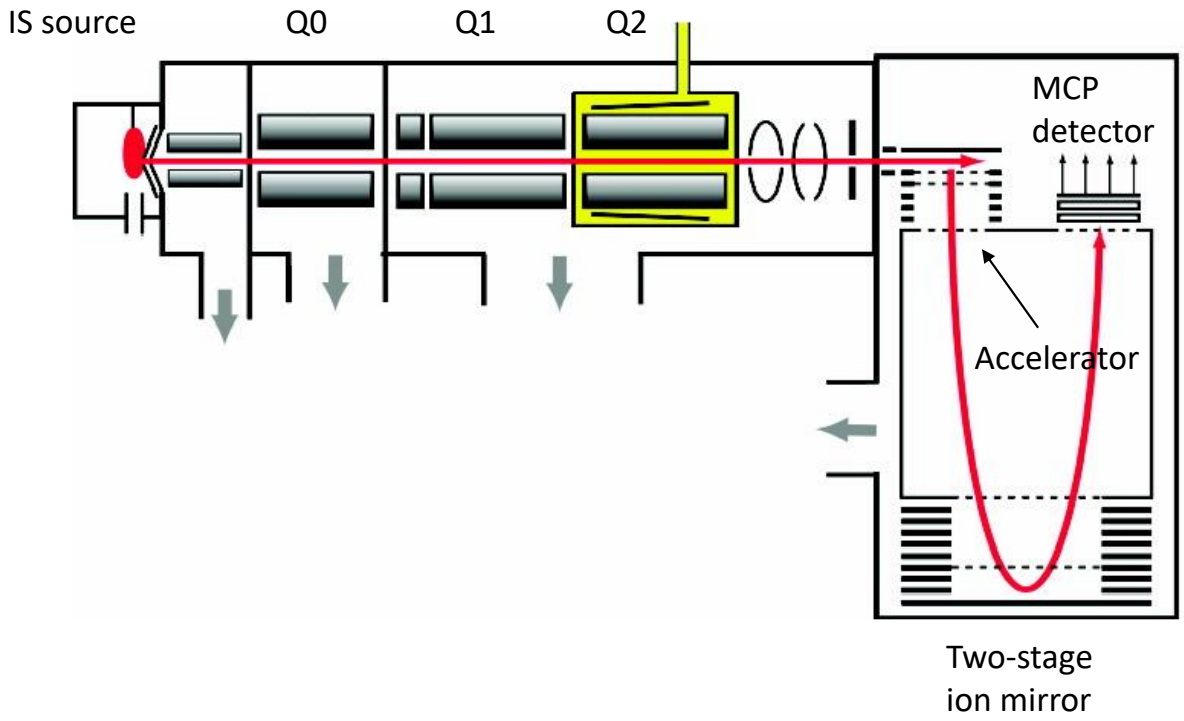


Figure 2.4 Schematic of the ABSciex 5600 mass spectrometer (Andrews et al. 2011).

The ions from the reflectron enter the detector. The detector transforms the ion count into a usable signal. The detector generates the signal proportional to the abundance of the incident ions. Metabolite profiling platform was generated using a combination of liquid chromatography and mass spectrometry based on Mass-to-charge ratio (*m/z*), Retention time (RT) and Intensity.

In this thesis, hybrid quadrupole-orthogonal acceleration time of flight mass spectrometry systems Sciex 5600 and Synapt G2 hybrid quadrupole-orthogonal acceleration-traveling

wave (Waters Corp., Milford, MA) were used for studies described in chapter 3, 4, 5 and 6. The schematic presentation of Synapt G2 instrument is depicted in Figure 2.6.

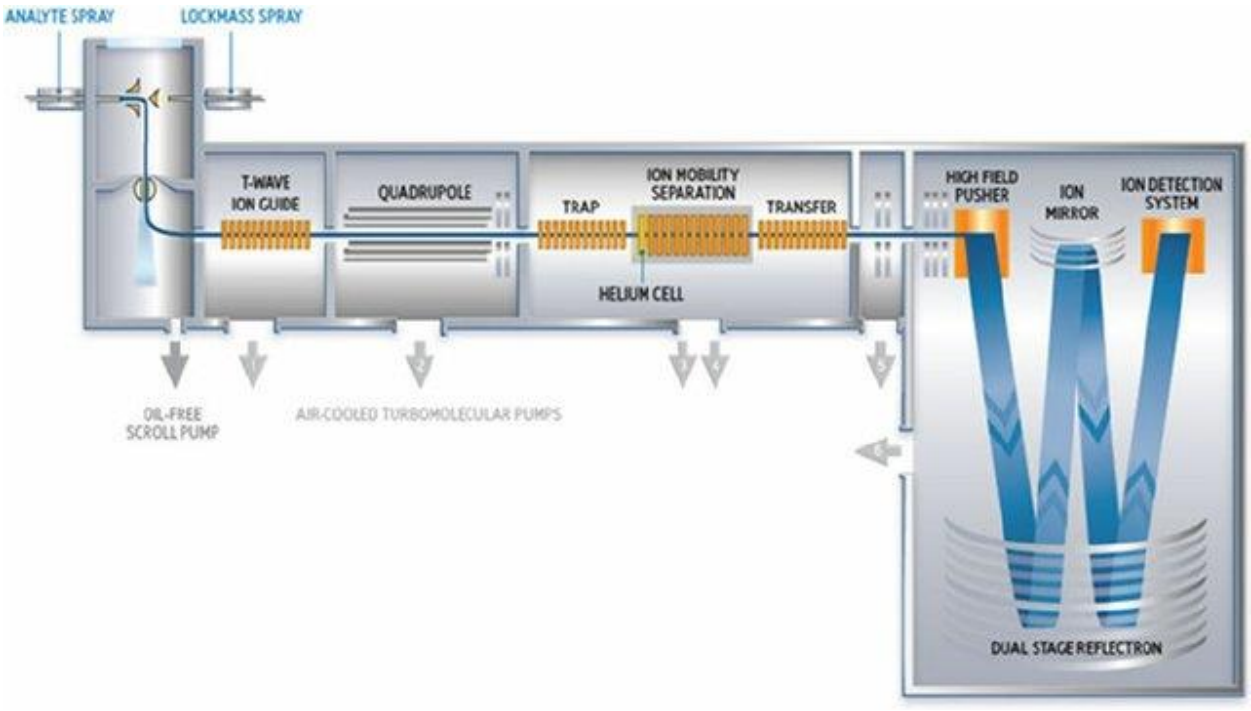


Figure 2.5 Schematic of the Synapt G2 mass spectrometer (http://www.waters.com/waters/).

The hybrid QToF instrument consisted of combination of a quadrupole, a collision chamber and a ToF analyzer to obtain accurate mass measurement, fragmentation experiments, and high-quality quantitation of metabolites and lipids. Sometimes these mixtures are complex and there are many compounds with the same elemental composition. Therefore, additional MS/MS or data independent acquisition or ion mobility information is used to make accurate annotations.

MS/MS experiments yield fragment ions. Fragment ion spectra are useful for level 2 annotation of metabolites using spectral databases. This allows annotation of metabolites based on accurate mass and fragment ion spectral similarity (X. Zhu, Chen, and Subramanian 2014). GC-MS/MS provides unique fragment fingerprints that are standard because the conditions in GC remain constant across all the instruments. However, in LC-

MS/MS the conditions, such as the type of mass spectrometry instrument, mode of ionization, and collision energy applied, cause variations in the fragmentation patterns.

LC-MS/MS method is used for untargeted metabolomics analysis. LC-MS/MS is usually conducted in conjunction with data dependent acquisition mode as the MS/MS fragmentation is dependent on intensity of the precursor ions. The mass spectrometry instrument carries out MS full-scan followed by MS/MS analysis in the DDA mode. DDA mode allows in obtaining both quantitative (obtained from the MS full-scan) and structural (obtained the MS/MS spectra) in the same analysis. Simultaneous data processing and metabolite identification is the primary advantage of DDA mode. DDA mode suffers from a drawback that the molecular features with low intensity will not be selected for fragmentation and will not have MS/MS data. The other limitation for MS/MS analysis is the amount of time allocated for MS full-scan is reduced in comparison to the acquisition time allocated to MS/MS spectra generation. This leads to decrease in signal intensity for MS1 features in DDA mode and is associated with the difficulty of detection and quantification of molecular features with low intensity (J. Guo and Huan 2020).

This led to the development of data independent acquisition technique to increase coverage of detected metabolites and decrease the false negative identifications (Tsugawa et al. 2015; Zhou et al. 2017).

Data-independent analysis enabled unbiased acquisition of all product ions for the precursor ions simultaneously. This allowed in obtaining near-complete coverage of metabolites and alleviates the possibility of type II error (Tsugawa et al. 2015; Zhou et al. 2017). Various kinds of data acquisition techniques have been used previously for high resolution mass spectrometry (Andrews et al. 2011; Tsugawa et al. 2015), SWATH (sequential window acquisition of all theoretical mass spectra), All ion fragmentation (AIF) (Gallart-Ayala et al. 2013; Naz et al. 2017) and (MS^E) (J. M. Castro-Perez et al. 2010).

SWATH MS acquires the precursor ions with fixed isolation windows in cyclic fashion to cover complete *m/z* range of precursor ions. This enables fragmentation of all metabolites in the sample and the fragment spectra information is used along with accurate mass, isotopic ratio to annotate the metabolites with the help of mass spectral libraries (J. M.

Castro-Perez et al. 2010). All ion fragmentation (AIF) utilizes higher energy Collisional Dissociation (HCD) fragmentation to fragment all precursor ions without performing mass filtering. AIF performs analysis using full scan HRMS at different collision energies (Naz et al. 2017). This allows AIF to overcome the limitation of MS/MS mode and perform qualitative analysis for low abundance metabolites and results in improvement in mass accuracy over MS/MS (Sentandreu et al. 2018).

The drawback of lower coverage of metabolome for MS/MS is addressed by MS^E. MS^E is a data acquisition technique in mass spectrometry for untargeted analysis of metabolites and lipids. In the process of data acquisition energy switches between low and high energy. During MS^E data acquisition in the first function MS1 Q1, first ring electrodes scans the precursors between 50-1200 m/z transfers it to second ring electrodes Q2 with low collision energy of 4 eV. The ions are then passed to ToF detector which detects ions with high resolution and mass accuracy. In the second function Q1 scans the same mass range however, energy in Q2 is switched to high collision energy between 30-65 eV. The high collision energy results in fragmentation of all precursor ions without any preselection. The primary advantage of MS^E analysis is associated with obtaining both information of the precursor and product ions in parallel alternating scans simultaneously in one analytical run (81,166). MS^E analysis was performed for chemical finger printing analysis of extracts of hot chili (Yap et al 2021). MSE example data for chemical standards of phytochemicals in chili extract are shown in Figure 2.7 (Yap et al. 2021).

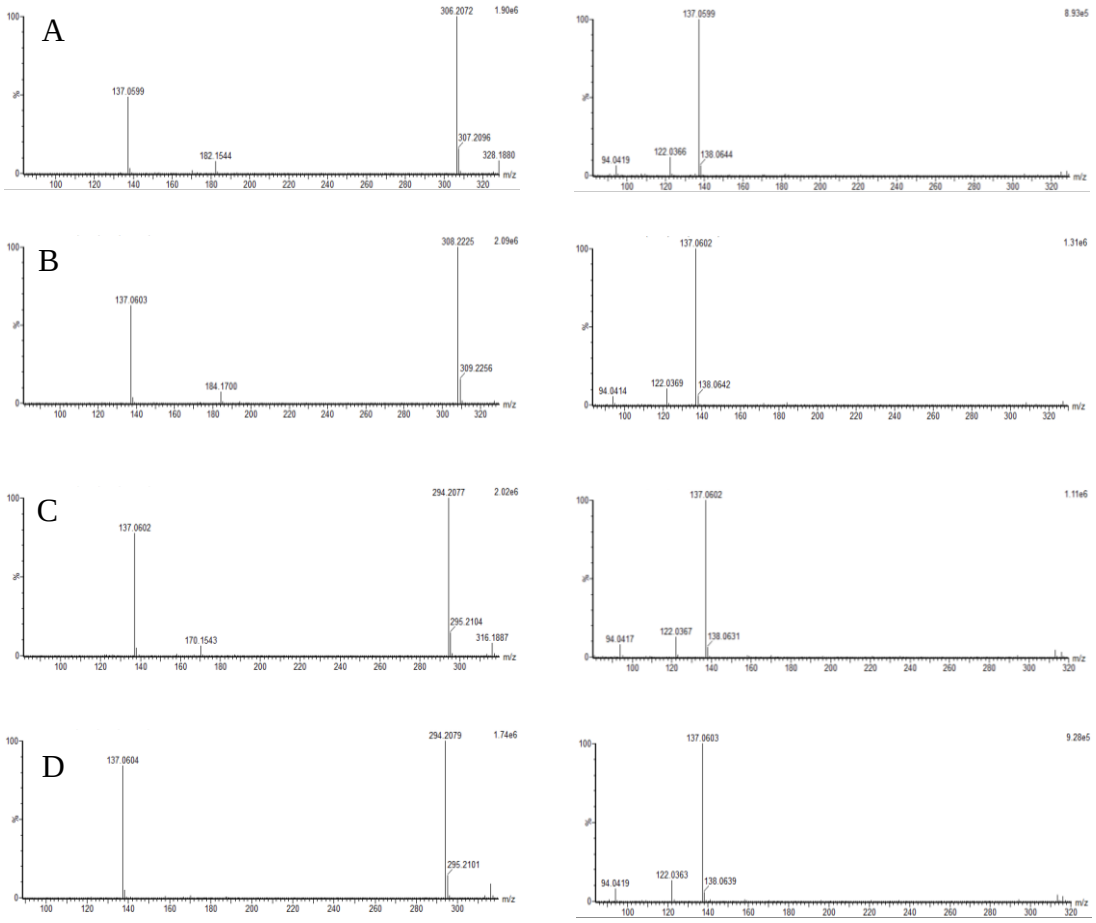


Figure 2.6 Mass spectra of standard compounds: (A) capsaicin, (B) dihydrocapsaicin, (C) nordihydrocapsaicin and (D) nonivamide. Left column: low collision energy (4 V) and right column: high collision energy (30 V) (Yap et al. 2021).

The ions are transferred to the time of flight (TOF) analyzer. TOF detects all the ions with high resolution and mass accuracy. The first scan function has the spectra for low energy in the collision cell whereas the second scan function contains the spectra for high energy which consists of fragment information. MS^E provides the advantage of analyzing product and precursor ions in the same analytical run. These are aligned based on the retention time and peak shape (Waters Corporation 2011). The data outputted from MS^E comprises of accurate mass, retention time, charge state and intensity. The schematic for MS^E is shown in Figure 2.8.

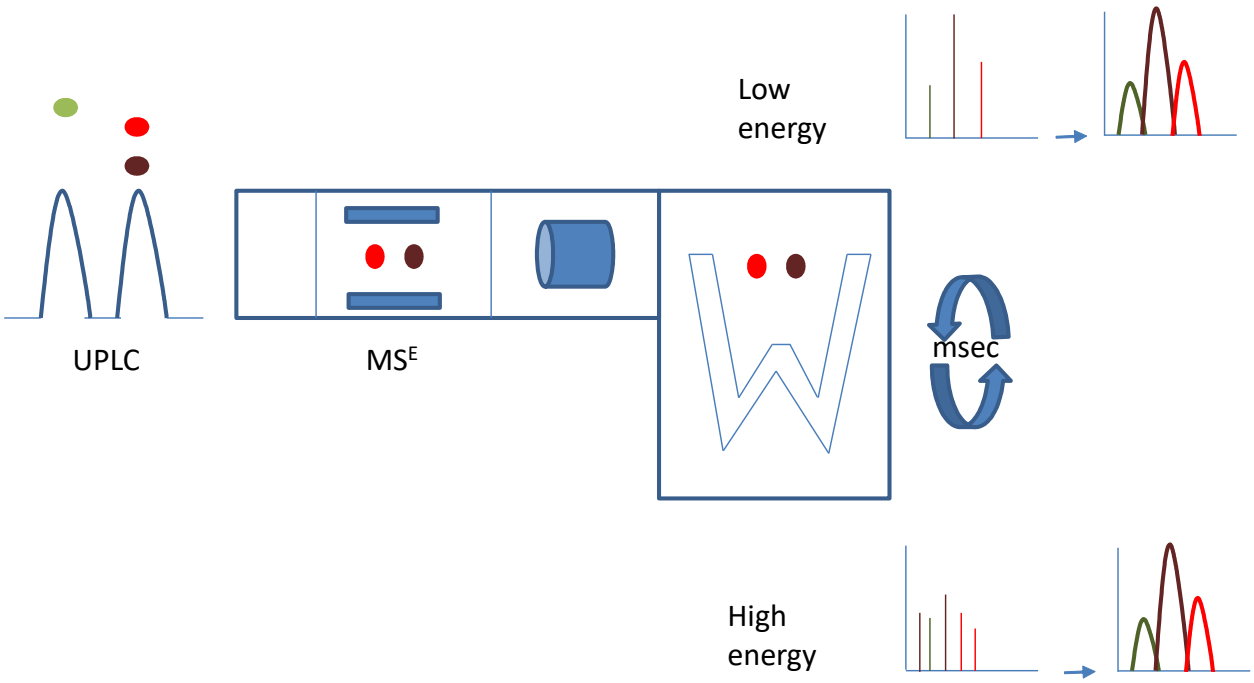


Figure 2.7 A schematic of UPLC-MSE analysis adapted from Plumb et al. (Plumb et al. 2006)

2.5 Ion mobility spectrometry – Mass Spectrometry (IMS-MS)

The other mass spectrometry technique frequently used in natural product discovery is ion mobility spectrometry (IMS-MS). Ion-mobility spectrometry is orthogonal gas phase separation technique for expanding the separation space and peak capacity.

Principle of ion mobility spectrometry

The ions move against the flow of the buffer gas because of the force exerted by the electric field. The velocity of the ions in this diffusion process is directly proportional to the electric field. The proportionality constant has relationship with the collision cross section (CCS) of the ion.

$$K = \frac{3q}{16N}\left(\frac{2pi}{kT}\right)^{\frac{1}{2}}\left(\frac{m+M}{mM}\right)^{\frac{1}{2}}\frac{1}{\Omega}$$

Q charge of the ion

N number density of buffer gas

k boltzmann constant

T absolute temperature

m mass of buffer gas

M mass of the ion

Omega collision cross section (Kanu et al. 2008)

Calibration of the instrument is performed to reduce the dependence of the collision cross section based on the IMS-MS platform and identify the optimal parameters for ion mobility separations using single charged polyalanine (0.1 mg/L) in both ESI positive and negative ion modes (Paglia et al. 2015). Calibration curve was constructed with normalized collision cross section was plotted with corrected drift time (J. Castro-Perez et al. 2011). The calibration curve was of the exponential form $y = Ax^b$. IMS-MS has applications in isomer, isobar and conformer separation, diminishing the chemical noise and clustering compounds into their chemical families. Ion mobility separation is carried out in ion mobility chamber. The separation of ions is performed based on size, shape, charge, and their behavior with buffer gas in the presence of electric field (Kanu et al. 2008; Wickramasekara et al. 2013). Ion mobility techniques can be mainly categorized into drift time IMS-MS, travelling wave IMS (TWIMS-MS), field asymmetric IMS-MS and trapped IMS-MS (Kanu et al. 2008).

TWIMS-MS

Separation was based on traveling wave ion mobility conducting mass spectrometry as ion mobility helped in separating based on structure and conformation. This technique utilizes travelling potential waveform and 6-fold rise in the operational pressure across drift section by integrating helium-filled ion entrance region in the TWIMS (Giles, Williams, and Campuzano 2011; May and McLean 2015). Synapt G2 HDMS Waters instrument (Figure 2.6) was used to conduct ion mobility experiments in chapter 3 and another collaborative

project which involved determining collision cross section for compounds in chili extracts (Yap et al. 2021).

The instrument comprises of tri-wave section which has three main parts traveling-wave (T-wave) ion guide regions including Trap Twave, ion mobility separation (IMS), and transfer T-wave region. The trap region accumulates the ions and releases them as packets into ion-mobility separation device. The ion mobility separation moves ions from the transfer region to orthogonal acceleration (oa) TOF analyzer (Maier et al. 2013). "Tri-wave" region comprises of a series of stacked ring ion guides (SRIG) that are filled with neutral background gas such as nitrogen. The mobility of ions through SRIG is performed by travelling wave potential generated by voltage pulses dynamically applied using stack of ring electrodes. Different ions interact differently to the travelling wave that forms basis of TWIMS separations (May and McLean 2013). The movement of ions in SRIG region depends upon the traveling wave height (V), the velocity of the traveling waves (m/s) and gas pressure (mbar). The dwell time of the ion in the SRIG is dependent on these factors and they are optimized to separate the ions based on the differences in collision cross sections.

The optimal parameters for A and b for the TWIMS platform obtained after calibration for the project identifying collision cross section of compounds in 'Super Hot' chili fruit (Capsicum annuum) after exposure to supplemental LED lights were 246.81 and 0.6628 (Yap et al. 2021). The traveling wave velocity of 550 m/s and wave height of 40 V were used. The pressure of the trap and transfer devices was maintained at 2.5×10^{-2} mbar. The helium gas flow for helium cell region was maintained at 180 ml/min to decrease the internal energy of the ions and reduce fragmentation. Nitrogen was used as the drift gas and flow rate for nitrogen was maintained at 90 mL/min. The mass spectrometry data was acquired in the MS^E mode with the trap device maintained at low collision energy of 4eV and the transfer device at high energy of 40 eV to fragment the precursor ions. The optimal parameters were used to perform further IMS measurements.

To summarize the size, shape, charge and collision cross section with drift gas N_2 and the travelling wave influence separation of ions (Claire et al. 2014). Ion mobility separation helps in identifying metabolites and lipid species in complex biological samples as it

provides a quick separation in milliseconds and provides increased peak capacity. The research studies in chapter 3 were performed using TWIMS for different structures and conformations of isomers of dicaffeoylquinic acids (Figure 2.8).

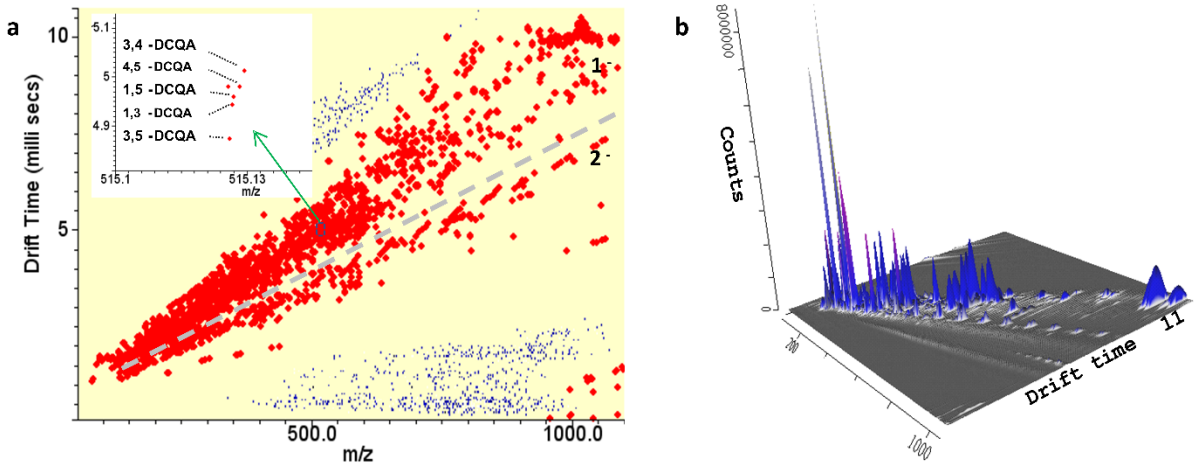


Figure 2.8 a LC-ESI q-IMS-MS/MS analysis of an aqueous extract of Centella asiatica (5 mg/mL of dry mass resuspended in methanol 70%, 10 ul injection). 2D map visualization of drift time versus m/z. The LC system was coupled to a Synapt G2 HDMS (Waters Corp., MA, USA) used for detection in negative ion electrospray mode. Nitrogen was used as carrier gas for ion mobility experiments. Data acquisition range was m/z 50-1200. The cone voltage was 20 V. The T-wave ion mobility cell was operated at 800 m/s and the wave amplitude was set at 35 V. Helium and nitrogen IMS carrier were both set at 80 ml/min. Each red dot represent a different drift time detected during the chromatographic run. b ESI q-IMS-MS/MS analysis of an aqueous extract of C. asiatica (5 mg/mL of dry mass resuspended in methanol 70%, 10 ul injection). 3D map visualization of drift time versus m/z. The LC system was coupled to a Synapt G2 HDMS (Waters Corp., MA, USA) used for detection in negative ion electrospray mode. Nitrogen was used as carrier gas for ion mobility experiments. Data acquisition range was m/z 50-1200. The cone voltage was 20

V. The T-wave ion mobility cell was operated at 800 m/s and the wave amplitude was set at 35 V. Helium and nitrogen IMS carrier were both set at 80 ml/min. DCQA- Di-caffeoylquinic acid

The spectra comprising of multidimensional format consisting of retention time, *m/z*, peak intensity, drift time obtained for IMS measurements were analyzed using vendor software Driftscope 2.1 to plot driftogram. Additional dimension of ion mobility in addition to the retention time and *m/z* that will allow better peak capacity and better analytical results.

Time Aligned Parellel Fragmentation

Mass spectral fragmentation technique that enables acquisition of additional structural information for the metabolite or lipid ions is by applying high collision energy in both the trap T-wave and transfer T-wave regions, which is referred to as time-aligned parallel (TAP) fragmentation. In this thesis TAP fragmentation was performed by maintaining both trap and transfer at the collision energy of 35eV (Yap et al. 2021). Argon as used as the collision gas and was maintained at the pressure of $9.11x10^{-3}$ mbar. The quadrupole Q1 selected the precursor ions with narrow *m/z* window of 1 Da. These ions are fragmented in the trap. The daughter ions were transferred to the ion mobility chamber where they are separated according to their collision cross section. The daughter ions are further fragmented in the transfer device to form grand-daughter ions. The daughter and grand-daughter ions are aligned based on the drift time (J. Castro-Perez et al. 2011).

2.6 Quality Control procedures in metabolomics

Mass spectrometry is a highly sensitive instrument and because of that it is prone to variations. The variations can be classified into biological and technical variations. The biological variations are seen because of variations in environment or genetic constitution. The technical variations can result from changes in state of local environment in the instrument. These can arise from contamination of ion source, column poisoning, sample degradation. To account for the systematic error, equal aliquots from all the samples are

pooled together to form quality control (QC) samples along with commercially available internal standards.

The internal standards are also used in QCs for monitoring the performance of the instrument platform and semi quantitative analysis. The choice of standards plays an important role in metabolic and lipid profiling, especially in the untargeted setting (H.P. et al. 2012). The reference standards provide quality assessments across different laboratories. The standards provided by NIST are widely accepted. SRM1950 (Simón-Manso et al. 2013) was used as the standard commonly used for plasma metabolomics to compare results across different laboratories. SRM1950 was used for research studies in chapter 5. Although expensive, these standards serve as a mechanism ensure long term stability of metabolites in the sample.

The optimal instrument performance and robustness, precision and accuracy of the analytical method was ensured and monitored by injecting QC (D. Broadhurst et al. 2018). QC sample injections were performed for conditioning the column before beginning the run. The research studies in the thesis were conducted in batches because of limitations of availability of instrument and time period elapsed since sample collection (Thonusin et al. 2017). The batch comprises of samples which are analyzed under same analytical conditions (Wehrens et al. 2016). The batch design comprises of blank, QC and sample injections.

Besides injecting in the beginning QCs are also injected after every 10 sample injections and at the end to determine analyte reproducibility and stability of the instrument (Dunn et al. 2011; Zelena et al. 2009). QC samples help in decreasing the biological variation as they are created by combining all the biological samples together in equal volumes (Zelena et al. 2009). Performance of the LC-MS platform is analyzed using the coefficient of variation (standard deviation / mean) in the intensity measurement for the feature ions in the QC samples. The features with the coefficient of variation (CV) greater than across the QC samples cannot be included because of lack of confidence in these features (Dunn et al. 2011). Similarly, blank sample injections comprising only solvent A without the sample were performed at the beginning, between samples and towards the end of analysis to evaluate any carryover or contaminations within a batch. The samples were randomized to

account for variations from batch effect (De Livera et al. 2015) and decrease carryover. Data preprocessing and normalization techniques are used to correct the batch effect and alleviate the variations arising from sample preparation and matrix effects (De Livera et al. 2015).

2.7 Data Preprocessing

There are computational challenges associated with analyses of metabolites and lipids in processing the signal obtained from mass spectrometer. Therefore, preprocessing ensures precision and accuracy in identification of features. The pre-processing steps involve peak detection, retention time correction and normalization based on a feature, QC or protein content. Preprocessing tools from vendor such as Progenesis QI (Nonlinear Dynamics) MarkerView (Sciex) and MassLynx (Waters) and programming languages like R and Python and open source tools such as XCMS and MZmine.

High resolution peak detection was performed using “continuous wavelet transform” (centWave) (Tautenhahn, Bottcher, and Neumann 2008). This algorithm evaluates if the peak fits in a filter function. The area of the peak is calculated by integration of the fitted function in the peak which is computed based on “region-of-interest” (ROI) algorithm. The retention time correction is performed to correct for shift in the peaks because of fluctuations in room temperature, temporal changes in compounds, degradation of column and other factors that impact separation of compounds (Patti, Tautenhahn, and Siuzdak 2012). These factors are non-linear and require non-linear retention time alignment algorithm. This algorithm searches for patterns in peaks and corrects them based on deviations in retention time with local regression model referred as “ordered bijective interpolated warping” (OBI-Warp) method (Patti, Tautenhahn, and Siuzdak 2012; Prince and Marcotte 2006). This model is preferred since it allows non-linear fitting and outlier detection. Once alignment is performed back-filling algorithm “fillpeaks” is used to integrate the spectral signal. The spectral signal is normalized for statistical analysis.

2.8 Normalization

The data driven normalization performs normalization under the assumption that most of metabolites do not change without looking at the identification of these metabolites (B.A. et al. 2013). Auto scaling, pareto scaling and log transformation are applied on data before performing statistical analysis. This helped in overcoming technical errors. This aided in correctly identifying the rank of importance for metabolites and identifying most abundant metabolites (van den Berg et al. 2006).

Log transformation is used to correct for non-equal variance in the dataset, transform multiplicative relationships into additive ones and convert skewed distributions into symmetric ones to apply normality assumptions. This can be applicable where there is a drastic difference in the magnitudes for two features and there is a need to reduce the larger quantities more than the smaller ones (van den Berg et al. 2006). Log (x+1) transformation is used in case of 0 values in the dataset. This avoids infinity or not defined values in the transformed dataset. In some cases, like the volcano plot logarithm function to the base 2 is used as it allows us to interpret the results in terms of no fold change or twice fold change.

Scaling is used to minimize the spread of data points. Autoscaling applies standard deviation as the scaling factor (van den Berg et al. 2006; Van Meer, Voelker, and Feigenson 2008). Pareto scaling performs scaling with square root of standard deviation as it allows to decrease the large fold changes more than the smaller fold changes. Furthermore, pareto scaling provides transformed data with units (van den Berg et al. 2006; Wenk 2005). On the other hand, method driven normalization utilizes internal standards and QC samples for normalization. Locally estimated scatterplot smoothing (LOESS) algorithm, uses polynomial regression model to fit the QCs and obtain the signal drift based on the injection order of the QCs. This drift correction is applied all the ion features detected in all the samples (Dunn et al. 2012). In some cases, the data acquisition can span across multiple years. It is a challenge to correct for the signal drift in the ion features during intra- and inter-batch data acquisition. Support vector regression normalization can be used to normalize the ion features if the relative standard deviation (RSDs) is less than 30% for

90% of the peaks and reduce undesired intra- and inter-batch variations (Sarafian et al. 2014).

2.9 Statistical Analysis

Once the spectral features were obtained after data processing and normalization, annotation of the features was performed at different levels followed by statistical data analysis to derive insights from LC-MS data.

2.9.1 Metabolites and lipids identifications and characterizations

There is a need for accurate identifications of metabolites and lipids from untargeted LC-MS profiling data to gain insight into essential biological information. The process of assigning identification or annotation of features with confidence is challenging and time-intensive because of the chemical diversity of biomolecules as it involves identification of few hundred metabolites or lipids from over 20,000 features in some cases (Alcazar Magana et al. 2020).

High level of confidence of annotation of metabolites and lipids is achieved by using accurate mass, retention time, and MS/MS fragmentation) and online databases (Kind and Fiehn 2013). The Metabolite Standard Initiative (MSI) has defined the minimum reporting requirements for metabolite/lipid annotation (Lloyd W. Sumner et al. 2007).

According to MSI the highest level (level 1) uses matching parameters *m/z*, retention time, isotopic pattern, MS/MS spectrum of an authentic standard utilized as a reference in the same condition as the analyte of interest. This match allows to ensure the accurate identification of analytes. However, in absence chemical standard if the compound identification has been performed using a reference library second level (level 2) of annotation referred to as putative annotation. The third level of identification comprises of compound classes characterized using spectral similarity or matching chemical properties to a known compound class. The last or fourth level of identification refers to the unknown compounds wherein there is only confidence about detection a real signal coming from analyte (Lloyd W. Sumner et al. 2007). In level 3 and level 4 cases mass spectrometry is usually supported by another orthogonal analytical technique such as NMR.

Databases

The mass spectral signal needs to be matched with signals in mass spectral databases. High resolution MS helps in achieving resolving power of (10,000-450,000 fullwidth half-height maximum, FWHM) with high mass accuracy and high sensitivity (J. Castro-Perez et al. 2011). Data analysis is performed using spectral alignment, feature extraction followed by annotation using metabolic databases. The precursor ions can form clusters, adducts and fragments which result in different *m/z* values but they still have the retention time (Mahieu, Genenbacher, and Patti 2016). Clustering algorithm is used to cluster the precursor ions with similar retention time and peak shape. The algorithm is applied to all the precursor ions to identify metabolites of interest (Kuhl et al. 2012) instead of false positives. Clustering can be performed in Progenesis QI or in R with the help of Pearson correlation in CAMERA (Collection of Algorithms for MEtabolite pRofile Annotation) software package. Databases usually evaluate confidence score in annotation based on the deviation in the *m/z* of the precursor ions (based on the resolution tolerance level of 10 ppm is selected), MS/MS fragmentation similarity and isotopic pattern. The Progenesis QI software allowed us to annotate metabolites with confidence when the score of the annotation was greater than 50. The score greater than 50 was achieved when the similarity of isotopic pattern was above 80%, MS/MS score was above 50% and the accurate mass deviation from exact mass was less than 10 ppm. The MS/MS fragments were compared with databases such as LipidBlast, Human Metabolic Database for lipid annotation in Progenesis QI. In future this will be performed autonomously wherein data acquisition will be performed simultaneously with automated identification with METLIN database (Tautenhahn et al. 2012).

Statistical methods were applied to the unannotated or annotated metabolomics and lipidomics datasets to identify patterns, determine whether there are significant differences between the groups under analysis, whether the significantly different features were correlated to each other, what models can be used to predict the dependent variables, what is the importance of the independent variables in predicting the dependent variable. The models used to predict the dependent variables can range from linear classical statistical models to machine learning models.

2.9.2 Classical statistical methods

Univariate statistical tools

Univariate statistical tools are part of classical statistical methods which are used to compare whether differences between the groups were significant or not significant. The comparison for independent features was performed using t-tests for comparisons across 2 groups and ANOVA for comparisons across more than two groups (Vinaixa et al. 2012). The student t test is used to perform pairwise comparison across two independent groups. The null hypothesis states that there is no difference between the means of two groups whereas the alternate hypothesis states that there is difference between them. In this thesis a significance level of 5% was chosen which means if the probability p-value is less than or equal to 5% the null hypothesis is rejected else the null hypothesis is accepted. Student t tests are used when the distribution of data is in the normal and the variance across the groups is equal. The normality assumption is validated using Shapiro Wilk test (Vinaixa et al. 2012). If the groups bore relationship with each other, then the paired test was used. Welch t test was used in case the variance across the two groups was not equal. If the normality assumption is not satisfied, the non-parametric tests were used. Mann Whitney U Test was used for independent groups and Wilcoxon signed rank test was used for paired tests when both the groups were sampled from the same individual or the groups bore relationship with each other.

The p-value indicates there is a chance of an event occurring given there is no difference between the groups. In other words when a cut-off of 0.05 is selected there is a probability of 5% that an incorrect decision is made. If multiple tests are performed this would result in a high number of false positives because of chance. Therefore, multiple test correction is applied to correct for false positive rates (Vinaixa et al. 2012). Benjamini, Krieger and Yekutieli method in Graphpad v8 was used to determine false discovery rate adjusted p-value to prevent false positives. Analysis of variance (ANOVA) is used to compare the differences across multiple groups with normal data distribution (greater than 2) with one factor. Kruskal-Wallis is the alternate nonparametric test used to check differences across multiple groups and can be used for all other types of data distribution.

Multivariate statistical approach

Multivariate statistical approach also belongs to classical statistical method which utilizes all or some of the features concurrently to determine relationship between the features. Multivariate statistical approaches can be classified into supervised and unsupervised approaches (D. I. Broadhurst and Kell 2006; Gowda et al. 2008). The unsupervised methods do not use training and test subsets as they identify similarity patterns across the samples without accounting for their corresponding groups. However, the supervised analysis the dataset in split into training and test dataset. Supervised analysis utilizes the training dataset to identify the relationship between group or scores and predicting the group or scores of the test dataset using that relationship (D. I. Broadhurst and Kell 2006).

Unsupervised methods

The unsupervised approach is primarily useful for classification of the datasets to identify patterns such as whether the treatment group differs from the control or not. The unsupervised approaches such as principal component analysis, heatmaps are used in untargeted LC-MS metabolomics and lipidomics to look for clusters in the dataset to derive insights about whether any or a particular molecular pathway is impacted across the different groups.

Principal component analysis (PCA)

Principal component analysis reduces the dimensions of the data by transforming the correlated variables into orthogonal uncorrelated principal components. These principal components are linearly related to each other. PCA maximizes the variance obtained across the samples. The scores are assigned to each sample in the scores plot which are calculated using linear combination of the original variables. The scores plot in Figure 3.2a in chapter 3 shows that the variability in the different accessions. The samples belonging to the same accessions are clustered together however there was high variability in the samples belonging to different accessions. The biplot in Figure 2.10 shows the importance of 11 marker compounds (loading variables) on variability of accessions of *Centella asiatica* (CA) water extracts.

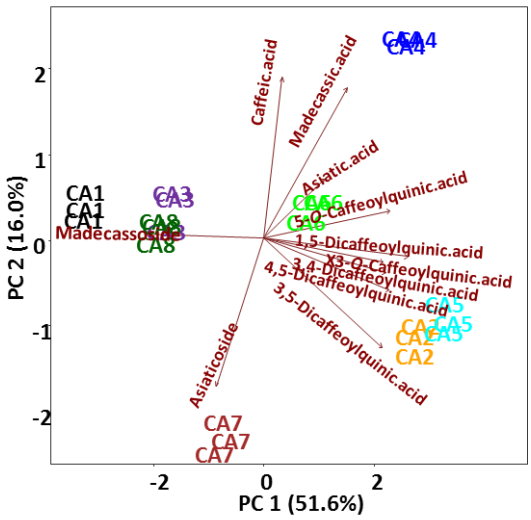


Figure 2.9 Principal component analysis (PCA) was performed using R software. PCA bi-plot represents importance of 11 marker compounds in CA water extract on variability across eight different accessions (CA1 to CA 8). PCA was performed using 5512 m/z features (Alcazar Magana et al. 2020).

Heatmaps and dendrogram

Heatmap is used visualize the differences across the different groups using gradients of colors. The upregulation and downregulation across metabolomics and lipidomics can be visualized using contrasting colors in a heatmap. The level of upregulation or downregulation is seen with a gradient of colors. The color gradient corresponds to the z-score that indicates the number of standard deviations of the value from the mean. In metabolomics and lipidomics heatmap is used to show relative abundances of *m/z* features or annotated metabolites and lipids across different groups.

Similarities and differences across metabolites and lipids and the samples in the heatmap can be visualized using a dendrogram. Dendrogram uses hierarchical clustering to cluster related features or samples into different groups. Clustering is based on a similarity measure such as Euclidean distance or Manhattan distance metric on the parameter of interest to create a tree. The most similar features were closest to each other and connected with a smallest branch whereas the most dissimilar features were farthest from each other

and were at the ends of the longest branch. The level of upregulation or downregulation of marker compounds in CA extracts are visualized in heat map in Figure 3.2d (Alcazar Magana et al. 2020), and the degree of similarity in the structure of marker compounds and CA accessions were visualized using dendrogram on x axis and y axis respectively. Pearson correlation was used to identify correlation between total ion chromatograms of CA accessions in Figure 3.2c from chapter 3. The Pearson correlation between total ion chromatograms of CA 2 and CA 6 is low on account of the two peaks appearing in CA 6 around 22 mins (Figure 3.2c). These two peaks can correspond to Madecassoside and Asiaticoside based on the heatmap in Figure 3.2d from chapter 3.

Supervised approach

The supervised approach utilizes the data to supervise and train the model using independent variables to accurately make the prediction on the test dataset. The accuracy of the model was obtained by comparing the predicted values and existing values of the dependent variable. If the dependent variables are categorical this is referred as classification and if the dependent variables are continuous this is referred as regression. In this thesis Partial least squares method has been used classical statistical supervised approach whereas machine learning algorithms such as elastic net, survival curves and random forest have been used to make model predictions.

PLS

Partial least squares use a linear model to predict the dependent variable based on the independent variable when the dependent variables are collinear. PLS utilizes dimension reduction which helps in obtaining the variables with higher importance in making accurate predictions. The principal components in PLS are linear combination of dependent variables (Tobias 1995). PLS regression is a multivariate regression technique that uses the reduced principal components to undertake least squares regression to make predictions about the dependent variable. Partial least squares discriminant analysis (PLS-DA) is a special case of PLS regression the dependent variable is categorical where PLS technique maximizes the separation between cluster of observations by rotation of PCA components that aids in understanding the importance of variables responsible for separation between

different categorical variables. PLS technique is used in biological settings to determine whether a clinical technique can accurately determine the diseased condition (Yamamoto et al. 2018) or to determine the most important independent variables for predicting the dependent variable in chapter 5.

2.9.3 Machine learning

Prediction modeling was performed by splitting the dataset into training and test dataset. To decrease the bias due to overfitting k-fold cross validation is applied which involves re-sampling of training dataset k times. Training, and test dataset were created for machine learning algorithms. All the observations were randomly sampled 70% into training and 30% into test dataset (Weng et al. 2017). The total number of observations need to be proportionately divided into training and test datasets, to prevent omission of a group accidentally in either training or test data which could result in sampling bias (train_test_split Vs StratifiedShuffleSplit n.d.). This was achieved using the StratifiedShuffleSplit function of sklearn.model_selection module in scikit-learn 0.23.1 package (Arnold et al. 2011) in chapter 5.

Cross-validation is performed to reduce overfitting of the training dataset. Overfitted model is not able to make accurate predictions on new dataset. This is resolved by using k portions of the observations successively to split the training dataset to training and test dataset to train the model k times each time on a different dataset and testing them on the final test dataset. Cross validation score is obtained using the function cross_val_score on the test dataset from sklearn.model_selection module in scikit-learn 0.23.1 package. The score obtained is greater for a model with lower overfitting. In addition to cross validation other techniques that deal with overfitting of the model are regularization where penalties proportional to the weights of learned parameters are added to the evaluation function in the process of training, obtain additional observations to train the data or use simpler models such as PLS-DA (Géron 2017).

Elastic net

Elastic net is a linear regression technique that helps in estimating a relationship between a set of dependent and independent variables by performing variable selection automatically and shrinking the coefficients of variables to select correlated dependent

variable clusters like a fishing net. In case of simple linear regression ordinary least squares method (OLS) is used to reduce the sum of the squared error between the actual and the predicted value. Elastic net technique is used in case of correlation between the independent variables. In case of correlation between the dependent variables the standard error of the coefficients is high which render the variables statistically insignificant. This problem is addressed by penalizing the OLS by using bias referred to L1 regularization and L2 regularization. Elastic net utilizes penalties of both L1 and L2 regularization (Deol 2019) and was used in chapter 3 of the thesis.

Survival curves

Kaplan Meier curves also referred to as survival curves are used to evaluate the outcome of patient post treatment or clinical intervention after different intervals of time also referred to as follow-up. The survival analysis is typically used in clinical trials to determine the effect of treatment. The time of survival analysis is referred as survival time and is defined as time taken until the occurrence of an event such as death. The survival analysis is impacted if subjects refuse to participate or do not provide accurate responses or if communication with them is lost, this category of observations are referred as censored observations and are handled by Kaplan Meier estimates (Kishore, Goel, and Khanna 2010). Kaplan-Meier estimate is the product of the probabilities of occurrence of event different intervals of time. Survival analysis also allows to distinguish subjects based on statistical difference in their outcomes (Hazra and Gogtay 2017).

During the study at each follow-up the survival probability decreases based on the number of participants experiencing an event until the study is complete. Kaplan Meier analysis handles the censored data by assuming the censored subject did not experience the event until the end of study.

Survival analysis also suffers from another drawback. The impact of censored event can be large if the number of subjects at risk are small. This effect can be visualized using 95% confidence intervals (CI). The confidence intervals become wider with the decrease in number of participants. Moreover the 95% CI also increases with time because of rise in number of censored subjects (Paemel 2019).

Random Forest

Random forest is a non-linear statistical technique that allows us to use existing features or independent variables to make a prediction on the dependent variable. Random forest is made of ensemble of decision trees. The decision tree is a non-linear model and it uses a target function created using the training set to predict the categories for all sample instances. The leaves of the decision tree (gini = 0) in other words no impurity, specify the category labels and the branches specify the combinations of the features that result in classification. The random forest algorithm carries out voting to identify the optimal solution by selecting the prediction of the decision tree that has the highest vote. The random forest algorithm utilizes gini importance to compute the importance of each feature used in prediction and was utilized in chapter 5.

Optimization of hyperparameters in machine learning

Hyperparameters need to be optimized for elastic net and random forest models. The hyperparameters can be optimized using grid and random search on the validation dataset (D. Chicco 2017). Root mean squared error score is computed in both grid and random search to identify the error associated with fitting of the model, the lower the value of RMSE the better is the model fit. Grid search will be performed using GridSearchCV from Scikit-Learn's model_selection module. The grid will have all the parameters that will be used for optimizing the model for GridSearchCV. The best parameters from grid for the test dataset will be determined using grid_search.best_params_. The hyperparameter search space is restricted in grid search it can be expanded with RandomizedSearchCV. This function will be able to search for optimum hyperparameters from wide range of values instead of being restricted to the values in the parameter grid and this function was utilized in chapter 5. The number of parameters that the model tests can be varied by changing the number of iterations and the number of iterations can be decided based on the computing resources available (Géron 2017).

2.9.4 Evaluation of performance of the predictive models

AUC-ROC

Area under the curve - receiver operating characteristics curve is important in evaluating performance of a classification model. ROC represents the probability curve whereas AUC denotes the degree of difference or in other words suggests how good the model is good at separating the dependent variable (categories). AUC of 1 represents ideal separation between the two categories whereas 0 represents that the model is making incorrect predictions. AUC of 0.5 indicates that the model cannot separate between the categories (Sarang 2018).

AUC is also widely used in machine learning to evaluate performance of machine learning models. However, machine learning models go further beyond than binary classifications. Classical AUC-ROC technique was designed for binary classification. In a collaborative project (Le et al 2021) AUC-ROC curves were used to identify biomarkers for diagnosis of number of diseased coronary arteries in adult patients with coronary artery diseases (Figure 2.11). AUC-ROC curves were also used to predict survival of adult patients with coronary artery diseases across a follow-up period of 5 years (Figure 2.12) (Le et al. 2021).

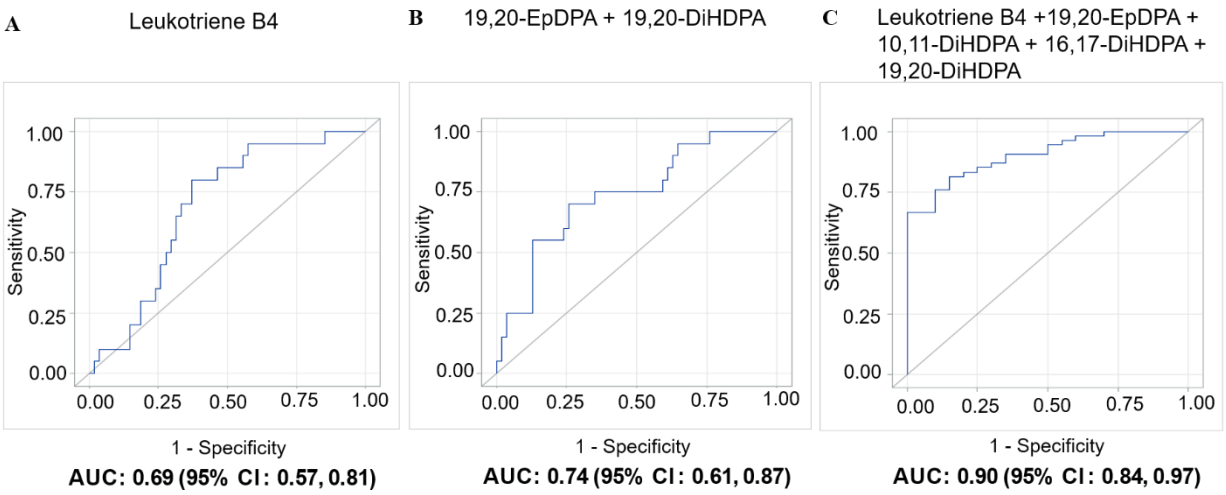


Figure 2.10 Diagnosis of number of diseased coronary arteries in adults with diseased coronary arteries (≥70% stenosis; n=74), as shown by receiver operating characteristic (ROC) curves: A) best single oxylipin model; B) best single oxylipin group model; C) smallest oxylipin panel model achieving AUC≥0.90 (Le et al. 2021).

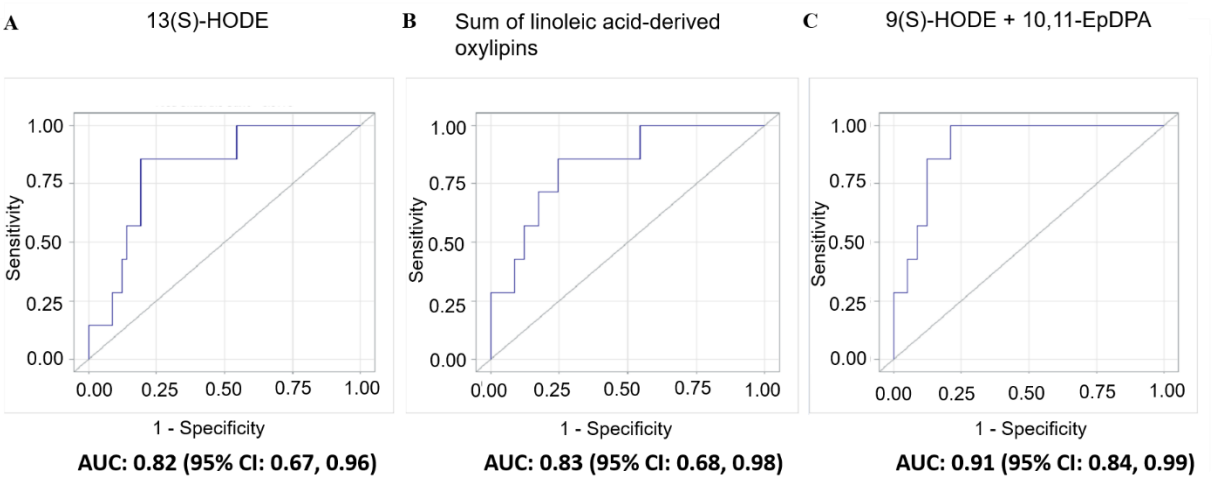


Figure 2.11 Prediction of survival during 5-year follow up in adults with diseased coronary arteries (≥70% stenosis; n=64), as shown by receiver operating characteristic (ROC) curves: A) best single oxylipin model; B) best single oxylipin group model; C) smallest oxylipin panel model achieving AUC≥0.90 (Le et al. 2021).

The need for multi-class classification in our case with similar complexity as compared to classical AUC was met with multiclass AUC (Till and Hand 2001). This method utilizes pairwise comparison to reduce multiclass AUC to binary AUC. This method was used in multiclass.roc function for multiclass comparisons in pROC V 1.17.0.1 R package (Turck et al. 2011). This function performed multiclass AUC as defined by (Till and Hand 2001) which computed mean of several AUC and could not be plotted. The multiclass.roc function was able to handle both uni-variate and multi-variate datasets. However, sometimes in clinical studies there is a need to evaluate multiple biomarkers for an assay to improve diagnostic accuracy. In case of evaluating biomarkers in plasma for coronary artery diseases (Le et al. 2021) there was need to identify panel of multiple biomarkers. CombiROC is an opensource graphic interface that utilizes combinatorial analysis to test multiple biomarker panels and determine their sensitivity, specificity, and AUC (Mazzara et al. 2017). Therefore, CombiROC method was optimized and used in this project to identify the combination of biomarkers with required sensitivity, specificity, and AUC above the threshold of 0.9.

Confidence interval

Statistical methods use a confidence interval comprising of range of values for the mean of the population instead of using a point estimate. Confidence interval was computed on metrics such as AUC-ROC, accuracy, sensitivity and F1-score. Confidence level of 95% was used for all the metrics. Confidence intervals are created with percentile and standard error method. In the percentile method 2.5^{th} and 97.5^{th} percentiles are calculated to obtain 95% percentile method. If the distribution is normal then interval ranges to 1.96 standard deviations away from the mean. Bootstrapping is defined as resampling with replacement. Initially the variable or metric for which the confidence interval needs to be generated is specified. This is followed by the process of generating 100 or 1000 replicates to create population. This process is followed by computation of summary statistics such as mean. This is followed by the final step of computing the confidence interval by percentile or standard error method. This will enable us to have confidence that accuracy of the statistical models will be present in the true population of accuracies 95% of the time (Chapter 8 Bootstrapping and Confidence Intervals | Statistical Inference via Data Science n.d.). The bootstrapping method was utilized to compute evaluation metrics of prediction models in chapter 5 of the thesis.

Evaluation of performance of ML models

Classification accuracy is defined as a ratio of accurate predictions of the category in test dataset to the total number of predictions of categories in test dataset. Accuracy is reduced because of two types of errors, type I error or false positives and type II error or false negatives. The false positives occur when the prediction denotes the positive category, but the dependent variable does not belong to that category. Similarly, false negative occurs when a prediction denotes a negative category, but the dependent variable does not belong to that category. Recall is defined as the number of right predictions for the positive category and precision represents the correct positive category predictions. Since recall and precision are both important metrics in classification it is difficult to make a choice between higher recall and lower precision or vice versa. Therefore F-measure which is defined as harmonic mean between the two is used to penalize the extreme values. Machine learning classifiers are evaluated based on classification accuracy, precision, recall and F1-score.

Confusion matrix is used to visualize the classifications made in the test dataset. Confusion matrix provides information about whether the test dataset is balanced in terms of number of observations. It enables understanding about the kind of errors the classification model is making. The confusion matrix compares true and predicted categories in the test dataset. The squares on the left diagonal correspond to correct predictions and all the other squares correspond to incorrect predictions. The confusion matrix is shown to compare the accuracy and misclassifications for machine learning model.

The machine learning models were combined with network integration to determine the bioactive compounds in *Centella Asiatica* extracts and evaluation metrics were utilized to identify the best suited models for prediction of bioactivity in chapter 3.

3. Botanical extracts - Fingerprinting and natural product drug discovery.

Part 1: Integration of mass spectral fingerprinting analysis with precursor ion (MS1) quantification for the characterization of botanical extracts: application to extracts of Centella asiatica (L.) Urban

Authors:

Armando Alcazar Magana[1,5], Kirsten Wright[2], Ashish Vaswani[1], Maya Caruso[2], Ralph L. Reed[4,5], Conner F. Bailey[1], Thuan Nguyen[6], Nora E. Gray[2], Amala Soumyanath[2], Joseph Quinn[2,3], Jan F. Stevens[4,5], and Claudia S. Maier[1*]

[1]Department of Chemistry, Oregon State University, 153 Gilbert Hall, Corvallis, Oregon 97331

[2]Department of Neurology, Oregon Health and Science University, Portland, Oregon

[3]Department of Neurology and Parkinson's Disease Research Education and Clinical Care Center (PADRECC), VA Portland Healthcare System, Portland, Oregon

[4]Department of Pharmaceutical Sciences, Oregon State University, 1601 SW Jefferson way, Corvallis, Oregon 97331

[5]Linus Pauling Institute, Oregon State University, 2900 SW Campus way, Corvallis, Oregon 97331

[6]OHSU-PSU School of Public Health, Oregon Health & Science University, Portland, Oregon

* claudia.maier@oregonstate.edu

Alcazar Magana, A., Wright, K., Vaswani, A., Caruso, M., Reed, R. L., Bailey, C. F., ... & Maier, C. S. (2020). Integration of mass spectral fingerprinting analysis with precursor ion (MS1) quantification for the characterisation of botanical extracts: application to extracts of Centella asiatica (L.) Urban. *Phytochemical Analysis*, *31*(6), 722-738.

3.1 Integration of mass spectral fingerprinting analysis with precursor ion (MS1) quantification for the characterization of botanical extracts: application to extracts of Centella asiatica (L.) Urban

3.1.1 Abstract

The metabolites synthesized in plant (also referred to as phytochemicals) possess biological activity that provides numerous potential health benefits. However, there are challenges associated with reproducibility of laboratory studies and clinical trials and identification of new metabolites as the present classical way of bioassay-guided fractionation, requires extensive purification until the isolation of a pure compound which may result in the rediscovery of already known bioactives and is also extremely time consuming. A quadrupole time-of-flight analyzer in conjunction with an optimized HPLC separation was used for in-depth untargeted fingerprinting for characterization of plant extracts and post-acquisition precursor ion quantification for determining levels of distinct phytochemicals for product integrity studies.

3.1.2 Introduction

There are more than 60,000 secondary metabolites in plants (Iason, Dicke, and Hartley 2012). These are referred as phytochemicals. They are involved in communication, signaling, defense or primary metabolic pathway regulation. There is a huge structural diversity on account of adaptations or functions performed due to evolution (Hartmann 1996).

Medicines are derived primarily from phytochemicals in large proportion of developing countries (Hartmann 1996). Phytochemicals are used by 80% of indigenous populations (Mahady 2001). Their popularity is increasing in the Western countries (R. van Breemen 2015; Mollaoğlu and Aciyurt 2013; Neiberg et al. 2011). Owing to the variations in the bioactive compounds ingestion of different doses has varying effects on the health of the patient (Dietz et al. 2017). In addition to medicines these are also used as botanical supplements to improve health. There is a huge supply of botanical medicines and supplements owing to the expanding demand. But there is an uncertainty regarding the

quality and safety of the products because of the increasing cases of toxicity with botanical supplements (Dietz et al. 2017).

Botanical supplements consist of plant metabolites or phytochemicals. There is a significant variation in the phytochemical profiles across batches because of differences in geography, plant materials, genetics, ontogenetic stage and post-harvest processing methods (Alqahtani et al. 2015; Bruni and Sacchetti 2009). The phytochemical composition has an impact on biological and pharmacological activity of botanical extracts that affects reproducibility of clinical trials. The botanical supplements comprise of botanical extracts which suffer from variability owing to the variation in the methods of extraction (Howard 2008). Extractions with use of chemicals and heat may lead to phytochemical degradation (Magana et al. 2015). The variation of composition and levels of phytochemicals in botanical supplements may impact the bioactivity of the extracts.

US Food and Drug Administration (FDA) treats the botanical extracts without claim for drug efficacy as food. Therefore, strict regulations of safety are not applied (R. van Breemen 2015). The increase in the efforts to make the characterization of unregulated over-the-counter botanical extract better, there has been advancement in the strategies to ensure authentication and consistency of plant derived supplements (R. B. Van Breemen, Fong, and Farnsworth 2008; R. van Breemen 2015; Fong et al. 2006).

There has been a shift in the trend of analyzing phytochemical preparations from thin layer chromatography or liquid chromatography coupled with UV spectrophotometric or fluorometric detection to high performance liquid chromatography coupled with mass spectrometry. This shift has happened because of high-resolution power and accurate mass measurements provided by mass spectrometry platforms along with collision-induced dissociation techniques for analyzing the chemical structures and quantification of phytochemicals in botanical extracts. The combination of high-resolution mass spectrometry and tandem mass spectrometry will create chemical fingerprint for the diverse phytochemicals used in botanical products.

Phytochemicals in the medicinal plant, *Centella asiatica* (*C. asiatica*) belonging to Apiaceae family were used as a proof-of-concept for chemical fingerprinting with the high-resolution mass spectrometry-based workflow along with annotation using phytochemical

databases. The workflow for phytochemical analysis was developed that utilized untargeted fingerprinting of aqueous extract from *C. asiatica* for accurate quantification of 8 caffeoylquinic acids, 7 flavonoids, 5 hydroxycinnamic acids and 4 pentacyclic triterpenoids. The analytical measures such as limit of detection, limit of quantification, dynamic range and reproducibility were determined for the quantification method. This allowed fingerprinting of aqueous extracts of *C. asiatica* from different accessions.

3.1.3 Experimental

Chemicals

LC-MS grade methanol and water were purchased from EMD Millipore. Formic acid ACS reagent was from Fisher Chemicals. The following certified standard compounds were used: 5-*O*-caffeoylquinic acid (1), epigallocatechin (2), catechin (3), dihydrocaffeic acid (4), 4-*O*-caffeoylquinic acid (5), 3-*O*-caffeoylquinic acid (6), caffeic acid (7), epicatechin (8), 1,5-dicaffeoylquinic acid (9), 1,3-dicaffeoylquinic acid (10), rutin (11), dihydroferulic acid (12), 3,4-dicaffeoylquinic acid (13), 3,5-dicaffeoylquinic acid (14), ferulic acid (15), 4,5-dicaffeoylquinic acid (16), naringin (17), isoferulic acid (18), quercetin (19), madecassoside (20), asiaticoside (21), kaempferol (22), madecassic acid (23) and asiatic acid (24). Compounds 2,3,6,12,15 and 18 were from Sigma Aldrich; 1,11,17,19 were from TCI America; 13 and 16 were from ChromaDex (Irvine, CA) and the rest of the compounds were from Toronto Research Chemicals.

Caffeoylquinic acids are prone to degradation or isomerization under certain conditions of pH, light exposure, and temperature (Dawidowicz and Typek 2011, 2015; Xue et al. 2016). To protect compounds from degradation, all standards and samples were prepared in methanolic solutions complemented with 0.1 % v/v formic acid and kept in the dark at -20 ˚C until analysis.

Plant materials and preparation of aqueous extracts of *Centella asiatica*

The identity of plant material was confirmed by the Department of Neurology, Oregon Health & Science University (Soumyanath et al. 2012). The preparation of the *C. asiatica* water extracts were reported previously (BOITEAU and RATSIMAMANGA 1956; Dawidowicz and Typek 2011, 2015; Gohil, Patel, and Gajjar 2010; Gray et al. 2014; Long, Stander, and Van Wyk 2012; Ramesh et al. 2014; Soumyanath et al. 2012; Xue et al. 2016). In brief, dried *C. asiatica* extracts were prepared by refluxing aerial parts of the plant (80

g per 1 liter of water) for 1.5 hours followed by cooling for 30 minutes and filtering the solution to remove plant debris. For all subsequent work the aqueous extracts were freeze-dried and stored at -80 ˚C.

For quantification of the individual compounds, a stock solution was prepared as follows. Ten mg of each freeze-dried extract powder was resuspended in 10 mL of aqueous methanol (70 % v/v with 0.1% v/v of formic acid) by sonication (30 min, 25 ˚C) and filtered with 0.22-µm PVDF Whatman filters before analysis. This procedure was used to prepare extracts from eight different accessions of the plant materials labeled from CA1 to CA8. Aliquots from each sample were mixed to generate a quality control sample (QC) used for evaluating LC-MS/MS platform performance.

Fingerprinting of Centella asiatica extracts by untargeted data-dependent analysis

For the untargeted chemical profiling analyses, high-performance liquid chromatography (HPLC) high resolution mass spectrometry (HRMS) in the data-dependent acquisition (DDA) mode was conducting using a Shimadzu Nexera UPLC system connected to an AB SCIEX TripleTOF® 5600 mass spectrometer equipped with a TurboSpray electrospray ionization source. All analyses were conducted using electrospray ionization in the negative ionization mode (ESI-). Chromatographic separation was achieved using an Inertsil Phenyl-3 column (4.6 x 150 mm). The injection volume was 10 µL and three technical replicates were carried out. A gradient with two mobile phases (A, water containing 0.1 % v/v formic acid; B, methanol containing 0.1 % v/v formic acid) was used with a flow rate of 0.4 mL min^{-1} in a 30 min run as follows: an initial one minute at 5% B, followed by 5 to 30% B from 1 to 10 minutes, then 30 to 100% B from 10 to 20 minutes, hold at 100% B from 20 to 25 minutes, and then return to 5% B from 25 to 30 minutes.

For annotating compounds in CA extract pooled CA sample were used (QC sample). Data-dependent MS/MS acquisitions in high sensitivity mode and in both ionization, modes was conducted for this purpose. For detecting negative ions, the mass spectrometer was operated using the following parameters: spray voltage -4,200 V; source temperature 550 °C; period cycle time 950 ms; accumulation time 100 ms; *m/z* scan range 100–1200, collision energy 35V with collision energy spread (CES) of 15 V. To expand the coverage of metabolites in the untargeted approach, a QC sample was also analyzed using the

positive ion electrospray mode (ESI+) and the instrument settings were as follows: spray voltage 4,500 V; source temperature 550 °C; period cycle time 950 ms; accumulation time 100 ms; *m/z* scan range 100–1200, collision energy 35V with a CES of 15 V. The mass spectrometer was equipped with a calibrant delivery system. Mass calibration was automatically performed after every fifth LC run.

Method development for quantification of selected phytochemicals in extracts

Targeted quantitative analysis was conducted for twenty-four compounds (Figure A.F.3.4) using same chromatographic conditions as mentioned for untargeted analysis. The mass spectrometric analysis was conducted in the negative ion mode (ESI-) with the following operating conditions:

spray voltage -4,200 V; source temperature 550 °C; period cycle time 950 ms; accumulation time 100 ms; *m/z* scan range 100–1200 (as described above).

External calibration was performed for 24 authentic compounds by computing area under the calibration curve of the precursor ion (MS1-based). The calibrations were performed with increasing concentrations of all compounds, 0.0, 0.005, 0.01, 0.05, 0.10, 0.50, 1.00, 5 and 10 mg/L, prepared in 70 % v/v methanol with 0.1 % v/v of formic acid. This approach was used for characterization of CAW extract as well as quantification of DCQA's and generation of MS/MS spectral data as input to create the GNPS network in part 2. SCIEX MultiQuant™ V3.0.2 analysis software was used for quantitative analysis, to calculate the peak areas under the curve for precursor ions.

Accuracy and recovery experiments

The accuracy of the MS1 method was validated by testing with three standard mixtures of known concentrations (low, 0.05 mg/L; medium, 0.50 mg/L; and high, 5.00 mg/L). Standard addition of authentic standards was carried out followed by recovery experiments for CA extracts using precursor ions wherein the quality control samples (QCs) were spiked at two different concentration levels (0.25 ng and 5 ng on-column for each compound) with standards for 24 compounds. In addition to that 1 mL of standard mix containing 0.0, 0.05 or 1.0 mg/L of each authentic compound was added to 1.0 mL of the pooled sample (200 mg dried CAW powder/L).

Application of precursor ion [MS1] quantification method for plant extracts

MS1 precursor ion quantification using the same chromatographic runs for untargeted analysis for characterization of plant extracts and product integrity studies.

Data processing and annotation of plant metabolites

The workflow shown in Figure A.F.3.1 was applied to gather putative annotations with a high degree of confidence. Apart from the typical parameters of exact mass, MS/MS spectra and isotopic pattern which are utilized to obtain level 2 annotations (Lloyd W. Sumner et al. 2007; Viant et al. 2017) manual curations were also performed that involved 1) Evaluation of the metabolite structures based on their suitability of the ionization mode in which a compound was detected (i. e., basic sites in a molecule that can be protonated in ESI+ or labile protons for ESI-), 2) Retention time for annotated features was interrogated to determine whether it matched with polarity of the molecule to avoid noise arising from compounds originating from in-source fragmentation, and 3) Selection of compounds that were previously reported in plants (listed in Table 3.1) because isolation for NMR analysis was out of the scope for this study.

The raw data processing was performed using Progenesis QI^{TM} software with Metlin^{TM} plugin V1.0.6499.51447 (NonLinear Dynamics, United Kingdom). Peak picking and tentative metabolite identifications with a specified level of confidence was conducted on raw mass spectral dataset. The annotations were performed by searching of mass spectral data against Metlin (Guijas et al. 2018), Human Metabolome Database (HMDB) (Wishart et al. 2018) an in-house compound library consisting of IROA standards (IROA Technology, Bolton, MA) and other commercially available standards (650 total). In addition to these libraries KNApSAcK (Afendi et al. 2012), PlantMAT (Qiu et al. 2016) and Chemical Entities of Biological Interest (ChEBI) (Hastings et al. 2016) online libraries were used for phytochemical annotations. The criteria used for putative annotations was score > 50 using the Progenesis QI workflow. Score greater than 50 represents isotopic pattern similarity above 80 %, MS/MS spectral data similarity > 50% and the deviation of the accurate mass from the exact mass lower than 5 ppm (Lloyd W Sumner et al. 2007). Twenty-two annotated features having high peak intensity in extracted ion chromatogram in positive ionization mode (Figure AS2) whereas twenty-four annotated high intensity mass spectral features in negative ion mode (Figure AS2).

Chemical similarity network and clustering

The chemical space network based on the compounds annotated with high confidence was built. In these networks, nodes represent compounds and edges exemplify similarity relationships based on 2D chemical structures. The Tanimoto algorithm was used for calculating measures of similarity. The derived Tanimoto coefficient represents an associative coefficient with a value ranging from 0-1, numerically expressing the structural similarity between a two-dimensional binary comparison (0 being no similarity and 1 being complete similarity) (Bero et al. 2018; Todeschini et al. 2012). The algorithm utilizes the simplified molecular input line-entry system (SMILES) as a fingerprint for the molecules being compared. The fingerprints for molecules A and B are then put into the Tanimoto algorithm: $T(A,B) = A \cap B / A \cup B$, often referred to as intersection over union. A Tanimoto coefficient greater than 0.68 indicates that the compounds being compared are structurally similar and statistically significant at the 95% confidence interval (Kim, Bolton, and Bryant 2012). A Cytoscape network was created using the Tanimoto coefficients for assigned compounds detected in an aqueous extract of *C. asiatica.*

3.1.4 Results and discussion

Untargeted fingerprinting analysis of CA extracts

Phenyl-bonded phase was used for chromatographic separation of phytochemicals in the botanical samples because the phenyl groups directly bonded to the silica surface will have π-π interactions with phenolic scaffolds in many secondary metabolites. The method utilized 30 minutes for each chromatographic run. Liquid chromatographic separation led to suitable resolution and peak capacity that was combined with DDA acquisitions for chemical fingerprinting. Quantification was conducted using the precursor ion peak are (MS1) with sufficient reproducibility by minimizing peak suppression and matrix effects.

Figure 3.1 shows a typical total ion chromatogram (TIC) for a *C. asiatica* water extract and the most intense molecular features fragmented in the DDA experiment acquired in negative ion mode. From over 20,000 *m/z*-features detected, 117 compounds were annotated after applying the workflow outlined in Figure AS1. To our knowledge, this analysis includes 87 compounds that were reported in plants but detected for the first time in *C. asiatica* extracts (Azerad 2016; Brinkhaus et al. 2000; Chandrika and Prasad Kumara

2015; Devkota et al. 2010; Fong et al. 2006; Gohil, Patel, and Gajjar 2010; Gray et al. 2018; James and Dubery 2009; Long, Stander, and Van Wyk 2012; Mustafa et al. 2010; Oyedeji and Afolayan 2005; Ramesh et al. 2014; Sangwan et al. 2013; Yoshida et al. 2005).

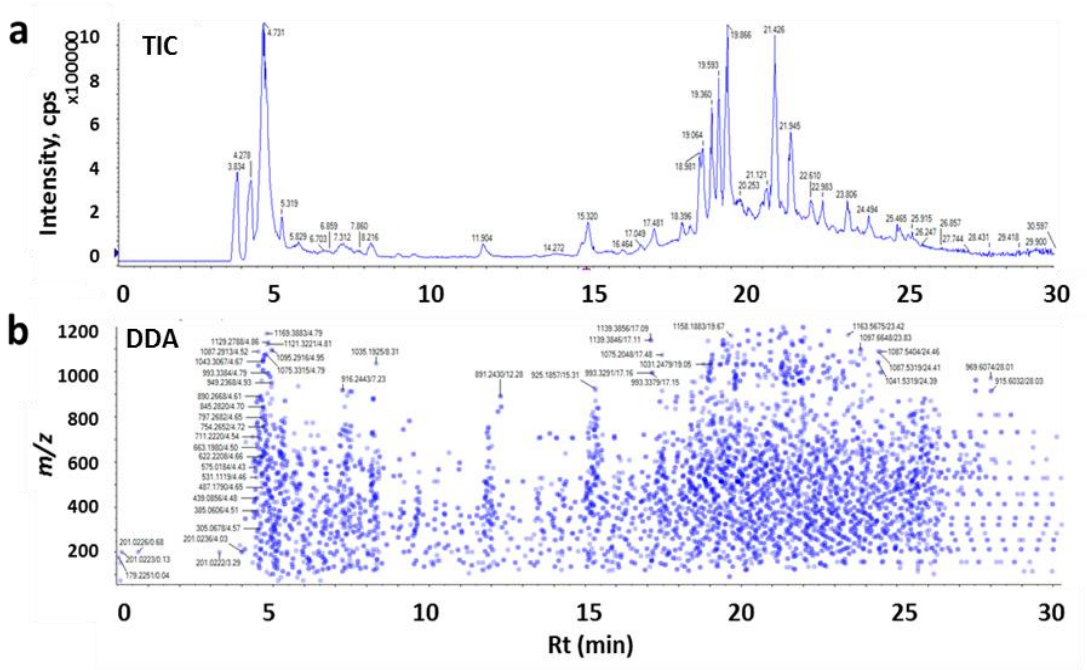


Figure 3.1 Examples of typical data obtained by untargeted analysis of a pooled Centella asiatica (CA) water extract using the data dependent acquisition mode a Total ion chromatogram (ESI−) (10 µL injection, 1 mg/L). b Distribution map of precursor ions submitted to collision induced dissociation along the elution period. The y-axis provides m/z information for the precursor ion; the x-axis represents the elution times for each precursor ion. Each one of the 5512 dots contain a fragmentation spectrum. The intensity of the blue color represents the ion abundance of the precursor ion. TIC, total ion chromatogram; DDA, data dependent acquisition.

MS/MS spectra of newly detected compounds in CA are provided in Figure AS7. Some of the most abundant compounds include six di-caffeoylquinic acid isomers, quinic acid, mono-caffeoylquinic acids, and several glycosides, such as asiaticoside, madecassoside and quercetin 3-O-glucoside. It is noteworthy that the current chromatographic separation

conditions resolved DCA isomers (Table AS1, Figure AS5). Analytical parameters namely *m/z*, retention time, detected adducts and molecular formulas for annotated compounds are shown in detail in Table S1. When compounds were detected in both ion modes, the one with highest signal to noise ratio (signal/noise) was included. A summary of annotated compounds is shown in Table 3.1, including five hydroxycinnamic acids, nine mono- and di-caffeoylquinic acids, twelve terpenoids, thirteen flavonoids, eleven glucosides among many other phytochemicals.

Table 3.1 Summary of detected compounds in C. asiatica extracts (pooled sample) by extensive querying and comparison with spectral and compound libraries (Metlin, KNApSAcK, HMDB, PantMAT, ChEBI and our in-house library) using Progenesis QI^{TM} and applying the workflow shown in Figure AS1. Compounds are labeled with their respective PubChem CID. Additional parameters are shown in Table AT3.1. Category was assigned according to structural similarity using Tanimoto algorithm, and they may correspond to more than one compound class. Compounds confirmed using authentic standards are shown in italics, all other compounds correspond to level 2 annotations. Eighty-seven compounds that were detected for the first time in C. asiatica extracts are denoted with a *.

Hydroxycinnamic acids	**CID**	**Phenolic compounds**	**CID**
Caffeic acid	689043	3-Hydroxy-2-oxo-3-phenylpropanoic acid	71581094
*Iso Ferulic acid**	736186	1-Caffeoyl-5-feruloylquinic acid*	121225501
*Dihydrocaffeic acid**	348154	3,4-Dihydroxybenzaldehyde*	8768
*Dihydroferulic acid**	14340	3,5-Dihydroxy-2-methylphenyl beta-D-glucopyranoside*	46184089
Ferulic acid	445858	3-Hydroxycoumarin	13650
Dicaffeoylquinic Acids		4-Hydroxybenzaldehyde*	126
1,3-Dicaffeoylquinic acid	6474640	5-Methoxysalicylic acid*	75787
1,4-Dicaffeoylquinic acid*	12358846	8-Acetoxy-4'-methoxypinoresinol 4-glucoside*	73830447
1,5-Dicaffeoylquinic acid	5281769	Aesculin*	5281417
3,4-Dicaffeoylquinic acid	5281780	*Catechin*	9064
3,5-Dicaffeoylquinic acid	6474310	*Coumarin*	323
4,5-Dicaffeoylquinic acid	6474309	*Epicatechin*	72276
Monocaffeoylquinic Acids		*Epigallocatechin**	72277
3-O-Caffeoylquinic acid	1794427	Folinic acid*	6006

4-O-Caffeoylquinic acid	9798666
5-O-Caffeoylquinic acid	5280633
Terpenoids	**CID**
26-(2-Glucosyl-6-acetylglucosyl]-1,3,11,22-tetrahydroxyergosta-5,24-dien-26-oate*	131752817
Asiatic acid	119034
Asiaticoside	24721205
Dysolenticin B*	56601655
Gentiopicroside*	88708
Madecassic Acid	73412
Madecassoside	91885295
Sambacin*	131752486
Swertiamarin*	442435
Tsangane L 3-glucoside*	73981648
b-Chlorogenin 3-[4'-(2'-glucosyl-3'-xylosylglucosyl)galactoside]	74193143
Shanzhiside*	11948668
Purine Derivatives	**CID**
2'-O-Methyladenosine*	102213
Adenine*	190
Adenosine*	60961
cAMP*	6076
5'-Deoxy-5'-(methylsulfinyl)adenosine*	165114
Succinoadenosine*	126969142
Guanosine*	6802
Others	**CID**
Cytosine*	597
Longicamphenylone*	91747202
Longifolenaldehyde*	565584
Uric acid*	1175
6-Docosenamide*	44584605
Deoxyfructosazine*	73452
Ginsenoyne K*	15736266
Fatty acid derivatives	**CID**
Caprylic Acid*	379
Palmitic acid*	985
Tetradecanedioic acid*	13185
16-Hydroxypalmitic acid*	10466
Traumatic Acid*	5283028
12-Oxodihydrophytodienoic acid*	5716902
Ginkgoic acid*	5281858
Kuwanon Y*	14334307
Kynurenic acid*	3845
N1,N5,N10,N14-Tetra-trans-p-coumaroylspermine*	9810941
Phlorin*	476785
Tropic acid*	10726
Xanthurenic acid*	5699
Amino acid derivatives	**CID**
2-Pyrrolidone-5-carboxylic acid	499
1-beta-D-Glucopyranosyl-L-tryptophan*	11772967
2,6-Piperidinedicarboxylic acid*	557515
4-Guanidinobutanoic acid*	25200642
5-Methoxy-L-tryptophan*	151018
6-Amino-9H-purine-9-propanoic acid*	255450
6-Oxo-2-piperidinecarboxylic acid*	3014237
L-Arginine*	28782
N-(1-Deoxy-1-fructosyl)phenylalanine*	101039148
N-Acetyl-L-glutamic acid*	70914
Niacin (Nicotinic acid)*	938
Pantothenic Acid	6613
Succinyl-L-proline*	194156
Vincosamide*	10163855
Organic Acids	**CID**
Citric acid*	19782904
L-Ribulose*	644111
Malate*	20130941
Succinate*	1110
Glucoside derivatives	**CID**
Stachyose*	439531
Carlosic acid methyl ester*	122391261
Daucic acid*	5316316
Digalacturonate*	439694
Dihydroactinidiolide*	27209
Furaneol 4-(6-malonylglucoside)*	131750900
Isovalerylglucuronide*	137383
Linustatin*	119301
Purgic acid B*	16091605
Flavonoids	**CID**

Nomilinic acid 17-glucoside*	444212	3,5-Dihydroxyphenyl 1-O-(6-O-galloyl-β-D-glucopyranoside)	131752603
Choline derivatives	**CID**	6-C-α-L-arabinosyl-8-C-β-L-arabinosylapigenin	122391238
Betaine*	247	Apimaysin*	101920411
Choline*	305	Astragalin	5282102
Choline O-Sulfate*	486	*Kaempferol*	5280863
Phosphocholine	1014	Mangiferin*	5281647
Amino sugar derivatives	**CID**	*Naringin*	442428
Enicoflavine*	5281564	Pelargonidin 3-O-glucoside*	443648
Muramic acid*	433580	*Quercetin*	5280343
N-Acetyl-D-glucosamine*	899	Quercetin 3-(6'-acetylglucoside)*	44259187
Soyacerebroside I*	131751281	Quercetin 3-O-glucoside*	5280804
D-1-[(3-Carboxypropyl)amino]-1-deoxyfructose*	131752417	*Rutin*	5280805
		Glabraoside A*	102393599

Principal component analysis (PCA) was used to obtain the similarities and differences in metabolite composition across eight available CA accessions (Figure 3.2) using the same *m/z* features containing MS/MS spectral information (presented in Figure 3.2b). PCA showed significant differences across the CA accessions. The PCA scores plot allowed visualization of different clusters for CA preparations. The PCA loading plots show the constituents with more variability among the *C. asiatica* accessions. Some of the marker compounds in *C. asiatica* are highlighted in Figure 3.2b. Di-caffeoylquinic acids and triterpenes are among compounds that showed high variation between the different accessions.

For additional contrasting of CA accessions, we also used as tools a correlation matrix and a heatmap based on area under the curve for extracted ion peaks. The correlation matrix aids in evaluating similarities and dissimilarities of extracts based on the correlation score (Figure 3.2c). Higher correlation scores (between 0.75 and 1) are indicated by red, scores between 0.74 and 0.51 are indicated by white and scores between 0.5 and 0.25 are represented by purple and lower correlation scores (less than 0.25) are represented by blue. The Pearson correlation value calculated between CA6 sample and CA2 sample is 0.27 which indicates that there is a little linear relationship between CA2 and CA6. The Pearson correlation value calculated between CA2 and CA1 samples is 0.48 which also indicates

fairly less amount of linear relationship between CA2 and CA1 samples. The Pearson correlation between CA6 and CA4 samples is 0.87 which indicates that these two samples are linearly related indicating similarities in metabolite contents for CA4 and CA6 extracts. The heatmap with hierarchical clustering (Figure 3.2d) visualizes the precursor ion peak area for 14 compounds evaluated in the CA extracts. Peak areas were averaged across three replicates. The dendrogram on the y-axis indicates the degree of similarity or differences between the CA compound levels in the CA accessions, e.g. CA3 and CA8 are closer in the clustering tree and C6 is separated.

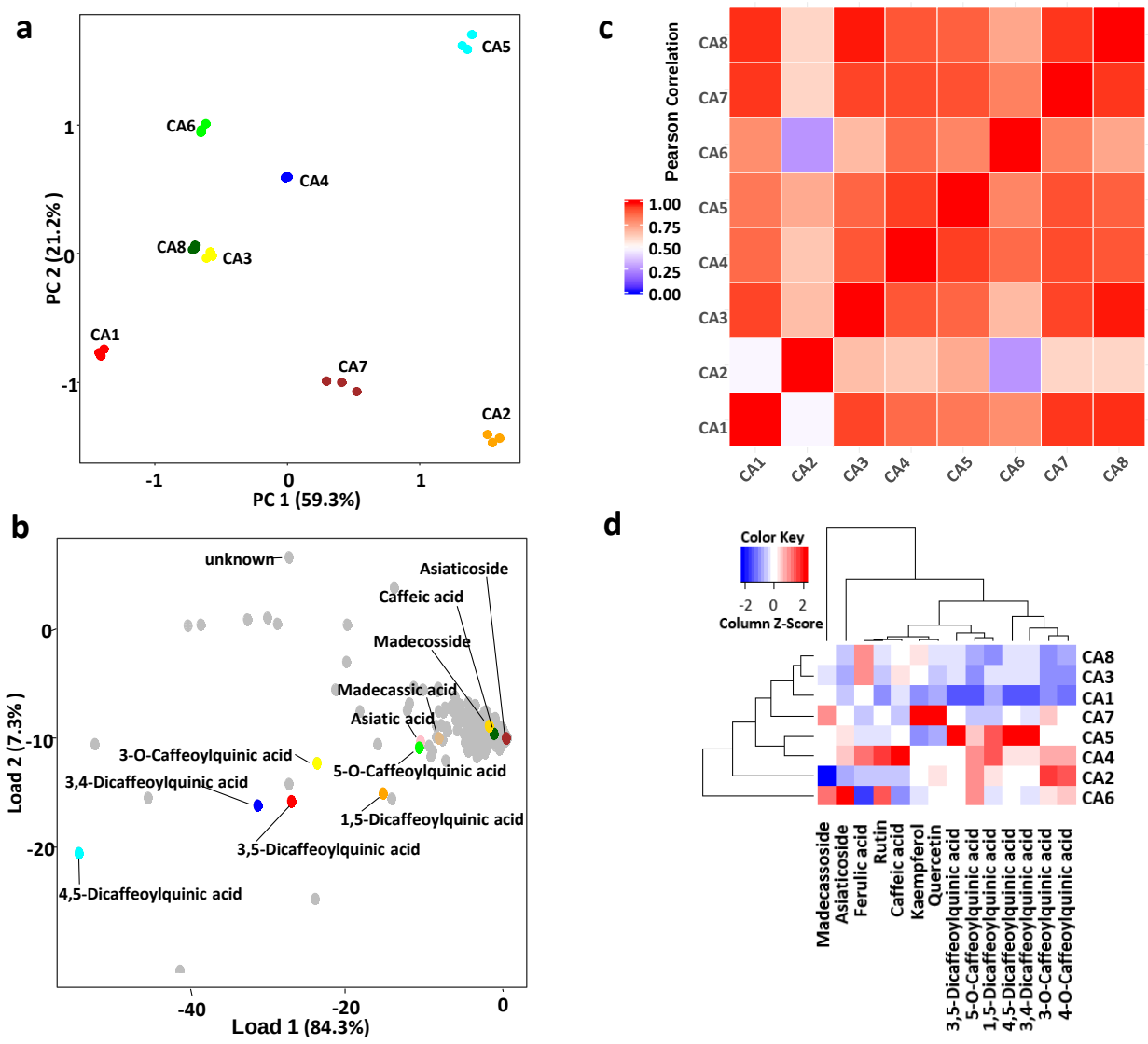


Figure 3.2 Analysis of similarities and dissimilarities of Centella asiatica (CA) water extracts. (A) Scores plot, each set of dots (technical triplicates) represents a CA extract from eight different accessions (CA1–CA 8). (B) Loadings plot indicating 11 selected compounds with higher concentration across quantified phytochemicals. Di-caffeoylquinic acids and triterpenes change across accessions. Principal component analysis (PCA) was performed using 5512 m/z features that provided MS/MS information (negative ion mode) and were consistently found in all CA extracts. (C) Correlation matrix between different C. asiatica accessions based on 5512 m/z features as the PCA. (D) Heatmap visualizing area under the curve for the chromatographic peaks of the compounds. The area under the curve has been averaged across three replicates. The colors in the heatmap indicate the z-

score which was calculated by subtracting the mean of the peak areas for a metabolite across different samples and dividing it by the standard deviation of the metabolite across all the samples. The red color indicates positive zscore, the white color indicates zero z-score, whereas the blue color indicates negative z-score. Higher intensity of the color in the scale indicates a higher magnitude of the z-score. The dendrogram on the x-axis indicates the degree of similarity between the metabolites, the closer the metabolites the higher the level of similarity in them and the metabolites have been clustered using hierarchical clustering. Similarly, the dendrogram on the y-axis indicates the degree of similarity between the different samples (different CA accessions), the closer the samples the higher the level of similarity in them and they have been clustered using hierarchical clustering (Ward, Euclidean distance). PCA was performed using MetaboAnalyst V4.0.

Structural similarity network

Figure 3.3 shows a 2D structural similarity network of 117 assigned compounds found consistently in the aqueous extract of all eight CA accessions, that was created using the Tanimoto similarity score (Todeschini et al. 2012).

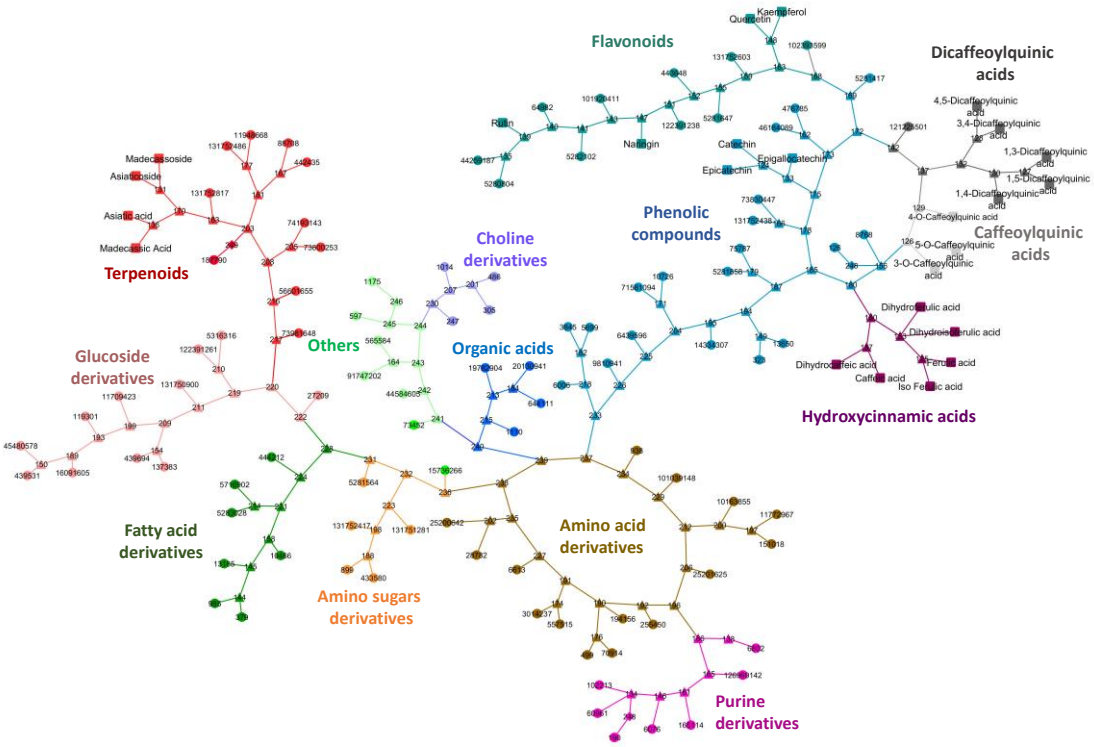


Figure 3.3 Cytoscape network for 117 assigned compounds in Centella asiatica described in Table 3.1. Cytoscape network for 117 assigned compounds in Centella asiatica described in Table 3.1. The clustering relationship is based on the molecular input line-entry system (SMILES) as a fingerprint for the molecules being compared (Tanimoto coefficient). A Tanimoto coefficient greater than or equal to 0.68 indicates that the compounds being compared are structurally similar and statistically significant at the 95% confidence interval. In this Cytoscape network, compounds are indicated by circular terminal nodes and labelled with their respective PubChem ID. Identified compounds, i.e. compounds for which authentic standards were available (Level 1 annotations), are indicated by a square node. Tanimoto scores greater than or equal to 0.68 are represented by triangular branch nodes, while scores less than 0.68 are depicted by diamond shaped branch nodes.

Compounds were arranged in 14 interconnected clusters that are structurally similar at the boundary nodes at the 95% confidence level. Compounds fall into the following clusters: fourteen amino acid derivatives, five amino sugar derivatives, four choline derivatives, six di-caffeoylquinic acids, seven fatty acid derivatives, 13 flavonoids, nine glucoside

derivatives, five hydroxycinnamic acids, three mono-caffeoylquinic acids, four organic acids, 21 phenolic compounds, seven purine derivatives, 12 terpenoids, and seven others compounds. It is noteworthy that classification was established according to structural similarity (Tanimoto algorithm), consequently some compounds may belong to more than one compound class.

The enzymatic machinery required to produce secondary metabolites in most plants is largely uncharacterized (Moghe and Last 2015). The lack of intermediate secondary metabolites along the metabolic pathways makes this characterization more complex. The detection of secondary metabolites never reported in plant extracts, in conjunction with a Cytoscape network clustered according to the structural similarity (Tanimoto algorithm), can potentially help with the characterization of new metabolic pathways by searching for potential enzymes responsible for the interconversion of metabolites clustering together. While *C. asiatica* is most known as a rich source of pentacyclic triterpenoids (Brinkhaus et al. 2000; Long, Stander, and Van Wyk 2012), relatively high percentages of caffeoylquinic acids and flavonoids have also been identified (Brinkhaus et al. 2000; Devkota et al. 2010). These secondary metabolites, specifically phenylpropanoid derivatives, have been associated with *C. asiatica's* anti-inflammatory, antioxidant, or other biological activities (Korkina LG 2007; Wink 2013).

Accurate quantification of phytochemicals in extracts using precursor ion (MS1) quantification

The quantification of phytochemicals was performed with a suitable chromatographic separation in conjunction with HPLC-MS/MS DDA acquisition for screening of compounds, combined with the use of molecular ion extraction for quantification of selected compounds in the same chromatographic run. From over 20,000 recorded *m/z* features, 5512 and 6906 most prominent *m/z* features acquired in negative (Figure 3.1) and positive ion mode, respectively, were fragmented in the DDA experiment. This untargeted approach provides a spectral library for thousands of potential compounds that can be mined in future applications. From the 117 assigned compounds (Table 3.1), 24 phytochemicals were selected as a proof of concept for quantification, including three mono-caffeoylquinic acids, five di-caffeoylquinic acids, seven flavonoids, five hydroxycinnamic acid derivatives and four triterpenes (structures shown in Figure. AS4).

Method validation for selected compounds

For accuracy, precision, repeatability, linearity, limit of detection (LOD), limit of quantification (LOQ) and range, the proposed method followed the typical validation procedure in accordance with the ICH Harmonized Tripartite Guideline (ICH 2012).

Figure AS5 shows LC-MS extracted ion chromatograms (XIC) obtained from authentic standards of these 24 compounds. Analytical parameters, namely $[M-H]^-$ *m/z*, retention time, accuracy for three different concentration levels, LOD, LOQ and inter-day coefficient of variation (RSD) were established for 24 phytochemicals using precursor ion extraction (Table 3.2). The analytical accuracy for three known concentration samples at the low (0.05 ppm), medium (0.50 ppm) and high (5.0 ppm) calibration curve intervals ranged from 87-125% (Table 3.2). The RSD was measured for a solution of 1 mg/L and ranged from 6.8 to 24% for 9 repetitions measured in a span of 6 months (Table 3.2).

Table 3.2 Analytical parameters for authentic standards. Exact m/z used for XIC, retention times, detection and quantification limits, % of accuracy for three concentrations, and % of relative standard deviation for 24 selected compounds. Compounds are sorted by retention time.

Compound	[M-H]⁻ [a]	Rt [b] (min)	LOD [c] (µg/L)	LOQ [d] (µg/L)	Low QC [e] 0.05 mg/L	Medium QC [e] 0.5 mg/L	High QC [e] 5 mg/L	%RSD [f]
5-*O*-Caffeoylquinic acid	353.0867	11.85	0.078	0.260	109	108	93	13.11
Epigallocatechin	305.0656	13.48	0.146	0.485	101	104	99	21.70
Catechin	289.0707	14.48	0.018	0.060	123	106	98	11.54
Dihydrocaffeic acid	181.0495	14.85	0.005	0.015	109	117	93	11.18
4-*O*-Caffeoylquinic acid	353.0867	15.08	0.040	0.134	125	116	93	12.13
3-*O*-Caffeoylquinic acid	353.0867	15.30	0.040	0.134	122	116	97	17.33
Caffeic acid	179.0339	15.83	0.319	1.064	111	110	99	11.71
Epicatechin	289.0707	16.80	0.013	0.045	107	100	98	12.52
1,5-Dicaffeoylquinic acid	515.1184	17.49	0.061	0.202	110	104	97	15.32
1,3-Dicaffeoylquinic acid	515.1184	17.49	0.061	0.202	110	104	97	15.32
Rutin	609.145	18.96	0.016	0.052	107	102	100	11.65
Dihydroferulic acid	195.0652	19.02	0.027	0.089	123	104	96	8.98
3,4-Dicaffeoylquinic acid	515.1184	19.11	0.169	0.562	105	99	100	15.30
3,5-Dicaffeoylquinic acid	515.1184	19.45	0.166	0.552	106	100	102	15.25
Ferulic acid	193.0495	19.55	0.001	0.004	111	105	93	6.79
4,5-Dicaffeoylquinic acid	515.1184	19.92	0.108	0.358	101	99	102	12.09
Naringin	579.1708	20.07	0.025	0.083	97	100	88	10.66
Iso Ferulic acid	193.0495	20.42	0.001	0.003	94	102	103	16.39
Quercetin	301.0342	21.10	0.068	0.227	111	100	97	12.15

Madecassoside	973.5003	21.47	0.008	0.025	110	97	87	16.94
Asiaticoside	957.5054	21.97	0.001	0.004	102	99	98	18.87
Kaempferol	285.0390	22.01	0.028	0.092	112	103	97	10.46
Madecassic Acid	503.3367	23.74	0.009	0.031	97	100	97	11.41
Asiatic acid	487.3418	24.41	0.005	0.015	96	95	92	12.61

[a]Exact mass, negative Ionization. Mass error detected <5 ppm
[b]Retention time
[c]Calibration detection limit evaluated as S/N ratio 3:1
[d]Calibration quantification limit evaluated as S/N ratio 10:1
[e]Quality control sample
[f]%RSD Measured for 1 mg/L. 9 measurements along 6 months span

The high-resolution accurate mass data led to comprehensive fingerprinting for botanical extracts for further interrogation post acquisition to obtain accurate quantification of phytochemicals by extracting the precursor ions and using the area under the peak for quantification [M1 quantification] in the same analytical run. Quantified compounds showed good linearity over three orders of magnitude (0.005 to 5.0 mg/L, r>0.990, Table AF3.2).

Matrix effect is a common drawback frequently observed in analytical procedures analyzing complex natural product mixture sample. The matrix effects were evaluated in the CA extracts, by pooling CA extract samples and spiking with the 24 available standards. A total ion chromatogram obtained for a CA extract and the same sample after standard addition is shown in Figure AS6. For plant extracts, recoveries of individual compounds ranged from 71 to 144% and 91 to 132% for 0.25 and 5.0 ng on-column, respectively (Table 3.3), confirming the feasibility of the proposed procedure for quantitative analysis of CA extracts.

Table 3.3 Recovery experiment. Recovery percentage and mean concentration of individual quantified compounds measured in a pooled CA sample (100 mg/L) using precursor ions with respective standard deviations obtained without standard addition and after addition of a mixture of 24 standards in two different concentration levels (0.25 and 5 ng of each standard on-column). All measurements are given in nanograms.

Compound	QC (ng on column)	QC +0.25 ng standards	% Recovery	QC +5 ng[a] standards	% Recovery

5-O-Caffeoylquinic acid	1.02±0.01	1.28±0.01	102	6.53±.46	110
Epigallocatechin	<LOQ[a]	0.18±0.02	71	4.55±0.20	91
Catechin	<LOQ	0.33±0.09	134	5.88±0.32	117
Dihydrocaffeic acid	<LOQ	0.28±0.01	113	5.34±0.27	107
4-O-Caffeoylquinic acid	0.88±0.02	1.17±0.03	115	6.31±0.09	109
3-O-Caffeoylquinic acid	2.45±0.03	2.79±0.09	133	7.68±0.10	105
Caffeic acid	0.67±0.07	0.97±0.03	123	6.06±0.43	108
Epicatechin	<LOQ	0.30±0.06	119	5.57±0.14	111
1,5-Dicaffeoylquinic acid	0.38±0.01	0.68±0.06	121	5.91±0.28	110
1,3-Dicaffeoylquinic acid	0.38±0.01	0.68±0.06	121	5.91±0.28	110
Rutin	0.04±0.01	0.31±0.02	106	5.44±0.35	108
Dihydroferulic acid	<LOQ	0.25±0.01	101	5.71±0.17	114
3,4-Dicaffeoylquinic acid	3.23±0.02	3.57±0.02	136	8.86±0.17	113
3,5-Dicaffeoylquinic acid	3.78±0.02	4.14±0.03	144	8.94±0.50	103
Ferulic acid	0.11±0.01	0.40±0.04	118	5.59±0.11	109
4,5-Dicaffeoylquinic acid	3.93±0.01	4.16±0.05	92	9.16±0.10	105
Naringin	<LOQ	0.29±0.03	117	5.86±0.84	117
Iso Ferulic acid	0.18±0.01	0.52±0.04	135	6.35±0.21	123
Quercetin	0.29±0.22	0.56±0.17	109	5.94±0.38	113
Madecassoside	25.5±0.95	25.77±0.07	108	31.6±2.20	110
Asiaticoside	10.74±0.12	11.0±0.16	114	16.1±0.55	108
Kaempferol	0.31±0.02	0.62±0.01	121	5.79±0.49	109
Madecassic Acid	1.35±0.01	1.65±0.03	120	7.58±0.24	124
Asiatic acid	0.62±0.02	0.97±0.28	138	7.25±0.53	132

[a]Calibration quantification limit evaluated as S/N ratio 10:1

A range of three orders of magnitude is typical for TOF analyzers, which is one of the disadvantages when we compare with triple quadrupole analyzers which usually feature a linear dynamic range that extends over six orders of magnitude. Nevertheless, the high resolution allows us to obtain the fingerprint and untargeted characterization in the same analytical run, saving instrument time, solvents and avoiding sample degradation.

For the developed quantification method, the combination of an optimized separation method with a high-resolution q-TOF mass spectrometer allowed the detection and quantification of phytochemicals in plant extracts at sub-parts per billion levels (except caffeic acid; LOQ 1.06 µg/L) with minimum sample processing. Modern Q-TOF mass spectrometers possess sensitivities typically associated with MS/MS-based selected reaction monitoring (SRM) methods. Reported values for 15 phenolic acids and 17 flavonoids (range from 3.4 to 228 µg/L) (Yilmaz et al. 2018) acquired using SRM in a

triple quadrupole mass spectrometer are comparable with our limit of detection. In addition, contemporary q-TOF instruments offer mass resolving power typically ≥25,000 (FWHM) at *m/z* 195. These q-ToF platforms obtain highly accurate mass measurement with high resolution for precursor ions and fragment ions. The high-quality data provided structural characterization and quantification of phytochemicals in complex mixtures with high confidence.

Quantification of phytochemicals in CA extracts from different sources

In previous studies, comparison of secondary metabolite production of *C. asiatica* was limited to four triterpenoids (asiatic acid, madecassic acid and their glycosides asiaticoside and madecassoside) (Das and Mallick 1991; Devkota et al. 2010; Madhusudhan, Neeraja, and Devi 2014; Müller et al. 2013; Randriamampionona et al. 2007; J. Singh et al. 2015; Yuan et al. 2015). Some other compounds (flavonoids and caffeoyl esters) were analyzed by LC-MS (Gray et al. 2014) and HPLC-DAD (Alqahtani et al. 2015; Devkota et al. 2010; Inamdar et al. 1996; S. Singh et al. 2014). A comparative study of nutrient content and yield performance of *C. asiatica* at different harvesting periods was reported. The study focused on yield measured by dry weight of leaves and some nutrients were compared (P,K, S, Ca, Mg, Zn, Cu, Fe, Mn and N) (Ramesh et al. 2014).

Our study shows that CA extracts were particularly rich in mono-caffeoylquinic acids (CQA), such as 3-,4- and 5-CQA, and di-caffeoylquinic acids (DCQA), such as 1,3; 1,5; 3,4; 3,5 and 4,5-DCQA, as well as some triterpenoids such asiaticoside, madecassoside and their aglycones (Table 3.3, Figure 3.4 and 3.5). CA extracts also contained several flavonoids and hydroxycinnamic acid derivatives. Figure 3.4 shows extracted ion chromatograms for 18 selected compounds quantified in CA water extract using area under the curve of the precursor ion acquired in DDA mode.

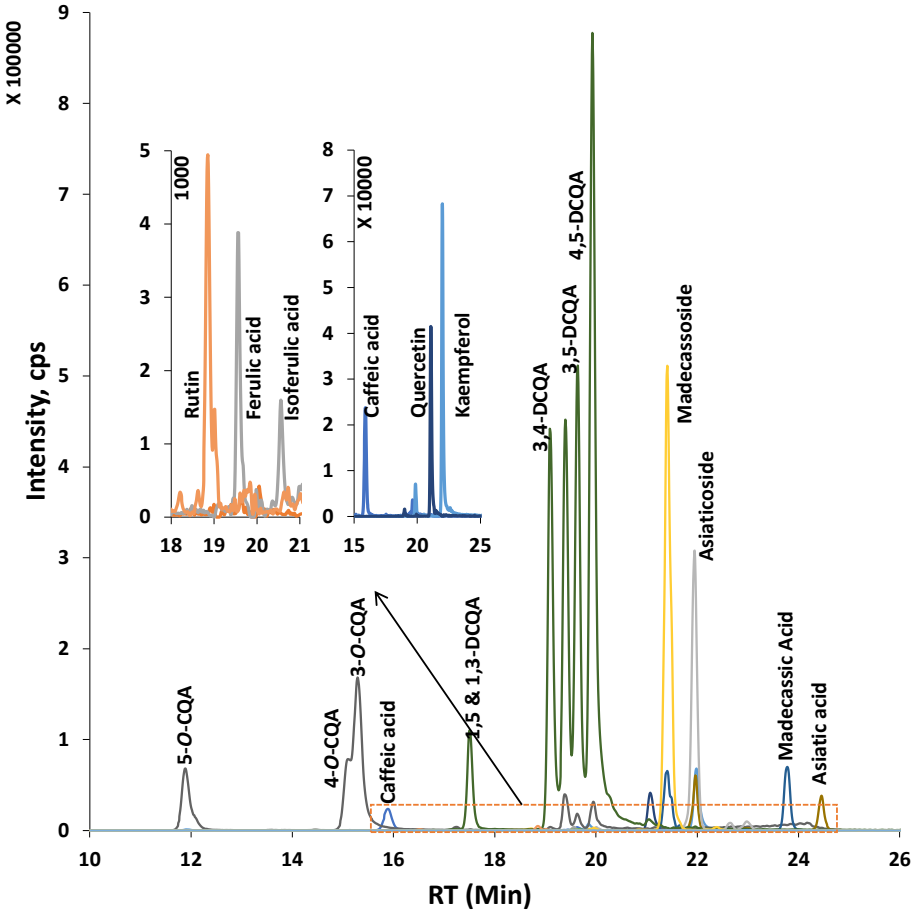


Figure 3.4 Extracted ion chromatograms (XICs) of 18 compounds that were used for precursor ion (MS1) quantification. Individual analytical parameters are shown in Table 3.2. XICs were obtained using the data-dependent acquisition (DDA) (ESI−) mode obtained for a pooled Centella asiatica (CA) sample.

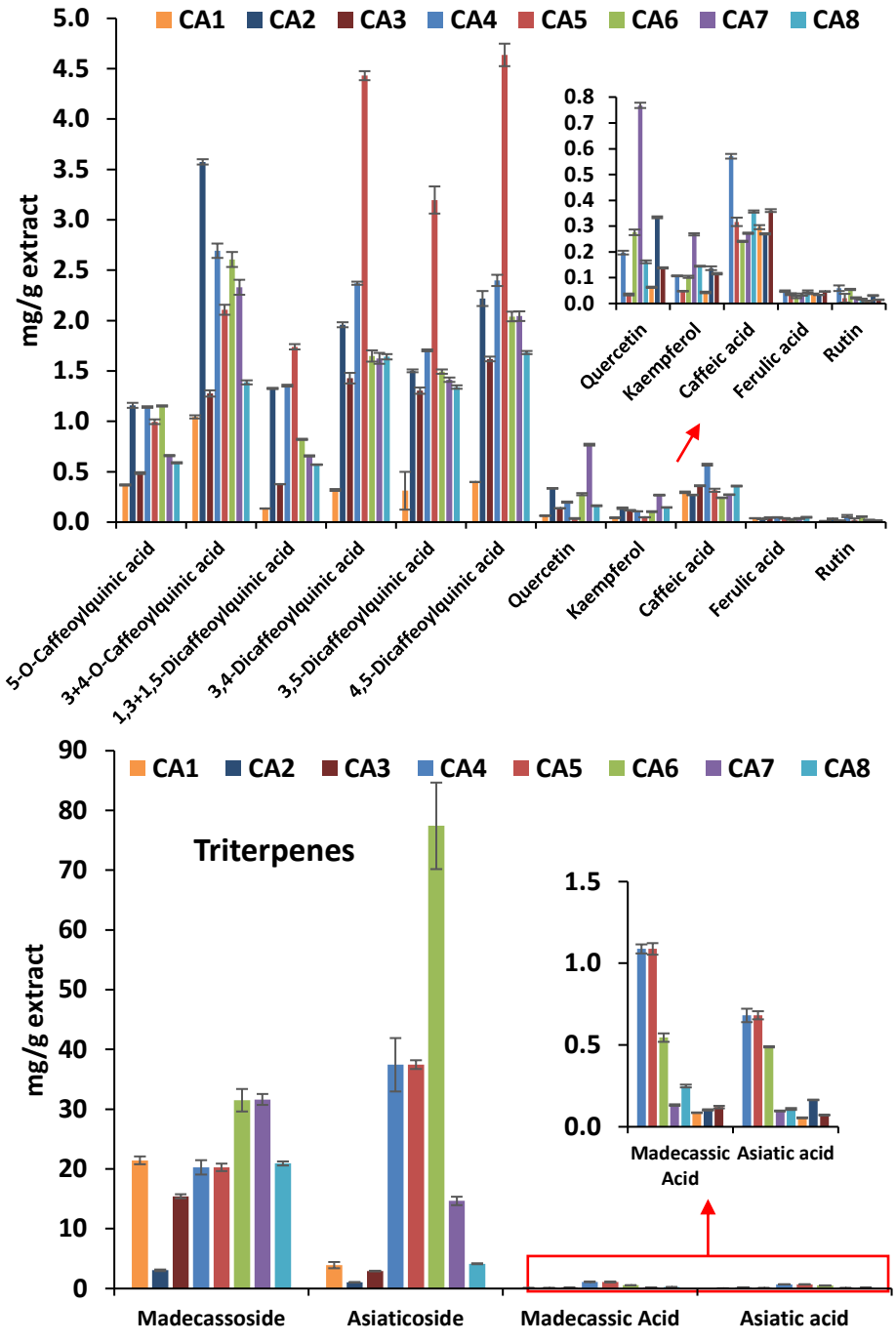


Figure 3.5 Analysis of eight different C. asiatica accessions (water extract) by precursor ion (MS1) quantification (SE from triplicates). Results are presented as mg/g of dry extract. A principal component analysis from the same analytical runs are shown in Figure 3.2.

Under the HPLC conditions used, 1,5 and 1,3-DCQA co-elute since they are *cis-trans* isomers (structures shown in Figure AS4). For all other compounds quantified in this study, the combination of suitable separation conditions with extraction of ion chromatograms at individual *m/z* values (Figure AS5, Table 3.2) enabled detection limits in the low nanomolar to picomolar range for 24 phytochemicals using precursor ions (Tables 3.2 and AS2). The optimized analytical procedure minimized interferences by improving the chromatographic separation of isomers with similar fragmentation patterns and thereby also optimized detection limits.

In this study eight different accessions of *C. asiatica* were quantified using precursor ion (MS1) quantification. We detected differences in concentration close to twenty-fold change for some compounds across the analyzed accessions. For instance, if we compare samples CA6/CA1, asiaticoside fold change was 19.9, asiatic acid 9.1, and madecassoside 1.5. In the case of di-caffeoylquinic acids, the ratio of CA5/CA1 for 4,5-dicaffeoylquinic was eleven-fold change. Mono-caffeoylquinic acids presented less variation across the accessions. For 5-*O*-caffeoylquinic acid, the highest difference was CA2/CA1 with a 2.7-fold change (Figure 3.5). This emphasizes the importance of establishing rigorous analytical procedures for botanical extracts and supplements to ensure product integrity and batch to batch reproducibility.

To conclude, the combination of suitable separation conditions with mass spectral data acquired with high resolving power using DDA acquisition enables the extraction of high-resolution accurate mass precursor ions with exact *m/z* values presenting deviations smaller than 5 ppm that in turn allows accurate quantification of phytochemicals with limits of quantification at 1.06 µg/L or lower. The described method was validated for the quantification of a) seven flavonoids, b) three structural isomers of caffeoylquinic acids, c) five di-caffeoylquinic acids, d) five caffeic acids derivatives and e) four terpenoids. The phytochemical composition of metabolites across different CA accessions was substantial, demonstrating that standardization and detailed characterization of plant extracts are prerequisites for reliable and reproducible studies aiming to determine the biological activity of CA and botanical extracts in general.

Overall, the first part of the study underscored the need for methods to efficiently analyze highly complex plant extracts to support the standardization of botanicals destined for

preclinical studies, clinical trials, and commercial products. The second part dealt with identification of neuroprotective bioactive in *Centella Asiatica* water (CAW) extract.

Part 2: High-throughput screening for plant bioactives using computational methods and LC-MS spectral networks. Application to Centella asiatica extracts.

Authors:

Armando Alcazar Magana[1], Ashish Vaswani[1], Kevin S. Brown[1], Wenbin Wu[1], Luce M. Mattio[1], Cristobal L. Miranda[1], Yuan Jiang[1], Paul Cheong[1], Amala Soumyanath[2], Jan F. Stevens[1], Claudia S. Maier[1*]

1. Oregon State University, Corvallis, OR
2. Oregon Health & Science University, Portland, OR

Planning to submit the manuscript to Journal of natural products.

3.2 High-throughput screening for plant bioactives using computational methods and LC-MS spectral networks. Application to Centella asiatica extracts.

3.2.1 Abstract

On the other hand, rapid screening of botanical extracts for novel natural products was performed using fractionation approach in conjunction with flow-injection mass spectrometry for obtaining chemical fingerprints of each fraction allowing the correlation of the relative abundance of molecular features representing individual phytochemicals with reads out of bioassays. A proof of concept study was conducted using *Centella asiatica* (*C. asiatica*), an Ayurvedic herb which has been associated with improving mental health and cognitive function for natural product discovery studies. In addition to that, subfractions of *C. asiatica* were exposed to human neuroblastoma MC65 cells to evaluate protective benefit derived from these subfractions against amyloid β-cytotoxicity. The % viability score of the subfractions was used in conjunction with molecular features obtained from mass spectrometry in computational models ElasticNet, PLS-DA and random forest to derive relationship of peak intensity of molecular features and % viability. The correlation of mass spectral features with MC65 protection and their presence in different sub-fractions were also visualized using GNPS molecular networking. All the models unequivocally identified di-caffeoylquinic acids to provide the highest MC65 protection which was also validated by classic reductionistic approach.

3.2.2 Introduction

The challenge of working with natural products in botanical extracts is finding the potential bioactives against specific diseases, in the traditional way —bioassay-guided fractionation— and in such a huge diversity is a tedious and time-consuming task (Nothias et al. 2018; Weller 2012). Bioassay-guided fractionation leads to separation of certain metabolites based on physicochemical properties —such as polarity, and assessing the bioactivity in a step-by-step fashion. There is also a possibility that after sequential steps of purification and assaying for finding the bioactive (Abbas-Mohammadi et al. 2018; Nothias et al. 2018; Shine et al. 2020; Stagliano et al. 2010; Weller 2012), a compound identified has already been discovered. There are risks associated with traditional

exhaustive fractionation up until the isolation of pure compound as this could lead to compound degradation and loss of synergies among compounds (Atanasov et al. 2015).

Aqueous extracts of *C. asiatica* are known to make improvements in memory and mental health (Brinkhaus et al. 2000; G.K., Muralidhara, and M.S. Bharath 2011; Gray et al. 2018; Kapoor 2018). Use of *C. asiatica* preparations in complementary medicine has been associated with improvement of memory in cognitive decline owing to aging and Alzheimer's disease (Gray et al. 2018; Kumar, Dogra, and Prakash 2009; Soumyanath et al. 2012). The objective of this study was to expedite bioactive identification that inhibit amyloid β-cytotoxicity primarily responsible for Alzheimer's disease in extracts from *C. asiatica* (Gray et al. 2014) utilizing orthogonal analytical techniques assisted by bioinformatics and statistical tools. In this effort, it is imperative to support reliable dereplications tools to optimize resources and prevent the rediscovering of known bioactives. One of the solutions identified to ease these concerns, was Global Natural Product Social Molecular Networking (GNPS) platform (Aron et al. 2019; Mingxun Wang et al. 2016) that can assist in the dereplication and annotation of specialized metabolites. GNPS is an online open-access mass spectrometry repository that aims the organization and sharing of spectral data and annotations. Moreover, GNPS provides access to online dereplication prioritization (Lang et al. 2008) and automated molecular networking analysis (Sidebottom et al. 2013; J. Y. Yang et al. 2013).

In addition to GNPS, computational approaches, Partial-least-square- Selectivity ratio (PLS-SR), ElasticNet (EN) were used to determine top bioactive candidates (Abbas-Mohammadi et al. 2018; Stagliano et al. 2010). Partial least squares (PLS) utilizes the spectral information to predict bioactive metabolites in complex natural product (Ali et al. 2013). The selectivity ratio enabled identification of bioactivity compounds by determining the correlation between molecular features identified and bioactivity levels in natural products (Kellogg et al. 2016). Elastic net was selected as it allowed accurate computation of the contribution of each bioactive phytochemical towards the total bioactivity of the fraction without limiting the number of phytochemicals being used for the prediction (Kirkpatrick et al. 2017).

3.2.3 Methods

Associating chemical diversity of *C.Asiatica* with % viability from MC65 bioassay

Centella Asiatica water extract was prepared using procedure discussed in part 1. For natural product discovery liquid:liquid extraction was performed followed by LH-20 column chromatography with methanol. The chemical diversity of compounds was generated by fractionation of DCM (7.8 g) and *n*-butanol (15.9 g) fraction of the aqueous extract on a Sephadex LH-20 column (Figure 3.6). Methanol was utilized to obtain 21 subfractions of CAW extract generated by solvent:solvent partitioning on the LH-20 column.

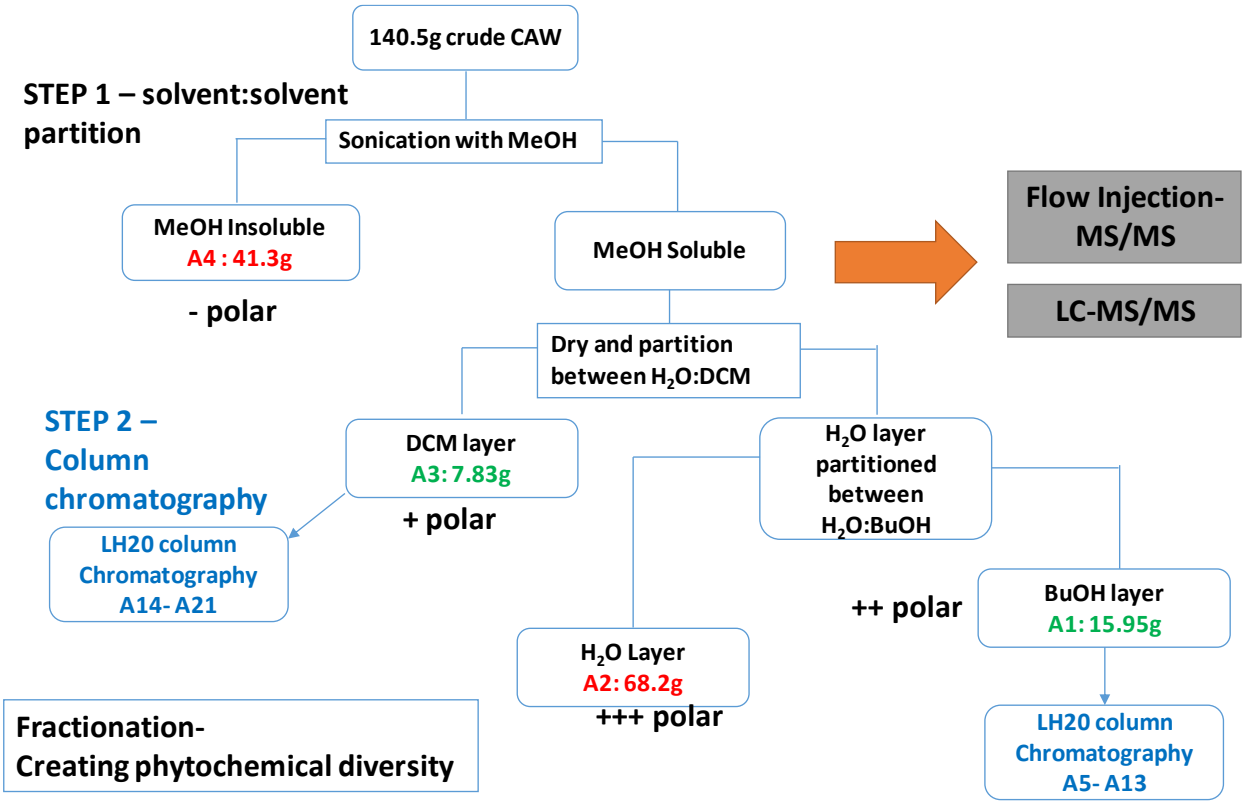


Figure 3.6 Fractionation scheme. 21 subfractions of CAW extract generated by solvent:solvent partitioning and LH-20 column chromatography. We analyzed each subfraction by flow-injection TOF positive ion mode and correlated the features found with cytoprotective activity in MC65 neuroblastoma cells exposed to Aβ toxicity.

Biological activity in MC65 cellular line

MC65 cells were used because of the ability to conditionally express the C-terminal fragment of amyloid precursor protein (APP CTF) (Sopher et al. 1996). In absence of tetracycline the cells generate endogenous Aβ that results in cell death within 3 days. There has been evidence that links Aβ aggregates and resulting cytotoxicity with oxidative stress (Woltjer et al. 2007). The maintenance of MC65 cells was performed in MEMEα supplemented with 10% FBS (Gibco-BRL, Carlsbad, CA) and 1 μg/mL tetracycline (Sigma-Aldrich, St. Louis, MO) using the procedure described in (Woltjer et al. 2003, 2007). Confluent cells were treated with trypsin followed by washing in PBS. The cells were later resuspended in OptiMEM without phenol red (Gibco/BRL, Carlsbad, CA). The cells from suspension were then plated at 25,000 cells/well in 48-well plates. GKW was used as a medium for growing these cells in without the presence of tetracycline. Measurement of cell viability was performed at 2.75 days with CellTiter 96 Aqueous Non-Radioactive Cell Proliferation Assay (Promega Corporation, Madison, WI). For statistical significance and repeatability, the experiments were performed in triplicate wells for each of the CAW sub-fractions and repeated 1-2 times (Soumyanath et al. 2012).

Natural product discovery using loop-injection-MS

Loop-injection (LI-MS) combined with high resolution accurate mass spectrometry (LC-TOF) was conducted using a Shimadzu Nexera UHPLC system connected to an AB SCIEX TripleTOF® 5600 (Concord, Ontario, Canada) mass spectrometer equipped with a Turbo V ionization source operated in positive and negative electrospray ion mode. Flow infusion (or flow injection) electrospray mass spectrometry was used for high throughput studies. As the sample analysis time is reduced to less than 5 min in absence of chromatography analysis of 1000 samples within a week became feasible. Retention time alignment is omitted. Prerequisite for the efficient application of FI-MS approaches is access to at least high-resolution mass spectrometry equipment. Ultrahigh resolution mass spectrometry platforms in conjunction with flow infusion approaches was used to achieve high throughput metabolomic analysis (Emmett and Lichti 2017; J. Han et al. 2008; Wood 2021; Y. Zhu et al. 2021), however it also had some drawbacks, in particular the difference in ionization efficiencies of analytes results in ion suppression for ions with low ionization (Draper et al. 2013). Chromatographic techniques allow better separation of the isobaric metabolites based on the difference in polarities (Want et al. 2013).

For negative ion mode acquisition, the following parameter settings were used to operate the mass spectrometer: spray voltage -4,200 V; source temperature 550 °C and a period cycle time of 150 ms was used. For ESI+ acquisitions, the instrument settings were the same as used in the negative ion mode except that the spray voltage was set to 4,500 V. The mass spectrometer was equipped with a calibrant delivery system. Mass calibration was automatically performed after every two hours. Mass spectrometry data was acquired in MS/MS or data dependent acquisition mode.

Statistical modeling to predict neuroprotective bioactivity

ElasticNet and PLS-DA approaches were applied to identify active specialized metabolites from the molecular features obtained from mass spectrometric analysis of aqueous extract of *Centella asiatica*. The 21 fractions were tested for their neuroprotective effect in the MC65 cell culture model of Aβ toxicity, which uses % cell viability as a measure of cell protection.

Elastic Net penalized logistic regression multivariate analysis was performed to determine the phytochemicals responsible for neuroprotective effect. The linear_model package from scikit-learn library v 0.24.2 was used to fit a regression model across 85 molecular features obtained from 21 fractions and CAW crude extract based on certain bound on the value of coefficients. Elastic net offered the advantage of shrinking of the some of the parameters to zero offering variable selection during the model fitting step. The penalization factor for elastic net was chosen using 10000 iterations and 5-fold cross-validation. The highest lambda from minimum standard error was selected for each iteration and median of the 10000 lambda values was computed to determine the final penalization factor.

The molecular features identified as leads are queried by their exact mass in online databases such as Human Metabolome Database (HMDB) (http://www.hmdb.ca/) and the Metlin database (https://metlin.scripps.edu) with exact mass in the range of 5 ppm. The annotations for molecular features were validated by comparing MS/MS fragment within range of 100 ppm with the online databases. PLS-DA was performed using *ropls* Bioconductor package (Thevenot 2017) to identify the significant contributors to neuroprotective activity. MCQs and DCQs were determined to have permutation test p-value < 0.05. The selectivity ratio was computed using getSelectivityRatio from *mdatools* package v 0.11.5 (Kucheryavskiy 2020) in R to identify discriminating *m/z*

molecular features. The selectivity ratios for the molecular features were plotted using Excel.

Molecular Networking

The MS/MS spectral data was used to create the GNPS network. MS/MS data have been deposited to the GNPS repository (http://gnps.ucsd.edu). The *Centella* fractions' chemical diversity and associated % viability was used to create molecular networks using the online workflow described for Global Natural Products Social molecular networking. MSCluster was used to cluster the identical MS/MS spectra into single spectrum. The precursor and fragment ions the spectra were compared to the spectral libraries with 0.1 Da for the precursor ions and 0.5 Da for fragment ions. The cosine score was used to compare similarities and differences of spectra with spectral libraries. The cosine score of 0.7 was used as a threshold for spectral match with libraries and the threshold for minimum matching peaks for annotating the spectral peaks was set at 6. The network was later imported and visualized using Cytoscape version 3.7.

3.2.4 Results and Discussion

The correlation of HRMS profiles of 21 CA subfractions with neuroprotective activity was obtained from the MC65 protection assay. Each subfraction was analyzed using flow-injection MS positive and negative ion mode and the correlation between cytoprotective activity of each subfraction in MC65 neuroblastoma cells that conditionally express Aβ precursor protein was determined. This allowed validation of relative concentration of previously identified CQAs was in positive correlation with neuroprotective activity (Figure 3.7).

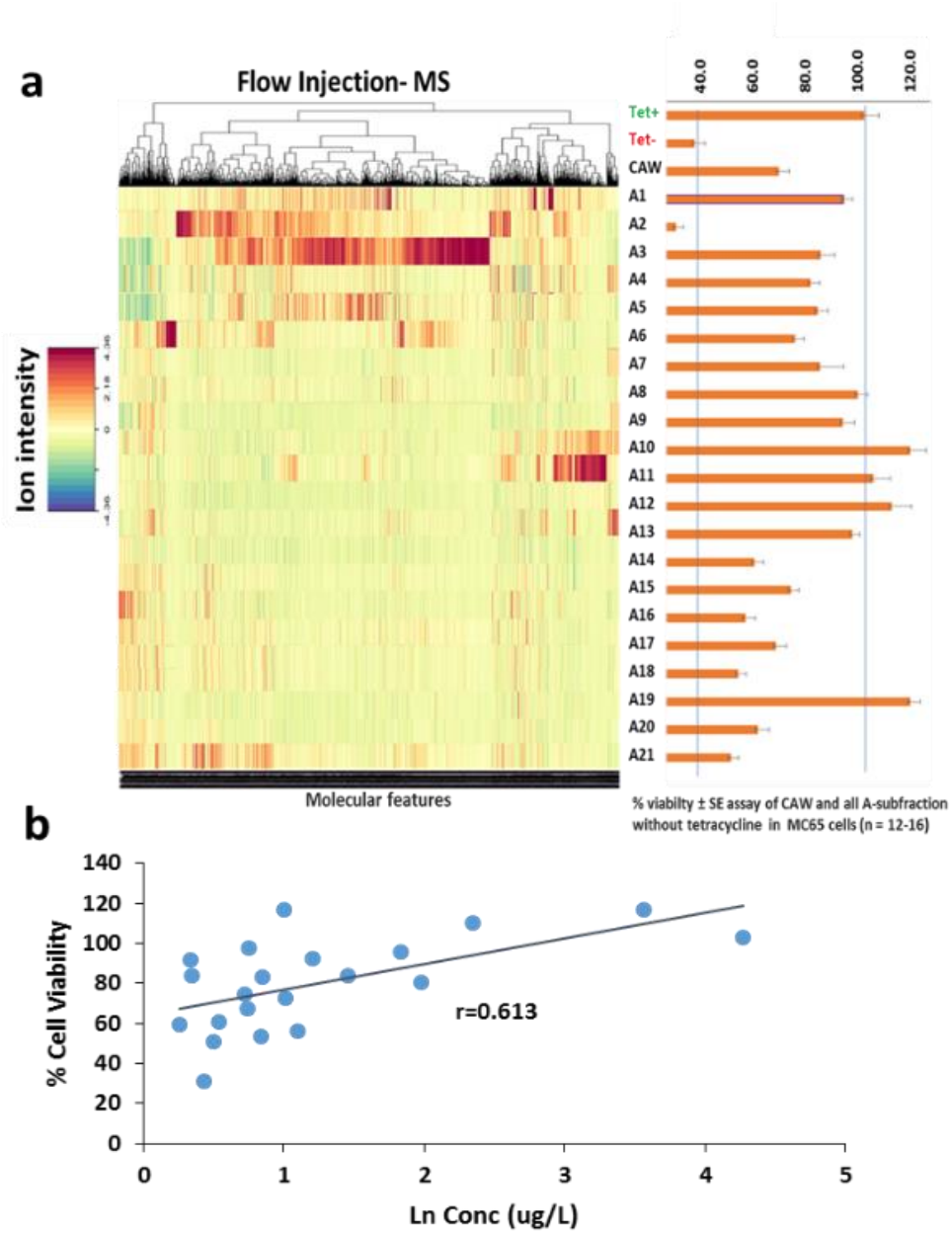


Figure 3.7 Creating phytochemical diversity to detect correlations of individual phytochemicals with biological activity. a. Left - Flow injection-TOF positive ion mode analysis of 21 CA subfractions. After data processing, over 1500 molecular features were aligned according to their molecular masses. Right - % Cell viability as an index of protection against Aβ toxicity. Bars represent % viability ± standard error of CAW extract and all subfractions of the CAW extract tested in MC65 cells in the presence of Aβ. b. Correlation of % cell viability with the concentration of diCQAs (sum of isomers) ([M-H]-, m/z 515.12) present in the 21 CA subfractions (each blue dots represents a subfraction).

% viability ± standard error assay of CA and all subfraction without tetracycline in MC65 cells (n = 12-16).

Computational approaches

The mass spectral feature, *m/z* 353.0874 and *m/z* 515.1191 were identified as bioactive phytochemicals responsible for neuroprotective effect using multivariate regression Elastic net model (Table 3.4).

Table 3.4 Selectivity ratio and elastic net hits for the experiment flow injection-TOF acquisition ion correlated with MC65 bioactivity assay.

Compound	SR	Variable Imp in ElasticNet model	Annotation	Ion mode
1.38_303.0502m/z	2.89		**Quercetin**	**POS**
1.62_257.0554m/z	1.88		**N/A**	**NEG**
1.41_353.0874m/z	1.78	1.58	**MCQs**	**NEG**
1.79_516.1262n	1.66	2.44	**DCQA's**	**NEG**
1.78_163.0385m/z	1.58	0.8	**Hydroxycoumarin**	**POS**
1.55_461.0720m/z	1.57		**Myricetin 3-glucoside**	**NEG**
1.41_179.0351m/z	1.57		**Caffeic Acid**	**NEG**
1.50_539.1153m/z	1.55		**N/A**	**POS**
1.41_537.1012m/z	1.54		**N/A**	**NEG**
1.45_605.0894m/z	1.53	0.73	**Bisdihydroquercetin**	**NEG**
1.52_513.1034m/z	1.52	0.69	**N/A**	**NEG**
1.66_477.0674m/z	1.52	0.66	**Quercetin 7-glucuronide**	**NEG**

Furthermore, the selectivity ratio plot of molecular features obtained from CAW extract using the PLS-DA model is shown in Figure 3.8.

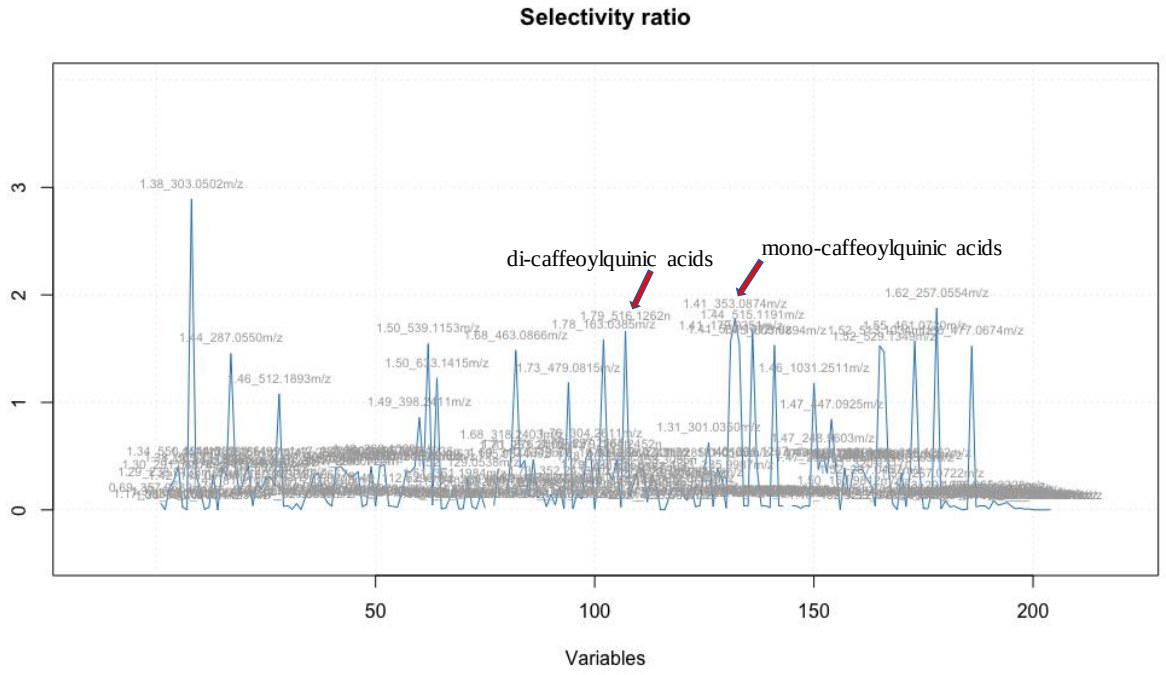


Figure 3.8 PLS selectivity ratio. Di-caffeoylquinic acids selectivity ratio was 1.79, and it was in their top 5-hit list of correlated compounds.

The *m/z* features with selectivity ratio greater than 1, indicating that 50% of original variance is explained by the molecular features that can be translated into potential leads for neuroprotective activity. The *m/z* 353.0874 $[M-H]^-$ had PLS selectivity ratio 1.78 and *m/z* 515.1191 $[M-H]^-$had PLS selectivity ratio 1.67 (Figure 3.8). By using our *in-house* Oregon Natural Products (ONAP) MS library containing 331 plant NPs, the *m/z* 353.0874 ion was assigned to mono-caffeoylquinic acid(s) and the *m/z* 515.1191 ion to di-caffeoylquinic acid(s) and verified their identities by LC-MS/MS comparison with authentic standards. In an earlier study using a classic reductionistic approach, we identified several mono- and di-caffeoylquinic acids as neuroprotective principles of *Centella asiatica*. This example shows the power of the overall approach outlined.

Analysis of clusters in the molecular network

The chemical diversity of CAW fractions' was mapped by creating molecular networks using the online workflow described for Global Natural Products Social molecular networking (Mingxun Wang et al. 2016). MS/MS spectra obtained from CAW crude extract and 21 fractions using high-resolution 5600 Sciex LC-QToF mass spectrometer in

the data-dependent acquisition mode were converted into mzXML format, networked with the Spectral Networking algorithm, and the resulting molecular networks viewed in Cytoscape (Excoffier et al. 2017) (Figure 3.9A). Bioactivity information was overlaid on the GNPS molecular network using Cytoscape. The abundance of molecular features was represented in a pie chart with sub-fractions and their associated bioactivity level occupying larger area for higher abundance. To map the MC65 bioassay results onto the nodes, each bioactivity level measured as % cell viability ranging from 38% (no protection against Aβ) to 100% (full protection) to 117% (cell viability > 100% is considered as cell proliferation) was assigned a specific color tag (21 in total), and this color tag was applied to all nodes in the molecular network (Figure 3.9B). By using our *in-house* database (Alcazar Magana et al. 2020), we were able to quickly find the corresponding network cluster representing the di-caffeoylquinic acids (Figure 3.9C). The color mapping of the nodes representing an *m/z* value with the 'pie slices' representing the proportion of the chromatographic fractions having a specific activity level (37-117%) allows visualization of bioactive compounds and their distribution across fractions. As expected, di-caffeoylquinic acids were primarily distributed over fractions with bioactivity levels in the range 103-117% (Figure 3.9D). We also identified the triterpene glycosides, madecassoside and asiaticoside, however, they were found in fractions having a lower bioactivity level as well (Figure 3.9E).

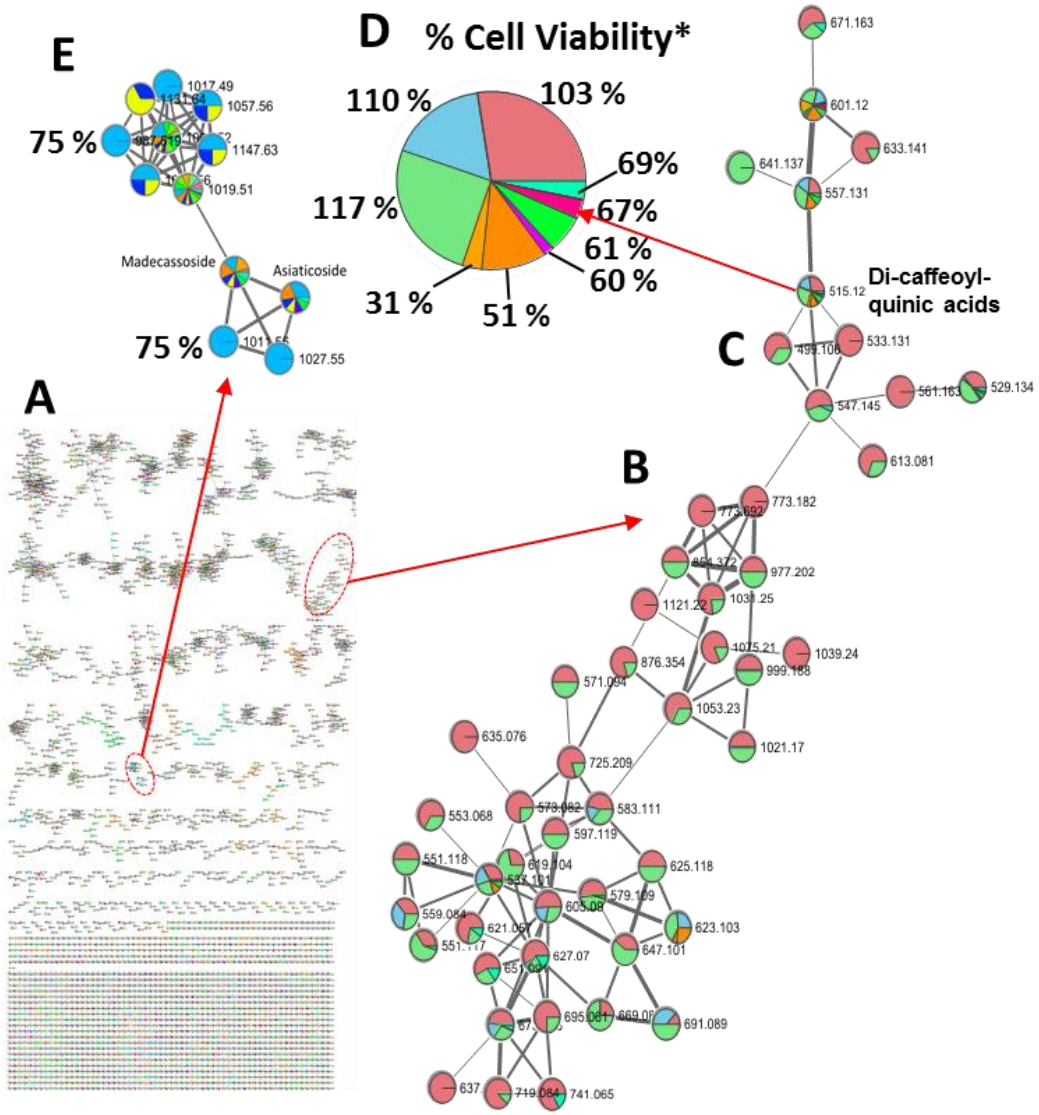


Figure 3.9 A) Generation of a massive molecular network from LC-TOF MS/MS analysis of 21 chromatographic fractions of aerial parts of Centella asiatica. (B, C) Bioactivity mapping shows that di-caffeoylquinic acids were found in fractions providing complete protection against Aβ toxicity (D), whereas triterpene glycosides (cluster E) were found in fractions providing partial protection (bioactivity level, 75%). GNPS constructs the molecular networks by aligning MS/MS spectra of the parent ions. The edges are constructed between nodes based on cosine score, which represents the similarity of the two nodes with each other. The nodes with the cosine score of zero are completely unrelated whereas the ones that have the cosine score of 1 are identical. We chose a cut-off value of 0.70 for identifying similar nodes. The spectral matching was performed using MS^2 fragmentation information.

This study demonstrated that computational models could be used to associate bioactivity to specialized metabolites in crude chromatographic fractions. This study also provided visualizations of association of bioactivity with the gradient of concentration in plant extract fractions without intensive purification. The integration of the ElasticNet approach and PLS selectivity ratio with the Global Natural Products Social Molecular Networking allowed in narrowing down the number of potential bioactive leads. In the case of *Centella asiatica* extracts, computational approaches successfully identified bioactives such as Di-caffeoylquinic acids using LC-MS/MS comparison with *in-house* library of natural products.

4. Comparative LC–MS/MS lipidomic analysis of macaque heart tissue flash frozen or embedded in optimal cutting temperature polymer (OCT): Practical considerations

Authors

Ashish Vaswani [1,†,] Armando Alcazar Magana [1,2,†], Eric Zimmermann [3], Wohaib Hasan [4], Jaishankar Raman [5,] Claudia S. Maier [1,2*]

[1] Department of Chemistry at Oregon State University, Corvallis, OR, USA.

[2] Linus Pauling Institute, Oregon State University, Corvallis, OR, USA.

[3] Oregon Health Science University, Portland, OR, USA.

[4] Cedars Sinai, Los Angeles, CA, USA.

[5] University of Melbourne, Melbourne, Parkville, VIC, Australia.

†Co-first authors

*Corresponding author

Claudia S. Maier, Department of Chemistry, Oregon State University, 2900 SW Campus Way, Corvallis, OR 97331, USA. Email: claudia.maier@oregonstate.edu

Vaswani, A., Alcazar Magana, A., Zimmermann, E., Hasan, W., Raman, J., & Maier, C. S. (2021). Comparative LC–MS/MS lipidomic analysis of macaque heart tissue flash frozen or embedded in optimal cutting temperature polymer (OCT): Practical considerations. *Rapid Communications in Mass Spectrometry*, e9155.

4.1 Abstract

Rationale: Biobanks of patient tissues have emerged as essential resources in biomedical research. Optimal cutting temperature (OCT) blends have shown to provide stability to the embedded tissue and is compatible with spectroscopic methods, such as infrared (IR) and Raman spectroscopy. Data derived from omics-methods are only useful if tissue damage caused by storage in OCT is minimal and well understood. In this context, we investigated the suitability of OCT storage for heart tissue destined for LC-MS/MS lipidomic studies.

Methods: To determine the compatibility of OCT storage with LC-MS/MS lipidomics studies. The lipid profiles of Macaque heart tissue snap-frozen in liquid nitrogen or stored in OCT were evaluated.

Results: We have evaluated a lipid extraction protocol suitable for OCT-embedded tissue that is compatible with LC-MS/MS. We annotated and evaluated the profiles of 306 lipid species from tissues stored in OCT or liquid nitrogen. For most of the lipid species (95.4%), the profiles were independent of the storage conditions. However, 4.6% of the lipid species; mainly plasmalogens, were affected by the storage method.

Conclusion: This study shows that OCT storage is compatible with LC-MS/MS lipidomics of heart tissue, facilitating the use of biobanked tissue samples for future studies.

4.2 Introduction

Tissue evaluation by biopsy is a cornerstone of pathology and is the foundation of disease diagnosis (Ziv, Durack, and Solomon 2016). Traditional methods of tissue processing include the use of formaldehyde, embedding in paraffin, prior to staining, so that the tissue can be sectioned for microscopic examination (Dietel et al. 2013). Flash freezing techniques are workarounds to avoid interfering noise from chemical fixatives, such as formaldehyde, in the spectroscopic signal (Jurowski et al. 2017).

Optimal cutting temperature compound (OCT), a blend of glycols and resins, is used as one of the primary cryopreservation methods by tissue biorepositories. OCT is used primarily for stabilizing tissue samples, preventing them from drying out, and allowing long-term preservation of the tissue samples, while allowing it to be sectioned easily. OCT storage also allows preservation of structural integrity of the tissue with the polymers, polyvinyl alcohol and polyethylene glycol, providing the supporting matrix (W. Zhang et al. 2015). However, the polymers in OCT cause ion suppression in mass spectrometry (MS) (Schwartz, Reyzer, and Caprioli 2003). MS-based techniques are frequently used for measuring lipids that potentially can serve as biomarkers associated with diseased tissues. However, the utilization of OCT-biobanked samples for LC-MS/MS lipidomics has been hindered because the potential effects of OCT storage on measuring profiles of lipid species in tissue samples have not been systematically studied so far.

Here, we report on a protocol for removal of OCT from OCT-embedded tissues that makes it possible to analyze lipids in OCT-embedded tissues using LC-ESI-MS/MS, allowing identification and quantification of lipid species in OCT-embedded heart tissues.

4.3 Experimental

4.3.1 Materials and reagents

SPLASH® LIPIDOMIX® Mass Spec Standard (Avanti Polar Lipids, Inc.) was added to mark the elution windows of 14 major lipid classes and is composed of the following deuterium-labeled lipid species: PC 15:0-18:1(d7), PE 15:0-18:1(d7), PS 15:0-18:1(d7), PG 15:0-18:1(d7), PI 15:0-18:1(d7), PA 15:0-18:1(d7), LPC 18:1(d7), LPE 18:1(d7), MG 18:1(d7), DG 15:0-18:1(d7), TG 15:0-18:1(d7)-15:0, SM 18:1(d9), Cholesterol-d7, and cholesterol ester 18:1(d7). HPLC-MS grade methanol, chloroform, acetonitrile, isopropanol, and water were purchased from EMD Millipore. Formic acid ACS reagent was purchased from Fisher Chemicals.

4.3.2 Sample collection and storage

After epicardial fat was removed, transmural left ventricular samples were procured from two different macaques H6 and H10 from the Oregon National Primate Research Center with no underlying heart conditions. H6 (fetal macaque FMH6) was a male at G131 (necropsy 06/09/2014) and H10 (FMH10) was female at G134 (necropsy 08/14/2014). One group (LN) was snap frozen in liquid nitrogen and the other group (OCT) was embedded in a cryomold with OCT (optimal cutting temperature) compound, then frozen on an isopropanol/dry ice slurry before storing it at -80^{o}C. The OCT-embedded tissue was stored for four years prior to LC-MS/MS analysis. The tissues were obtained based on the guidelines from the institutional review board.

4.3.3 OCT removal and lipid extraction

Macaque heart tissues snap frozen in OCT or liquid nitrogen (LN), were obtained from Oregon Health & Science University. Samples were processed as previously reported (Loken and Demetrick 2005; W. Zhang et al. 2015) with some modifications. In brief, the tissues were washed with aqueous ethanol (70% v/v) two times followed by water to remove OCT from the tissue. After removal of OCT, 20 milligrams of tissue were placed with 300 μL of water:methanol 1:1 solvent mixture and 5 μL of SPLASH® LIPIDOMIX® in a 2 mL reinforced tube with ceramic beads (2.8 mm) along with 0.2 mg/mL of butylated hydroxytoluene (BHT) to prevent lipid oxidation (Rubbosi et al.

1994; Watson et al. 1997). The tubes were placed in a bead beater homogenizer (Precellys^{TM} 24, Bertin Technologies, USA) and tissues were homogenized using six repeated cycles (20 s, 6000 rpm with 30 s of ice cooldown in between). The homogenate was mixed with 500 μL chilled chloroform (containing 0.02% of BHT), vortexed and mixed with 300 μL of chilled water. This homogenate mixture was centrifuged at 10,000 *g* (4˚C, 10 min). Centrifugation separated the suspension into three phases. The top phase was comprised of mainly of polar compounds, the bottom phase contained most of the non-polar metabolites (used for lipidomics) and the protein pellet was formed in the interphase (Snyder et al. 2013). After SpeedVac drying, lipid extracts were created by re-suspending the residue in 200 μL of acetonitrile:isopropanol mixture (50 % v/v) and stored at -20 ^{o}C until analysis.

4.3.4 UPLC-MS/MS analysis

Lipidomic analysis was conducted in the data-dependent acquisition (DDA) mode using an AB SCIEX TripleTOF® 5600 mass spectrometer (AB SCIEX, Concord, Canada) coupled to Shimadzu Nexera UHPLC system. Data was acquired in the positive and negative electrospray ionization mode. The autosampler was maintained at 6 ^{o}C. Lipids were separated using an Acquity CSH C18 (2.1 x 150 mm, 1.7 μm) column (Waters Co.) A linear gradient with two mobile phases was used as previously reported by Cajka *et al.* (Cajka and Fiehn 2016b) with some modifications. Briefly, for acquiring data in the positive ion mode, phase A (60:40 (v/v) acetonitrile/water containing 10 mM ammonium formate and 0.1% formic acid (FA)) and B (90:10 (v/v) isopropanol/acetonitrile containing 10 mM ammonium formate and 0.1% formic acid) were used. The elution gradient was as follows: 0 min, 15% B; 2 min, 30% B; 2.5 min, 48% B; 11 min, 80% B; 11.5 min, 100% B; 12 min, 100% B; and 12-15 min, 15% B. The flow rate was maintained at 0.6 mL/min.

Electrospray voltage was set at 5500 V; and the source temperature was set at 550 ˚C. Period cycle time of 700 ms; accumulation time of 100 ms; *m/z* scan range of 100–1700; and collision energy of 35 V with collision energy spread (CES) of 15 V were used. Mass calibration was automatically performed after every fifth LC run.

For the negative ion mode, the same gradient and instrument parameters were used but with the following changes: Mobile phase A and B used 10 mM ammonium acetate without formic acid and electrospray voltage was set to -4500 V.

Three replicates were injected for each of the tissue sample. The injection order was randomized for all samples with QCs and blanks injected after every five analytical runs.

The mass spectral data was normalized by tissue weight after OCT removal (Rohrbach et al. 2020). The DDA data were processed with Progenesis QI^{TM} software (NonLinear Dynamics, United Kingdom). For facilitating the comparative and differential evaluation of the datasets, we used "all compound normalization" in Progenesis QI, which includes alignment of spectral data and collapse to single compound ions, the later based on deconvolution of $[M-H]^-$, $[M+FA-H]^-$, $[M-H_2O-H]^-$, and $[M+Na-2H]^-$ adducts in the negative ion mode and $[M+H]^+$, $[M+Na]^+$, $[M+2Na-H]^+$, and $[M+CH_3OH+H]^+$ in the positive ion mode. The aligned and deconvoluted molecular features were matched against lipid databases (HMDB, METLIN, and LipidBlast) for obtaining annotations for distinct lipid species. If the Progenesis score was 50 or higher, the lipid species annotation was considered a solid level 2 annotation and used for statistical analysis. A score higher than 50 is typically reached when isotope pattern similarity is greater than 80%, MS/MS match is > 50% and accurate mass is 10 ppm or better (Housley et al. 2018; Magana et al. 2020). An annotation was included in the resulting dataset if the coefficient of variation for abundance for the ions was lesser than or equal to 25%. If a molecular feature was detected in positive and negative ion mode, the one with greater signal-to-noise ratio was chosen. Relative levels of the lipid species were obtained by computing their peak areas (Magana et al. 2020).

4.3.5 Statistical analysis

Annotated lipid species were compared among OCT-embedded heart tissue samples and liquid nitrogen (LN) flash frozen tissues using univariate and multivariate statistical methods. The normality assumption of the lipid species distribution across the replicates was validated with Shapiro-Wilk normality test. The profiles of lipid species across the OCT and LN groups were compared using Welch unpaired t-test for normal distribution.

Hierarchical clustering analysis (HCA) bottom-up approach was used to analyze whether the effect of storage condition or the biological variations was more profound on the lipid profile of the primate heart tissues. Multivariate principal component analysis (PCA) analysis was conducted on 306 lipid species (Table S4.1) to find trends in their levels across biological specimens and storage conditions and identify the outliers.

False discovery rate (FDR) was used to adjust *p*-values to account for 'by chance' assignment of features to be significant in the population. FDR has been widely used in lipidomics studies for multiple hypothesis testing to alleviate type I errors (false positives). FDR was computed using the Benjamini and Hochberg method (Benjamini and Hochberg 1995) as the number of features was greater than the number of samples (D. I. Broadhurst and Kell 2006). The scarcity of the animal model and the associated small sample size made us address the issue of limited statistical power by selecting FDR at 10% significance level for identification of truly significant features. The p.adjust function from R package **stats** version 3.6.2 ({R Core Team} 2020) was used with "fdr" argument for adjusting *p*-values of 306 lipids obtained from H6 and H10 tissues stored in OCT and LN storage conditions for evaluating the impact of OCT storage.

4.4 Results and Discussion

4.4.1 Results

Methanolic extracts of heart tissue samples after removal of OCT with ethanol were analyzed by UPLC-high resolution MS/MS with the objective to evaluate the impact of OCT storage on the levels of lipid species that are typically detected in heart tissue samples. We annotated 224 molecular species in ESI- data and 132 in the ESI+ data. Thirty-five annotated lipids were recorded in both ionization modes, in such case, annotations with less CV on quality control (QCs) samples were kept. Of the 321 molecular species, 306 were lipids, that were used for statistical comparison to determine to what extent OCT storage affects the detection and quantification of lipid species. The annotated lipid species are described in detail in Supplementary table S4.1.

We used univariate statistical analysis to evaluate the effect of storage conditions for lipid species belonging to following lipid classes: lyso-phospholipids (LysoPL),

phosphatidylcholines (PC), phosphatidylethanolamines (PE), plasmalogens, sphingomyelins (SM), diacylglycerols (DG) and triacylglcerols (TG). Using Welch's t-test we determined that levels of saturated PCs, saturated PEs and SMs were not significantly different across OCT and LN storage conditions. Specifically, the lipid classes, PC and lysoPC, for the two storage conditions did not statistically differ from each other (*p*-values = 0.61 and 0.66, respectively). The ratio of PC and lysoPC across the two different storage conditions was not statistically significant (*p*-value = 0.73) across the different storage conditions. The lipid class PE and SM also did not account for statistical differences (*p*-values of 0.71 and 0.13) between the two storage conditions. The lipid class DG and ratio of levels of polyunsaturated (more than two double bonds) DG to di-unsaturated DG also did not exhibit statistical differences (*p*-values 0.63 and 0.93) across the two storage conditions.

The plasmalogen lipid class (plasmenyl phosphatidylcholine (pPC), plasmenyl phosphatidylethanolamine (pPE) and plasmenyl lyso phosphatidylcholine (pLPC)) composed of unsaturated and saturated plasmalogens, were not statistically different (*p*-values = 0.42 and 0.94) between the two storage conditions. However, the ratio of unsaturated plasmalogens to the saturated plasmalogens displayed statistically significant differences (*p*-value=0.074) between OCT and LN storage conditions based on Welch's t-test. Of the significantly different lipid species the majority (29%) belonged to plasmalogens. The 5% FDR threshold seemed too stringent because none of the lipid species would have been recognized as showing significant differences owing to OCT storage, therefore 10% FDR threshold was selected to obtain a more robust evaluation of observed disparate lipid species dependent on storage conditions.

Overall, fourteen (out of 306) lipid species, all belonging to unsaturated phospholipids and unsaturated (PC and PE) plasmalogens, with a statistically significant (meeting 10% FDR threshold) decrease in levels associated with OCT storage.

The dendrogram, heatmap and PCA (Figure 4.1) multivariate analysis was performed on the lipid profiles of heart tissues stored in LN or embedded in OCT and flash frozen. Euclidean distance measure across 306 lipids was used to generate the dendrogram. The 30 most dissimilar lipids are represented in a heatmap. The PCA scores plot revealed two

clusters for the two groups, H6 and H10, without any overlap between their 95% confidence interval.

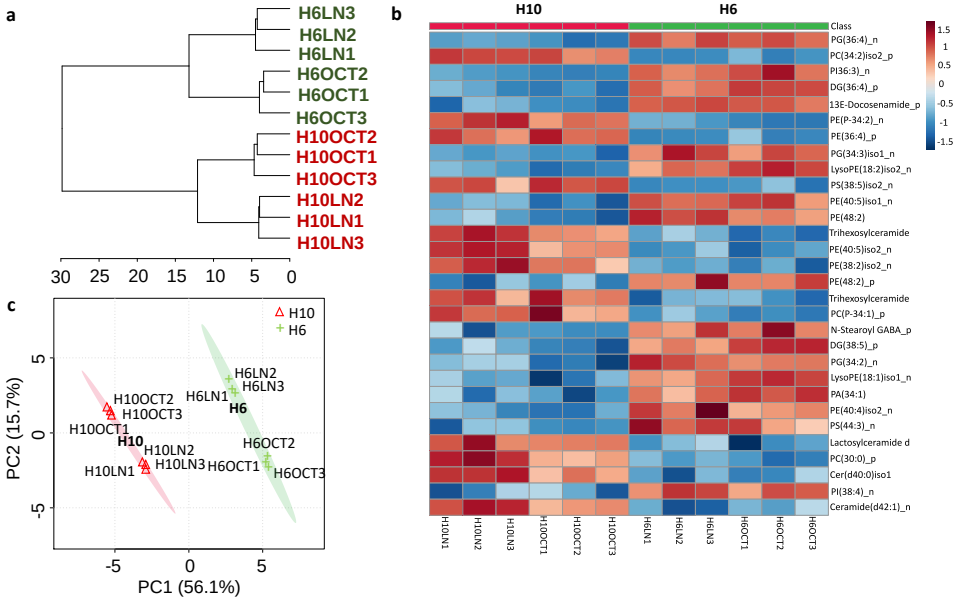


Figure 4.1 Evaluation of storage conditions based on 306 lipid species detected in methanolic extracts of Macaque heart tissues H6 and H10 (biological samples). Heart tissue was embedded in OCT, frozen on an isopropanol/dry ice slurry and stored at -80oC or directly flash-frozen in liquid nitrogen (LN). When lipids were detected in both positive and negative ion modes, the one with lesser variation in the quality control sample was kept. a) Hierarchical clustering dendrogram constructed using Euclidean distances and Ward algorithm. b) Heat map of the 30 most dissimilar lipid species from 306 lipid species obtained from Macaque heart tissues stored in OCT or liquid nitrogen. The color scale represents z-score of the annotated lipid species with red representing positive z-scores (higher peak areas) and blue representing negative z-scores (lower peak areas). c) Principal component analysis scores plot obtained from lipid species of heart tissues H6 and H10 (biological samples) stored in OCT and liquid nitrogen showed clustering of the biological specimens H6 and H10 using 306 lipids into two distinct groups.

4.4.2 Discussion

The goal of this study was to evaluate whether there were differences in lipid profiles of heart tissues stored by embedding in OCT, frozen on an isopropanol/dry ice slurry and

stored at -80^{o}C or directly flash-frozen in liquid nitrogen (LN). OCT was removed prior to LC-MS/MS analysis by rinsing with ethanol (Loken and Demetrick 2005; W. Zhang et al. 2015).

Initially, we used multivariate analysis methods to explore which variable (biological background or storage condition) is predominant. The dendrogram, heatmap and PCA (Figure 4.1) were used to assess the effect of OCT storage on lipid profile of heart tissues. HCA (dendrogram) and heatmap showed distinct separation based on the biological samples H10 and H6 suggesting that impact of the biological samples is important rather than the effect of storage condition.

PCA scores plot revealed that variability observed due to the LC-MS analysis was minimal. Although there were few differences in the lipid profiles caused by the storage conditions, the variations caused by differences in biological specimens were more significant. Multivariate statistical analysis of tissues stored under OCT and liquid nitrogen suggest that OCT-embedded tissues could be used for LC-MS analysis after removing OCT from the tissue with ethanol and water.

Previously, Rohrbach et al. (Rohrbach et al. 2020) validated a procedure to analyze sphingolipid alterations in OCT embedded human tissues. Our findings agree with Rohrbach's results, but importantly, it offers greater coverage of lipid classes. The lipid species comparison using volcano plot (Figure BS1) across OCT and liquid nitrogen indicated that less than 5% of the lipids were significantly different at 10% FDR threshold.

The effect of OCT storage on the lipid profile of heart tissues was quantified using correlation and Forest plots (Figure 4.2). The correlation analysis reveals that the lipid species levels showed a strong linear correlation and minimal deviation from the line of best fit indicating that storage conditions has little or no effects on the detected levels. Coefficient of determination R^2 was greater than 99% across the two storage conditions for both biological replicates H6 and H10, demonstrating reproducibility of the analytical method. Our analysis also showed that the two storage conditions had similar effect on majority of the lipid species as the slope of the lipid species profile for the two storage conditions for H6 and H10 was 0.99 and 0.92 across the replicates (Shah et al. 2015). The forest plot demonstrates the differences between major classes of lipids obtained from heart

tissues stored under OCT and liquid nitrogen storage conditions. The percentage difference across the lipid classes also revealed that some lipid species showed increased ion signals (negative difference percentage) or decreased ion signals (positive difference percentage) because of OCT storage. The significant differences associated with OCT storage were only linked to decreased ion signals for lipids.

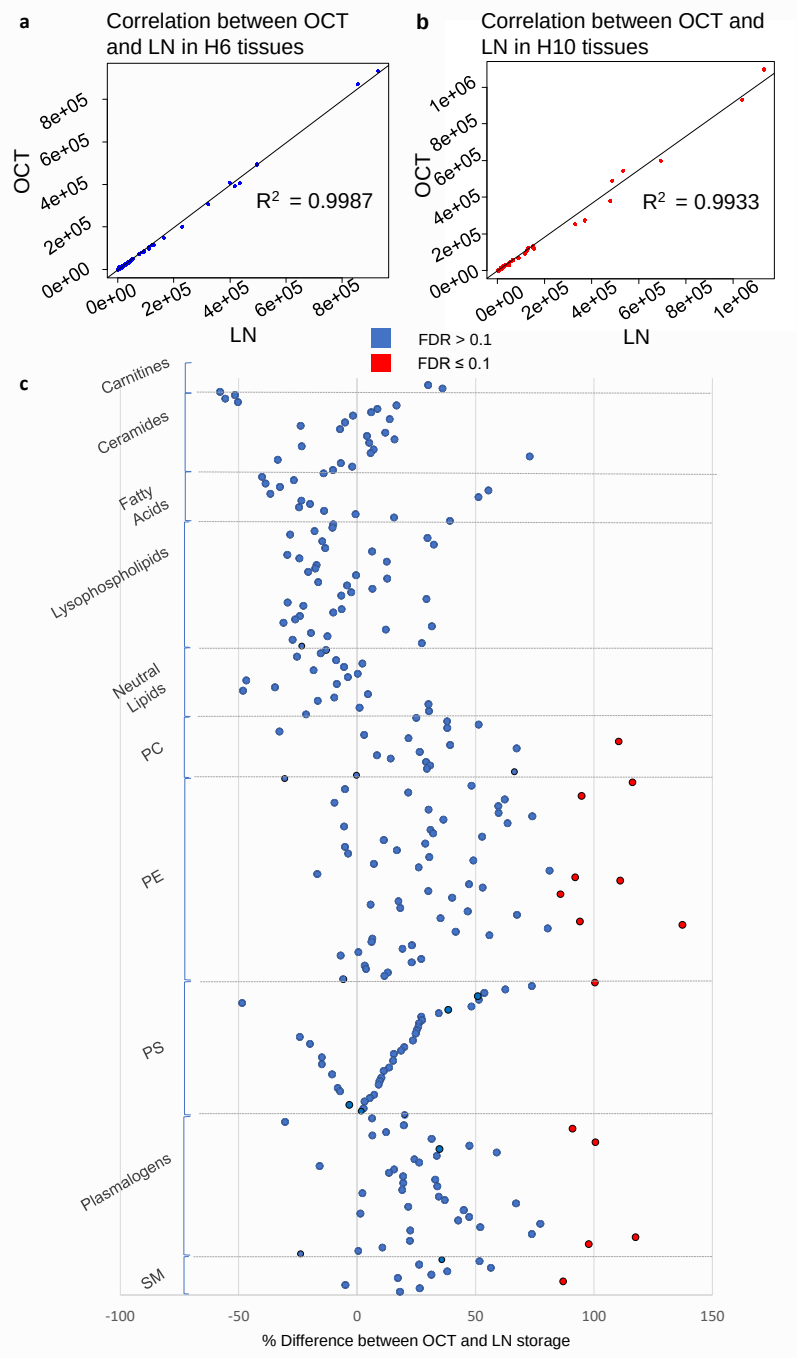


Figure 4.2 Untargeted lipidomic analysis a) Pearson correlation was computed across 306 lipid species annotated in ESI positive and ESI negative ion mode between OCT storage and liquid nitrogen storage (Table BS1) for H6 heart tissues, b) Pearson correlation was computed across 306 lipid species annotated in ESI positive and ESI negative ion mode between OCT storage and liquid nitrogen storage (Table BS1) for H10 heart tissues, c) Plot displaying percentage difference between LN and OCT storage conditions for 268 lipid

species belonging to the following lipid classes: carnitines, ceramides, fatty acids (FA), lyso-phospholipids, phosphatidylcholines (PC), phosphatidylethanolamines (PE), phosphatidylserine, plasmalogens, sphingomyelins (SM) and neutral lipids (monoacylglycerols (MG), diacylglycerols (DG) and triacylglycerols (TG)). The percentage difference between OCT and LN storage conditions are displayed on x-axis and the lipid classes are displayed on y-axis. The individual lipids (Level 2 annotations) indicated as red circles showed significant differences across LN and OCT storage condition (FDR corrected p-value ≤ 0.1), whereas ones in blue indicate no significant differences (FDR corrected p-value > 0.1).

Understanding the chemical response of diverse lipids species in different storage conditions, such as OCT and LN, is essential for clinical studies. Our method covers major lipid classes relevant to cardiac metabolism. The long-chain acylcarnitines and free fatty acids are associated with energy production in the heart tissue (Chokshi et al. 2012). Upregulation in the levels of TG in heart tissue suggests impairment in the cardiovascular function owing to reduced fusion in mitochondria and dysregulation of autophagy (Eum et al. 2020; Zhao et al. 2014). Elevation in the levels of DGs and ceramides in heart tissue has been reported in cardiomyopathy (Chokshi et al. 2012), along with a decrease in contractility of the heart (Braz JC, Gregory K, Pathak A 2004; Drosatos et al. 2011). The deregulation of polyunsaturated DG has also been associated with ischemia in the heart tissue (Gysembergh et al. 2000). OCT storage did not have significant effect on lipid species levels belonging to carnitines, fatty acids, TG and DG (Table S1). Alterations of sphingomyelin (SM) levels in the serum have been associated with atherosclerosis, ischemia and myocardial reperfusion injury after ischemia (Hannun and Obeid 2018; Reforgiato et al. 2016; Wong et al. 2000). Dysregulation of PE in heart tissue is associated with reduced contractility (Vecchini et al. 2000) and ischemia (Post, Bijvelt, and Verkleij 1995). The lysophospholipids in plasma are linked with atherosclerosis and ischemia (Abdel-Latif et al. 2015). Elevation of lysoPCs in heart tissues has been observed in tachycardia and arrhythmias resulting in sudden cardiac death (Giffin et al. 1988; Lou et al. 2020; Rigoni et al. 2007; Sedlis et al. 1983). The elevation in the ratio of PC to LPC in

the serum has been linked to arterial stiffness (Paapstel et al. 2018) and alterations in the level of plasmalogens (PC(P)s, PE(P)s and LysoPC(P)s) in plasma are associated with coronary artery disease (Meikle et al. 2011).

Comparison of SMs, PEs, lysophospholipids, PC to LPC ratio, unsaturated and saturated plasmalogens levels revealed no significant impact of OCT storage., which in turn suggests, the suitability of OCT storage for biobanking of heart tissue samples for future studies. However, OCT storage did affect the ratio of unsaturated plasmalogens to saturated plasmalogens. Plasmalogens have a vinyl ether bond at the sn-1 position and polyunsaturated fatty acids at the sn-2 position. The hydrogen atoms adjacent to the position of the vinyl ether bond are weakly bound, making this class of lipids susceptible to oxidation (Bourdillon et al. 2014). OCT storage over long periods might expose heart tissues to oxidative degradation, resulting in changes in the levels of plasmalogens across OCT and liquid nitrogen storage conditions (Clark 1986; Southard et al. 1991).

Taken together, the lipidomic analysis indicated that after a typical extraction (Folch Jordi, Lees M, and Stanley G H Sloane 1957) the storage with OCT did not majorly impact the lipid profile in macaque hearts. Thus, we conclude, that the described procedure can be utilized for profiling the major lipid classes (except for unsaturated plasmalogens) of OCT-biobanked heart tissue samples.

4.5 Conclusion

The results of our study suggest that the described protocol allows determining lipid species profiles of OCT-embedded tissues by ESI LC-MS/MS. Caution needs to be taken when evaluating plasmalogens in OCT-embedded tissue, likely due to sensitivity of plasmalogens to oxidation This study demonstrated the effective analysis of Macaque heart tissue stored in OCT, while not using precious human specimens. From a clinical perspective, this is valuable information because OCT storage does not alter lipidomic signatures. Therefore, the ability to have tissue stored in OCT and available for metabolic assessment using analytical methods, such as mass spectrometry, makes it easier to access

a wider pool of cardiac tissue. This applies to tissues that are stored in biobanks and biorepositories.

Acknowledgments

The authors acknowledge the technical support provided by the mass spectrometry center at Oregon State University.

Funding

Funding information

We acknowledge NIH grant S10RR027878 and support from the Oregon National Primate Research Center (ONPRC), and departmental funds from OHSU Division of Cardiac Surgery.

5. A cautionary tale: Metabolomics reveals that doxycycline, used widely in inducible gene silencing experiments, causes metabolic dysregulation in breast cancer cells

Authors

Ashish Vaswani [1,] Armando Alcazar Magana [1,2], Stanislau Stanisheuski [1]; Patrick Reardon [1]; Ronn Leon [2]; Sara Courtneidge [2]; Claudia S. Maier [1*]

[1] Oregon State University, Corvallis, OR, USA.

[2] Oregon Health Science University, Portland, OR, USA.

*Corresponding author

Claudia S. Maier, Department of Chemistry, Oregon State University, 2900 SW Campus Way, Corvallis, OR 97331, USA. Email: claudia.maier@oregonstate.edu

Planning to submit the manuscript to Metabolites.

5.1 Abstract

An integrated analytical method was developed to determine if a doxycycline (DOX)-dependent gene expression knockdown system is a viable strategy in the context of metabolomic studies of breast cancer cells for studying the biological effects of targeted gene silencing. The frequently employed triple negative breast cancer cell lines, MDA-MB-231 and HS578T, with and without DOX-inducible scrambled RNA constructs were used to determine impact of doxycycline.

Polar metabolites typically associated with tricarboxylic acid and lipid contents of these cells were assessed. Comparative metabolomic analysis was conducted on the cells, using combination of NMR and, hydrophilic interaction chromatography (HILIC) method in conjunction with high resolution tandem mass spectrometry (HRMS/MS). Comparative lipidomic analysis was conducted using reversed phase ultra-performance lipid chromatography (RPLC) coupled with a high-resolution mass spectrometry to comprehensively access the impact of the DOX-inducible gene silencing strategy. Metabolite differences between cells with doxycycline treatment and control cells were mapped to central carbon, one carbon, and methionine and glutathione metabolic pathways. Lipid classes that were significantly affected (FDR ≤ 0.05) between doxycycline treatment and control groups mainly comprised of phosphatidylcholines, phosphatidylethanolamines and plasmalogens. Overall, the differences in comparative metabolomics and lipidomics suggested that DOX-based gene expression knockdown strategies unexpectedly affected metabolic pathways in the breast cancer cell lines tested under the conditions applied.

5.2 Introduction

Gene silencing strategies are popular experimental tools to study biological mechanisms in cell cultures and animal models. One of the popular gene silencing techniques is the usage of short hairpin RNA (shRNA) expression vectors. This mechanism uses single lentivirus vector system, for knockdown of the target gene by RNA interference (Achenbach, Brunner, and Heermeier 2003). RNA-interference controlled by drugs enables conditional

endogenous gene expression, but this method has such drawbacks as ineffective knockdown, potential irreversibility and complex design involving multiple vectors.

The doxycycline induced knockdown of a target gene with a single lentivirus was found to provide specific gene knockdown, effective control of gene expression, and was found to be fully reversible (Aagaard et al. 2007; Matsukura 2003; Szulc et al. 2006). Doxycycline is a member of the tetracycline family and has been used as an antibiotic for decades and more recently against a variety of inflammatory diseases (Krakauer and Buckley 2003).

Inhibition of apoptosis (Li et al. 2014), alterations of energy metabolism (Weinhouse et al. 1956), membrane fluidity (Eiriksson et al. 2018) and epigenetic modifications (Rubinek et al. 2012) have been associated with breast cancer cells. Rapid cell growth of cancer cells results in an increased demand for cell membrane components. Phosphatidylcholines (PCs) are required for biosynthesis of cell membranes. Due to the fact that polyunsaturated PCs are more likely to be oxidized which may potentially cause cell damage and death, higher levels of saturated PC and lower levels of polyunsaturated PC in cancer cells are favored in cell membranes (S. Guo et al. 2014; Hilvo et al. 2011). Sphingomyelins (SMs) and phosphatidylethanolamines (PE) have been linked with apoptosis through caspase activation and specific PE-binding that helps in identification of apoptotic cells respectively (Ségui et al. 2006; X. Wang et al. 2004). The plasmalogens are responsible for inhibiting epithelial-mesenchymal transformation (EMT) that is responsible for imparting membrane fluidity to the breast cancer cells.

Epigenetic silencing of the tumor suppressor E-cadherin gene, responsible for inhibiting EMT, from DNA methylation of 5'CpG regions results in increase in proliferation and metastases of cancer cells (Widschwendter and Jones 2002). Doxycycline has been known to target E-cadherin gene that leads to the arrest of breast cancer cell migration (Zhong et al. 2017). The migratory breast cancer cells also require energy metabolism which is linked with increase in lactate production and utilization of glutamate in the mitochondria (Peiris-Pagès et al. 2016). Doxycycline has been known to adversely impact cellular proliferation and mitochondrial function in cancer cells (Saikali and Singh 2003; Su et al. 2013) which, in turn, decreases tumor cell survival (L. Zhang et al. 2017).

MDA-MB-231 and HS578T were studied widely (Chavez, Garimella, and Lipkowitz 2010) and were chosen since they were triple negative breast cancer cell lines, cell lines that were devoid of the estrogen receptors (ER2), progesterone receptors (PR), and were devoid of amplified production of human epidermal growth factor 2 (HER2). MDA-MB-231 and HS578T are convenient in vitro models for studying mechanism associated with aggressive cancer cell phenotypes that undergo metastasis with high propensity. In addition to metastasis, MDA-MB-231 breast cancer cell line also exhibit lack of differentiation.

This study aimed at assessing whether doxycycline based knockdown strategy along with scrambled RNA constructs as they did not have the potential to induce specific gene knockdown. impacted the metabolic pathways associated with proliferation and mitochondrial function using the metabolomics approach. We observed differences in metabolite and lipid levels caused by the response of these model cells to doxycycline treatment compared to untreated cells.

5.3 Materials & Methods

5.3.1 Chemicals

Polar metabolite standards were purchased from IROA Technologies. The SPLASH® LIPIDOMIX® Mass Spec Standard was used as the internal standards mixture. The SPLASH® LIPIDOMIX® Mass Spec Standard comprised of (Lyso PC (19:0), PC (17:0/17:0), PE (17:0/17:0), PG (17:0/17:0), TG (17:0/17:0/17:0), Tridecanoic acid (FA 13:0), Pentadecanoic acid (FA 15:0), Heptadecanoic acid (FA 17:0) lipids and was purchased from Avanti Lipids. HPLC-MS grade methanol, chloroform, acetonitrile, isopropanol and water were purchased from EMD Millipore. Formic acid ACS reagent was purchased from Fisher Chemicals.

5.3.2 Metabolite Extraction for LC-MS/MS and ^{1}H NMR

Samples were initially washed with PBS to maintain structural integrity of the cells. The cells were homogenized in methanol by rigorous shaking with ceramic beads. The homogenizer was operated at 5000x g with a pulse of 20 s for 2 mins. Metabolites, lipids and proteins were extracted from the cells using methanol, water and chloroform in the

ratio 6:6:5.4 (Snyder et al. 2013). The extract was centrifuged at 10000x *g* for 10 mins at 4ºC which separated the extract into 3 phases, namely polar, non-polar and protein. The top layer of liquid extract comprising the polar fraction contained predominately polar metabolites, while the bottom layer contained the non-polar metabolites, mainly lipids. The polar layer was aspirated and separated from other two phases. This was followed by aspiration of non-polar phase all the while ensuring that the protein pellet was left undisturbed. The polar phase was divided into two fractions; 90% was reserved for NMR analysis and 10% was used for the MS analysis. Fractions were dried in a centrifugal evaporator.

The NMR-destined dried polar fraction was reconstituted in 1mM PBS and NMR standard comprising of sodium trimethylsilylpropanesulfonate (DSS) and deuterated water (D_2O). The MS-destined polar fraction was reconstituted with 10 mM ammonium acetate in 95/5 acetonitrile/water. For extraction of lipids the method by Bligh and Dyer (Bligh and Dyer 1959) was followed. The lipid portion was reconstituted using the extraction solvent comprising of isopropanol and acetonitrile (10% v/v acetonitrile) and 0.02% of butylated hydroxytoluene (BHT) and 1µg/ml IS mixture. The lyophilized protein pellets from all cell samples were stored at -80 ºC prior to protein quantification by Bradford assay.

5.3.3 Protein Estimation

Protein pellets were dissolved in 400 µl of 0.1 % Rapigest solution in 50 mM ammonium bicarbonate and protein concentration was determined by Bradford assay. To dissolve completely, it was vortexed and heated up to 65 ºC for one hour. After getting transparent homogeneous solution 10 µl aliquots were taken and diluted 50 times in 50 mM ammonium bicarbonate.

20 ul of 50 times diluted protein solution were transferred to one well of 96-well plate. 20 ul of BSA standard solutions having protein concentrations from 0 to 0.2 µg/µg were used instead of the sample for calibration curve. Then, 180 µl of Quick Start™ Bradford 1x Dye Reagent were added to every well, the plate was gently mixed on the autoshaker for 5 minutes, and absorbance at 450 nm was measured. For each Sample/BSA standard three wells were assigned, analysis was performed in triplicate and the measurements were averaged.

5.3.4 HILIC LC-MS Method Optimization

Metabolites are commonly represented by a few isomers in human cells. This elevates the importance of chromatographic separation for identification and quantification of these metabolites. Optimization of chromatographic separation would also help in minimizing ion suppression. The mechanism of separation in HILIC chromatography involves hydrophilic partitioning of polar metabolites in stationary polar phase. The analytes are retained using dipole-dipole interactions, hydrogen bonding and ion exchange with the stationary phase in the column. Therefore, the mobile phase gradient was optimized to prevent non-polar metabolite elution in dead volume and peak tailing. The column was tested with 3 different gradients, 80%, 90% and 95% of acetonitrile. At least 5% water was maintained in the mixture during the process of gradient selection to improve the life of the column (Snyder et al. 2013). The gradient with 90% B was selected to obtain sharp peaks of metabolites without losing any metabolite information. Once the HILIC method was optimized, a library was created from authentic standards for the HILIC-MS method that helped in metabolite assignments based on matching, retention time and mass spectral data with the curated spectral information.

5.3.5 HILIC LC-MS analysis

Polar metabolites were separated by a hydrophilic interaction liquid chromatographic (HILIC) method using SeQuant® ZIC®-HILIC (3.5 µm, 200 Å) 150 x 1 mm column with a constant flow rate of 0.1 ml/min. Mobile phase A was composed of 5% acetonitrile solution with 10 mM ammonium acetate and mobile phase B comprised of 95% acetonitrile solution with 10 mM ammonium acetate (Snyder et al. 2013). The injection volume was 1 µL and it comprised of metabolites extracted from approximately 5000 cells per injection. The concentration gradient was set at initially at one minute at 90% B, followed by 5 to 30% B from 1 to 18 minutes, then 30% to 90% B from 18 to 20 minutes, then gradient was held at 90% B from 20 to 30 minutes. The mass spectrometry analysis was performed on AB SCIEX TripleTOF 5600 mass spectrometer which is a quadrupole time-of-flight mass spectrometer. This instrument was operated in information dependent acquisition with high sensitivity using an acquisition time of 0.2 s and a full scan range of 50–1000 Da. The instrument was operated with electrospray ionization source in positive (ESI+) and

negative (ESI-) ion modes. The mass spectrometer was operated using the following parameter settings: spray voltage 4500 V (positive mode) -4000 V (negative mode); source temperature 500 °C, period cycle time 950 ms, accumulation time 100 ms, *m/z* scan range between 100 and 1200 Da, collision energy 35 V with collision energy spread (CES) of 15 V. Each MS/MS scan had an accumulation time of 0.20 s and a range of 40–1000 Da.

5.3.6 RPLC-MS analysis

Lipid profiling was conducted using HSS T3 column 150Å, 1.8 µm, and 2.1mm × 100mm (Waters, Milford, MA, USA) on an Acquity UPLC (Waters, Milford, MA, USA). The column temperature was maintained at 55°C, auto-sampler was held at 6°C. The mobile phase for lipidomics in ESI+ mode comprised of solvent A consisting of aqueous solution of acetonitrile (60% acetonitrile %v/v) along with 10 mM ammonium formate and 0.1% formic acid acting as a buffer and solvent B consisting of mixture of acetonitrile and isopropanol (10% acetonitrile %v/v) and 10 mM ammonium formate and 0.1% formic acid. The flow rate was maintained at 0.4mL/min. Concentration of solvent B was set at 15% initially and it was modulated according to the following gradient:

0-0.3 min 15 % (B); 0.3–2.2 min 30 % (B); 2.2–9 min 50 % (B); 9–9.30 min 80 % (B); 9.30–11.8 min 100 % (B); 11.8–16 min 15 % (B). This gradient was used for both ESI+ and ESI- modes.

UPLC setup is connected to a Synapt G2 HDMS mass spectrometer (Waters, Manchester, U.K.). The mass spectrometer was operated using the capillary voltage and cone voltage of ± 2.5 KV and ± 35 V, respectively, for positive and negative ion modes respectively. Mass spectrometry analysis was performed using ESI+ mode with a source temperature of 100 °C, a spray voltage of 4500 V, period cycle time of 274 ms, accumulation time of 250 ms, *m/z* scan range between 80 and 1500 and collision energy of 45 V with CES of 10 V. Certain lipids such as fatty acids cannot be detected using ESI+ so the Synapt G2 instrument was also operated in ESI- mode. The mobile phase for lipidomics in ESI- mode comprised of solvent A consisting of aqueous solution of acetonitrile (60% acetonitrile %v/v) along with 10 mM ammonium acetate acting as a buffer and solvent B consisting of mixture of acetonitrile and isopropanol (10% acetonitrile %v/v) and 10 mM ammonium acetate. The ESI- mode was operated with a source temperature of 100 °C, spray voltage

of 2500 V, period cycle time of 274 ms, accumulation time of 100 ms, *m/z* scan range between 80 and 1500 and collision energy of 35 V with CES of 10 V. Mass spectrometry analysis for both ESI+ and ESI- modes was performed in data-independent and multi-parallel collision-induced dissociation MS^E acquisition mode, the precursor ion data was collected using a low collision energy setting of 5 eV, while for the fragment ion data collection the collision energy ramped from 15eV to 55 eV applied in the trap region of the Tri-Wave device. Argon gas was used for the collision-induced dissociation.

5.3.7 NMR

NMR data was collected on OSU's 800 MHz NMR spectrometer equipped with a triple resonance cryogenic probe.

5.3.8 Experimental Design

The experiment comprised triplicates of parental cells of doxycycline treatment (cells with DOX-based RNA construct) and control groups (cells with RNA construct without doxycycline) for both MD-MBA 231 and HS578T cell lines. The LC-MS experimental design comprised of two sample injections for each of these biological replicates. In addition to the sample injections, quality control (QC), created by pooling together 20 μL of each extract was also injected to account for deviations in data acquisition. The sample injections were randomized, and blanks and QC injections after every five sample injections to check for the performance of the method. NMR injections were carried out only for the polar extract of the biological replicates for both the cell lines.

LC-MS and NMR data were collected in three biological replicates of doxycycline treatment (cells with DOX-based RNA construct) and control groups (cells with RNA construct without doxycycline) for both MD-MBA 231 and HS578T cell lines. The LC-MS experimental design comprised of two sample injections for each of these biological replicates. In addition to the sample injections, quality control (QC), created by pooling together 20 μL of each extract was also injected to account for deviations in data acquisition. The sample and qc injections were randomized, and blanks and QC injections after every five sample injections to check for the performance of the method. NMR injections were carried out only for the polar extract of the biological replicates for both the cell lines.

5.3.9 Data Analysis

Peakview software was used to process UPLC-MS data for metabolomics and MassLynx software (version 2.2; Waters, Milford, MA, USA) was used to acquire and process UPLC-MS data for lipidomics. Extracted ion chromatograms were generated using MS data in these programs. The retention times obtained from these extracted ion chromatograms were compared with retention times of standards from the library.

The retention times of samples deviated within a window of two minutes with the standards. So automatic feature alignment could not be used. Therefore, Multiquant (version 3.0.2) was used for integration of peak areas of metabolites manually. IROA metabolite standards library, along with Human metabolic database (HMDB) and Metlin database was used for feature identification. Progenesis QI^{TM} software with Metlin^{TM} plugin (version 1.0.6499.51447; NonLinear Dynamics, United Kingdom) was used to align reversed phase chromatographic peaks, to perform feature detection with peak picking function, features were then matched with lipid databases for lipid identifications and compare the lipid levels between doxycycline treatment and control using intensity of fragments of lipids. The Progenesis used Human Metabolic Database and LipidBlast databases for lipid annotation. Lipid annotations with confidence were made for lipids that obtained score ≥ 50. The score ≥ 50 represented similarity of isotope pattern greater than 90%, MS/MS score higher than 50% and less than 5 ppm mass deviation from accurate mass. The product ions were annotated based on their elemental composition compared to the *m/z* values obtained with an accuracy of 10 ppm. The small polar metabolites were identified and quantified with NMR using the Chenomx (version 8.42; Chenomx, Edmonton, AB, Canada) software suite. Peak areas of analytes obtained using mass spectrometry and concentrations obtained using NMR were normalized by protein concentration to account for variation of metabolite and lipid levels across biological replicates.

The Welch's t-test with false discovery rate correction and unsupervised multivariate statistical methods were used to compare metabolites and lipids obtained from HRMS and NMR techniques across scr + and scr – cells. MetaboAnalyst (version 4.0) and RStudio (version 1.1) were utilized for data processing and analysis. Metabolites which were found to be differential by either MS or NMR method (>1.5 fold change, FDR<0.05) were

imported into Pathway analysis and Enrichment analysis tools provided by MetaboAnalyst 5.0 website.

5.4 Results

Metabolite assignments were performed by comparing the retention time, exact mass, isotope similarity, MS/MS fragmentation to the in-house library which resulted in 166 metabolite assignments in the ESI+ mode and 153 metabolite assignments in the ESI- mode (Table CS1 and A.T6.2). There were 254 metabolite assignments in both ESI+ and ESI- mode in mass spectrometry which were substantially higher than the number of metabolites identified using NMR (29). These metabolites and their levels were imported into Pathway analysis module of MetaboAnalyst 5.0 website to find out what metabolic pathways were affected by doxycycline. Minor metabolic pathways mentioned above were classified into four major metabolic pathways, namely central carbon, one carbon, methionine and glutathione metabolism. Tricarboxylic acid cycle (Figure CS1 a), glutamate and glutamine metabolism (Figure CS1 b), alanine-aspartate and glutamate metabolism (Figure CS1 c) can be classified under central carbon metabolism. The central carbon metabolism had significant hits (false discovery rate corrected p value ≤ 0.05) (Table CS3) in comparison with expected and total hits. Similarly, purine and pyrimidine metabolism, can be combined into one carbon metabolism (Clare et al. 2019; Suh, Choi, and Friso 2016) had significant hits FDR adjusted p value ≤ 0.05) (Table CS3) across MDA-MB-231 cells. Although glutathione (Figure CS1 d) and methionine metabolism did not have significant hits due to lack of standards in the IROA library, they showed a high impact (> 0.1). Fifty-three metabolites (Table CS4) belonging to these pathways were selected because of their reproducibility across both NMR and MS platforms and the statistical significance of these metabolites on the major biological pathways (Table CS3). 32 of these metabolites were selected using ESI positive ion mode whereas 21 metabolites were selected in ESI negative ion mode. Spectra information for 15 metabolites was available in both positive and negative modes. NMR was used to quantitatively determine concentration levels of 17 metabolites of which 9 were found using MS.

5.4.1 Comparative Metabolomics Analysis in MDA-MB-231 cells and HS578T cells

The results of comparison of 53 metabolites between doxycycline treatment (scr+) and control cells (scr-) in MDA-MB-231 cell line are listed in Table CS4. These results were analyzed to validate if the metabolic profile of doxycycline interferes with metabolic pathways involved in the breast cancer cells. Twelve metabolites were selected in the central carbon metabolism, eight metabolites from one carbon metabolism, eight metabolites from methionine metabolism and three metabolites from glutathione metabolism. The fold changes for the metabolites across central carbon, one carbon, methionine, glutathione metabolism was quantified using mass spectrometry and NMR between scr+ and scr- cells of the MDA-MB-231 cell line. False discovery rate (FDR) was obtained for statistical comparison of the metabolite levels across scr+ and scr- cells using Metaboanalyst 4.0 website. The metabolomics results comprising the fold change of metabolites between scr+ and scr- cells and their statistical significance (FDR) in each of the four metabolic pathways are shown using bar plots (Figure 5.1). Bar plots showed that the doxycycline treatment interfered with the oncometabolites in MDA-MB-231 cells (Table AT6.4) in the central carbon metabolism (Figure 5.1a), one carbon (Figure 5.2b), methionine (Figure 5.2c), glutathione metabolism (Figure 5.2d). Glutamate, glycine, choline, taurine, 5'-MethylThioAdenosine (5'MTA), NAD+ were upregulated significantly, whereas aspartate was downregulated significantly between scr+ and scr- cells of MDA-MB-231 cell line.

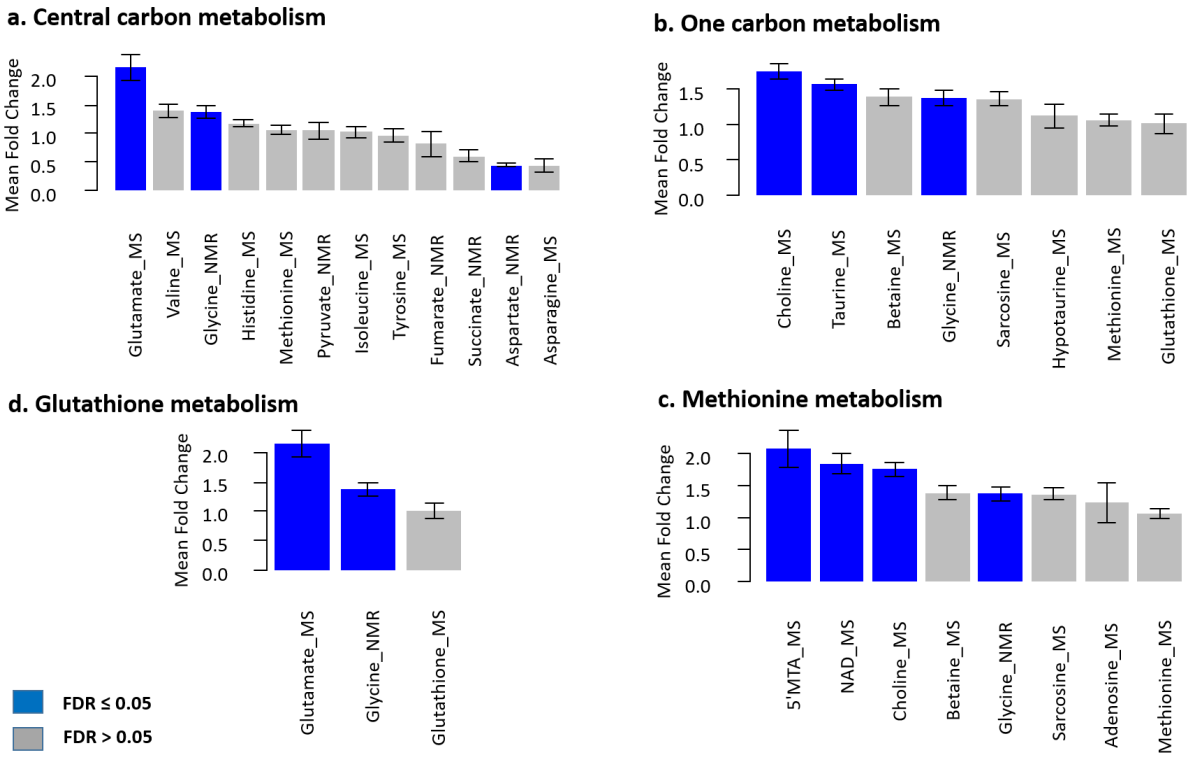


Figure 5.1 Bar plot with standard errors visualizing fold change for metabolite levels across super-pathways for the comparison of scr+ versus scr- MDA-MB-231 cells along with their statistical significance and corresponding acquisition method (NMR or MS) (a) Central carbon metabolism (b) One carbon metabolism (c) Methionine metabolism (d) Glutathione metabolism.

Comparative metabolomics was also performed on HS578T cells to validate whether doxycycline only impacted one type of cells or whether the effect of doxycycline was more profound, and whether it impacted multiple cell lines. NMR was used to detect and quantify fifteen metabolites and the fold changes across the doxycycline treatment (scr+) and control (scr-) cells of HS578T cell line. The statistically significant metabolites were obtained using false discovery rate adjusted t-test in Metaboanalyst 4.0 website. The fold changes and statistical significance are depicted in bar plot in Figure 5.2a and are listed in Table CS5. Lactate was significantly downregulated in scr+ cells as compared to scr- cells. Although not statistically significant individually, the depletion of energy metabolism

intermediates group that consisted of NAD+, pyruvate, glutamate and glycine was also observed to be statistically significant (p-value = 0.05) and with lactate the energy metabolism was statistically significant (p-value = 0.002) with substantial evidence across scr+ and scr- cells. Principal component analysis (PCA) (Figure 5.2b) was applied using these fifteen metabolites to identify the variation in the HS578T cells because of doxycycline treatment. PCA results showed clear separation between the scr+ (doxycycline treatment) and scr- cells (control).

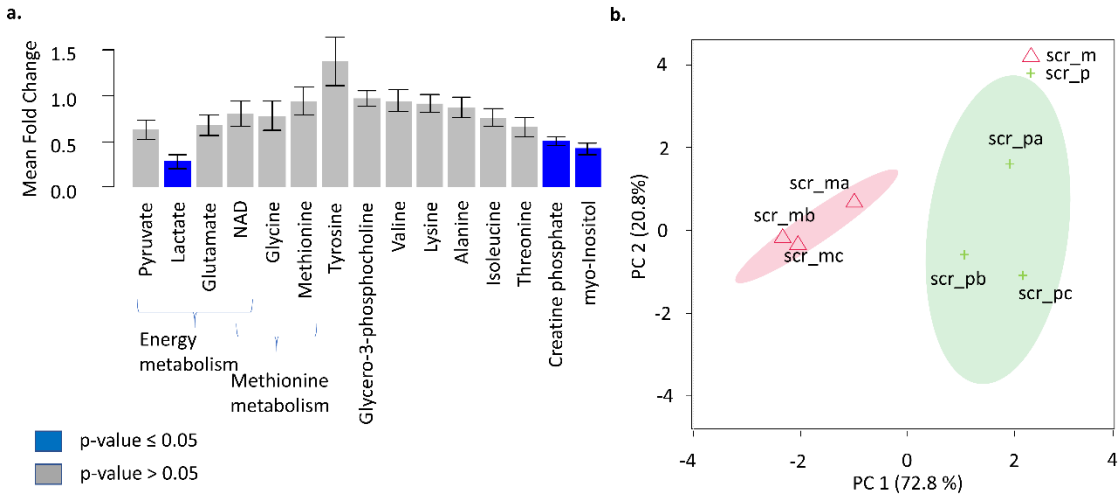


Figure 5.2 NMR Metabolomics to study the effect of doxycycline on HS578T cell line. (a) Bar plot with standard errors to visualize fold changes across polar metabolites identified in scr+ and scr- cells along with their statistical significance. (b) Score plot for principal component analysis of polar metabolites across three biological replicates of scr+ and scr- cells.

5.4.2 Comparative Lipidomics Analysis in MDA-MB-231 and HS578T cells

The classes of lipids that were involved in cancer proliferation, membrane fluidity, mitochondrial impairment, cell signalling and apoptosis were selected. Therefore, 32 different lipids (Table CS6) for MDA-MB-231 cells and (Table CS7) for HS578T cells

were selected from phosphatidylcholines (PC), phosphatidylethanolamine (PE), plasmalogens, SM (sphingomyelins) and cardiolipins (CL) lipid classes from a list of 69 annotated lipids in case of MDA-MB-231 cells (Table CS8) and amongst the list of 78 annotated lipids in case of HS578T cells (Table CS9) using RPLC-MS in both ESI+ and ESI- modes. The selected lipids had different number of carbons and a varying level of unsaturation.

Thirty two different lipids (Table CS6) for MDA-MB-231 cells and (Table CS7) for HS578T cells were selected from phosphatidylcholines (PC), phosphatidylethanolamine (PE), plasmalogens, SM (sphingomyelins) and cardiolipins (CL) lipid classes from a list of 69 annotated lipids in case of MDA-MB-231 cells (Table CS8) and from the list of 78 annotated lipids in case of HS578T cells (Table CS9) using RPLC-MS in both ESI+ and ESI- modes as these lipids were associated with cancer proliferation, membrane fluidity, mitochondrial impairment, cell signalling and apoptosis. These lipids selected had different number of carbons and a varying level of unsaturation.

The volcano plots in Figure 5.3a and Figure 5.4a shows the lipidomic profile for doxycycline treatment for MDA-MB-231 and HS578T cells respectively. Most of the lipids considered, were significantly upregulated in cells that received the doxycycline treatment (Figure 5.4a) in MDA-MB-231 cells. However, there was no clear trend observed for lipids in HS578T cells but most of lipids considered were significantly dysregulated (Figure 5.4a). The lipids in both MDA-MB-231 and HS578T cells clustered into two distinct groups for the 32 selected lipids between the doxycycline treatment with scrambled RNA construct (scr+) and scrambled RNA construct without doxycycline (scr-) as seen in PCA plot in Figure 5.3b and Figure 5.4b respectively.

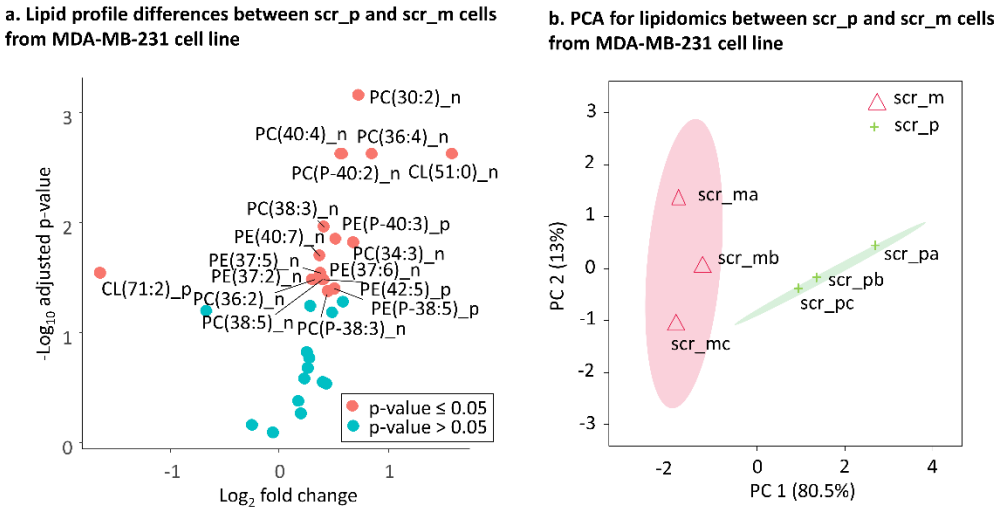


Figure 5.3 Results of comparative lipidomics between scr+ (scr_p) and scr- (scr_m) of MDA-MB-231 cell line. (a) Volcano plot comparing the differences in the lipid profile caused due to doxycycline treatment in MDA-MB-231 cells. The adjusted p-value and fold change were obtained using false discovery rate (FDR) and the mean ratio of lipid levels respectively across scr+ and scr- cells. The lipids with adjusted p-value ≤ 0.05 (statistically significant) are represented in red whereas the lipids with adjusted p-value > 0.05 are represented in blue. Lipids are represented as LipidFamily(NC:NB)_p, where lipid family specifies which lipid family does the lipid of interest belong to, NC represents number of carbons and NB denotes the number of unsaturated bonds. '_p' represents ESI positive mode of data acquisition or '_n' mode shows ESI negative mode of data acquisition. The significant lipids (p-value ≤0.05) are denoted by *. (b) Score plot for principal component analysis of lipids across 3 biological replicates of scr+ and scr- cells for all lipids detected.

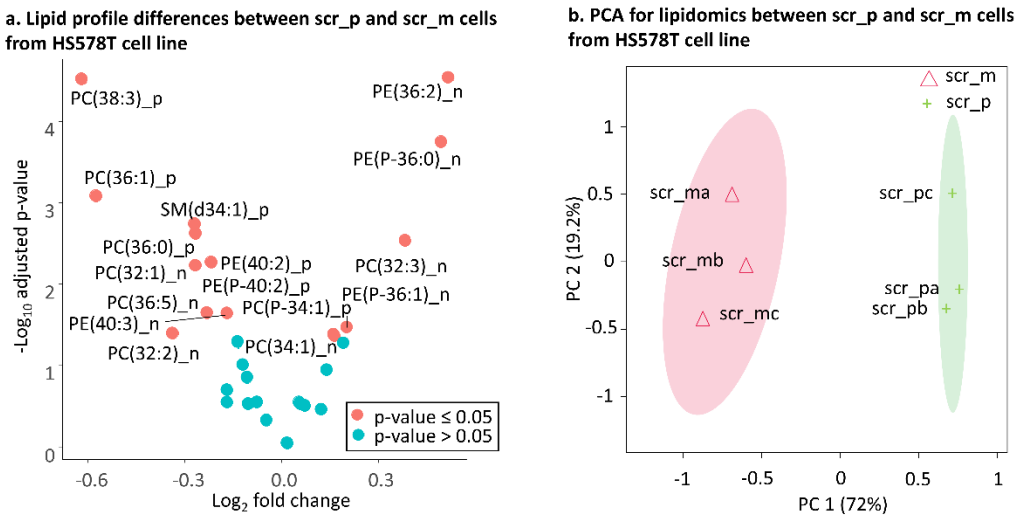


Figure 5.4 Results of comparative lipidomics between scr+ (scr_p) and scr- (scr_m) of HS578T cell-line. (a) Volcano plot comparing the differences in the lipid profile caused due to doxycycline treatment in HS578T cells. The adjusted p-value and fold change were obtained using false discovery rate (FDR) and the mean ratio of lipid levels respectively across scr+ and scr- cells. The lipids with adjusted p-value ≤ 0.05 (statistically significant) are represented in red whereas the lipids with adjusted p-value > 0.05 are represented in blue. Lipids are represented as LipidFamily(NC:NB)_p, where lipid family specifies which lipid family does the lipid of interest belong to, NC represents number of carbons and NB denotes the number of unsaturated bonds. '_p' represents ESI positive mode of data acquisition or '_n' mode shows ESI negative mode of data acquisition. The significant lipids (p-value ≤ 0.05) are denoted by *. (b) Score plot for principal component analysis of lipids across 3 biological replicates of scr+ and scr- cells for all lipids detected.

Significantly higher levels of unsaturated PEs, PE (37.2), PE (37.5) and PE (37.6) were observed in the scr+ cells of MDA-MB-231 cell line (Figure 5.3a). However, the HS578T cell line showed both upregulation and downregulation trends in unsaturated PEs as PE (36:2) was upregulated while PE (40:2) and PE (40:3) were downregulated significantly in

the scr+ cells (Figure 5.4a). Similarly unsaturated PCs, PC (30:2), PC (34:3), PC (36:2), PC (36:4), PC (38:3), PC (38:5) and PC (40:4) were also upregulated significantly in the scr+ cells of the MDA-MB-231 cell line (Figure 5.4a). The HS578T cell line did not exhibit a clear downregulation or upregulation trend in unsaturated PCs as there was significant increase in the levels of PC (32.3) and PC (34.1) and a significant decrease in the levels of PC (32:1), PC (32:2), PC (36:0), PC (36:1) and PC (36:5) in the scr+ cells (Figure 5.4a). Unsaturated sphingomyelin, SM (d34:1) was significantly downregulated in the scr+ cells of HS578T cell line (Figure 5.4a).

Significant upregulation of plasmalogens PC (P-38:3), PC (P-40:2), PE (P-38:5) and PE (P-40:3) were observed in the scr+ cells of the MDA-MB-231 cell line (Figure 5.3a). However, the HS578T cell line exhibited significantly higher levels of PE (P-36:60) and significantly lower levels of PE (P-36:0) (Figure 5.4a). There was a significant decrease in CL(i-71:2) and on the other hand, CL(i-51:0) showed significant increase (Figure 5.3a) in scr + vs scr – cells in MDA-MB-231 cell line.

5.5 Discussion

5.5.1 Correlation between metabolite levels detected using NMR and HILIC-MS

Both analytical techniques, NMR and LC-MS have their own set of advantages and disadvantages (Nagrath et al. 2011). The combination of NMR and LC-MS/MS methods resulted in an expanded number of metabolites detected due to the complementarity of the two methods (Figure 5.5a). The metabolites found in both NMR and LC-MS were alanine, aspartate, choline, glutamate, isoleucine, lactate, phenylalanine, succinate, taurine, threonine, tyrosine, and valine. The metabolites were classified into three groups based on their coefficient of variation which was calculated using lm function in R v4.0.1 in R Studio Version 1.2.5033. The metabolites with coefficient of determination (R^2) greater than or equal to 0.7 were classified into the high correlation group. The high correlation group was composed of glutamate, phenylalanine, tyrosine, and taurine. Glutamate had the highest coefficient of determination of 0.95 (Figure 5.5b).

Tyrosine and taurine had coefficient of determination of 0.83 and 0.8 respectively followed by phenylalanine with coefficient of determination of 0.73. The R^2 value for correlation for

tyrosine is comparable with R^2 value reported previously (Nagana Gowda et al. 2018). The metabolites that had the R^2 value between 0.4 and 0.7 were classified into the medium correlation group. The medium correlation group encompasses alanine, aspartate, choline, threonine, and valine. The correlation values for threonine, alanine, choline, and aspartate are 0.69, 0.66, 0.52 and 0.47 respectively. The metabolites that had correlation between 0.3 and 0.4 were classified into the low correlation group. R^2 values for isoleucine and lactate were 0.37 and 0.33 respectively and they showed low correlation with of NMR with LC-MS. Glutamine was classified into poor correlation group. The possible explanation for poor correlation in glutamine is cyclization (Nagana Gowda et al. 2018). The other possible explanation for lower coefficient of determination for the polar metabolites in this study as compared to the previously reported correlations (Nagana Gowda et al. 2018) is the lower sample size and lower concentration in this study.

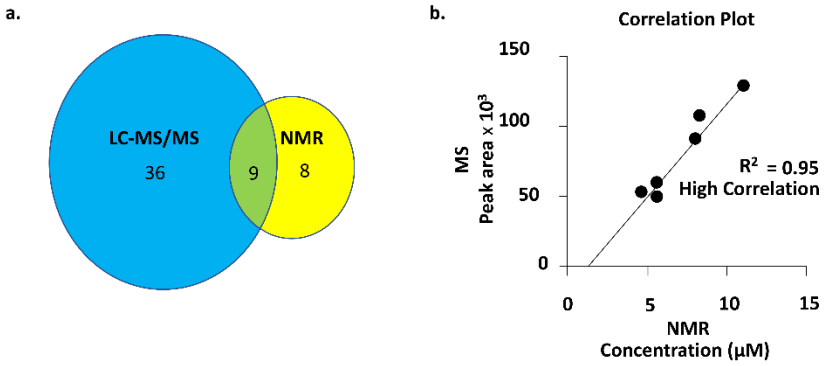


Figure 5.5 Correlation between NMR and LC-MS for polar metabolites detected across scr+ and scr- MDA-MB-231 cells (a) Venn diagram representation of metabolites observed using NMR & LC-MS in ESI+ and ESI– (b) Correlation between glutamate levels in NMR and LC-MS.

5.5.2 Effects of Doxycycline on Metabolites in MDA-MB-231 cells and HS578T cells

Glutamate, glutathione, glycine and serine metabolism and Warburg effect have been previously reported to be significantly affected in MCF-7 breast cancer cell line (Semreen et al. 2019) . The central carbon metabolism was found to be significantly affected (Figure 5.1a). Mitochondrial function is linked to metabolic pathways including glycolysis, pentose phosphate pathway, tri-carboxylic acid (TCA) cycle and glutaminolysis. These metabolic pathways grouped together form the central carbon super pathway which is known to be affected in the cancer cells (C. Yang et al. 2008). Aspartate is synthesized from oxaloacetate, which is a part of TCA cycle. Aspartate is also one of metabolites involved in biosynthesis of nucleic acids. The levels of aspartate are lower in the cancer cells because of increased requirement of *de novo* synthesis of nucleic acids and aspartate-malate shuttle for mitochondrial respiration in the breast cancer cells (Cheng et al. 2018; Xie et al. 2015). The significant decrease (p-value = 0.03) in levels of aspartate between control and doxycycline treatment groups suggests that mechanism of proliferation or mitochondrial function is possibly impacted by doxycycline.

The glutamate levels were significantly upregulated in the scr+ cells as they were more than two times the level in the scr- cells. The glutamate is synthesized from glutamine and is converted to alpha-ketoglutarate that enters the TCA cycle. The depletion of glutamine and significant upregulation in glutamate (p-value = 0.02) which could be seen in both NMR and LC-MS suggests that energy metabolism is negatively impacted by doxycycline. The decline in energy production from doxycycline treatment suggests a decrease in the proliferation of the cancer cells. (Uifălean et al. 2016) . The cells proliferating rapidly require glycine because the demand for glycine is much more than rate at which it is synthesized in the rapidly proliferating cells (Braun et al. 2012). The significant upregulation of glycine (FDR = 0.02) in scr+ cells suggest a decrease in their proliferation rate. The significant upregulation of glutamate and glycine and downregulation of aspartate between scr+ and scr- suggests deregulation of central carbon metabolism (Figure 5.1a).

Glutamate, glycine and glutathione are intermediates of glutathione metabolism (Locasale 2013) and were upregulated in scr+ cells in case of MDA-MB-231 cell line (Figure 5.1d). Upregulation of glutamate (FDR = 0.02) and glycine (FDR = 0.02) was statistically

significant. The transsulfuration pathway produces glutathione utilizing glutamate and glycine, plays an important role in establishing redox balance by removing reactive oxygen species (ROS) and regulating protein function by S-glutathionylation (Locasale 2013) (Xiong et al. 2011). Glutathione metabolism is associated with maintenance of redox balance. Glutathione metabolism scavenges reactive oxygen species (ROS), therefore deregulation of glutathione metabolism could suggest deregulation of inhibition of apoptosis in the scr+ cells (Acharya et al. 2010).

In addition to glycine, choline and taurine levels were also upregulated in scr+ in comparison with scr- cells of the MDA-MB-231 cell-line. This suggests deregulation of the one carbon metabolism, which is responsible for metabolic integration of one-carbon units from amino acid inputs towards cellular and nuclear biosynthesis, substrates for epigenetic gene silencing and signal transduction (Locasale 2013). Upregulation of glycine can adversely impact the proliferation of the cancer cells by preventing calcium signaling (Rose, Cattley, et al. 1999; Rose, Madren, et al. 1999). The upregulation of taurine metabolism also suggests downregulation of nucleotide metabolism (Mehrmohamadi et al. 2014) in scr+ cells which also suggests the proliferation of cancer cells could be negatively impacted because of decrease in *de-novo* synthesis of nucleotides. Choline provides one carbon units to betaine which is involved in the formation of methionine and deregulation of choline affects both one carbon and methionine metabolism. Choline levels have also been negatively correlated with breast cancer which could indicate that upregulation of choline could negatively impact the scr+ cells (Xu et al. 2008).

The methionine pathway seems to be deregulated since there are discriminating differences in the levels of NAD, choline, glycine and 5'-methylthioadenosine between scr + and scr – cells (Figure 5.1c). The methionine pathway responds to alterations in TCA pathway with epigenetic modifications in mitochondrial DNA associated with DNA methylation. TCA pathway regulates the mitochondrial DNA (mtDNA) depletion. DNA methylation is also impacted in response to mtDNA depletion resulting in epigenetic modifications in gene expression for methionine metabolism (Lozoya et al. 2018). The dysregulation in the steady state levels of 5'-methylthioadenosine and choline across scr+ and scr- cells indicated alteration in the methionine salvage pathway in MDA-MB-231 cells (Albers

2009; Lozoya et al. 2018). NAD+ helps in rescuing methionine salvage pathway and averts changes in DNA methylation (Lozoya et al. 2018). The NAD+ is also linked with purine nucleotide synthesis and other signaling pathways (Chiarugi et al. 2012). Nucleotide biosynthesis is essential for proliferation of tumor cells and upregulation of NAD+ levels in scr+ cells suggests that the doxycycline treatment will have an effect on DNA methylation and proliferation of scr+ cells.

The cancer cells have higher levels of lactate as they obtain energy majorly through aerobic glycolysis or Warburg effect (Weinhouse et al. 1956). The significant downregulation of lactate (p-value = 0.04) in scr+ cells of HS578T cell line (Figure 5.2a) suggests inhibition of glycolysis. In addition to glycolytic intermediate lactate, other energy metabolism intermediates, NAD+, pyruvate, glutamate and glycine, were also downregulated. Their downregulation was statistically significant as a group of energy metabolism intermediates, that suggested energy production in the scr+ cells was affected.

Decrease in energy production in the cancer cells also suggest a decrease in the proliferation and migratory potential. Therefore, doxycycline is linked with deregulation of energy metabolism of HS578T cells. Although the effect of doxycycline is not the same in MDA-MB-231 and HS578T cells doxycycline has significant impact on one or more super-pathways in these cells. Different cell lines also exhibited different sensitivity to doxycycline (Zhong et al. 2017). Difference in metabolite levels in both triple negative cell lines led us to believe that dysregulation of metabolites in scr + cells in the aggressive triple negative cancer cell line could be attributed to doxycycline.

5.5.3 Effects of Doxycycline on Lipids in MDA-MB-231 and HS578T cells

As may be seen from PCA plots presented on Figure 5.4b and 6.5b, the lipidomic profiles are different for MDA-MB-231 and HS578T cells. Levels of unsaturated phospholipids are higher in scr+ as compared with scr- cells (Figure 5.3) from MDA-MB-231 cell line. Phospholipids such as PC, SM, PE are an important component of cellular membranes in presence of doxycycline the polyunsaturated phospholipids, PC (36:4), PC (38:5), PC (40.4), PE (37:5), PE (37:6), PE (40:7) PE (42:5) and PE (46:5) and appear to be significantly upregulated in scr+ cells of MDA-MB-231 breast cancer cell line. Plasmalogens have been associated with membrane fluidity (Eiriksson et al. 2018) and CL

has been associated with mitochondrial function (Dória et al. 2013; Hardy et al. 2003). Breast cancer cells have lower levels of polyunsaturated phospholipids (Cífková et al. 2015) as polyunsaturated phospholipids are more likely to be oxidized that has potential to cause cell damage. As a result cancer cells favor higher levels of saturated phospholipids in cell membranes and lower levels of polyunsaturated phospholipids in cancer cells (Hilvo et al. 2011). The results indicate upregulation of unsaturated phospholipids indicating increased susceptibility of scr+ cells of MDA-MB-231 breast cancer cell line to cellular damage and apoptosis.

Significantly higher levels of PC (32:1) and PC (32:2) were observed (bar plot Figure 5.4a) in doxycycline treated cells of HS578T cell line. These lipid species are involved in *de-novo* fatty acid synthesis used in cellular membrane synthesis. Their upregulation of these lipid species is associated with the decreased survival of triple negative breast cancer cells with statistical significance (Hilvo et al. 2011). PCs can be converted to SMs (Sphingomyelins). The possible mechanism associated with the higher levels of PCs is apoptosis in the scr+ cells as significantly lower SM levels have been found for SM (34:1) in scr+ HS578T cells which has been previously linked with inhibition of apoptosis (Cífková et al. 2015); (Ségui et al. 2006). However, downregulation of SMs association with inhibition of apoptosis could not be confirmed due to the lack of sphingomyelin annotations. Differences in the lipidomic profile of these lipids in the tumor cells suggests that the doxycycline treatment induced apoptosis in the triple negative breast cancer cells.

The migratory potential of the breast cancer cells through epithelial to mesenchymal transition EMT is also associated with decrease in phosphatidylcholine- and phosphatidylethanolamine plasmalogens. However, there was a significant increase in the levels of PC (P-38:3), PC(P-40:2), PE(P-38:5) and PE(P-40:3) in the MDA-MB-231 cell line. The other plasmalogens in MDA-MB-231 also showed upregulation trend in the scr+ cells. The vinyl ether bond in plasmalogens decreases the fluidity of the membranes that decreases the migratory potential of breast cancer cells (Eiriksson et al. 2018). Although majority of plasmalogens in the HS578T cell line did not exhibit upregulation or downregulation trend, PC (P-34:1) and PE (P-36:0) were significantly deregulated between scr+ and scr- cells in HS578T cell line. Clinical studies using plasmalogen based cancer

biomarkers (Smith et al. 2008) would result in erroneous result in identifying lipid signatures associated with cancer cells (Cífková et al. 2015) because of deregulation of plasmalogen lipid biomarkers. This would lead to ineffective drugs as it would falsely estimate higher efficacy for drug of interest.

Migration of metastatic cancer cells is also regulated by cardiolipins which are associated with activity of proteins in the inner membrane of mitochondria (A. J. Chicco and Sparagna 2007; Dória et al. 2013). The downregulation of CL(i-71:2) and upregulation of CL(i-51:0) seen in Figure 5.4a in doxycycline treated MDA-MB-231 cells points at two different effects of doxycycline on cardiolipins with different lengths and saturation levels. Decrease in the cardiolipin levels has been associated with change in mitochondrial function leading to apoptosis in breast cancer cells (Hardy et al. 2003). This suggests that doxycycline treatment seems to alter saturated and unsaturated cardiolipins in a different way or that doxycycline treatment impacts shorter chain cardiolipins differently as compared to the longer chains. The downregulation of cardiolipins could not be fully linked with apoptosis in breast cancer cells because of lack of cardiolipin annotations found in these cells.

5.6 Conclusion

There were significant differences in 53 metabolites detected and quantified in MDA-MB-231 cell line associated with central carbon, one carbon, methionine, glutathione metabolism. Similar trends were also observed in HS578T cell line. These effects were attributed to Warburg effect caused by doxycycline treatment. The oncogenesis in the cancer cells is impacted in the breast cancer cells because of deregulation of the intermediates in these metabolic pathways. Similarly, comparable differences were observed between doxycycline treatment and control in both MDA-MB-231 and HS578T cells in the lipid species, PCs, PEs, plasmalogens and SMs. The alterations of the lipidomic profile suggested that the biosynthesis of cellular membranes, migration potential and inhibition of apoptosis in the breast cancer cells were deregulated. The metabolomics and lipidomics results allowed us to conclude that some of the important characteristics of breast cancer cells such as elevated migratory potential, potential of rapid cell division and energy production are restricted in cancer cells because of doxycycline treatment. Use of doxycycline in chemotherapies also suggests the role of doxycycline to inhibiting the

metastasis of breast cancer cells. Therefore, caution must be exercised when using this system, because the doxycycline itself can have significant effects on cell metabolism. The results also outline the mechanisms by which doxycycline inhibits invasion of cancer cells utilizing the metabolomics approach.

6. Summary, Conclusion and Perspectives

This chapter describes the summary of the insights obtained by conducting computationally assisted metabolomics. Metabolomics can be conducted using wide range of hyphenated techniques that comprise of LC-IR, GC-MS, GC-UV, GC-FID, LC-MS and LC-NMR. GC works well with volatile compounds, and it can be annotated well with online databases but suffers from the drawback of lesser coverage of metabolites. The mass spectrometry instrument is not easily available because it is very expensive. In these situations, other detector such as UV/visible or flame ionization detector (FID) can be used with GC. However, these instruments do not offer wide coverage of metabolites and their sensitivity performance is lower in terms of detection limit (Bai et al. 2018). LC-NMR is less sensitive as compared to mass spectrometry but provides more structural information than mass spectrometry. LC-IR is slower as compared to mass spectrometry and is less sensitive however the instrument is less expensive in comparison to mass spectrometry instrument. The strengths of LC-MS include wide metabolome coverage, quantitative measurements, detection limit in the picomole range. These analytical techniques are complementary to each other with each one having their own strengths and limitations.

This thesis was focused on method development in both aspects of metabolomics, untargeted LC-MS/MS analysis, and computational tools, that when combined enhanced in statistical analysis, functional interpretation, and data visualization for deriving insights in the biology of cells and tissues, as well as the characterization of botanical extracts.

Chapter 3 describes the development of an anlytical strategy that combins LC-MS metabolomics with computational tools to advance the characterization of the phytochemical composition of plant material and faciliate the discovery of bioactive compounds in complexes matrices. As such, this porject played a key role in understanding bioactivity of a botanical extract and potential health benefits. The robustness of the LC-MS metabolomics methods enables testing of product integrity and consistency to improve reproducibility of laboratory and preclinical studies and. most importantly, clinical trials. The primary focus of this study was to develop an efficient method for obtaining chemical

fingerprints of botanical extracts and perform quantitative measurement of marker compounds commonly associated with health benefits that can be used to determine product integrity of botanical extracts.

Centella asiatica (*C. asiatica*), an Ayurvedic herb with potential applications in enhancing mental health and cognitive function (Soumyanath et al. 2012), was used as a case study. *C. asiatica* water extract was prepared by refluxing aerial parts of the plant followed by filtering. The extracts were suspended in aqueous methanol and separation was performed using HPLC with Phenyl-3 column utilizing π-π interactions. The HPLC analysis was coupled with HRMS using quadrupole time-of-flight analyzer in data dependent acquisition (DDA) mode. This enabled the characterization of the extracts and the putative identification of 117 compounds. 2D structural similarity network was created using these compounds with the Tanimoto similarity score that displayed clusters which are structurally similar at the boundary nodes at the 95% confidence level. Precursor ions (MS1) in conjunction with authentic standards and external calibration curves were used for the quantification of selected marker compounds. In this study it was demonstrated that chemical fingerprinting of botanical extracts can be performed using combination of liquid chromatography, high resolution mass spectrometry in the DDA mode with less than 5 ppm deviation in *m/z* values for precursor ions (Alcazar Magana et al. 2020). This technique is robust and will result in accurate quantification of phytochemicals with very low limits of quantification (1.06 µg/L). This analysis revealed diverse metabolite composition in different *C. asiatica* accessions. This provided evidence to warrant the need for standardization and chemical detailed characterization of plant extracts for reliability and reproducibility in studies looking to determine biological activity of CA and botanical extracts. In addition to that characterization is also required to demonstrate reliability in efficacy and safety of botanical extracts in laboratory studies and clinical trials conducting studies on their health benefits.

Another aspect of this project was the development and implementation of computational approaches to identify medicinal activity of the plant. The potential health benefits that plant materials offer are due to the metabolites produced in the plants. The usual way to search for these metabolites is to use bioassay-guided fractionation that requires multiple steps of purification that takes a long time. This study aimed at reducing the time needed

for discovery of these bioactive metabolites using simple fractionation approaches in conjunction with bioassays, flow injection mass spectrometry and computational tools. This linking the medicinal activity (bioactivity) of each fraction with the chemical composition of each fraction. *Centella asiatica* water (CAW) extract was used as a case study. Correlation score between neuroprotective bioactivity of CAW extract and the subfractions for human neuroblastoma MC65 cells against amyloid β-cytotoxicity (Copenhaver et al. 2011) and relative abundance of metabolites identified in CAW extract and the subfractions. Computational approaches were evaluated to predict metabolites contributing towards neuroprotective bioactivity. ElasticNet and PLS-DA selectivity ratio were used to predict the neuroprotective bioactivity of the molecular (mass spectral) features.

Additionally, Global Natural Product Social Molecular Networking (GNPS) platform (Mingxun Wang et al. 2016) was utilized for the purposes of dereplication to facilitate annotation of plant metabolites and the annotation of bioactivity to metabolites. GNPS generates molecular networks based on spectral similarity (MS/MS spectral similarity), to derive insights about the structure of metabolites, hitherto unknown compounds, and to suggest undiscovered metabolic mechanisms in plants, etc. In this study, the GNPS platform was leveraged to visualize correlation of molecular features to neuroprotective bioactivity of MC65 assay and the corresponding peak areas in the subfractions. This approach identified di-caffeoylquinic acids as molecular features from *Centella asiatica* which provided the highest protection against amyloid β-cytotoxicity. Results obtained from the computational approach were validated with a traditional reductionistic approach of finding the neuroprotective activity of di-caffeoylquinic acids (Gray et al. 2014).

The fourth chapter deals with assessing the impact of storage of tissues in OCT. This project establishes protocol to utilize tissues stored in OCT, since the biobanks usually store tissues in OCT as it provides stability to the embedded tissue and does not affect infrared (IR) and Raman spectroscopic methods. The study evaluated the compatibility of tissue storage in OCT by conducting LC-MS/MS lipidomic studies to compare lipid profile across OCT storage and storage by snap freezing in liquid nitrogen with Macaque heart tissue (Vaswani et al. 2021). The lipid extraction protocol was designed for OCT-

embedded tissue and was tested for compatibility. The lipid profile was obtained using LC-MS/MS analysis. The 306 annotated lipids were compared across both OCT and LN storage conditions. The comparison confirmed the hypothesis that 95.4% lipid species did not have an effect of storage condition. However, 4.6% of the lipid species mainly plasmalogens, were affected by the OCT storage method and their levels decreased in storage under OCT. The results of this study described standard operating procedure (SOP) to be followed in determining lipid species profiles of OCT-embedded tissues by ESI LC-MS/MS. However, caution needs to be taken when evaluating plasmalogens in OCT-embedded tissue, likely due to sensitivity of plasmalogens to oxidation. Therefore, the ability to have tissue stored in OCT and available for metabolic assessment using analytical methods, such as mass spectrometry, makes it easier to access a wider pool of cardiac tissue.

The fifth chapter outlines studies designed to determine if a doxycycline (DOX)-dependent gene expression knockdown system is a viable strategy in the context of metabolomic studies of breast cancer cells for studying the biological effects of targeted gene silencing. This study involved two widely used breast cancer cell lines, triple negative MDA-MB-231 and HS578T. Untargeted mass spectrometry metabolomics and lipidomics was used to determine metabolite and lipid differences in presence and absence of DOX. The metabolomics was conducted using hydrophilic liquid chromatography coupled with mass spectrometry (HILIC-MS) in data dependent acquisition mode. Since molecular pathways involved in energy metabolism comprise of polar metabolites NMR was also used along with HILIC-MS for metabolomics. HILIC chromatographic method was optimized to alleviate the problems of ion suppression and peak tailing. The lipidomics was conducted using reversed phase liquid chromatography couple with mass spectrometry in MS^E mode. The metabolomics allowed us to identify the effect of doxycycline on major metabolic pathways such as central carbon, one carbon, methionine and glutathione metabolism. The doxycycline effect was also apparent in deregulation of lipid metabolism belonging to phosphatidylcholines (PC), phosphatidylethanolamine (PE), plasmalogens, SM (sphingomyelins) associated with cancer proliferation, membrane fluidity, cell signalling and apoptosis. The lipid species of varying number of carbons and a varying level of unsaturation were used to assess this effect. Doxycycline

was found to affect energy metabolism in both cells. Doxycycline was also found to interfere with inhibition of apoptosis and impact methylation of DNA in promoter of tumor suppressor genes such as E cadherin and increase the PC and PE plasmalogen level that led to the arrest of the migration in MDA-MB-231 breast cancer cells. Therefore, this study would caution researchers from using doxycycline-basedgene silencing strategies in case that metabolomics and lipidomics analysis are the predominant experimental strategies to assess the effects of gene silencing.

To conclude, this thesis describes mass spectrometry methods in conjunction with computational tools to advance vastly different application areas, ranging from identifying disease mechanisms to biomarker discovery, and the characterization of botanicals including their bioactive compounds. Mass spectrometry has emerged as a primary measurement technology in "omics" (Girolamo et al. 2013)because it can detect a wide range of metabolites and lipid species even at low nano and pico-molar concentrations. The coverage of metabolites can be improved by combining mass spectrometry with gas phase separation techniques, such as ion mobility spectrometry platforms, to expand the analytical separation space and to distinguish isomers. Mass spectrometry methods are considered comprehensive methods as they have number of advantages such as high sensitivity, reproducibility, and wide coverage of compound detection. The biggest area of research using mass spectrometry is omics studies. Omics studies involving mass spectrometry generate big datasets. Analysis of these datasets is advanced by computational tools, such as complex machine learning and deep learning algorithms to predict disease onset or progression or in drug discovery. However, because of the projects described in this thesis involved smaller sized datasets, classical linear regression, and machine learning algorithms with lower tendency of overfitting were used.

The integration of mass spectrometry and computational tools has advance natural product research. Spectral databases integrating MS/MS spectra and ion mobility spectra could lead to rapid and more efficient dereplication of natural product mixtures. Moreover, integrating platforms like GNPS, that are working automated pipelines, will advance the integration of mass spectral data with secondary information, e.g. bioactivity, to facility the integration of information from different technologies. Additional computational tools

could be integrated into such platforms to determine relationship of metabolites with each other structurally and functionally and generate reports about these networks to enable more efficient discovery of natural products in the future.

Beyond the scope of this thesis, mass spectrometry methods have contributed to advancements in the food and agricultural sciences. Metabolomics has been used to improve agricultural practices aimed to improve the quantity and quality of agricultural produce. Mass spectrometry-based analyses are capable of determining optimal growth conditions, including, for example, how LED light impacts plant growth (Yap et al. 2021). High resolution mass spectrometry also finds application in identifying adulterants in food (Witjaksono and Alva 2019). Wine industry utilizes mass spectrometry in assessing the quality of wine specimens (Arapitsas et al. 2020).

The metabolomics and lipidomics field will also greatly aid in assisting traditional genomics and transcriptomics in personalized medicine and surgeries. Technological advances in instrumentation such as mass spec pen (Brown, Pirro, and Graham Cooks 2018) can be leveraged in minimal invasive surgeries to accurately identify cancer tissue in real time.

Bibliography

{R Core Team}. 2020. “R: A Language and Environment for Statistical Computing.” *R Foundation for Statistical Computing*. http://www.r-project.org.

Aagaard, Lars et al. 2007. “A Facile Lentiviral Vector System for Expression of Doxycycline-Inducible ShRNAs: Knockdown of the Pre-MiRNA Processing Enzyme Drosha.” *Molecular Therapy* 15(5): 938–45.

Abbas-Mohammadi, Mahdi et al. 2018. “Acetylcholinesterase-Inhibitory Activity of Iranian Plants: Combined HPLC/Bioassay-Guided Fractionation, Molecular Networking and Docking Strategies for the Dereplication of Active Compounds.” *Journal of Pharmaceutical and Biomedical Analysis* 158: 471–79.

Abdel-Latif, Ahmed, Paula M. Heron, Andrew J. Morris, and Susan S. Smyth. 2015. “Lysophospholipids in Coronary Artery and Chronic Ischemic Heart Disease.” *Current Opinion in Lipidology* 26(5): 432–37.

Abdullah, Laila et al. 1950. “Harmonizing Lipidomics: NIST Interlaboratory Comparison Exercise for Lipidomics Using Standard Reference Material 1950 – Metabolites in Frozen Human Plasma.” *Journal of L*.

Acharya, Asha, Ila Das, Des Chandhok, and Tapas Saha. 2010. “Redox Regulation in Cancer: A Double-Edged Sword with Therapeutic Potential.” *Oxidative medicine and cellular longevity* 3(1): 23–34.

Achenbach, Tatjana V., Bodo Brunner, and Kathrin Heermeier. 2003. “Oligonucleotide-Based Knockdown Technologies: Antisense versus RNA Interference.” *ChemBioChem* 4(10): 928–35.

Acin-Perez, Rebeca et al. 2009. “Cyclic AMP Produced inside Mitochondria Regulates Oxidative Phosphorylation.” *Cell Metabolism* 9(3): 265–76.

Afendi, Farit Mochamad et al. 2012. “KNApSAcK Family Databases: Integrated Metabolite-Plant Species Databases for Multifaceted Plant Research.” *Plant and Cell Physiology* 53(2).

Albers, Eva. 2009. “Metabolic Characteristics and Importance of the Universal Methionine Salvage Pathway Recycling Methionine from 5′-Methylthioadenosine.” *IUBMB Life* 61(12): 1132–42.

Alcazar Magana, Armando et al. 2020. “Integration of Mass Spectral Fingerprinting Analysis with Precursor Ion (MS1) Quantification for the Characterisation of Botanical Extracts: Application to Extracts of Centella Asiatica (L.) Urban.” *Phytochemical Analysis* 31(6): 722–38.

Ali, Kashif et al. 2013. “Identification of Bioactive Metabolites against Adenosine A1 Receptor Using NMR-Based Metabolomics.” *Metabolomics* 9(4): 778–85.

Alpert, Andrew J. 1990. “Hydrophilic-Interaction Chromatography for the Separation of Peptides, Nucleic Acids and Other Polar Compounds.” *Journal of Chromatography A* 499(C): 177–96.

Alqahtani, Ali et al. 2015. “Seasonal Variation of Triterpenes and Phenolic Compounds in Australian Centella Asiatica (L.) Urb.” *Phytochemical Analysis* 26(6): 436–43.

Álvarez-Sánchez, B., F. Priego-Capote, and M. D. Luque de Castro. 2010. “Metabolomics Analysis I. Selection of Biological Samples and Practical Aspects Preceding Sample Preparation.” *TrAC - Trends in Analytical Chemistry* 29(2): 111–19.

Alyass, Akram, Michelle Turcotte, and David Meyre. 2015. “From Big Data Analysis to Personalized Medicine for All: Challenges and Opportunities.” *BMC Medical Genomics* 8(1).

Andrews, Genna L. et al. 2011. “Performance Characteristics of a New Hybrid Quadrupole Time-of-Flight Tandem Mass Spectrometer (TripleTOF 5600).” *Analytical Chemistry* 83(13): 5442–46.

Arapitsas, Panagiotis et al. 2020. “Use of Untargeted Liquid Chromatography-Mass Spectrometry Metabolome to Discriminate Italian Monovarietal Red Wines, Produced in Their Different Terroirs.” *Journal of Agricultural and Food Chemistry* 68(47): 13353–66.

Arnold, Ken et al. 2011. “Scikit-Learn: Machine Learning in {P}ython.” *Journal of Machine Learning Research* 12(85): 2825–30. http://jmlr.org/papers/v12/pedregosa11a.html.

Aron, Allegra et al. 2019. “Reproducible Molecular Networking Of Untargeted Mass Spectrometry Data Using GNPS.”

Atanasov, Atanas G. et al. 2015. “Discovery and Resupply of Pharmacologically Active Plant-Derived Natural Products: A Review.” *Biotechnology Advances* 33(8): 1582–1614. https://linkinghub.elsevier.com/retrieve/pii/S0734975015300276.

Azerad, Robert. 2016. “Chemical Structures, Production and Enzymatic Transformations of Sapogenins and Saponins from Centella Asiatica (L.) Urban.” *Fitoterapia* 114: 168–87.

B.A., Ejigu et al. 2013. “Evaluation of Normalization Methods to Pave the Way towards Large-Scale LC-MS-Based Metabolomics Profiling Experiments.” *OMICS A Journal of Integrative Biology* 17(9): 473–85. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed11&NEWS=N&AN=2013565871.

Banerjee, Shibdas, and Shyamalava Mazumdar. 2012. “Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte.” *International Journal of Analytical Chemistry* 2012: 1–40.

Barallobre-Barreiro, Javier, Yuen-Li Chung, and Manuel Mayr. 2013. “Proteomics and Metabolomics for Mechanistic Insights and Biomarker Discovery in Cardiovascular Disease.” *Revista Española de Cardiología (English Edition)* 66(8): 657–61.

Bedair, Mohamed, and Lloyd W. Sumner. 2008. “Current and Emerging Mass-Spectrometry Technologies for Metabolomics.” *TrAC - Trends in Analytical Chemistry* 27(3): 238–50.

Beger, Richard. 2013. “A Review of Applications of Metabolomics in Cancer.” *Metabolites* 3(3): 552–74.

Benjamini, Yoav, and Yosef Hochberg. 1995. “Controlling the False Discovery Rate: A

Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57(1): 289–300. http://doi.wiley.com/10.1111/j.2517-6161.1995.tb02031.x.

van den Berg, Robert A. et al. 2006. "Centering, Scaling, and Transformations: Improving the Biological Information Content of Metabolomics Data." *BMC Genomics* 7.

Bero, Siti Asmah et al. 2018. "Weighted Tanimoto Coefficient for 3D Molecule Structure Similarity Measurement." http://arxiv.org/abs/1806.05237.

Bingol, Kerem. 2018. "Recent Advances in Targeted and Untargeted Metabolomics by NMR and MS/NMR Methods." *High-Throughput* 7(2).

Bligh, E. G., and W. J. Dyer. 1959. "A Rapid Method of Total Lipid Extraction and Purification." *Canadian Journal of Biochemistry and Physiology* 37(1): 911–17.

BOITEAU, P., and A. R. RATSIMAMANGA. 1956. "Asiaticoside Extracted from Centella Asiatica and Its Therapeutic Uses in Cicatrization of Experimental and Refractory Wounds (Leprosy, Cutaneous Tuberculosis and Lupus)." *Thérapie* 11(1): 125–49.

Bourdillon, Max T. et al. 2014. "Oxidation of Plasmalogen, Low-Density Lipoprotein and Raw 264.7 Cells by Photoactivatable Atomic Oxygen Precursors." *Photochemistry and Photobiology* 90(2): 386–93.

Bowden, John A. et al. 2017. "Harmonizing Lipidomics: NIST Interlaboratory Comparison Exercise for Lipidomics Using SRM 1950-Metabolites in Frozen Human Plasma." *Journal of Lipid Research* 58(12): 2275–88. https://linkinghub.elsevier.com/retrieve/pii/S0022227520335549.

Braun, J E et al. 2012. "Metabolite Profiling Identifies a Key Role for Glycine in Rapid Cancer Cell Proliferation." *Science* (May): 1040–45.

Braz JC, Gregory K, Pathak A, et al. 2004. "PKC-Alpha Regulates Cardiac Contractility and Propensity toward Heart Failure." *Nat Med.* 10(3): 248–54. doi:10.1038/nm1000.

van Breemen, RB. 2015. “Development of Safe and Effective Botanical Dietary Supplements.” *Planta Medica* 81(11).

Van Breemen, Richard B., Harry H.S. Fong, and Norman R. Farnsworth. 2008. “Ensuring the Safety of Botanical Dietary Supplements.” *American Journal of Clinical Nutrition* 87(2).

Brinkhaus, B., M. Lindner, D. Schuppan, and E. G. Hahn. 2000. “Chemical, Pharmacological and Clinical Profile of the East Asian Medical Plant Centella Asiatica.” *Phytomedicine* 7(5): 427–48.

Broadhurst, David et al. 2018. “Guidelines and Considerations for the Use of System Suitability and Quality Control Samples in Mass Spectrometry Assays Applied in Untargeted Clinical Metabolomic Studies.” *Metabolomics* 14(6): 72. http://link.springer.com/10.1007/s11306-018-1367-3.

Broadhurst, David I., and Douglas B. Kell. 2006. “Statistical Strategies for Avoiding False Discoveries in Metabolomics and Related Experiments.” *Metabolomics* 2(4): 171–96.

Brown, Hannah Marie, Valentina Pirro, and R. Graham Cooks. 2018. “From DESI to the MasSpec Pen: Ambient Ionization Mass Spectrometry for Tissue Analysis and Intrasurgical Cancer Diagnosis.” *Clinical Chemistry* 64(4): 628–30.

Bruni, Renato, and Gianni Sacchetti. 2009. “Factors Affecting Polyphenol Biosynthesis in Wild and Field Grown St. John’s Wort (Hypericum Perforatum L. Hypericaceae/Guttiferae).” *Molecules* 14(2): 682–725.

Buszewski, Bogusław, and Sylwia Noga. 2012. “Hydrophilic Interaction Liquid Chromatography (HILIC)-a Powerful Separation Technique.” *Analytical and Bioanalytical Chemistry* 402(1): 231–47.

Cajka, Tomas, and Oliver Fiehn. 2014. “Comprehensive Analysis of Lipids in Biological Systems by Liquid Chromatography-Mass Spectrometry.” *TrAC - Trends in Analytical Chemistry* 61: 192–206.

Cajka, Tomas, and Oliver Fiehn. 2016a. “Increasing Lipidomic Coverage by Selecting

Optimal Mobile-Phase Modifiers in LC–MS of Blood Plasma." *Metabolomics* 12(2): 1–11.

Cajka, Tomas, and Oliver Fiehn. 2016b. "Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics." *Analytical Chemistry* 88(1): 524–45.

Castro-Perez, Jose et al. 2011. "Localization of Fatty Acyl and Double Bond Positions in Phosphatidylcholines Using a Dual Stage CID Fragmentation Coupled with Ion Mobility Mass Spectrometry." *Journal of the American Society for Mass Spectrometry* 22(9): 1552–67.

Castro-Perez, Jose M. et al. 2010. "Comprehensive LC-MSE Lipidomic Analysis Using a Shotgun Approach and Its Application to Biomarker Detection and Identification in Osteoarthritis Patients." *Journal of Proteome Research* 9(5): 2377–89.

Chandrika, Udumalagala Gamage, and Peramune A.A.S. Prasad Kumara. 2015. "Gotu Kola (Centella Asiatica): Nutritional Properties and Plausible Health Benefits." *Advances in Food and Nutrition Research* 76: 125–57.

"Chapter 8 Bootstrapping and Confidence Intervals | Statistical Inference via Data Science." https://moderndive.com/8-confidence-intervals.html.

Chavez, Kathryn J., Sireesha V. Garimella, and Stanley Lipkowitz. 2010. "Triple Negative Breast Cancer Cell Lines: One Tool in the Search for Better Treatment of Triple Negative Breast Cancer." *Breast Disease* 32(1–2): 35–48.

Checa, Antonio, Carmen Bedia, and Joaquim Jaumot. 2015. "Lipidomic Data Analysis: Tutorial, Practical Guidelines and Applications." *Analytica Chimica Acta* 885: 1–16.

Cheng, Chun Ting et al. 2018. "Arginine Starvation Kills Tumor Cells through Aspartate Exhaustion and Mitochondrial Dysfunction." *Communications Biology* 1(1).

Chernushevich, Igor V., Alexander V. Loboda, and Bruce A. Thomson. 2001. "An Introduction to Quadrupole-Time-of-Flight Mass Spectrometry." *Journal of Mass Spectrometry* 36(8): 849–65.

Chester, Thomas L. 2013. “Recent Developments in High-Performance Liquid Chromatography Stationary Phases.” *Analytical Chemistry* 85(2): 579–89.

Chiarugi, Alberto, Christian Dölle, Roberta Felici, and Mathias Ziegler. 2012. “The NAD Metabolome — a Key Determinant of Cancer Cell Biology.” *Nature Reviews Cancer* 12: 741. https://doi.org/10.1038/nrc3340.

Chicco, Adam J., and Genevieve C. Sparagna. 2007. “Role of Cardiolipin Alterations in Mitochondrial Dysfunction and Disease.” *American Journal of Physiology - Cell Physiology* 292(1).

Chicco, Davide. 2017. “Ten Quick Tips for Machine Learning in Computational Biology.” *BioData Mining* 10(1): . 135. doi:10.1186/s13040-017-0155-3. PMC5721660.P.

Chokshi, Aalap et al. 2012. “Ventricular Assist Device Implantation Corrects Myocardial Lipotoxicity, Reverses Insulin Resistance, and Normalizes Cardiac Metabolism in Patients with Advanced Heart Failure.” *Circulation* 125(23): 2844–53.

Cífková, Eva et al. 2015. “Determination of Lipidomic Differences between Human Breast Cancer and Surrounding Normal Tissues Using HILIC-HPLC/ESI-MS and Multivariate Data Analysis.” *Analytical and Bioanalytical Chemistry* 407(3): 991–1002.

Claire, Eyers, Christopher J. Gray, Stephen Holman, and Francesco Lanucara. 2014. “The Power of Ion Mobility-Mass Spectrometry for Structural Characterization and the Study of Conformational Dynamics.” *Nature Chemistry* 6(4): 281–94.

Clare, Constance E., Amey H. Brassington, Wing Yee Kwong, and Kevin D. Sinclair. 2019. “One-Carbon Metabolism: Linking Nutritional Biochemistry to Epigenetic Programming of Long-Term Development.” *Annual Review of Animal Biosciences* 7: 263–87.

Clark, Ian A. 1986. “Tissue Damage Caused by Free Oxygen Radicals.” *Pathology* 18(2): 181–86.

Copenhaver, Philip F. et al. 2011. “A Translational Continuum of Model Systems for

Evaluating Treatment Strategies in Alzheimer's Disease: Isradipine as a Candidate Drug." *DMM Disease Models and Mechanisms* 4(5): 634–48.

Das, Anathbandhu, and Ranajit Mallick. 1991. "Correlation between Genomic Diversity and Asiaticoside Content in Centella Asiatica (L.) Urban." *Botanical Bulletin of Academia Sinica* 32(1): 1–8.

Dawidowicz, Andrzej L., and Rafal Typek. 2011. "The Influence of PH on the Thermal Stability of 5-O-Caffeoylquinic Acids in Aqueous Solutions." *European Food Research and Technology* 233(2): 223–32.

Dawidowicz, Andrzej L., and Rafal Typek. 2015. "Thermal Transformation of Trans-5-O-Caffeoylquinic Acid (Trans-5-CQA) in Alcoholic Solutions." *Food Chemistry* 167: 52–60.

Deol, Gurkamal. 2019. "An Introduction to Ridge, Lasso, and Elastic Net Regression." *Https://Hackernoon.Com/an-Introduction-To-Ridge-Lasso-and-Elastic-Net-Regression-Cca60B4B934F*: 1–8. https://hackernoon.com/an-introduction-to-ridge-lasso-and-elastic-net-regression-cca60b4b934f.

Dettmer, Katja, Pavel A. Aronov, and Bruce D. Hammock. 2007. "Mass Spectrometry-Based Metabolomics." *Mass Spectrometry Reviews* 26(1): 51–78. http://doi.wiley.com/10.1002/mas.20108.

Devkota, Anjana et al. 2010. "Centella Asiatica (L.) Urban from Nepal: Quali-Quantitative Analysis of Samples from Several Sites, and Selection of High Terpene Containing Populations for Cultivation." *Biochemical Systematics and Ecology* 38(1): 12–22.

Dietel, M. et al. 2013. "Predictive Molecular Pathology and Its Role in Targeted Cancer Therapy: A Review Focussing on Clinical Relevance." *Cancer Gene Therapy* 20(4): 211–21.

Dietz, Birgit M. et al. 2017. "DESIGNER Extracts as Tools to Balance Estrogenic and Chemopreventive Activities of Botanicals for Women's Health." *Journal of Natural Products* 80(8): 2284–94.

Dória, M. Luísa et al. 2013. “Lipidomic Analysis of Phospholipids from Human Mammary Epithelial and Breast Cancer Cell Lines.” *Journal of Cellular Physiology* 228(2): 457–68.

Draper, John, Amanda J. Lloyd, Royston Goodacre, and Manfred Beckmann. 2013. “Flow Infusion Electrospray Ionisation Mass Spectrometry for High Throughput, Non-Targeted Metabolite Fingerprinting: A Review.” *Metabolomics* 9(SUPPL.1): 4–29.

Drosatos, Konstantinos et al. 2011. “Cardiomyocyte Lipids Impair β-Adrenergic Receptor Function via PKC Activation.” *American Journal of Physiology - Endocrinology and Metabolism* 300(3).

Dunn, Warwick B. et al. 2011. “Procedures for Large-Scale Metabolic Profiling of Serum and Plasma Using Gas Chromatography and Liquid Chromatography Coupled to Mass Spectrometry.” *Nature Protocols* 6(7): 1060–83.

Dunn, Warwick B., Ian D. Wilson, Andrew W. Nicholls, and David Broadhurst. 2012. “The Importance of Experimental Design and QC Samples in Large-Scale and MS-Driven Untargeted Metabolomic Studies of Humans.” *Bioanalysis* 4(18): 2249–64.

Eiriksson, Finnur Freyr et al. 2018. “Altered Plasmalogen Content and Fatty Acid Saturation Following Epithelial to Mesenchymal Transition in Breast Epithelial Cell Lines.” *International Journal of Biochemistry and Cell Biology* 103: 99–104.

El-Aneed, Anas, Aljandro Cohen, and Joseph Banoub. 2009. “Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers.” *Applied Spectroscopy Reviews* 44(3): 210–30.

Emmett, Mark R., and Cheryl F. Lichti. 2017. “Ultrahigh-Resolution Lipid Analysis with Fourier Transform Ion Cyclotron Resonance Mass Spectrometry.” In , 21–43. http://link.springer.com/10.1007/978-1-4939-6946-3_3.

Eum, Jung Yong et al. 2020. “Aging-Related Lipidomic Changes in Mouse Serum, Kidney, and Heart by Nanoflow Ultrahigh-Performance Liquid Chromatography-Tandem Mass Spectrometry.” *Journal of Chromatography A* 1618.

Excoffier, Laurent et al. 2017. “Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks.” *Nucleic Acids Research* 13(16): e149–e149. http://genome.cshlp.org/content/13/11/2498.abstract%0Ahttps://dx.doi.org/10.1093/nar/gkx626.

Fahy, Eoin et al. 2005. “A Comprehensive Classification System for Lipids.” *Journal of Lipid Research* 46(5): 839–61. https://linkinghub.elsevier.com/retrieve/pii/S0022227520339687.

Folch Jordi, Lees M, and Stanley G H Sloane. 1957. “A Simple Method for the Isolation and Purification of Total Lipides from Animal Tissues.” *The Journal of Biological Chemistry* 226: 497–509. http://www.ncbi.nlm.nih.gov/pubmed/20589730.

Fong, Harry H. -S. et al. 2006. “Evidence-Based Herbal Medicine: Challenges in Efficacy and Safety Assessments.” : 11–26.

Franklin, Sarah, and Thomas M. Vondriska. 2011. “Genomes, Proteomes, and the Central Dogma.” *Circulation: Cardiovascular Genetics* 4(5): 576.

G.K., Shinomol, Muralidhara, and Muchukunte M.S. Bharath. 2011. “Exploring the Role of ‘Brahmi’ (Bacopa Monnieri and Centella Asiatica) in Brain Function and Therapy.” *Recent Patents on Endocrine, Metabolic & Immune Drug Discovery* 5(1): 33–49.

Gallart-Ayala, H. et al. 2013. “Versatile Lipid Profiling by Liquid Chromatography-High Resolution Mass Spectrometry Using All Ion Fragmentation and Polarity Switching. Preliminary Application for Serum Samples Phenotyping Related to Canine Mammary Cancer.” *Analytica Chimica Acta* 796: 75–83.

Géron, Aurélien. 2017. O’Reilly Media *Hands-on Machine Learning with Scikit-Learn and TensorFlow : Concepts, Tools, and Techniques to Build Intelligent Systems*.

Giffin, M., G. Arthur, P. C. Choy, and R. Y.K. Man. 1988. “Lysophosphatidylcholine Metabolism and Cardiac Arrhythmias.” *Canadian Journal of Physiology and Pharmacology* 66(3): 185–89.

Gika, Helen G., Georgios A. Theodoridis, Robert S. Plumb, and Ian D. Wilson. 2014. "Current Practice of Liquid Chromatography-Mass Spectrometry in Metabolomics and Metabonomics." *Journal of Pharmaceutical and Biomedical Analysis* 87: 12–25.

Giles, Kevin, Jonathan P. Williams, and Iain Campuzano. 2011. "Enhancements in Travelling Wave Ion Mobility Resolution." *Rapid Communications in Mass Spectrometry* 25(11): 1559–66.

Girolamo, Francesco, Isabella Lante, Maurizio Muraca, and Lorenza Putignani. 2013. "The Role of Mass Spectrometry in the 'Omics' Era." *Current Organic Chemistry* 17(23): 2891–2905.

Glass, Christopher K., and Jerrold M. Olefsky. 2012. "Inflammation and Lipid Signaling in the Etiology of Insulin Resistance." *Cell Metabolism* 15(5): 635–45.

Gohil, Kashmira J., Jagruti A. Patel, and Anuradha K. Gajjar. 2010. "Pharmacological Review on Centella Asiatica: A Potential Herbal Cure-All." *Indian Journal of Pharmaceutical Sciences* 72(5): 546–56.

Goldberg, Stanley. 2008. "Mechanical/Physical Methods of Cell Disruption and Tissue Homogenization." *Methods in molecular biology (Clifton, N.J.)* 424: 3–22.

Gowda, G. A.Nagana et al. 2008. "Metabolomics-Based Methods for Early Disease Diagnostics." *Expert Review of Molecular Diagnostics* 8(5): 617–33.

Gray, Nora E. et al. 2014. "Caffeoylquinic Acids in Centella Asiatica Protect against Amyloid-β Toxicity." *Journal of Alzheimer's Disease* 40(2): 359–73.

Gray, Nora E. et al. 2018. "Centella Asiatica: Phytochemistry and Mechanisms of Neuroprotection and Cognitive Enhancement." *Phytochemistry Reviews* 17(1): 161–94.

Griffiths, William J., and Yuqin Wang. 2009. "Mass Spectrometry: From Proteomics to Metabolomics and Lipidomics." *Chemical Society Reviews* 38(7): 1882–96.

Gross, J. H. 2011. *Mass Spectrometry: A Textbook*.

Guijas, Carlos et al. 2018. “METLIN: A Technology Platform for Identifying Knowns and Unknowns.” *Analytical Chemistry* 90(5): 3156–64.

Guo, Jian, and Tao Huan. 2020. “Comparison of Full-Scan, Data-Dependent, and Data-Independent Acquisition Modes in Liquid Chromatography-Mass Spectrometry Based Untargeted Metabolomics.” *Analytical Chemistry* 92(12): 8072–80.

Guo, Shuai, Yanmin Wang, Dan Zhou, and Zhili Li. 2014. “Significantly Increased Monounsaturated Lipids Relative to Polyunsaturated Lipids in Six Types of Cancer Microenvironment Are Observed by Mass Spectrometry Imaging.” *Scientific Reports* 4.

Gysembergh, A. et al. 2000. “Brief Preconditioning Ischemia Alters Diacylglycerol Content and Composition in Rabbit Heart.” *Basic research in cardiology* 95(6): 457–65.

H.P., Benton et al. 2012. “Intra- and Interlaboratory Reproducibility of Ultra Performance Liquid Chromatography-Time-of-Flight Mass Spectrometry for Urinary Metabolic Profiling.” *Analytical Chemistry* 84(5): 2424–32. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed10&NEWS=N&AN=22304021.

Han, Jun et al. 2008. “Towards High-Throughput Metabolomics Using Ultrahigh-Field Fourier Transform Ion Cyclotron Resonance Mass Spectrometry.” *Metabolomics* 4(2): 128–40. http://link.springer.com/10.1007/s11306-008-0104-8.

Han, Xianlin. 2016. Lipidomics: Comprehensive Mass Spectrometry of Lipids *Lipidomics: Comprehensive Mass Spectrometry of Lipids*.

Han, Xianlin, and Richard W. Gross. 2003. “Global Analyses of Cellular Lipidomes Directly from Crude Extracts of Biological Samples by ESI Mass Spectrometry: A Bridge to Lipidomics.” *Journal of Lipid Research* 44(6): 1071–79.

Hannun, Yusuf A., and Lina M. Obeid. 2018. “Sphingolipids and Their Metabolism in Physiology and Disease.” *Nature Reviews Molecular Cell Biology* 19(3): 175–91.

Hardy, Serge et al. 2003. “Saturated Fatty Acid-Induced Apoptosis in MDA-MB-231

Breast Cancer Cells. A Role for Cardiolipin." *Journal of Biological Chemistry* 278(34): 31861–70.

Hartmann, Thomas. 1996. "Diversity and Variability of Plant Secondary Metabolism: A Mechanistic View." *Proceedings of the 9th International Symposium on Insect-Plant Relationships*: 177–88.

Hastings, Janna et al. 2016. "ChEBI in 2016: Improved Services and an Expanding Collection of Metabolites." *Nucleic Acids Research* 44(D1): D1214–19. https://www.ebi.ac.uk/chebi/init.do.

Hazra, Avijit, and Nithya Gogtay. 2017. "Biostatistics Series Module 9: Survival Analysis." *Indian Journal of Dermatology* 62(3): 251–57.

Hilvo, Mika et al. 2011. "Novel Theranostic Opportunities Offered by Characterization of Altered Membrane Lipid Metabolism in Breast Cancer Progression." *Cancer Research* 71(9): 3236–45.

Hirayama, Akiyoshi et al. 2009. "Quantitative Metabolome Profiling of Colon and Stomach Cancer Microenvironment by Capillary Electrophoresis Time-of-Flight Mass Spectrometry." *Cancer Research* 69(11): 4918–25.

de Hoffmann, E., and V. Stroobant. 2007. *Mass Spectrometry: Principles and Applications*.

Hotamisligil, Gökhan S. 2006. "Inflammation and Metabolic Disorders." *Nature* 444(7121): 860–67.

Housley, Lauren et al. 2018. "Untargeted Metabolomic Screen Reveals Changes in Human Plasma Metabolite Profiles Following Consumption of Fresh Broccoli Sprouts." *Molecular Nutrition and Food Research* 62(19).

Howard, L. 2008. "Processing Techniques and Their Effect on Fruit and Vegetable Phytochemicals." *Improving the Health-Promoting Properties of Fruit and Vegetable Products*: 449–72.

Hyötyläinen, Tuulia, and Matej Orešič. 2015. "Optimizing the Lipidomics Workflow for

Clinical Studies - Practical Considerations." *Analytical and Bioanalytical Chemistry* 407(17): 4973–93.

Iason, Glenn R., Marcel Dicke, and Susan E. Hartley, eds. 2012. *The Ecology of Plant Secondary Metabolites*. Cambridge: Cambridge University Press. http://ebooks.cambridge.org/ref/id/CBO9780511675751.

ICH. 2012. "ICH Topic Q2 (R1) Validation of Analytical Procedures : Text and Methodology." *International Conference on Harmonization*.

Inamdar, P. K., R. D. Yeole, A. B. Ghogare, and N. J. De Souza. 1996. "Determination of Biologically Active Constituents in Centella Asiatica." *Journal of Chromatography A* 742(1–2): 127–30.

James, Jacinda T., and Ian A. Dubery. 2009. "Pentacyclic Triterpenoids from the Medicinal Herb, Centella Asiatica (L.) Urban." *Molecules* 14(10): 3922–41.

Johnson, Caroline H., Julijana Ivanisevic, H. Paul Benton, and Gary Siuzdak. 2015. "Bioinformatics: The next Frontier of Metabolomics." *Analytical Chemistry* 87(1): 147–56.

Jurowski, Kamil et al. 2017. "Comprehensive Review of Trends and Analytical Strategies Applied for Biological Samples Preparation and Storage in Modern Medical Lipidomics: State of the Art." *TrAC - Trends in Analytical Chemistry* 86: 276–89.

Kanu, Abu B. et al. 2008. "Ion Mobility-Mass Spectrometry." *Journal of Mass Spectrometry* 43(1): 1–22.

Kapoor, L. D. 2018. CRC Handbook of Ayurvedic Medicinal Plants *CRC Handbook of Ayurvedic Medicinal Plants*.

Kellogg, Joshua J. et al. 2016. "Biochemometrics for Natural Products Research: Comparison of Data Analysis Approaches and Application to Identification of Bioactive Compounds." *Journal of Natural Products* 79(2): 376–86.

Kim, Sunghwan, Evan E. Bolton, and Stephen H. Bryant. 2012. "Effects of Multiple Conformers per Compound upon 3-D Similarity Search and Bioassay Data

Analysis.” *Journal of Cheminformatics* 4(1).

Kind, Tobias, and Oliver Fiehn. 2013. “Advances in Structure Elucidation of Small Molecules Using Mass Spectrometry.” *Frontiers of Bioanalytical Chemistry*: 129–66.

Kirkpatrick, Christine L. et al. 2017. “The ‘PepSAVI-MS’ Pipeline for Natural Product Bioactive Peptide Discovery.” *Analytical Chemistry* 89(2): 1194–1201.

Kishore, Jugal, ManishKumar Goel, and Pardeep Khanna. 2010. “Understanding Survival Analysis: Kaplan-Meier Estimate.” *International Journal of Ayurveda Research* 1(4): 274.

Korkina LG, . 2007. “Phenylpropanoids as Naturally Occurring Antioxidants: From Plant Defense to Human Health.” *Cellular and Molecular Biology* 53(1): 15–25.

Krakauer, Teresa, and Marilyn Buckley. 2003. “Doxycycline Is Anti-Inflammatory and Inhibits Staphylococcal Exotoxin-Induced Cytokines and Chemokines.” *Antimicrobial Agents and Chemotherapy* 47(11): 3630–33.

Kucheryavskiy, Sergey. 2020. “Mdatools – R Package for Chemometrics.” *Chemometrics and Intelligent Laboratory Systems* 198: 103937. https://linkinghub.elsevier.com/retrieve/pii/S0169743919305672.

Kuhl, Carsten et al. 2012. “CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets.” *Analytical Chemistry* 84(1): 283–89.

Kumar, Anil, Samrita Dogra, and Atish Prakash. 2009. “Neuroprotective Effects of Centella Asiatica against Intracerebroventricular Colchicine-Induced Cognitive Impairment and Oxidative Stress.” *International Journal of Alzheimer’s Disease* 2009.

L., Denoroy, Zimmer L., Renaud B., and Parrot S. 2013. “Ultra High Performance Liquid Chromatography as a Tool for the Discovery and the Analysis of Biomarkers of Diseases: A Review.” *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* 927: 37–53.

http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L52390649%0Ahttp://dx.doi.org/10.1016/j.jchromb.2012.12.005.

L., Lai et al. 2014. “Energy Metabolic Reprogramming in the Hypertrophied and Early Stage Failing Heart a Multisystems Approach.” *Circulation: Heart Failure* 7(6): 1022–31. http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L603467782%0Ahttp://dx.doi.org/10.1161/CIRCHEARTFAILURE.114.001469.

Lang, Gerhard et al. 2008. “Evolving Trends in the Dereplication of Natural Product Extracts: New Methodology for Rapid, Small-Scale Investigation of Natural Product Extracts.” *Journal of Natural Products* 71(9): 1595–99.

Lawrence, Toby, Derek A. Willoughby, and Derek W. Gilroy. 2002. “Anti-Inflammatory Lipid Mediators and Insights into the Resolution of Inflammation.” *Nature Reviews Immunology* 2(10): 787–95.

Le, D. Elizabeth et al. 2021. “Plasma Oxylipins: A Potential Risk Assessment Tool in Atherosclerotic Coronary Artery Disease.” *Frontiers in Cardiovascular Medicine* 8.

Li, H et al. 2014. “Phosphatidylethanolamine-Binding Protein 4 Is Associated with Breast Cancer Metastasis through Src-Mediated Akt Tyrosine Phosphorylation.” *Oncogene* 33(37): 4589–98. https://doi.org/10.1038/onc.2013.408.

De Livera, Alysha M. et al. 2015. “Statistical Methods for Handling Unwanted Variation in Metabolomics Data.” *Analytical Chemistry* 87(7): 3606–15.

Locasale, Jason W. 2013. “Serine, Glycine and One-Carbon Units: Cancer Metabolism in Full Circle.” *Nature Reviews Cancer* 13: 572. https://doi.org/10.1038/nrc3557.

Loken, Steve D., and Douglas J. Demetrick. 2005. “A Novel Method for Freezing and Storing Research Tissue Bank Specimens.” *Human Pathology* 36(9): 977–80.

Long, H. S., M. A. Stander, and B. E. Van Wyk. 2012. “Notes on the Occurrence and Significance of Triterpenoids (Asiaticoside and Related Compounds) and Caffeoylquinic Acids in Centella Species.” *South African Journal of Botany* 82: 53–59.

Lou, Jia Qian et al. 2020. “Investigation of Heart Lipid Changes in Acute β-AR Activation-Induced Sudden Cardiac Death by Time-of-Flight Secondary Ion Mass Spectrometry.” *Analyst* 145(17): 5889–96.

Lozoya, Oswaldo A. et al. 2018. “Mitochondrial Nicotinamide Adenine Dinucleotide Reduced (NADH) Oxidation Links the Tricarboxylic Acid (TCA) Cycle with Methionine Metabolism and Nuclear DNA Methylation.” *PLoS Biology* 16(4).

Madhusudhan, N. Ch, P. Neeraja, and Prathibha Devi. 2014. “Comparative Analysis of Active Constituents in Centella Asiatica Varieties (Majjaposhak and Subhodak).” *International Journal of Pharmaceutical and Phytopharmacological Research* 4(2): 105–8.

Magana, Armando Alcazar et al. 2015. “Determination of Small Phenolic Compounds in Tequila by Liquid Chromatography with Ion Trap Mass Spectrometry Detection.” *Food Analytical Methods* 8(4): 864–72.

Magana, Armando Alcazar et al. 2020. “Vitamin c Activates the Folate-Mediated One-Carbon Cycle in C2C12 Myoblasts.” *Antioxidants* 9(3).

Mahady, G. B. 2001. “Global Harmonization of Herbal Health Claims.” *Journal of Nutrition* 131(3 SUPPL.).

Mahieu, Nathaniel G., Jessica Lloyd Genenbacher, and Gary J. Patti. 2016. “A Roadmap for the XCMS Family of Software Solutions in Metabolomics.” *Current Opinion in Chemical Biology* 30: 87–93.

Maier, Claudia S. et al. 2013. “Electrospray Ionization Traveling Wave Ion Mobility Spectrometry Mass Spectrometry for the Analysis of Plant Phenolics: An Approach for Separation of Regioisomers.” *50 Years of Phytochemistry Research*: 21–41.

Matsukura, S. 2003. “Establishment of Conditional Vectors for Hairpin SiRNA Knockdowns.” *Nucleic Acids Research* 31(15): 77e – 77.

May, Jody C., and John A. McLean. 2013. “The Influence of Drift Gas Composition on the Separation Mechanism in Traveling Wave Ion Mobility Spectrometry: Insight from Electrodynamic Simulations.” *International Journal for Ion Mobility*

Spectrometry 16(2): 85–94.

May, Jody C., and John A. McLean. 2015. “Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation.” *Analytical Chemistry* 87(3): 1422–36.

Mazzara, Saveria et al. 2017. “CombiROC: An Interactive Web Tool for Selecting Accurate Marker Combinations of Omics Data.” *Scientific Reports* 7.

Van Meer, Gerrit, Dennis R. Voelker, and Gerald W. Feigenson. 2008. “Membrane Lipids: Where They Are and How They Behave.” *Nature Reviews Molecular Cell Biology* 9(2): 112–24.

Mehrmohamadi, Mahya, Xiaojing Liu, Alexander A. Shestov, and Jason W. Locasale. 2014. “Characterization of the Usage of the Serine Metabolic Network in Human Cancer.” *Cell Reports* 9(4): 1507–19.

Meikle, Peter J. et al. 2011. “Plasma Lipidomic Analysis of Stable and Unstable Coronary Artery Disease.” *Arteriosclerosis, Thrombosis, and Vascular Biology* 31(11): 2723–32.

Menendez, Javier A., and Ruth Lupu. 2007. “Fatty Acid Synthase and the Lipogenic Phenotype in Cancer Pathogenesis.” *Nature Reviews Cancer* 7(10): 763–77.

Moghe, Gaurav D., and Robert L. Last. 2015. “Something Old, Something New: Conserved Enzymes and the Evolution of Novelty in Plant Specialized Metabolism.” *Plant Physiology* 169(3): 1512–23.

Mollaoğlu, Mukadder, and Alkar Aciyurt. 2013. “Use of Complementary and Alternative Medicine among Patients with Chronic Diseases.” *Acta Clinica Croatica* 52(2): 181–88.

Müller, Viola et al. 2013. “Centelloside Accumulation in Leaves of Centella Asiatica Is Determined by Resource Partitioning between Primary and Secondary Metabolism While Influenced by Supply Levels of Either Nitrogen, Phosphorus or Potassium.” *Journal of Plant Physiology* 170(13): 1165–75.

Mustafa, R. A., A. Abdul Hamid, S. Mohamed, and F. Abu Bakar. 2010. “Total Phenolic

Compounds, Flavonoids, and Radical Scavenging Activity of 21 Selected Tropical Plants." *Journal of Food Science* 75(1).

Nagana Gowda, G A et al. 2018. "NMR-Guided Mass Spectrometry for Absolute Quantitation of Human Blood Metabolites." *Analytical chemistry* 90(3): 2001–9.

Nagrath, Deepak, Christine Caneba, Thasni Karedath, and Nadege Bellance. 2011. "Metabolomics for Mitochondrial and Cancer Studies." *Biochimica et Biophysica Acta - Bioenergetics* 1807(6): 650–63.

Naz, Shama et al. 2017. "Development of a Liquid Chromatography-High Resolution Mass Spectrometry Metabolomics Method with High Specificity for Metabolite Identification Using All Ion Fragmentation Acquisition." *Analytical Chemistry* 89(15): 7933–42.

Naz, Shama, Maria Vallejo, Antonia García, and Coral Barbas. 2014. "Method Validation Strategies Involved in Non-Targeted Metabolomics." *Journal of Chromatography A* 1353: 99–105.

Neiberg, Rebecca H. et al. 2011. "Occurrence and Co-Occurrence of Types of Complementary and Alternative Medicine Use by Age, Gender, Ethnicity, and Education among Adults in the United States: The 2002 National Health Interview Survey (NHIS)." *Journal of Alternative and Complementary Medicine* 17(4): 363–70.

Nicholson, Jeremy K., and Ian D. Wilson. 2003. "Understanding 'global' Systems Biology: Metabonomics and the Continuum of Metabolism." *Nature Reviews Drug Discovery* 2(8): 668–76.

"NIST1950 Metabolites in Human Plasma." https://www.sigmaaldrich.com/catalog/product/sial/nist1950?lang=en&region=US.

Nothias, Louis Félix et al. 2018. "Bioactivity-Based Molecular Networking for the Discovery of Drug Leads in Natural Product Bioassay-Guided Fractionation." *Journal of Natural Products* 81(4): 758–67.

Nováková, Lucie, Hana Vlčková, and Solich Petr. 2012. "Evaluation of New Mixed-

Mode UHPLC Stationary Phases and the Importance of Stationary Phase Choice When Using Low Ionic-Strength Mobile Phase Additives." *Talanta* 93: 99–105.

Oyedeji, O. A., and A. J. Afolayan. 2005. "Chemical Composition and Antibacterial Activity of the Essential Oil of Centella Asiatica Growing in South Africa." *Pharmaceutical Biology* 43(3): 249–52.

Paapstel, K. et al. 2018. "Inverse Relations of Serum Phosphatidylcholines and Lysophosphatidylcholines with Vascular Damage and Heart Rate in Patients with Atherosclerosis." *Nutrition, Metabolism and Cardiovascular Diseases* 28(1): 44–52.

Paemel, Ruben Van. 2019. "Kaplan Meier Curves: An Introduction." : 1–11. https://towardsdatascience.com/kaplan-meier-curves-c5768e349479#:~:text=The visual representation of this,for the population under investigation.

Paglia, Giuseppe et al. 2015. "Ion Mobility-Derived Collision Cross Section As an Additional Measure for Lipid Fingerprinting and Identification." *Analytical Chemistry* 87(2): 1137–44. https://pubs.acs.org/doi/10.1021/ac503715v.

Patti, Gary J. 2011. "Separation Strategies for Untargeted Metabolomics." *Journal of Separation Science* 34(24): 3460–69.

Patti, Gary J., Ralf Tautenhahn, and Gary Siuzdak. 2012. "Meta-Analysis of Untargeted Metabolomic Data from Multiple Profiling Experiments." *Nature Protocols* 7(3): 508–16.

Patti, Gary J., Oscar Yanes, and Gary Siuzdak. 2012. "Innovation: Metabolomics: The Apogee of the Omics Trilogy." *Nature Reviews Molecular Cell Biology* 13(4): 263–69.

Peiris-Pagès, Maria et al. 2016. "Cancer Stem Cell Metabolism." *Breast Cancer Research* 18(1).

Plumb, Robert S. et al. 2006. "UPLC/MSE; a New Approach for Generating Molecular Fragment Information for Biomarker Structure Elucidation." *Rapid Communications in Mass Spectrometry* 20(13): 1989–94. https://onlinelibrary.wiley.com/doi/10.1002/rcm.2550.

Post, J. A., J. J.M. Bijvelt, and A. J. Verkleij. 1995. "Phosphatidylethanolamine and Sarcolemmal Damage during Ischemia or Metabolic Inhibition of Heart Myocytes." *American Journal of Physiology - Heart and Circulatory Physiology* 268(2 37-2).

Prince, John T., and Edward M. Marcotte. 2006. "Chromatographic Alignment of ESI-LC-MS Proteomics Data Sets by Ordered Bijective Interpolated Warping." *Analytical Chemistry* 78(17): 6140–52.

Qiu, Feng et al. 2016. "PlantMAT: A Metabolomics Tool for Predicting the Specialized Metabolic Potential of a System and for Large-Scale Metabolite Identifications." *Analytical Chemistry* 88(23): 11373–83.

Ramesh, B. N. et al. 2014. "Comparative Study on Anti-Oxidant and Anti-Inflammatory Activities of Caesalpinia Crista and Centella Asiatica Leaf Extracts." *Journal of Pharmacy and Bioallied Sciences* 6(2): 86–91.

Randriamampionona, Denis et al. 2007. "Comparative Analysis of Active Constituents in Centella Asiatica Samples from Madagascar: Application for Ex Situ Conservation and Clonal Propagation." *Fitoterapia* 78(7–8): 482–89.

Reforgiato, M. R. et al. 2016. "Inhibition of Ceramide de Novo Synthesis as a Postischemic Strategy to Reduce Myocardial Reperfusion Injury." *Basic Research in Cardiology* 111(2).

Rigoni, Michela et al. 2007. "Calcium Influx and Mitochondrial Alterations at Synapses Exposed to Snake Neurotoxins or Their Phospholipid Hydrolysis Products." *Journal of Biological Chemistry* 282(15): 11238–45.

Rohrbach, Timothy D. et al. 2020. "A Simple Method for Sphingolipid Analysis of Tissues Embedded in Optimal Cutting Temperature Compound." *Journal of lipid research* 61(6): 953–67.

Rose, Michelle L., Russell C. Cattley, et al. 1999. "Dietary Glycine Prevents the Development of Liver Tumors Caused by the Peroxisome Proliferator WY-14,643." *Carcinogenesis* 20(11): 2075–81.

Rose, Michelle L., Johnathan Madren, Hartwig Bunzendahl, and Ronald G. Thurman.

1999. “Dietary Glycine Inhibits the Growth of B16 Melanoma Tumors in Mice.” *Carcinogenesis* 20(5): 793–98.

Rubbosi, Homero et al. 1994. “Nitric Oxide Regulation of Superoxide and Peroxynitrite-Dependent Lipid Peroxidation.” *The Journal of Biological Chemistry* 269(October 21): 26066–75.

Rubinek, Tami et al. 2012. “Epigenetic Silencing of the Tumor Suppressor Klotho in Human Breast Cancer.” *Breast Cancer Research and Treatment* 133(2): 649–57.

Saikali, Zeina, and Gurmit Singh. 2003. “Doxycycline and Other Tetracyclines in the Treatment of Bone Metastasis.” *Anti-Cancer Drugs* 14(10): 773–78.

Sangwan, Rajender S. et al. 2013. “De Novo Sequencing and Assembly of Centella Asiatica Leaf Transcriptome for Mapping of Structural, Functional and Regulatory Genes with Special Reference to Secondary Metabolism.” *Gene* 525(1): 58–76.

Sarafian, Magali H. et al. 2014. “Objective Set of Criteria for Optimization of Sample Preparation Procedures for Ultra-High Throughput Untargeted Blood Plasma Lipid Profiling by Ultra Performance Liquid Chromatography-Mass Spectrometry.” *Analytical Chemistry* 86(12): 5766–74.

Sarang, Narkhede. 2018. “Understanding AUC - ROC Curve.” *Towards Data Science*: 4–11. https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5.

Schwartz, Sarah A., Michelle L. Reyzer, and Richard M. Caprioli. 2003. “Direct Tissue Analysis Using Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry: Practical Aspects of Sample Preparation.” *Journal of Mass Spectrometry* 38(7): 699–708.

Sedlis, S. P., P. B. Corr, B. E. Sobel, and G. G. Ahumada. 1983. “Lysophosphatidyl Choline Potentiates Ca2+ Accumulation in Rat Cardiac Myocytes.” *American Journal of Physiology - Heart and Circulatory Physiology* 13(1).

Ségui, Bruno et al. 2006. “Sphingolipids as Modulators of Cancer Cell Death: Potential Therapeutic Targets.” *Biochimica et Biophysica Acta - Biomembranes* 1758(12): 2104–20.

Semreen, Mohammad H. et al. 2019. “Comparative Metabolomics of MCF-7 Breast Cancer Cells Using Different Extraction Solvents Assessed by Mass Spectroscopy.” *Scientific Reports* 9(1).

Sentandreu, Enrique et al. 2018. “A Survey of Orbitrap All Ion Fragmentation Analysis Assessed by an R MetaboList Package to Study Small-Molecule Metabolites.” *bioRxiv*: 257147.

Shah, Punit et al. 2015. “Tissue Proteomics Using Chemical Immobilization and Mass Spectrometry.” *Analytical Biochemistry* 469: 27–33.

Shine, V. J. et al. 2020. “Bioassay Guided Fractionation of Cyclea Peltata Using in Vitro RAW 264.7 Cell Culture, Antioxidant Assays and Isolation of Bioactive Compound Tetrandrine.” *Journal of Ayurveda and Integrative Medicine* 11(3): 281–86.

Sidebottom, Ashley M. et al. 2013. “Integrated Metabolomics Approach Facilitates Discovery of an Unpredicted Natural Product Suite from Streptomyces Coelicolor M145.” *ACS Chemical Biology* 8(9): 2009–16.

Simón-Manso, Yamil et al. 2013. “Metabolite Profiling of a NIST Standard Reference Material for Human Plasma (SRM 1950): GC-MS, LC-MS, NMR, and Clinical Laboratory Analyses, Libraries, and Web-Based Resources.” *Analytical Chemistry* 85(24): 11725–31. https://pubs.acs.org/doi/10.1021/ac402503m.

Singh, Jyoti et al. 2015. “Enhanced Secondary Metabolite Production and Pathway Gene Expression by Leaf Explants-Induced Direct Root Morphotypes Are Regulated by Combination of Growth Regulators and Culture Conditions in Centella Asiatica (L.) Urban.” *Plant Growth Regulation* 75(1): 55–66.

Singh, Shrawan, D. R. Singh, V. Shajeeda Banu, and N. Avinash. 2014. “Functional Constituents (Micronutrients and Phytochemicals) and Antioxidant Activity of Centella Asiatica (L.) Urban Leaves.” *Industrial Crops and Products* 61: 115–19.

Škrášková, Karolina et al. 2013. “Fast and Sensitive UHPLC Methods with Fluorescence and Tandem Mass Spectrometry Detection for the Determination of Tetracycline Antibiotics in Surface Waters.” *Journal of Chromatography B: Analytical*

Technologies in the Biomedical and Life Sciences 927: 201–8.

Smith, Rosina E. et al. 2008. "A Reliable Biomarker Derived from Plasmalogens to Evaluate Malignancy and Metastatic Capacity of Human Cancers." *Lipids* 43(1): 79–89.

Snyder, Nathaniel W et al. 2013. "Untargeted Metabolomics from Biological Sources Using Ultraperformance Liquid Chromatography-High Resolution Mass Spectrometry (UPLC-HRMS)." *Journal of visualized experiments : JoVE* (75): e50433.

Sopher, Bryce L. et al. 1996. "Neurodegenerative Mechanisms in Alzheimer Disease: A Role for Oxidative Damage in Amyloid β Protein Precursor-Mediated Cell Death." *Molecular and Chemical Neuropathology* 29(2–3): 153–68.

Soumyanath, Amala et al. 2012. "Centella Asiatica Extract Improves Behavioral Deficits in a Mouse Model of Alzheimer's Disease: Investigation of a Possible Mechanism of Action." *International Journal of Alzheimer's Disease*.

Southard, J. H. et al. 1991. "The Role of Oxygen Free Radicals in Organ Preservation." *Klinische Wochenschrift* 69(21–23): 1073–76.

Stagliano, Michael C., Joshua G. DeKeyser, Curtis J. Omiecinski, and Daniel D. Jones. 2010. "Bioassay-Directed Fractionation for Discovery of Bioactive Neutral Lipids Guided by Relative Mass Defect Filtering and Multiplexed Collision-Induced Dissociation." *Rapid Communications in Mass Spectrometry* 24(24): 3578–84.

Su, Wenru et al. 2013. "Doxycycline-Mediated Inhibition of Corneal Angiogenesis: An MMP-Independent Mechanism." *Investigative Ophthalmology and Visual Science* 54(1): 783–88.

Suh, Eunkyung, Sang-Woon Choi, and Simonetta Friso. 2016. "One-Carbon Metabolism." In *Molecular Basis of Nutrition and Aging*, Elsevier, 513–22. https://linkinghub.elsevier.com/retrieve/pii/B9780128018163000364.

Sumner, L W, Z T Lei, and D V Huhman. 2011. "Mass Spectrometry Strategies in Metabolomics." *Journal of Biological Chemistry* 286(29): 25435–42.

Sumner, Lloyd W. et al. 2007. “Proposed Minimum Reporting Standards for Chemical Analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI).” *Metabolomics* 3(3): 211–21.

Sumner, Lloyd W et al. 2007. “Proposed Minimum Reporting Standards for Chemical Analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI).” *Metabolomics* 3(3): 211–21.

Szulc, Jolanta et al. 2006. “A Versatile Tool for Conditional Gene Expression and Knockdown.” *Nature Methods* 3(2): 109–16.

Tautenhahn, Ralf et al. 2012. “An Accelerated Workflow for Untargeted Metabolomics Using the METLIN Database.” *Nature Biotechnology* 30(9): 826–28.

Tautenhahn, Ralf, Christoph Bottcher, and Steffen Neumann. 2008. “Highly Sensitive Feature Detection for High Resolution LC/MS.” *BMC Bioinformatics* 9.

Thevenot, Etienne A. 2017. “Ropls : PCA , PLS (-DA) and OPLS (-DA) for Multivariate Analysis and Feature Selection of Omics Data.” (April): 1–11.

Thonusin, Chanisa et al. 2017. “Evaluation of Intensity Drift Correction Strategies Using MetaboDrift, a Normalization Tool for Multi-Batch Metabolomics Data.” *Journal of Chromatography A* 1523: 265–74.

Till, R. J., and D. J. Hand. 2001. “A Simple Generalisation of the Area under the ROC Curve for Multiple Class Classification Problems.” *Machine Learning* 45(2): 171–86.

Tobias, Randall D. 1995. “An Introduction to Partial Least Squares Regression.” *Proceedings of the twentieth annual SAS users group international conference*: 1250–57. http://www.sascommunity.org/sugi/SUGI95/Sugi-95-210 Tobias.pdf%5Cnpapers2://publication/uuid/0A1BFD33-B570-45AC-BC62-3E93AA77A6D2.

Todeschini, Roberto et al. 2012. “Similarity Coefficients for Binary Chemoinformatics Data: Overview and Extended Comparison Using Simulated and Real Data Sets.” *Journal of Chemical Information and Modeling* 52(11): 2884–2901.

"Train_test_split Vs StratifiedShuffleSplit." https://medium.com/@411.codebrain/train-test-split-vs-stratifiedshufflesplit-374c3dbdcc36.

Triebl, Alexander et al. 2017. "Lipidomics by Ultrahigh Performance Liquid Chromatography-High Resolution Mass Spectrometry and Its Application to Complex Biological Samples." *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* 1053: 72–80.

Tsugawa, Hiroshi et al. 2015. "MS-DIAL: Data-Independent MS/MS Deconvolution for Comprehensive Metabolome Analysis." *Nature Methods* 12(6): 523–26.

Turck, Natacha et al. 2011. "PROC: An Open-Source Package for R and S+ to Analyze and Compare ROC Curves." *BMC Bioinformatics* 8: 12–77. http://link.springer.com/10.1007/s00134-009-1641-y.

Uifălean, Alina et al. 2016. "The Impact of Soy Isoflavones on MCF-7 and MDA-MB-231 Breast Cancer Cells Using a Global Metabolomic Approach." *International Journal of Molecular Sciences* 17(9).

Vaswani, Ashish et al. 2021. "Comparative LC–MS/MS Lipidomic Analysis of Macaque Heart Tissue Flash Frozen or Embedded in Optimal Cutting Temperature Polymer (OCT): Practical Considerations." *Rapid Communications in Mass Spectrometry*. https://onlinelibrary.wiley.com/doi/10.1002/rcm.9155.

Vecchini, Alba et al. 2000. "Molecular Defects in Sarcolemmal Glycerophospholipid Subclasses in Diabetic Cardiomyopathy." *Journal of Molecular and Cellular Cardiology* 32(6): 1061–74.

Viant, Mark R., Irwin J. Kurland, Martin R. Jones, and Warwick B. Dunn. 2017. "How Close Are We to Complete Annotation of Metabolomes?" *Current Opinion in Chemical Biology* 36: 64–69.

Vinaixa, Maria et al. 2012. "A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data." *Metabolites* 2(4): 775–95.

Vinayavekhin, Nawaporn, and Alan Saghatelian. 2010. "Untargeted Metabolomics." *Current Protocols in Molecular Biology* (SUPPL. 90).

Vogenberg, F R, C I Barash, and M Pursel. 2010. “Personalized Medicine - Part 1: Evolution and Development into Theranostics.” *P and T* 35(10): 560-567+576. https://www.scopus.com/inward/record.uri?eid=2-s2.0-78149328298&partnerID=40&md5=8b7750e702ff1454a87ef5324eead9a9.

Wang, Miao, Chunyan Wang, and Xianlin Han. 2017. “Selection of Internal Standards for Accurate Quantification of Complex Lipid Species in Biological Extracts by Electrospray Ionization Mass Spectrometry—What, How and Why?” *Mass Spectrometry Reviews* 36(6): 693–714.

Wang, Mingxun et al. 2016. “Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking.” *Nature Biotechnology* 34(8): 828–37.

Wang, Xiaojian et al. 2004. “A Novel Human Phosphatidylethanolamine-Binding Protein Resists Tumor Necrosis Factor α-Induced Apoptosis by Inhibiting Mitogen-Activated Protein Kinase Pathway Activation and Phosphatidylethanolamine Externalization.” *Journal of Biological Chemistry* 279(44): 45855–64.

Want, Elizabeth J. et al. 2013. “Global Metabolic Profiling of Animal and Human Tissues via UPLC-MS.” *Nature Protocols* 8(1): 17–32.

Waters Corporation. 2011. “An Overview of the Principles of MSE, The Engine That Drives MS Performance.” *Waters Application Note*: 1–6.

Watson, Andrew D. et al. 1997. “Structural Identification by Mass Spectrometry of Oxidized Phospholipids in Minimally Oxidized Low Density Lipoprotein That Induce Monocyte/Endothelial Interactions and Evidence for Their Presence in Vivo.” *Journal of Biological Chemistry* 272(21): 13597–607.

Wehrens, Ron et al. 2016. “Improved Batch Correction in Untargeted MS-Based Metabolomics.” *Metabolomics* 12(5): 88. http://link.springer.com/10.1007/s11306-016-1015-8.

Weinhouse, Sidney, Otto Warburg, Dean Burk, and Arthur L. Schade. 1956. “On Respiratory Impairment in Cancer Cells.” *Science* 124(3215): 267–72.

Weller, Michael G. 2012. “A Unifying Review of Bioassay-Guided Fractionation, Effect-Directed Analysis and Related Techniques.” *Sensors (Switzerland)* 12(7): 9181–9209.

Weng, Stephen F. et al. 2017. “Can Machine-Learning Improve Cardiovascular Risk Prediction Using Routine Clinical Data?” *PLoS ONE* 12(4).

Wenk, Markus R. 2005. “The Emerging Field of Lipidomics.” *Nature Reviews Drug Discovery* 4(7): 594–610.

Wickramasekara, Samanthi et al. 2013. “Electrospray Quadrupole Travelling Wave Ion Mobility Time-of-Flight Mass Spectrometry for the Detection of Plasma Metabolome Changes Caused by Xanthohumol in Obese Zucker (Fa/Fa) Rats.” *Metabolites* 3(3): 701–17.

Widschwendter, Martin, and Peter A. Jones. 2002. “DNA Methylation and Breast Carcinogenesis.” *Oncogene* 21–35(3): 5462–82.

Wink, M. 2013. “Evolution of Secondary Metabolites in Legumes (Fabaceae).” *South African Journal of Botany* 89: 164–75.

Wishart, David S. et al. 2018. “HMDB 4.0: The Human Metabolome Database for 2018.” *Nucleic Acids Research* 46(D1): D608–17.

Witjaksono, Gunawan, and Sagir Alva. 2019. “Applications of Mass Spectrometry to the Analysis of Adulterated Food.” In *Mass Spectrometry - Future Perceptions and Applications*, IntechOpen. https://www.intechopen.com/books/mass-spectrometry-future-perceptions-and-applications/applications-of-mass-spectrometry-to-the-analysis-of-adulterated-food.

Wolf, Claude, and Peter J. Quinn. 2008. “Lipidomics: Practical Aspects and Applications.” *Progress in Lipid Research* 47(1): 15–36.

Woltjer, Randall L. et al. 2003. “Advanced Glycation Endproduct Precursor Alters Intracellular Amyloid- β/AβPP Carboxy-Terminal Fragment Aggregation and Cytotoxicity.” *Journal of Alzheimer’s Disease* 5(6): 467–76.

Woltjer, Randall L. et al. 2007. “Effects of Chemical Chaperones on Oxidative Stress and Detergent-Insoluble Species Formation Following Conditional Expression of Amyloid Precursor Protein Carboxy-Terminal Fragment.” *Neurobiology of Disease* 25(2): 427–37. https://linkinghub.elsevier.com/retrieve/pii/S0969996106002609.

Wong, Ma Li et al. 2000. “Acute Systemic Inflammation Up-Regulates Secretory Sphingomyelinase in Vivo: A Possible Link between Inflammatory Cytokines and Atherogenesis.” *Proceedings of the National Academy of Sciences of the United States of America* 97(15): 8681–86.

Wood, Paul L. 2021. “Flow Infusion ESI High Resolution Mass Spectrometry Metabolomics Analytical Platforms.” In , 1–7. http://link.springer.com/10.1007/978-1-0716-0864-7_1.

Xie, Guoxiang et al. 2015. “Lowered Circulating Aspartate Is a Metabolic Feature of Human Breast Cancer.” *Oncotarget* 6(32): 33369–81.

Xiong, Ying, Joachim D. Uys, Kenneth D. Tew, and Danyelle M. Townsend. 2011. “S-Glutathionylation: From Molecular Mechanisms to Health Outcomes.” *Antioxidants and Redox Signaling* 15(1): 233–70.

Xu, Xinran et al. 2008. “Choline Metabolism and Risk of Breast Cancer in a Population-Based Study.” *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 22(6): 2045–52.

Xue, Meng et al. 2016. “Stability and Degradation of Caffeoylquinic Acids under Different Storage Conditions Studied by High-Performance Liquid Chromatography with Photo Diode Array Detection and High-Performance Liquid Chromatography with Electrospray Ionization Collision-Induc.” *Molecules* 21(7).

Yamamoto, Tsunehisa et al. 2018. “Label-Free Evaluation of Myocardial Infarct in Surgically Excised Ventricular Myocardium by Raman Spectroscopy.” *Scientific Reports* 8(1).

Yan Lim, Jin, and Hiu Yee Kwan. 2020. “Roles of Lipids in Cancer.” *Advances in Lipid Metabolism*.

Yang, Chen, Adam D Richardson, Andrei Osterman, and Jeffrey W Smith. 2008. "Profiling of Central Metabolism in Human Cancer Cells by Two-Dimensional NMR, GC-MS Analysis, and Isotopomer Modeling." *Metabolomics* 4(1): 13–29. https://doi.org/10.1007/s11306-007-0094-y.

Yang, Jane Y. et al. 2013. "Molecular Networking as a Dereplication Strategy." *Journal of Natural Products* 76(9): 1686–99.

Yap, Esther Shiau Ping et al. 2021. "Plant Growth and Metabolic Changes in 'Super Hot' Chili Fruit (Capsicum Annuum) Exposed to Supplemental LED Lights." *Plant Science* 305.

Yilmaz, Mustafa Abdullah et al. 2018. "A Comprehensive LC–MS/MS Method Validation for the Quantitative Investigation of 37 Fingerprint Phytochemicals in Achillea Species: A Detailed Examination of A. Coarctata and A. Monocephala." *Journal of Pharmaceutical and Biomedical Analysis* 154: 413–24.

Yoshida, Miyako et al. 2005. "Antiproliferative Constituents from Umbelliferae Plants VII. Active Triterpenes and Rosmarinic Acid from Centella Asiatica." *Biological and Pharmaceutical Bulletin* 28(1): 173–75.

Yuan, Yongfang et al. 2015. "Biopharmaceutical and Pharmacokinetic Characterization of Asiatic Acid in Centella Asiatica as Determined by a Sensitive and Robust HPLC-MS Method." *Journal of Ethnopharmacology* 163: 31–38.

Zelena, Eva et al. 2009. "Development of a Robust and Repeatable UPLC−MS Method for the Long-Term Metabolomic Study of Human Serum." *Analytical Chemistry* 81(4): 1357–64. https://pubs.acs.org/doi/10.1021/ac8019366.

Zhang, Le, Liang Xu, Fengchun Zhang, and Erina Vlashi. 2017. "Doxycycline Inhibits the Cancer Stem Cell Phenotype and Epithelial-to-Mesenchymal Transition in Breast Cancer." *Cell Cycle* 16(8): 737–45.

Zhang, Wen et al. 2015. "Comprehensive Proteome Analysis of Fresh Frozen and Optimal Cutting Temperature (OCT) Embedded Primary Non-Small Cell Lung Carcinoma by LC-MS/MS." *Methods* 81: 50–55.

Zhao, Lin et al. 2014. “Evidence for Association of Mitochondrial Metabolism Alteration with Lipid Accumulation in Aging Rats.” *Experimental Gerontology* 56: 3–12.

Zhong, Weilong et al. 2017. “Doxycycline Inhibits Breast Cancer EMT and Metastasis through PAR-1/NF-κB/MiR-17/E-Cadherin Pathway.” *Oncotarget* 8(62).

Zhou, Juntuo et al. 2017. “Development of Data-Independent Acquisition Workflows for Metabolomic Analysis on a Quadrupole-Orbitrap Platform.” *Talanta* 164: 128–36.

Zhu, Xiaochun, Yuping Chen, and Raju Subramanian. 2014. “Comparison of Information-Dependent Acquisition, SWATH, and MS All Techniques in Metabolite Identification Study Employing Ultrahigh-Performance Liquid Chromatography-Quadrupole Time-of-Flight Mass Spectrometry.” *Analytical Chemistry* 86(2): 1202–9.

Zhu, Yanlong et al. 2021. “Ultrahigh-Resolution Mass Spectrometry-Based Platform for Plasma Metabolomics Applied to Type 2 Diabetes Research.” *Journal of Proteome Research* 20(1): 463–73. https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00510.

Ziv, Etay, Jeremy C. Durack, and Stephen B. Solomon. 2016. “The Importance of Biopsy in the Era of Molecular Medicine.” *Cancer Journal (United States)* 22(6): 418–22.

Appendices

Appendix A

Supplemental Material for Chapter 3.

Figure AS1 : Untargeted workflow approach. This workflow yielded 24 identified (L1 annotation) and 93 tentatively assigned (L2) compounds (117 in total). Compound description is detailed in Table AS1.

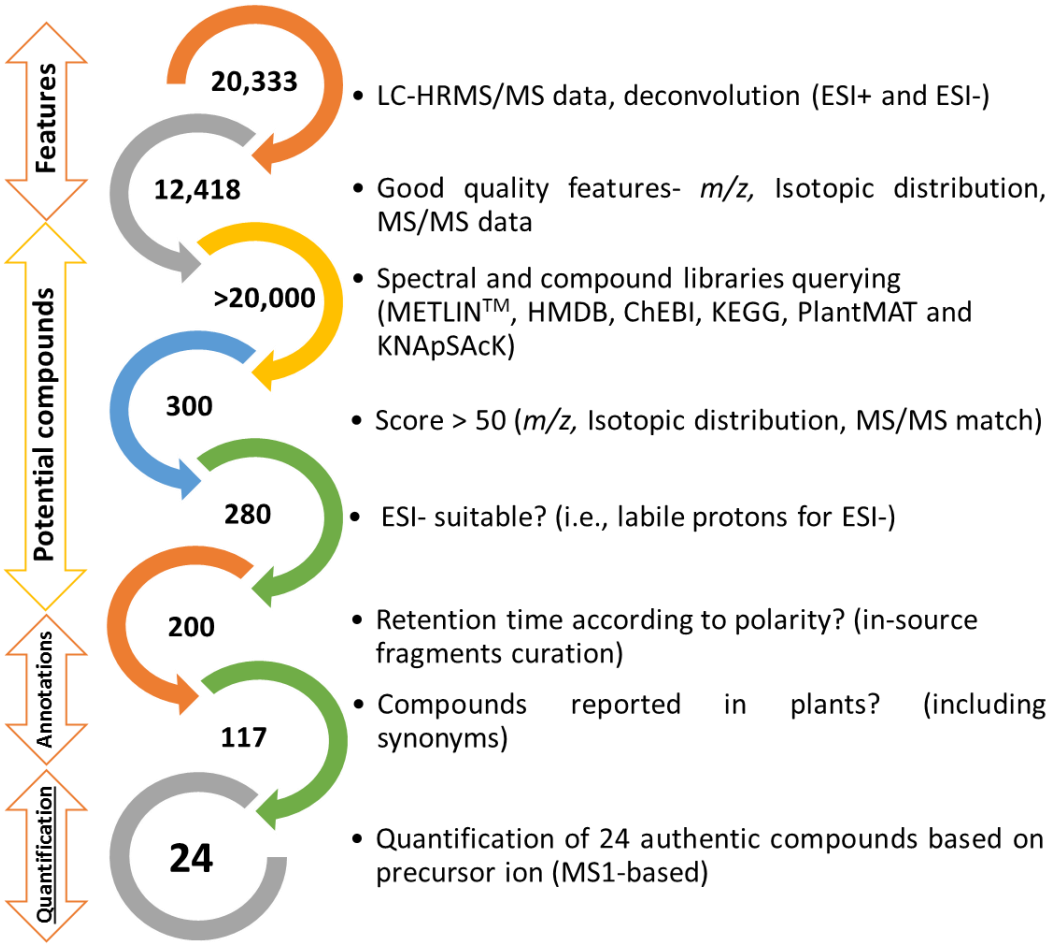


Figure AS2: Untargeted analysis of *C. asiatica* water extract (positive ion mode). Twenty-two annotated features were selected for extracted ion chromatograms. Additional

information namely retention time, *m/z*, molecular formula and detected adducts is shown on Table AS1.

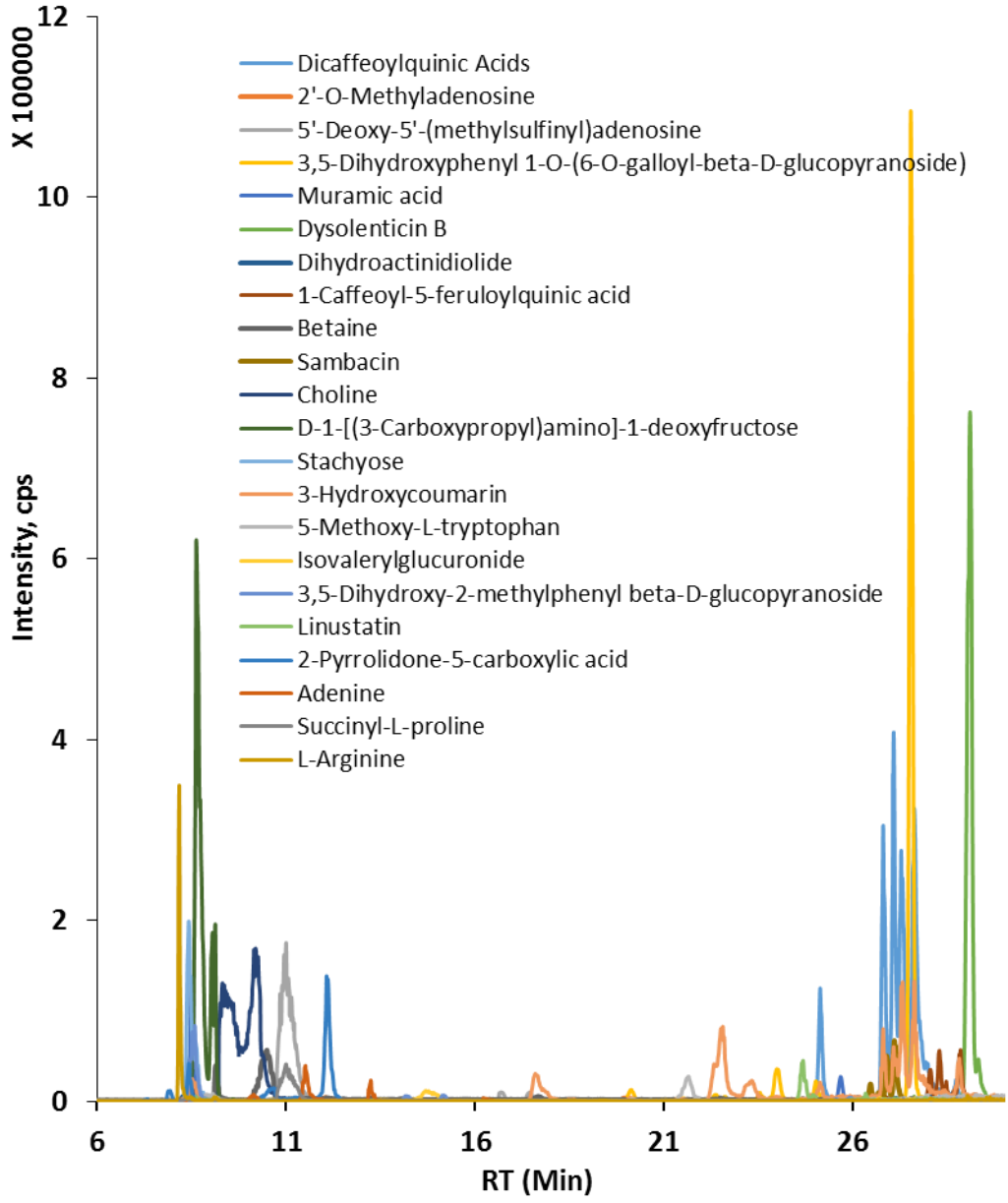


Figure AS3: Untargeted analysis of C. asiatica water extract (negative ion mode). 24 annotated features with higher intensity were selected for extracted ion chromatograms.

Additional information namely retention time, m/z, molecular formula and detected adducts is shown in Table AS1.

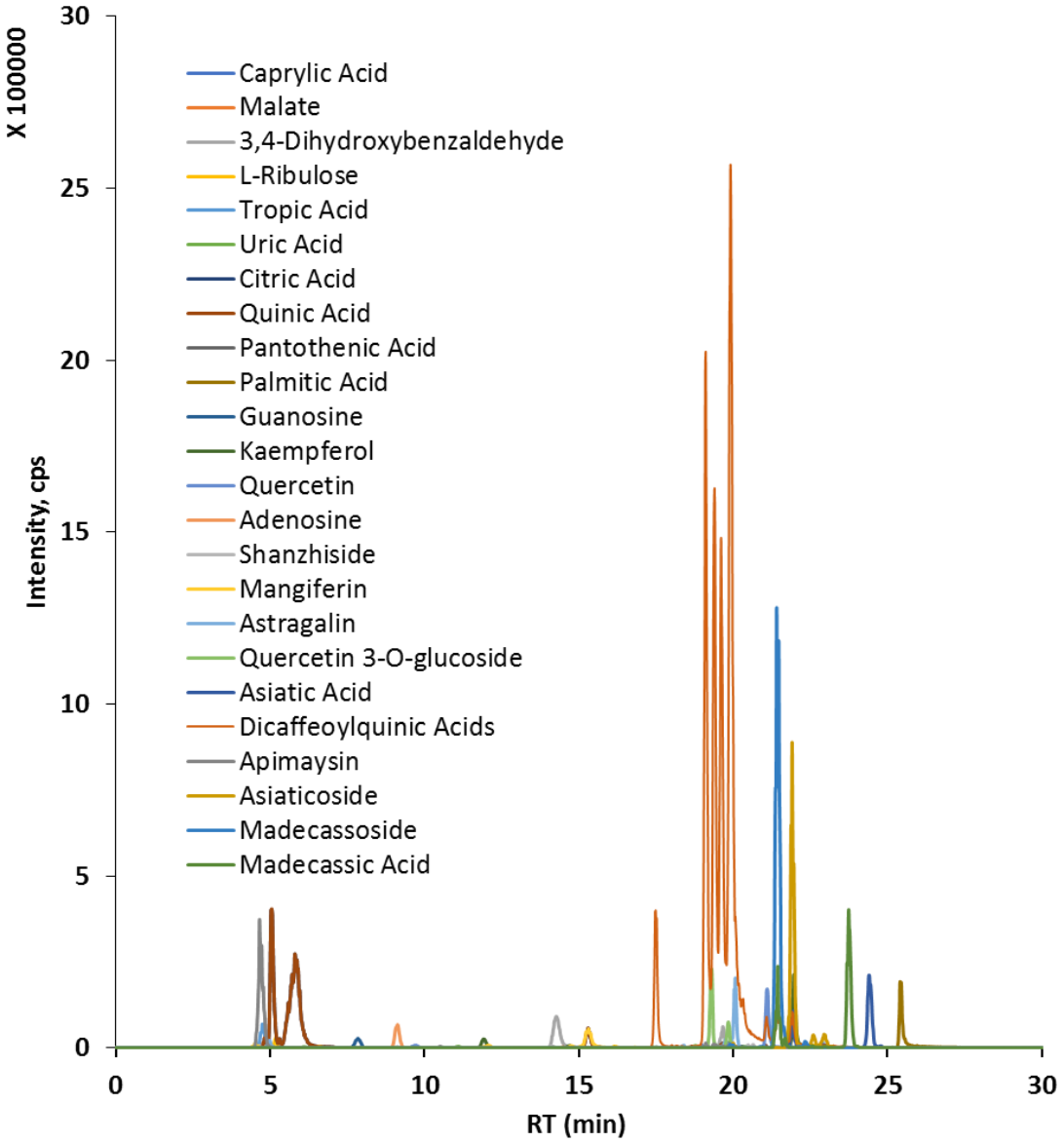
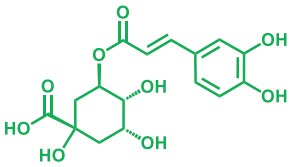
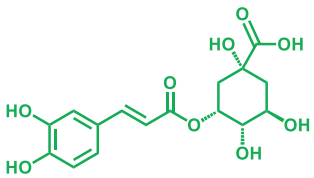


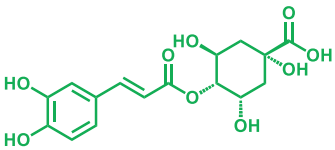
Figure AS4: Chemical structures of compounds for which authentic standards were available and which were selected as marker compounds for quantification. Three mono-caffeoylquinic acids (green), five di-caffeoylquinic acids (magenta), seven flavonoids (purple), five hydroxycinnamic acid derivatives (black) and four triterpenes (blue).



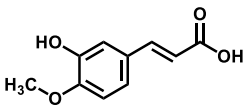
3-O-Caffeoylquinic acid



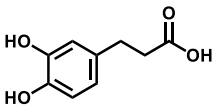
5-Caffeoylquinic acid



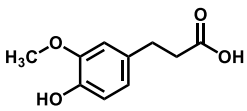
4-O-Caffeoylquinic acid



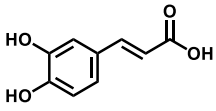
Isoferulic acid



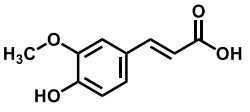
Dihydrocaffeic acid



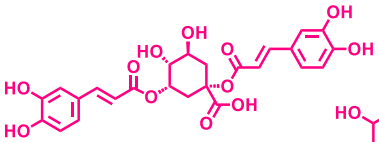
Dihydroferulic acid



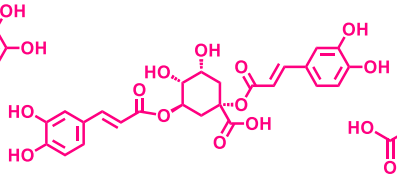
Caffeic acid



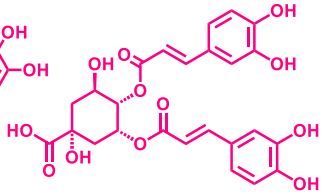
Ferulic acid



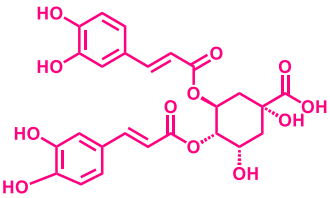
1,5-Dicaffeoylquinic acid



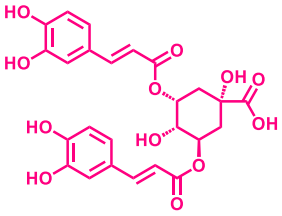
1,3-Dicaffeoylquinic acid



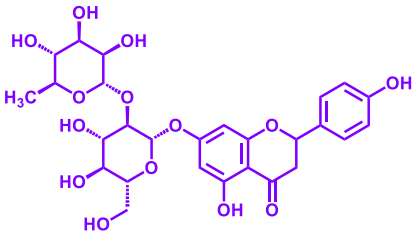
4,5-Dicaffeoylquinic acid



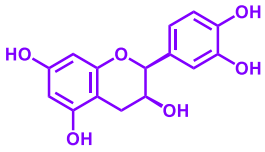
3,4-Dicaffeoylquinic acid



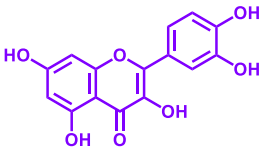
3,5-Dicaffeoylquinic acid



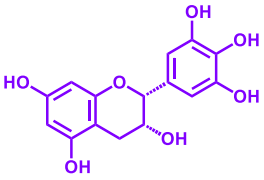
Naringin



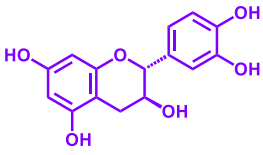
Epicatechin



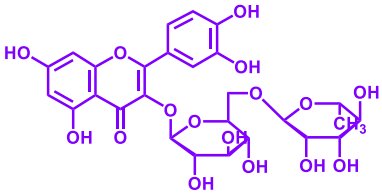
Quercetin



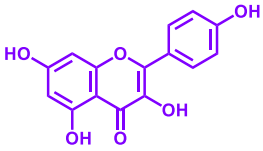
Epigallocatechin



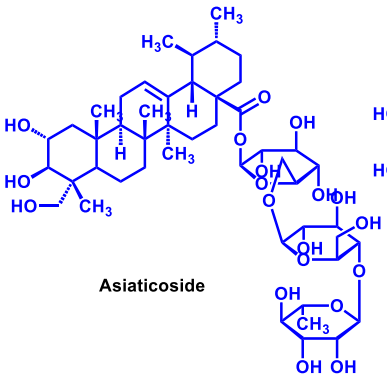
Catechin



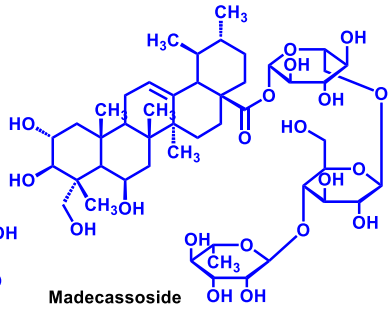
Rutin



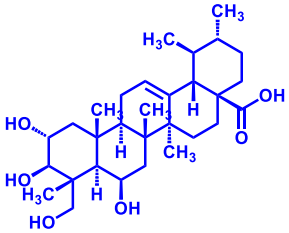
Kaempferol



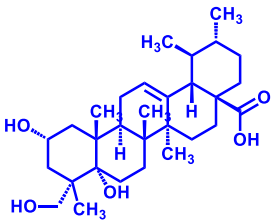
Asiaticoside



Madecassoside



Madecassic Acid



Asiatic acid

Figure AS5: Extracted ion chromatograms obtained for the calibration solution containing 24 compounds (1 mg/L each). Negative ion mode extracted *m/z* values are indicated in the Figure. Individual analytical parameters are shown in Table 3.2.

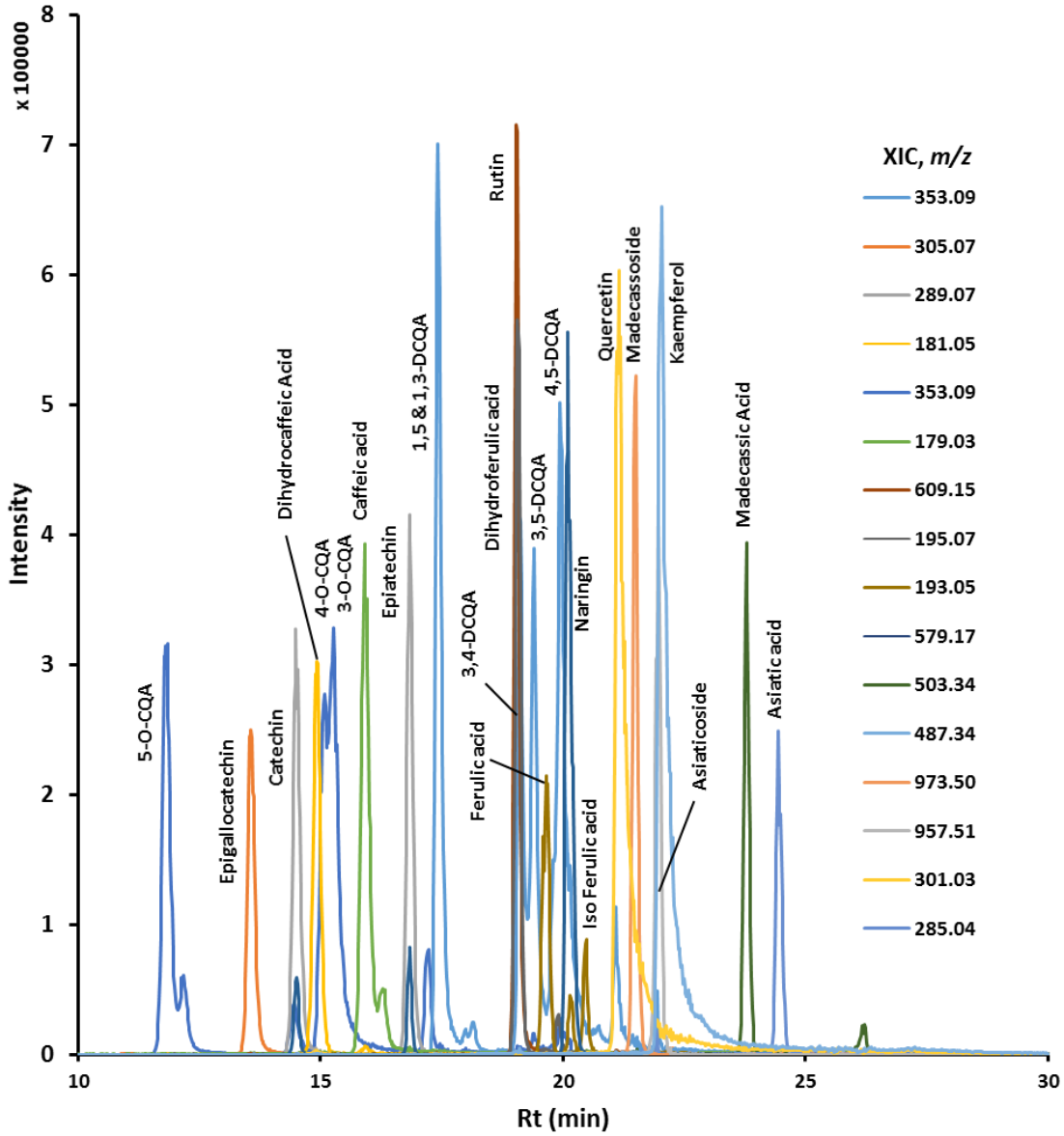


Figure AS6: Standard addition experiment. 1.0 mL of standard mix containing 1.0 mg/L of each compound was added to 1.0 mL of the pooled CA sample (200 mg/L). Total ion chromatogram (TIC) obtained for *C. asiatica* water extract (solid line, 10 µl injection) and same sample after standard addition (dotted line)

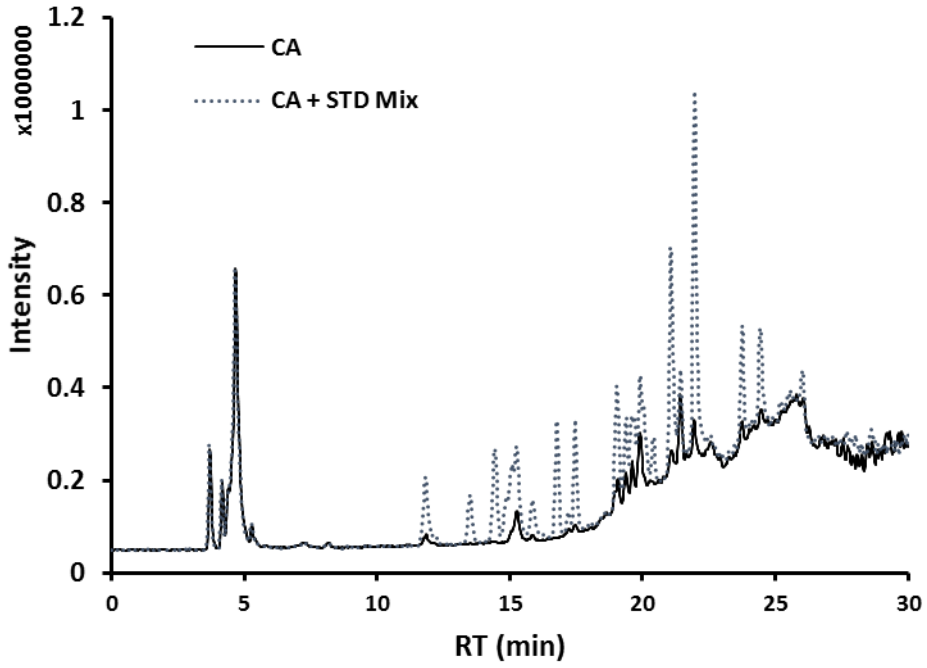
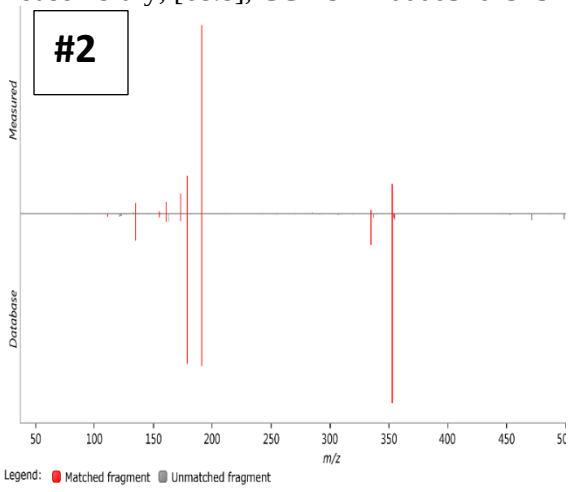
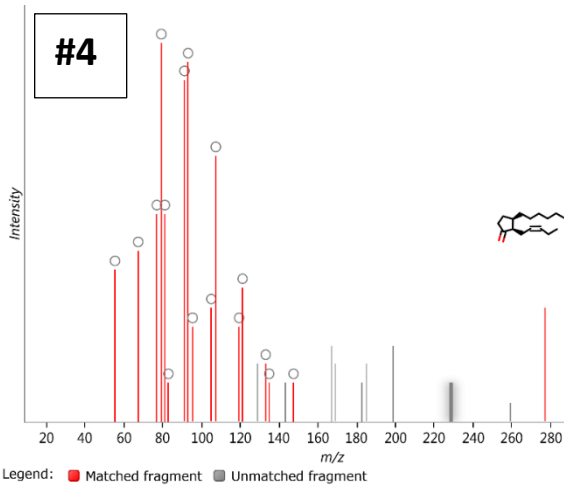

x1000000
1.2
1
0.8
0.6
0.4
0.2
0
Intensity
CA
CA + STD Mix
0
5
10
15
20
25
30
RT (min)

Figure AS7: MS/MS spectra of compounds in C. asiatica extracts (pooled CA sample) that were assigned tentatively (L2 annotations) by extensive querying and comparison with spectral libraries (including METLIN, our in-house library, ChEBI, and the Human Metabolite Database (HMDB)) using Progenesis QITM and applying the workflow shown in Figure S1. Red lines were matches against the databases. Eighty-seven compounds that were detected in C. asiatica aqueous extracts and tentatively assigned but have not been reported for C. asiatica as of to date are denoted with an * in Table 3.1. The number shown in the spectra matches the entry number # in Table AS1. MS/MS scores are indicated in [ ] and were obtained using Progenesis QI. GNPS identifiers are provided in {}.

1,4-Dicaffeoylquinic acid, CID: 12358846, in-house library; [68.8]; CCMSLIB00005467925

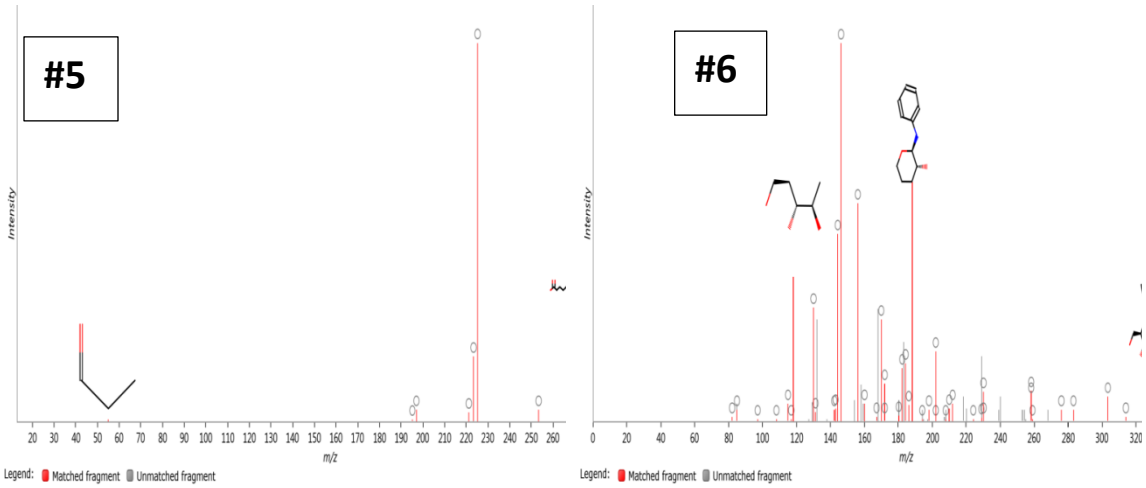


12-Oxodihydrophytodienoic acid, CID: 5716902, ChEBI; [31.0]; CCMSLIB00005467924

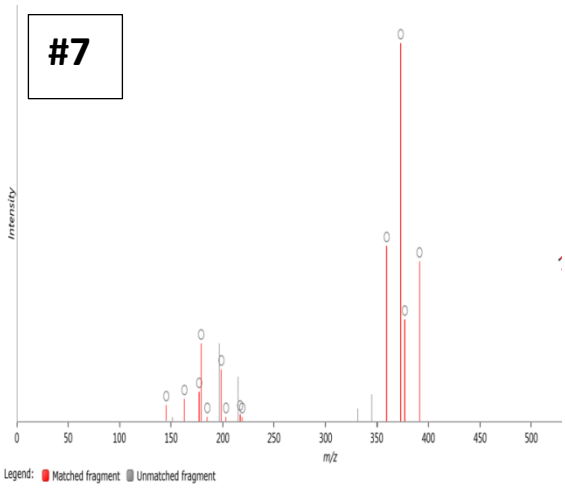


16-hydroxypalmitic acid, CID: 10466, ChEBI; [86.2]; {CCMSLIB00005467727}

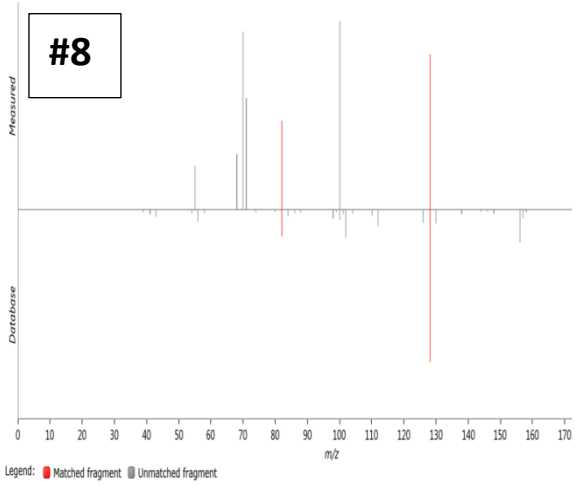
1-beta-D-Glucopyranosyl-L-tryptophan, CID: 11772967, ChEBI; [62.4]; {CCMSLIB00005467655}


#5
Intensity
m/z
Legend: Matched fragment Unmatched fragment
#6
Intensity
m/z
Legend: Matched fragment Unmatched fragment

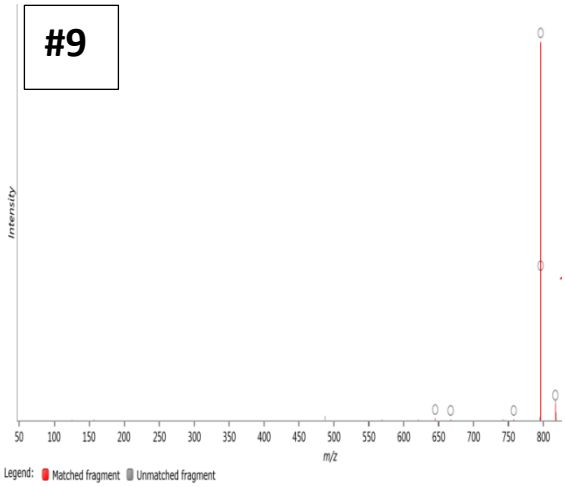
1-Caffeoyl-5-feruloylquinic acid, CID: 121225501, ChEBI; [49.6]; {CCMSLIB00005467656}



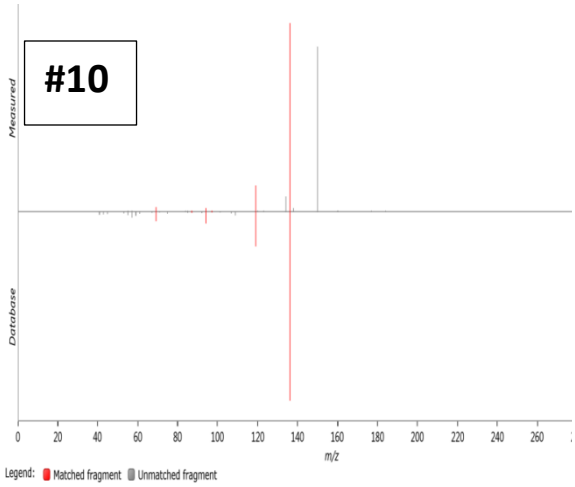
2,6-Piperidinedicarboxylic acid, CID: 557515, METLIN; [53.1]; {CCMSLIB00005467657}



26-(2-Glucosyl-6-acetylglucosyl]-1,3,11,22-tetrahydroxyergosta-5,24-dien-26-oate, CID: 131752817, HMDB; [43.5]; {CCMSLIB00005467797}

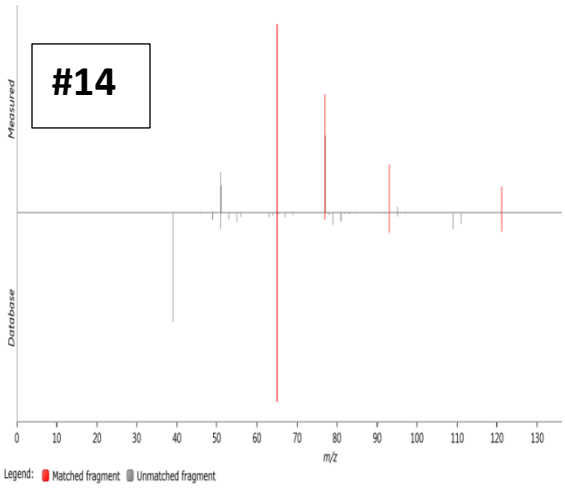


2'-O-Methyladenosine, CID: 102213, METLIN; [56.2]; {CCMSLIB00005467658}

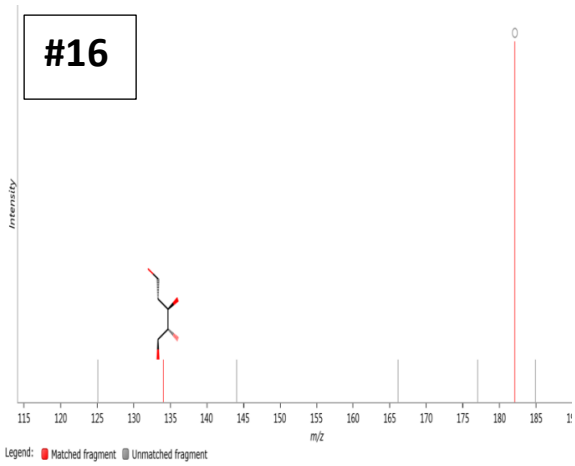


3,4-Dihydroxybenzaldehyde, CID: 8768, ChEBI; [63.5]; {CCMSLIB00005467663}

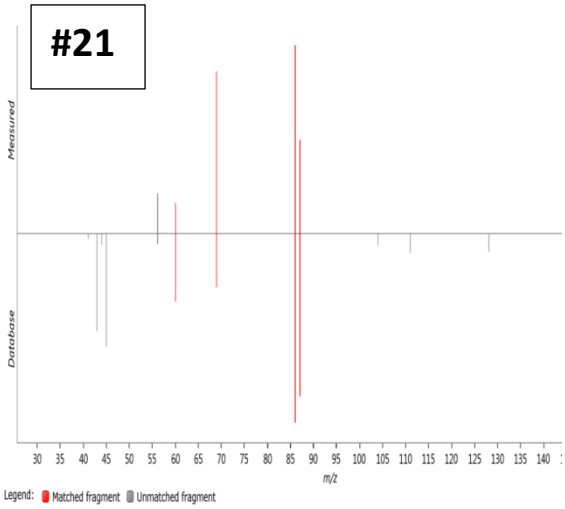
3,5-Dihydroxy-2-methylphenyl beta-D-glucopyranoside, CID: 46184089, ChEBI; [59.1]; {CCMSLIB00005467664}



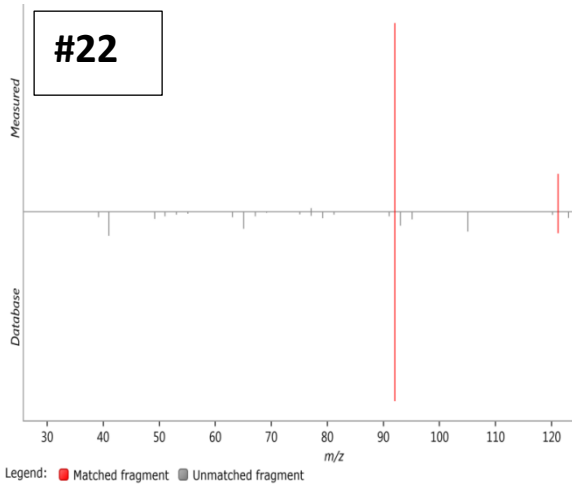
4-Guanidinobutanoic acid, CID: 25200642, METLIN; [79.4]; {CCMSLIB00005467667}



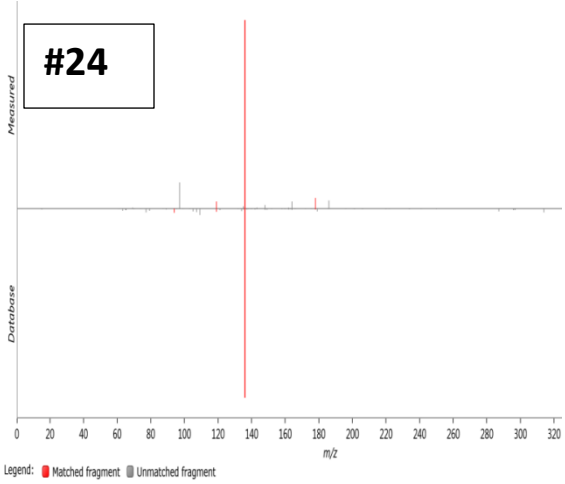
4-Hydroxybenzaldehyde, CID: 126, METLIN; [78.8]; {CCMSLIB00005467733}

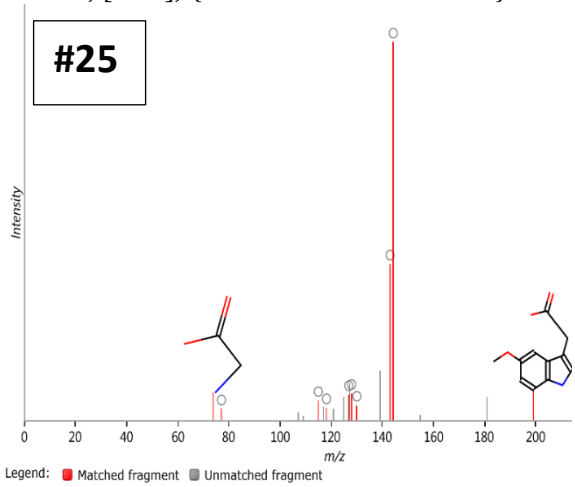


5'-Deoxy-5'-(methylsulfinyl)adenosine, CID: 165114, METLIN; [62.9]; {CCMSLIB00005467668}

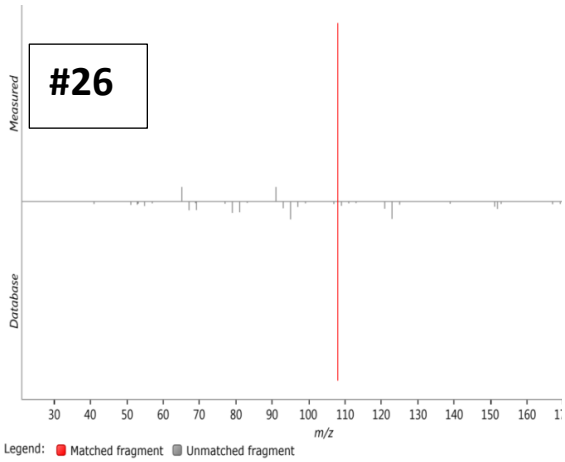


5-Methoxy-L-tryptophan, CID: 151018, ChEBI; [58.5]; {CCMSLIB00005467669}

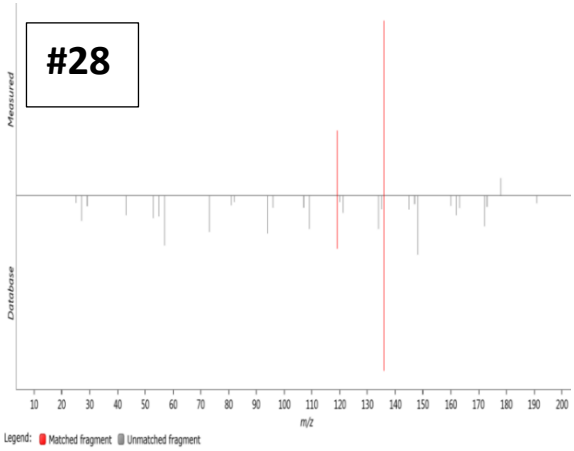




5-Methoxysalicylic acid, CID: 75787, KNApSAcK
; [69.2]; {CCMSLIB00005467734}

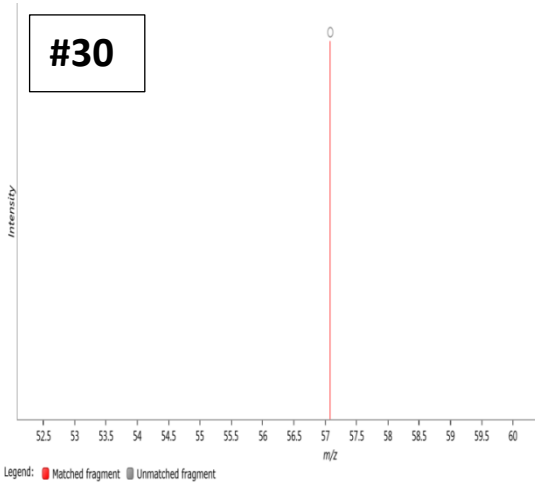


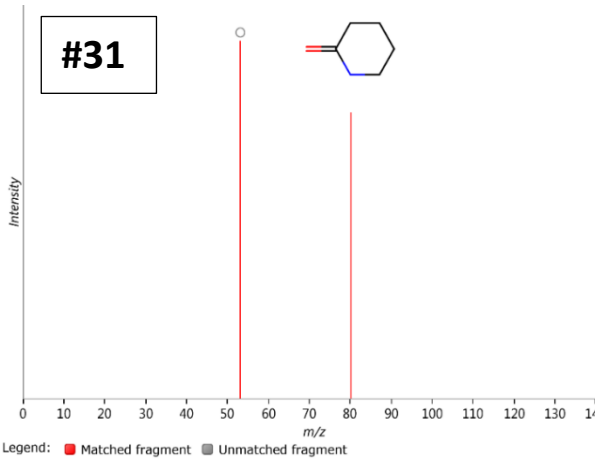
6-Amino-9H-purine-9-propanoic acid, CID: 255450, METLIN; [48.8]; {CCMSLIB00005467673}



6-Docosenamide, CID: 44584605, ChEBI; [98.7]; {CCMSLIB00005467675}

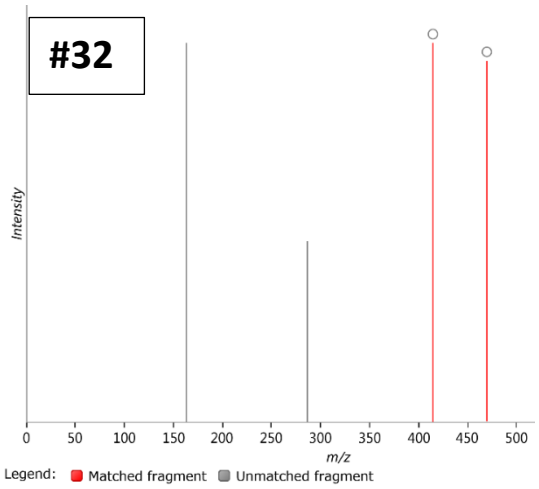
6-Oxo-2-piperidinecarboxylic acid, CID: 3014237, ChEBI; [89.9]; {CCMSLIB00005467676}



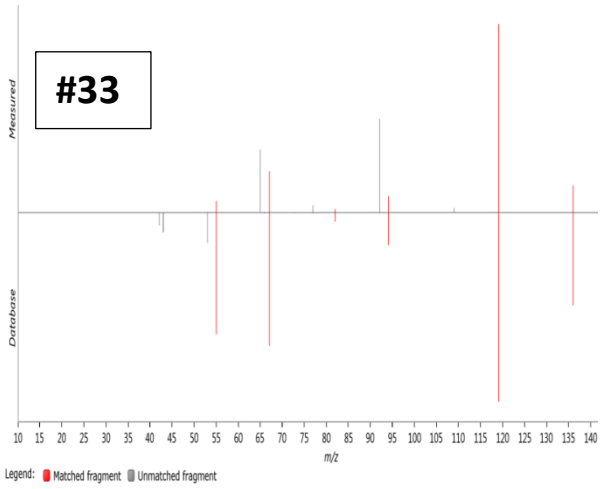


8-Acetoxy-4'-methoxypinoresinol 4-glucoside, CID: 73830447, HMDB; [52.9]; {CCMSLIB00005467677}

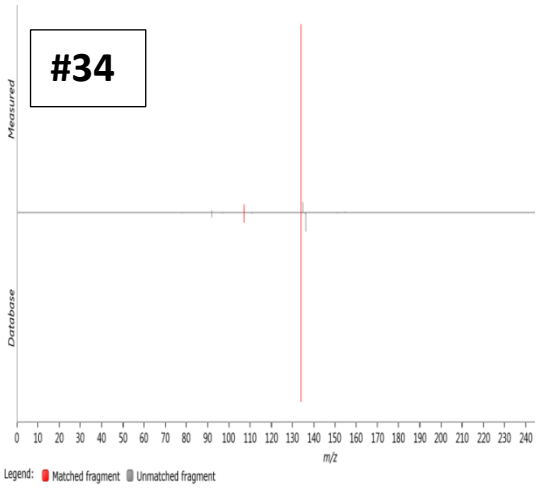
Adenine, CID: 190, METLIN; [88.6]; {CCMSLIB00005467678}



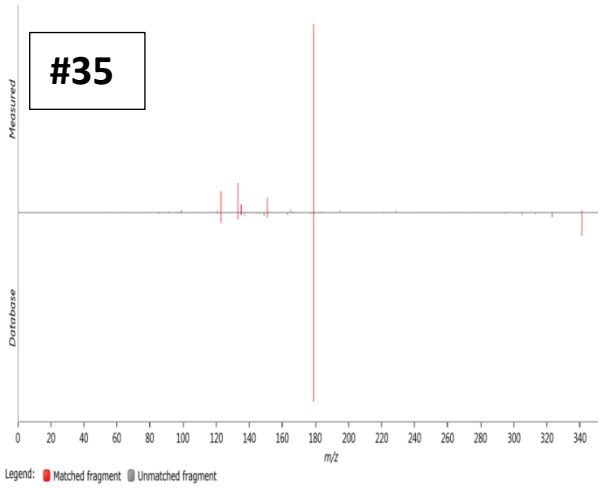
Adenosine, CID: 60961, in-house library; [91.6]; {CCMSLIB00005467736}



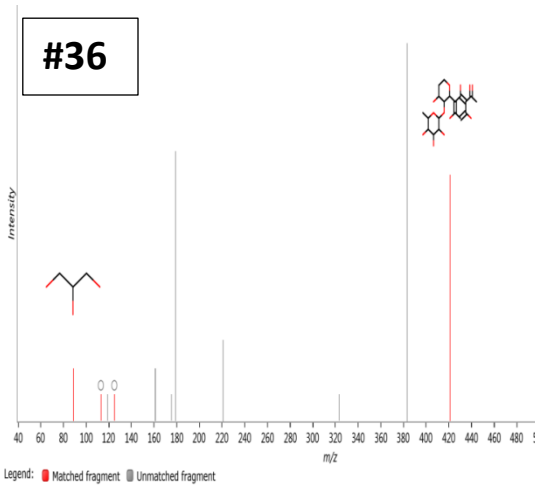
Aesculin, CID: 5281417, METLIN; [69.3]; {CCMSLIB00005467679}



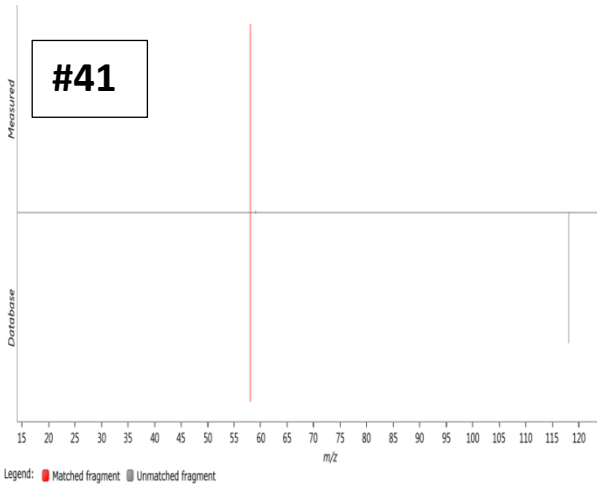
Apimaysin, CID: 101920411, HMDB; [45.4]; {CCMSLIB00005467737}



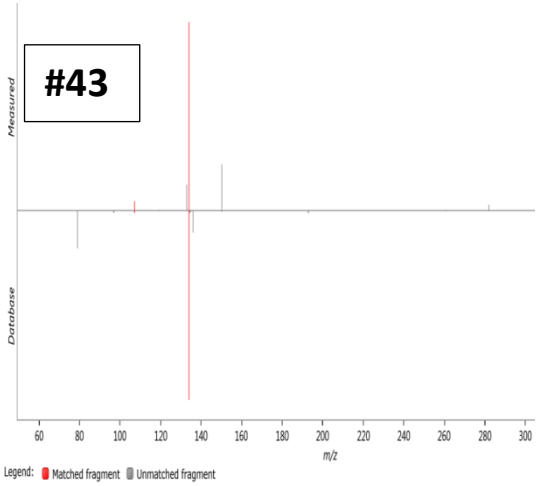
Betaine, CID: 247, METLIN; [66.5]; {CCMSLIB00005467650}



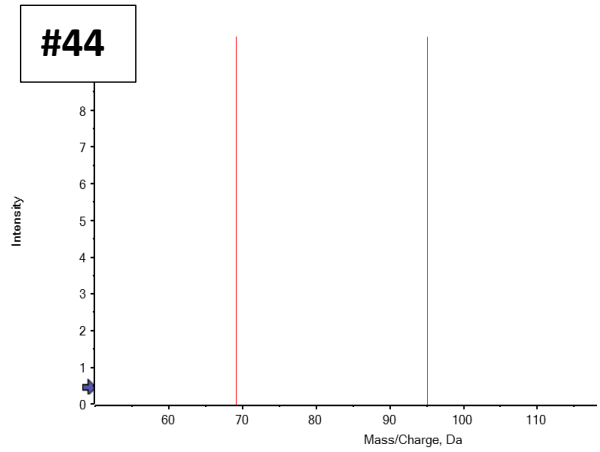
cAMP, CID: 6076, METLIN; [59.7]; {CCMSLIB00005467743}



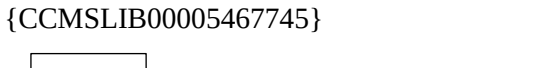
Caprylic acid, matched with in-house library, CID: 379, HMDB; [90.2]; {CCMSLIB00005467744}

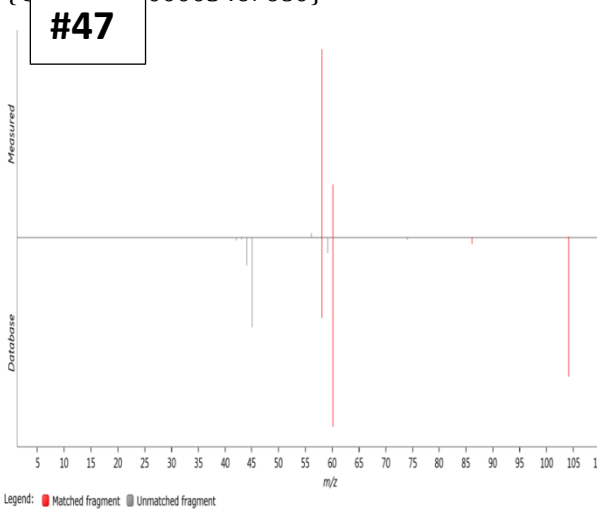


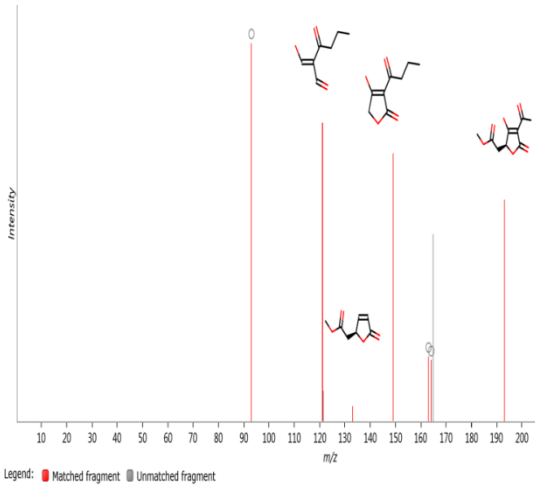
Carlosic acid methyl ester, CID: 122391261, ChEBI; [70.8]; {CCMSLIB00005467745}



Choline, CID: 305, METLIN; [75.6]; {CCMSLIB00005467680}

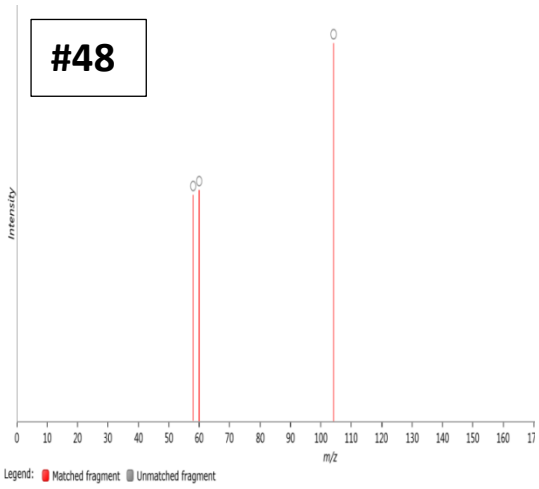


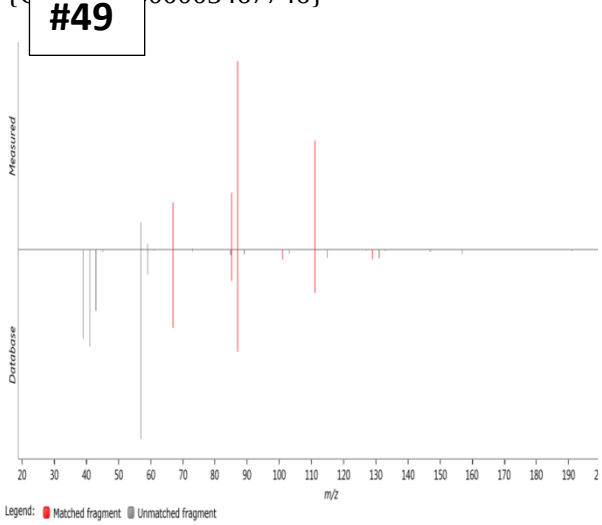




Choline O-Sulfate, CID: 486, ChEBI; [95.1]; {CCMSLIB00005467681}

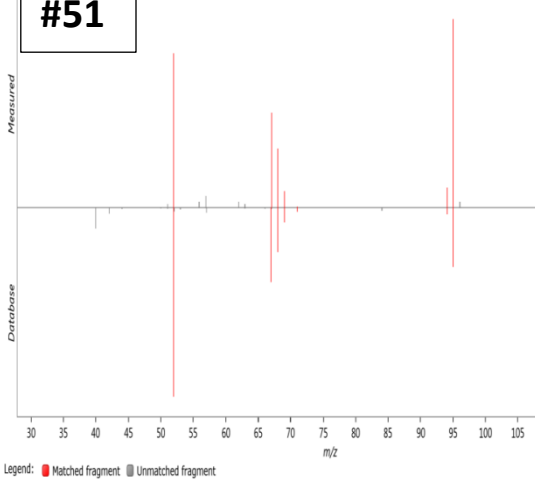
Citric acid, matched with in-house library, CID: 19782904, METLIN; [69.5]; {C 00005467746}

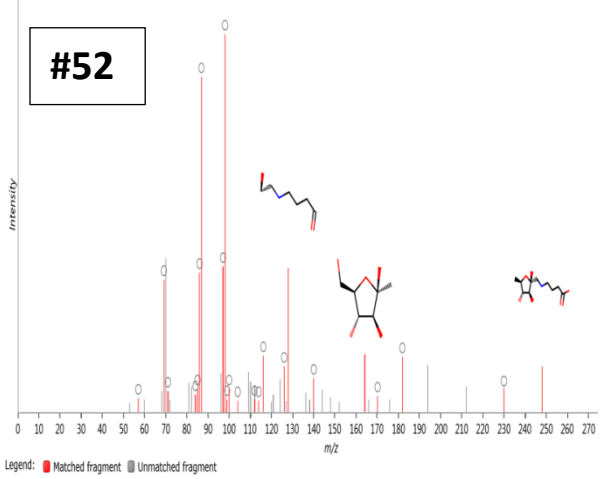




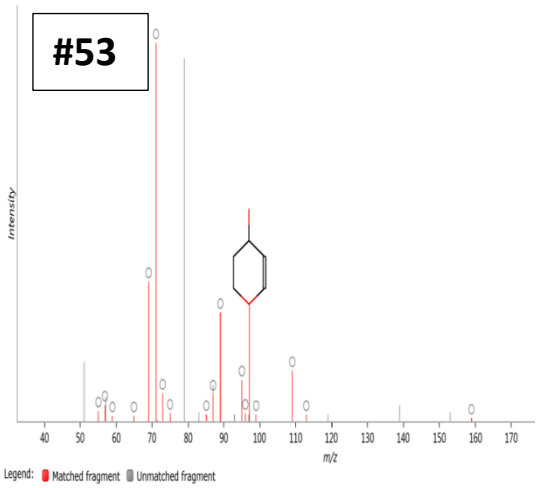
Cytosine, CID: 597, METLIN; [92.3]; {C B00005467683}

D-1-[(3-Carboxypropyl)amino]-1-deoxyfructose, CID: 131752417, HMDB; [53.6]; {CCMSLIB00005467684}

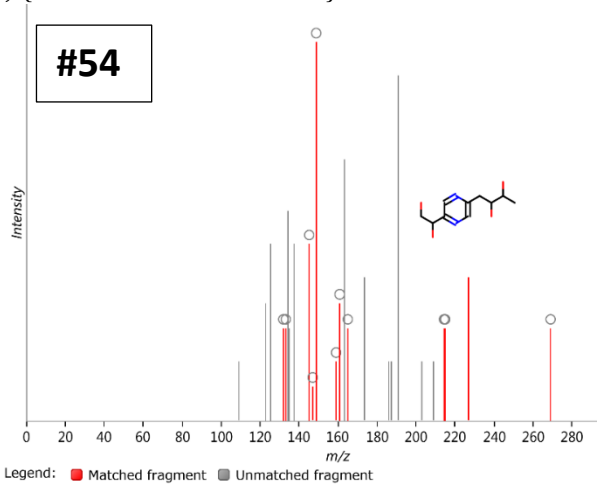




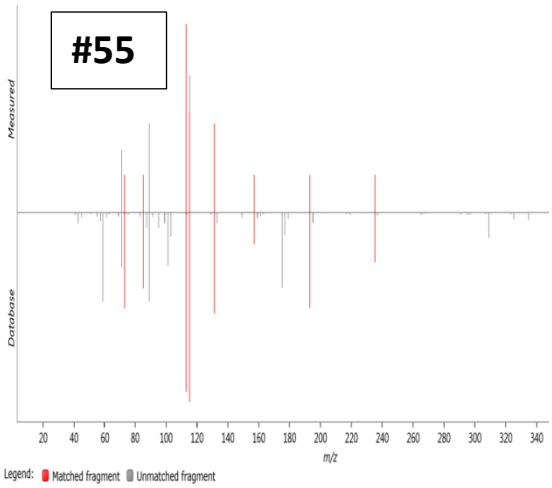
Daucic acid, CID: 5316316, HMDB; [47.7]; {CCMSLIB00005467747}



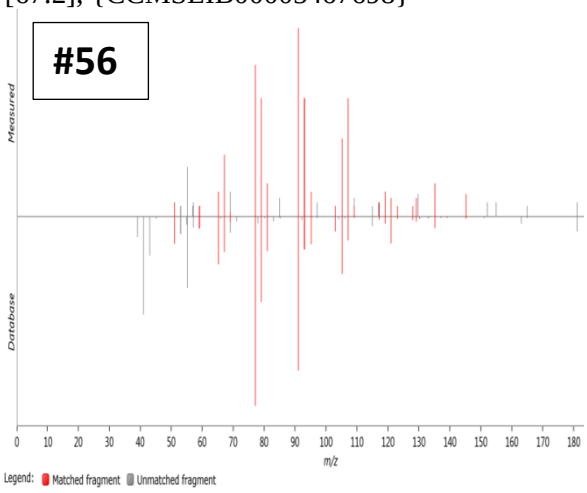
Deoxyfructosazine, CID: 73452, HMDB; [60.4] ; {CCMSLIB00005467697}



Digalacturonate, CID: 439694, METLIN; [40.6]; {CCMSLIB00005467748}

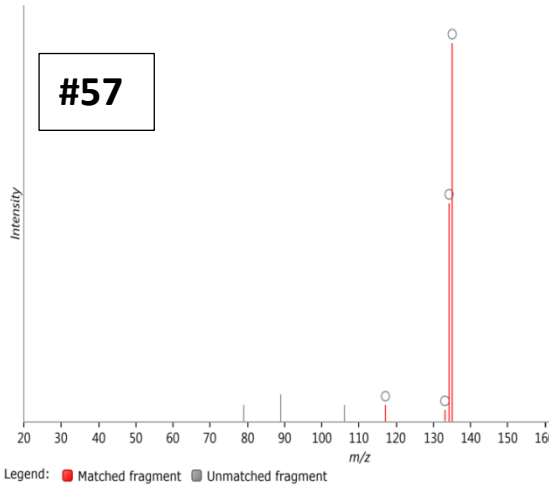


Dihydroactinidiolide, CID: 27209, METLIN; [67.2]; {CCMSLIB00005467698}

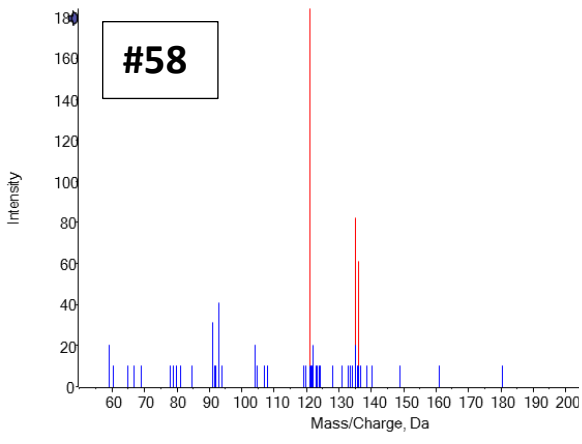


Dihydrocaffeic acid, CID: 348154, ChEBI; [65.1]

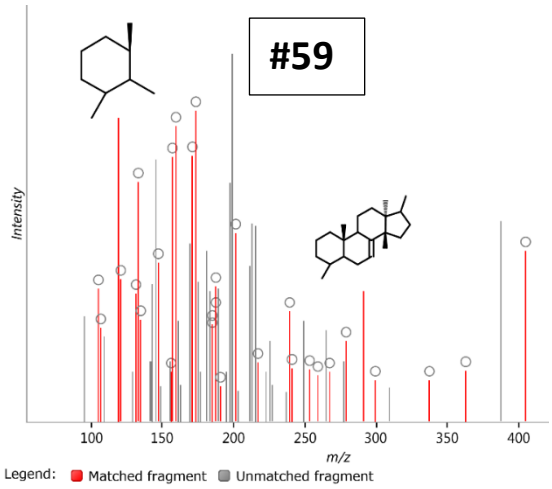
Dihydroferulic acid, matched with in-house library, CID: 14340; [55]

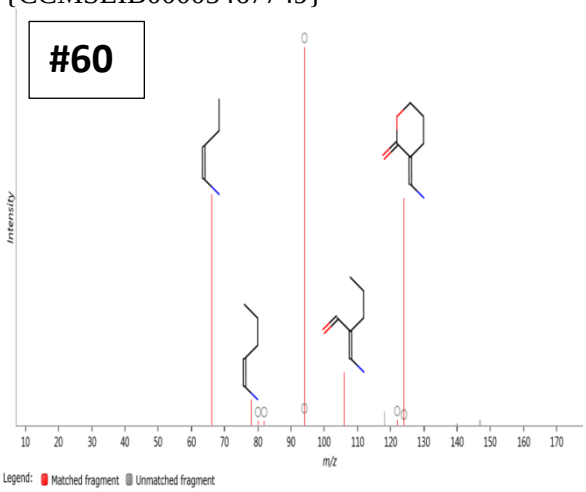


Dysolenticin B, CID: 56601655, ChEBI; [40.4]; {CCMSLIB00005467699}



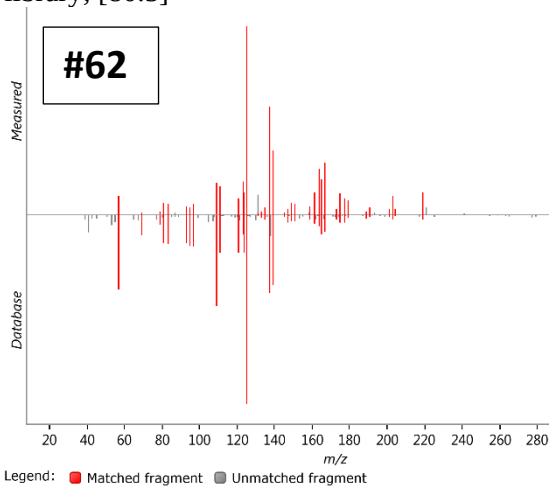
Enicoflavine, CID: 5281564, ChEBI; [71.7]; {CCMSLIB00005467749}

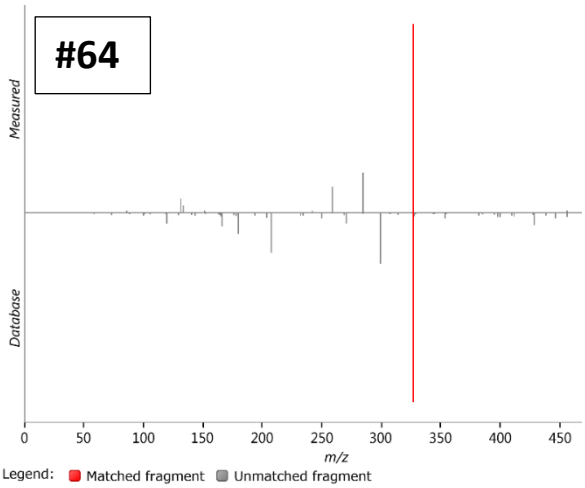




Epigallocatechin, CID: 72277, in-house library; [80.5]

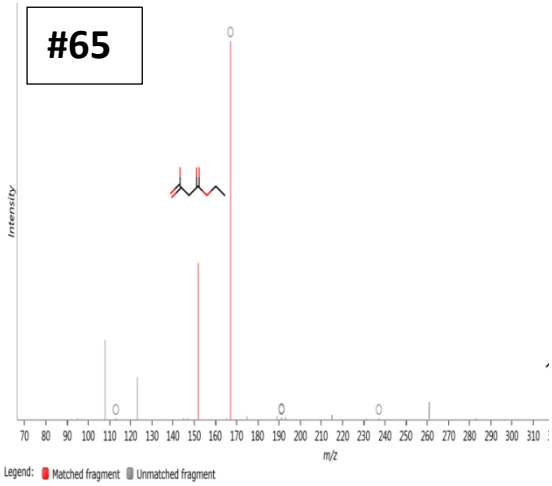
Folinic acid, CID: 6006, in-house library; [50.1]; {CCMSLIB00005467700}

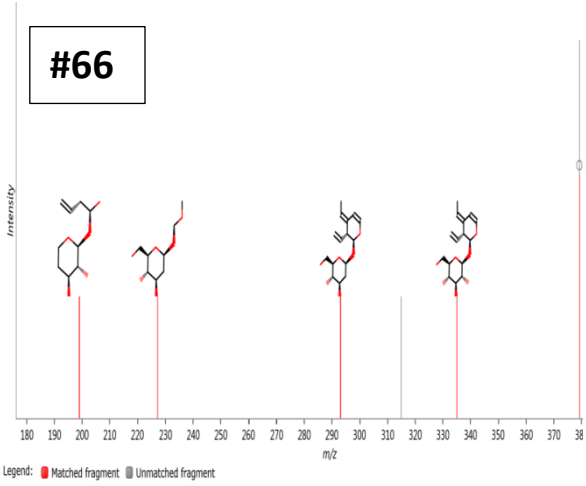




Furaneol 4-(6-malonylglucoside), CID: 131750900, HMDB; [44.1]; ; {CCMSLIB00005467750}

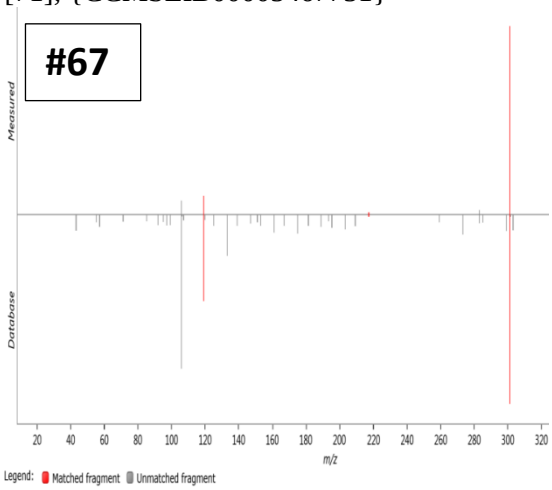
Gentiopicroside, CID: 88708, ChEBI; [34.4]; {CCMSLIB00005467701}

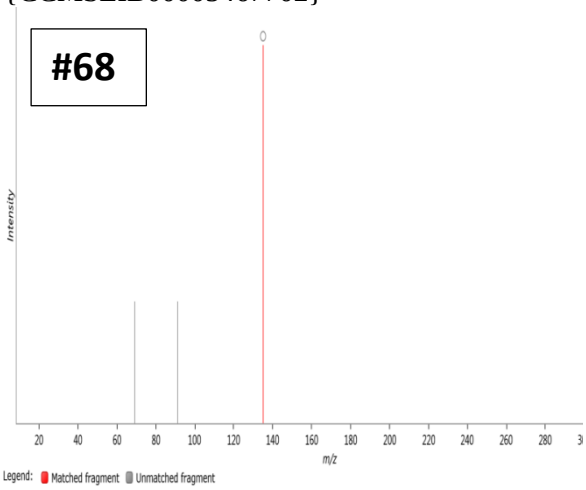




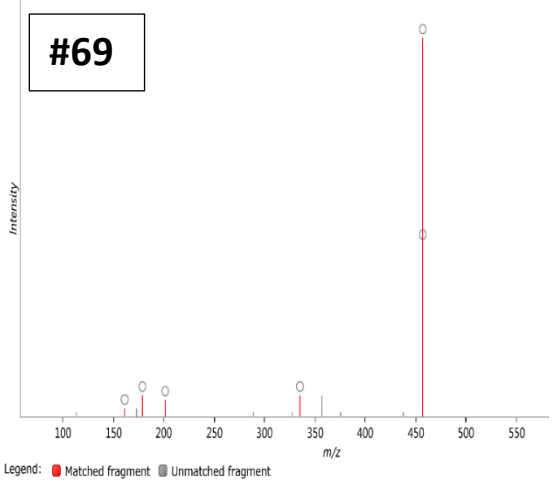
Ginkgoic acid, CID: 5281858, METLIN; [71]; {CCMSLIB00005467751}

Ginsenoyne K, CID: 15736266, HMDB; [73.3]; {CCMSLIB00005467702}

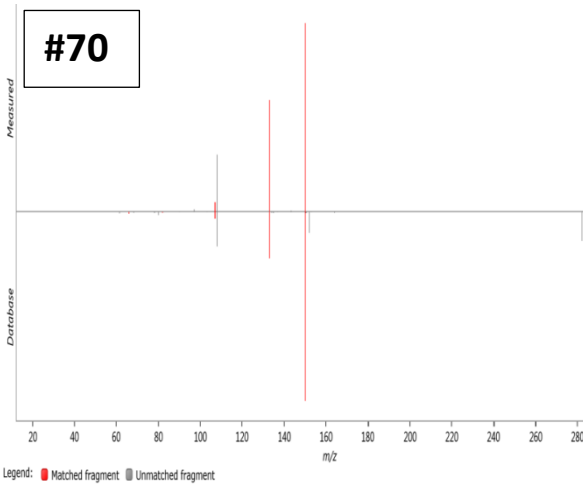




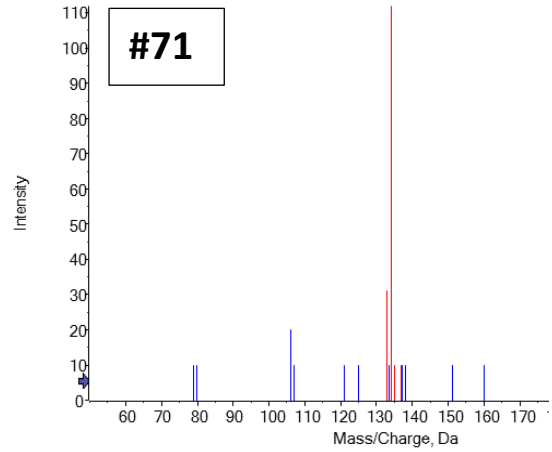
Glabraoside A, CID: 102393599, ChEBI; [36.3]; {CCMSLIB00005467752}



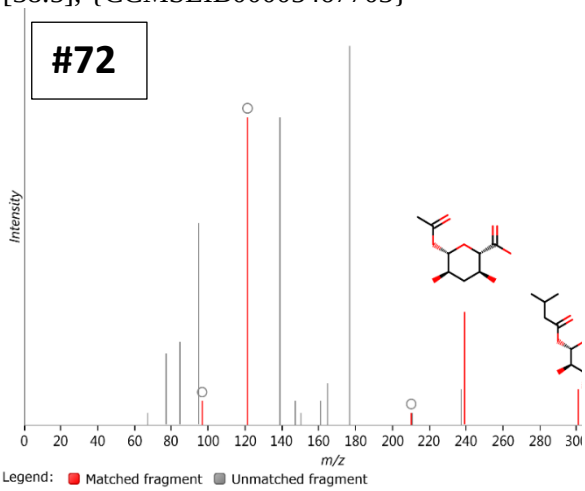
Guanosine, CID: 6802, METLIN; [86.8]; {CCMSLIB00005467753}



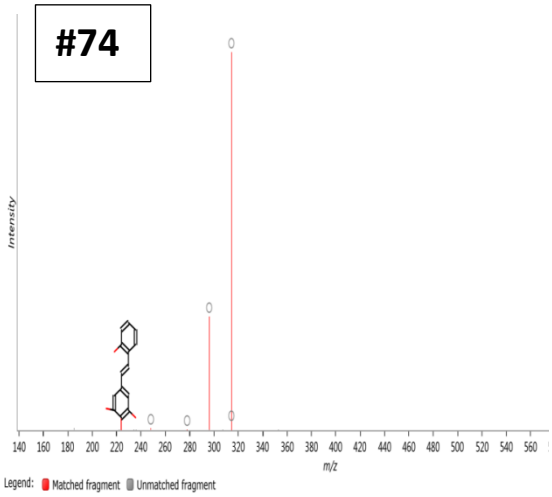
Isoferulic acid, matched with in-house library, CID: 736186, HMDB; [53.6]



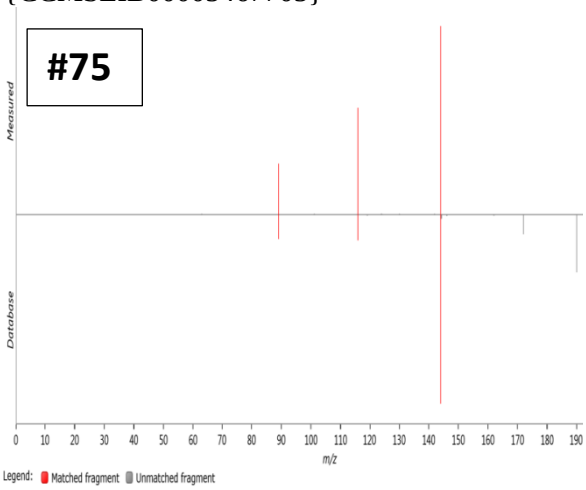
Isovalerylglucuronide, CID: 137383, HMDB; [38.5]; {CCMSLIB00005467703}



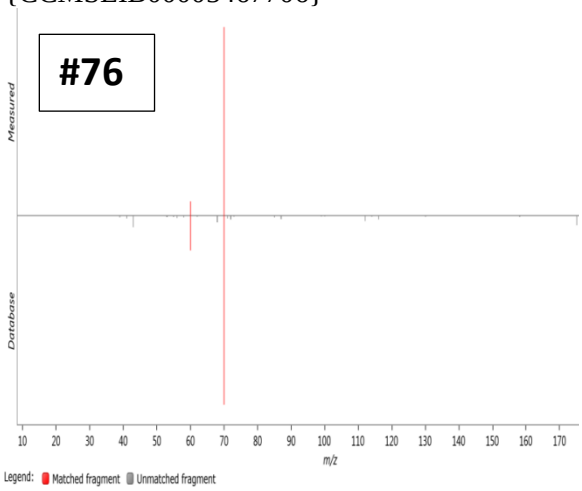
Kuwanon Y, CID: 14334307, HMDB; [86.2]; {CCMSLIB00005467704}



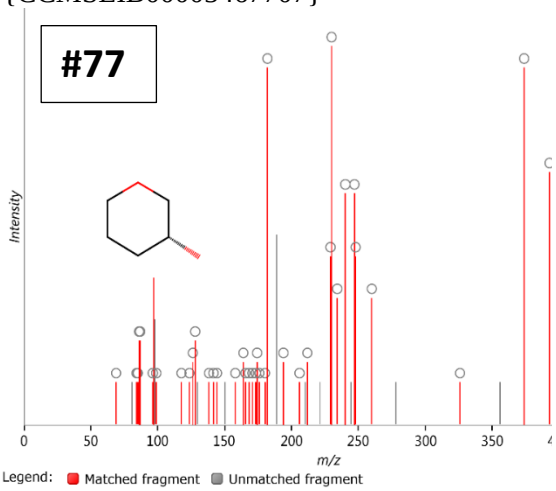
Kynurenic acid, CID: 3845, METLIN; [86]; {CCMSLIB00005467705}



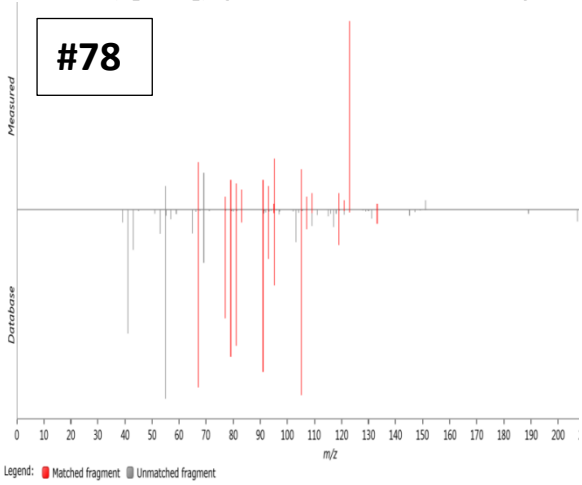
L-Arginine, CID: 28782, METLIN; [74.8]; {CCMSLIB00005467706}



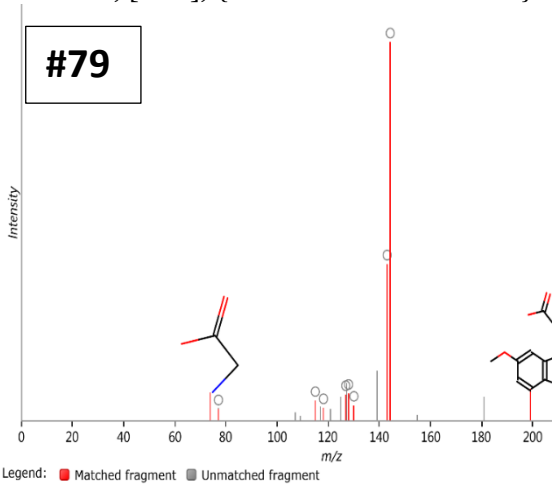
Linustatin, CID: 119301, METLIN; [59.8]; {CCMSLIB00005467707}



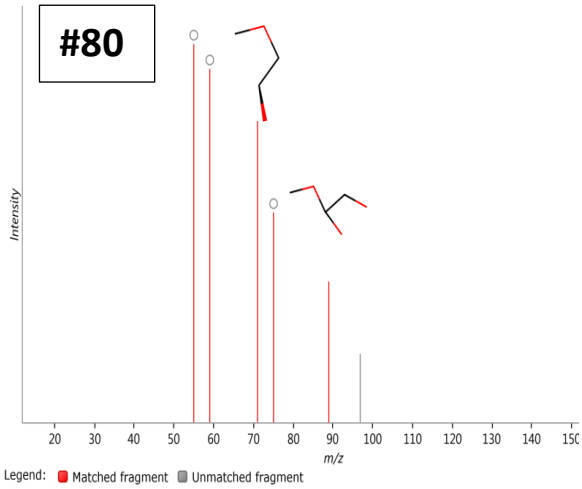
Longicamphenylone, CID: 91747202, METLIN; [51.4]; {CCMSLIB00005467708}



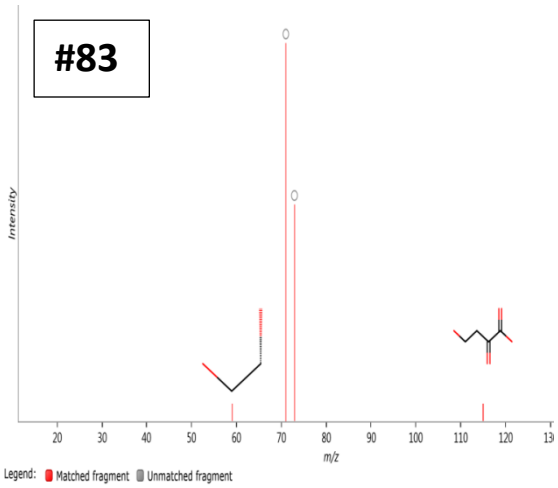
Longifolenaldehyde, CID: 565584, METLIN; [50.9]; {CCMSLIB00005467709}



L-Ribulose, CID: 644111, HMDB; [71.7]; {CCMSLIB00005467754}

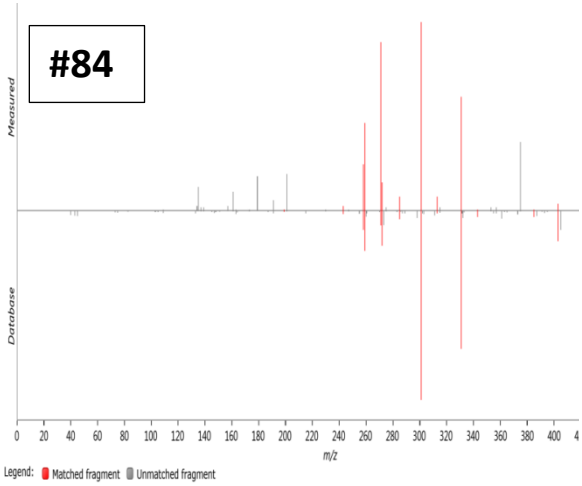


Malate, CID: 20130941, in-house library; [95.1]; {CCMSLIB00005467755}

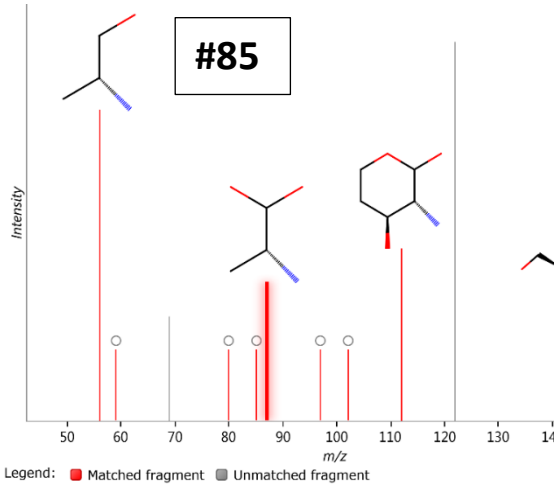


Mangiferin, CID: 5281647, METLIN; [58.2]; {CCMSLIB00005467756}

Muramic acid, CID: 433580, ChEBI; [90.7]; {CCMSLIB00005467710}



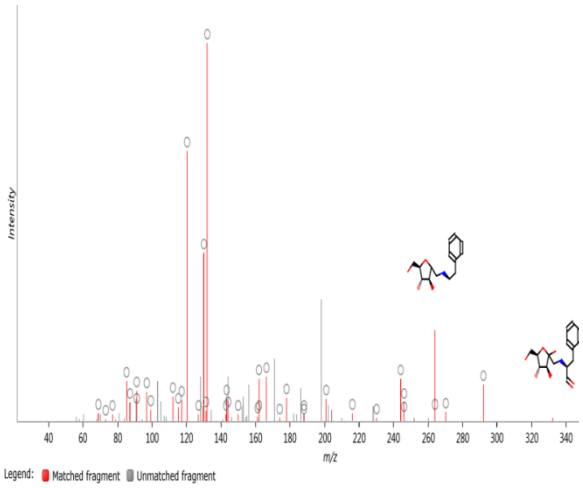
N-(1-Deoxy-1-fructosyl)phenylalanine, CID: 101039148, HMDB; [52.3]; {CCMSLIB00005467711}



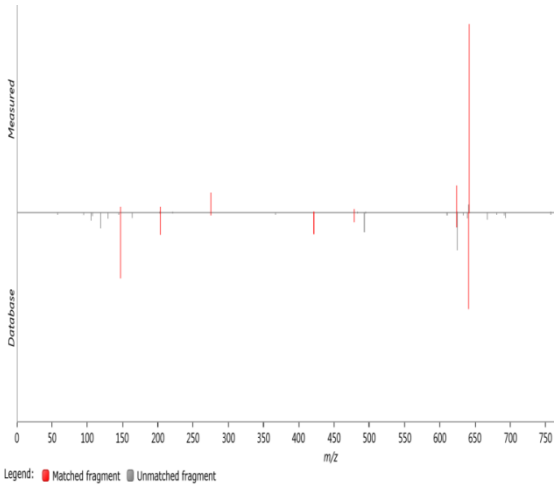
N1,N5,N10,N14-Tetra-trans-p-coumaroylspermine, CID: 9810941, METLIN; [48.9]; {CCMSLIB00005467712}

#86

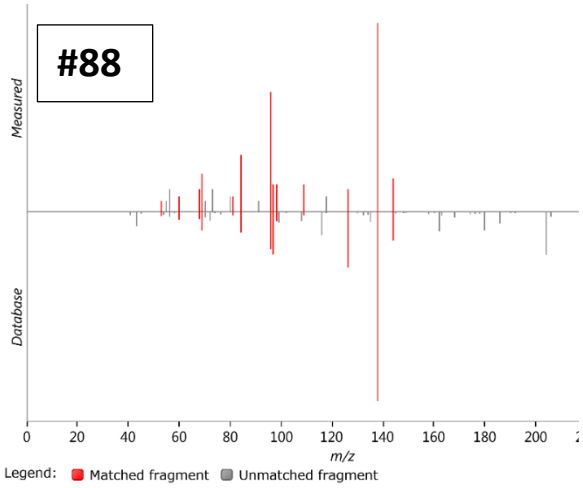
#87



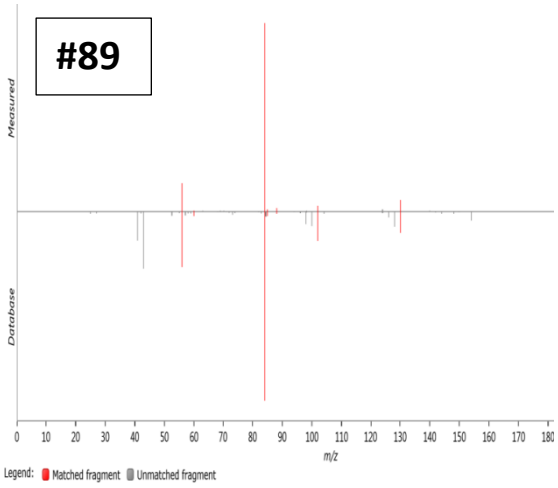
N-Acetyl-D-glucosamine, CID: 899, in-house library; [58.8]; {CCMSLIB00005467713}



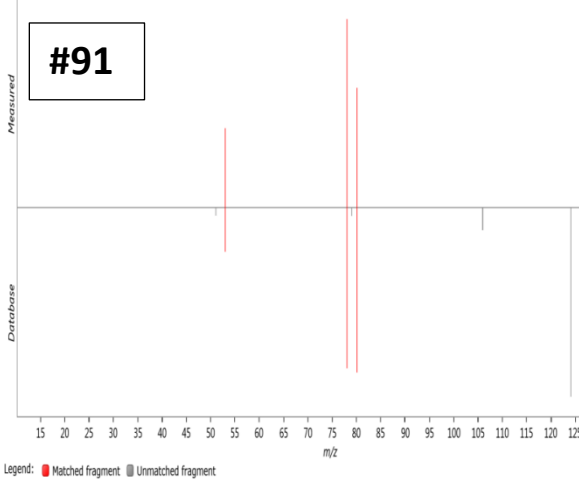
N-Acetyl-L-glutamic acid, CID: 70914, METLIN; [65]; {CCMSLIB00005467714}

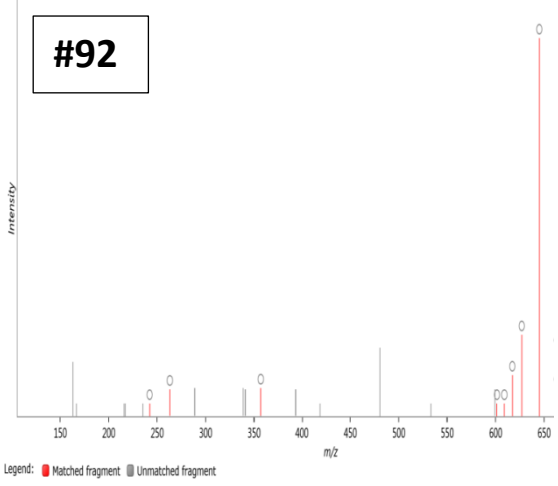


Niacin (Nicotinic acid), CID: 938, METLIN; [88.8]; {CCMSLIB00005467776}

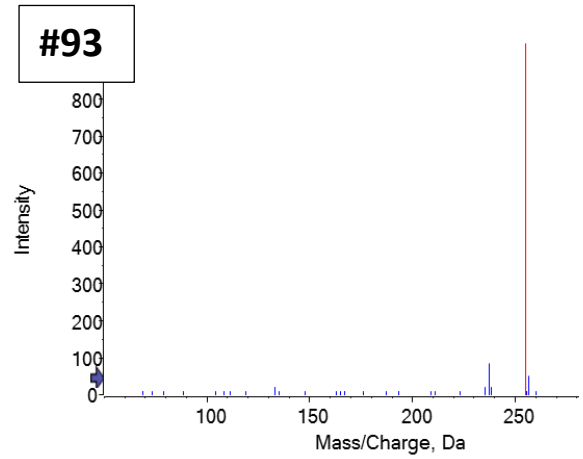


Nomilinic acid 17-glucoside, CID: 444212, HMDB; [45.5]; {CCMSLIB00005467716}

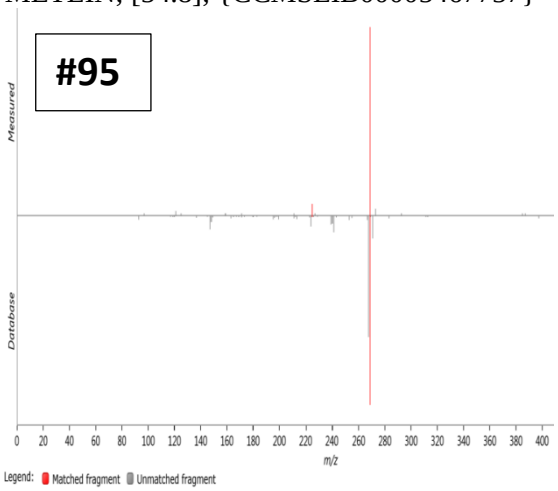




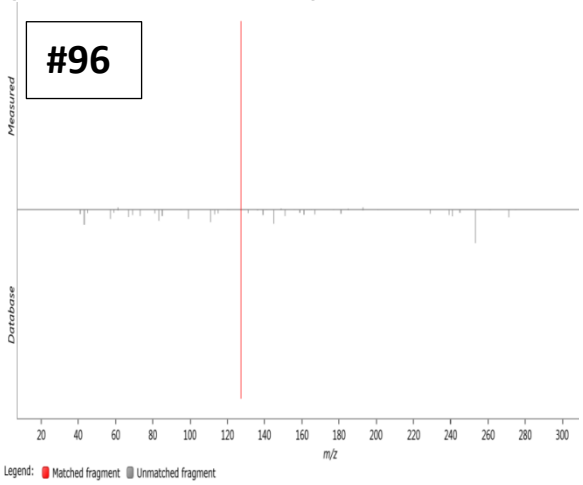
Palmitic acid, matched with in-house library, CID: 985, HMDB; [26.3]



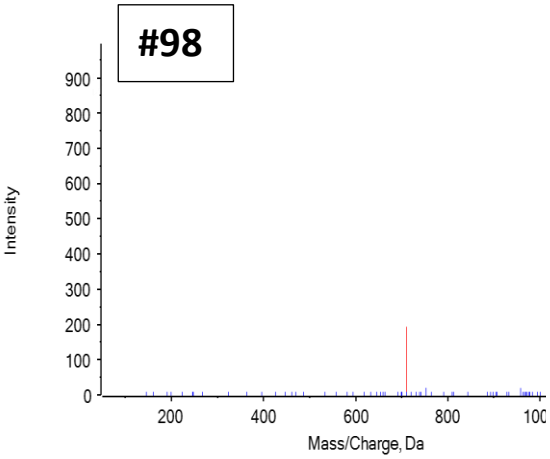
Pelargonidin 3-O-glucoside, CID: 443648, METLIN; [54.8]; {CCMSLIB00005467757}



Phlorin, CID: 476785, METLIN; [37.3]; {CCMSLIB00005467718}

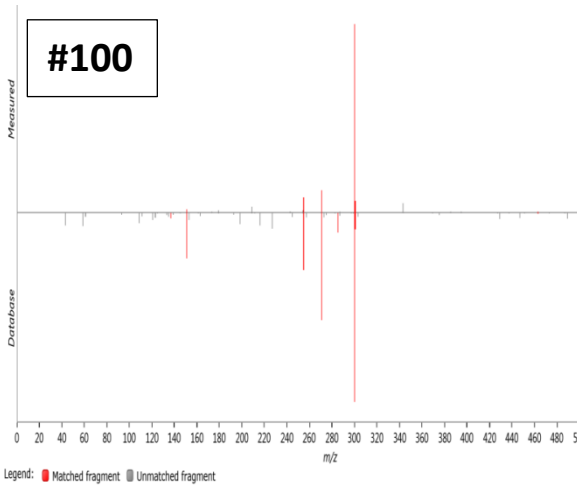


Purgic acid B, CID: 16091605, KNApSAcK; [33.4]; {CCMSLIB00005467775}

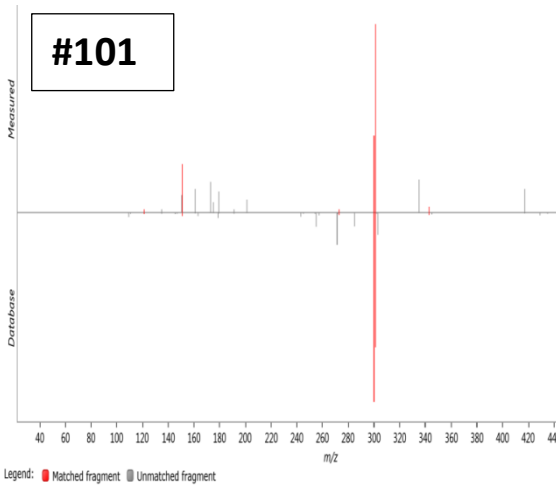


Quercetin 3-(6'-acetylglucoside), CID: 44259187, METLIN; [59.7]; {CCMSLIB00005467759}

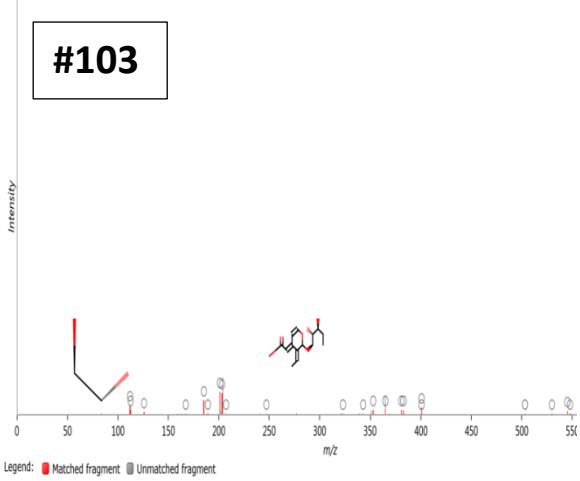
Quercetin 3-O-glucoside, CID: 5280804, METLIN; [62.5]; {CCMSLIB00005467759}



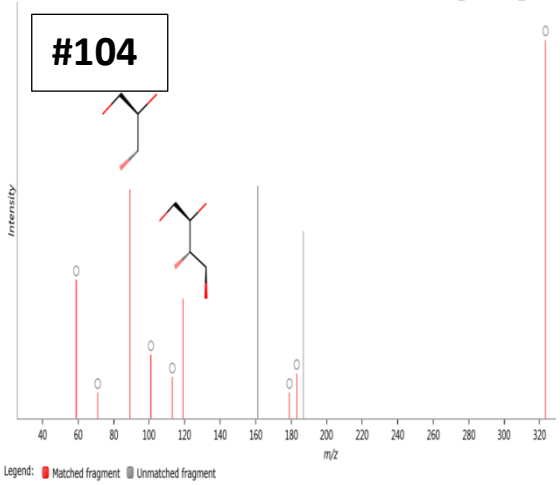
Sambacin, CID: 131752486, HMDB; [34.1]; {CCMSLIB00005467720}



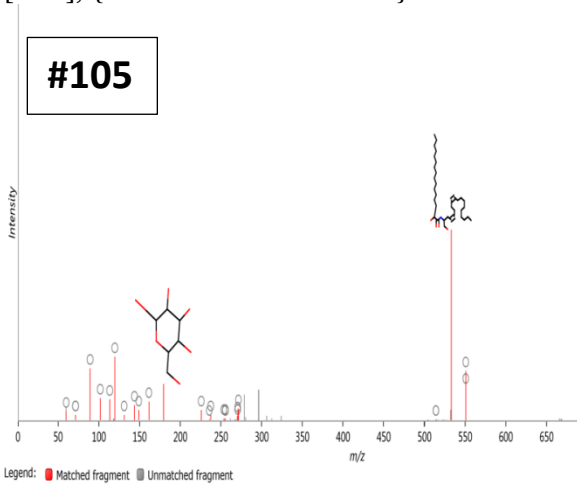
Shanzhiside, CID: 11948668, ChEBI; [81.3]

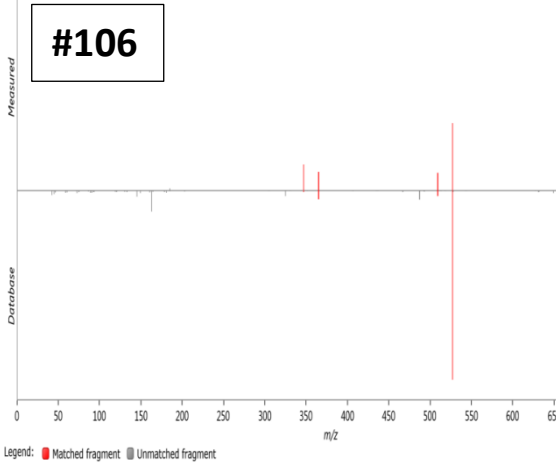


Soyacerebroside, CID: 131751281, HMDB; [37.7]; {CCMSLIB00005467760}



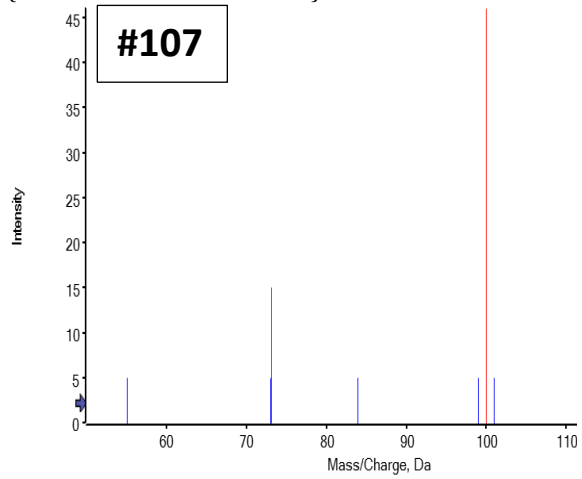
Stachyose, CID: 439531, METLIN; [68.6]; {CCMSLIB00005467721}

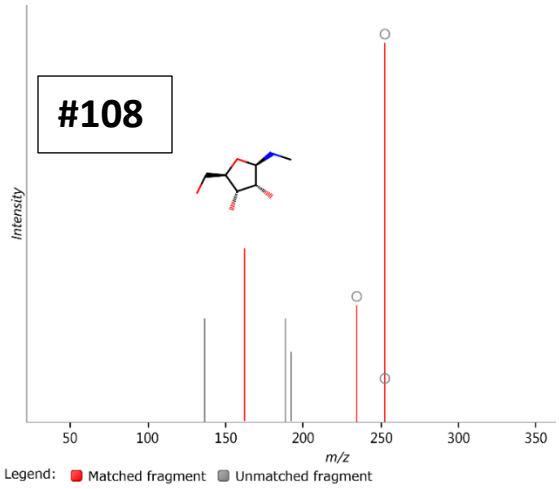




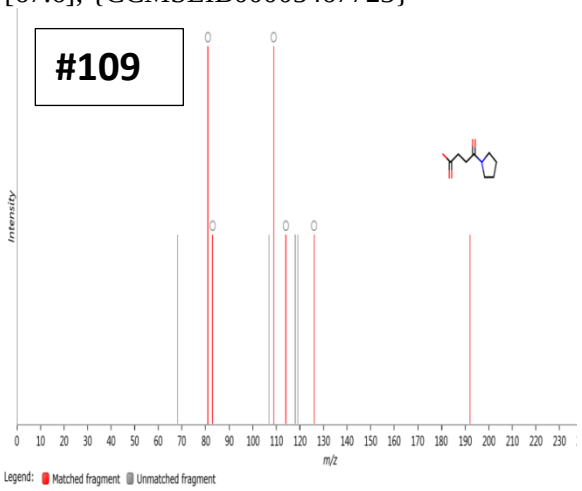
Succinoadenosine, CID: 126969142, METLIN; [57.0]; {CCMSLIB00005467722}

Succinate, matched with in-house library, CID: 1110, HMDB; [55.0]; {CCMSLIB00005467761}

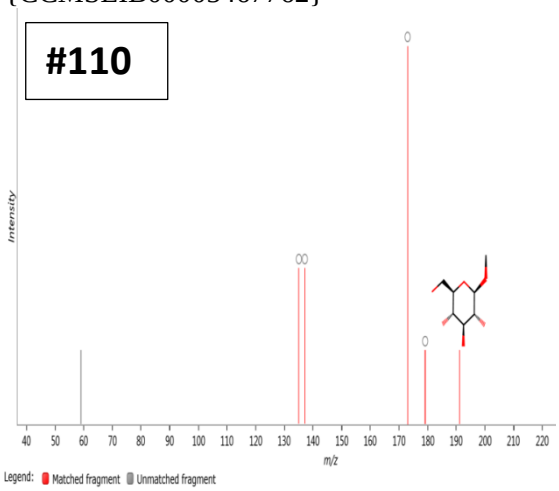




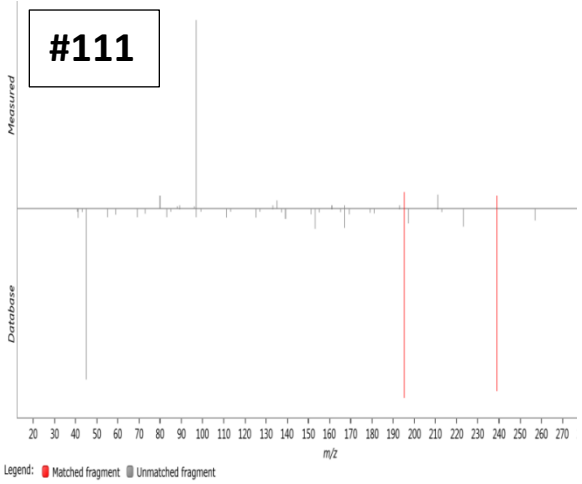
Succinyl-L-proline, CID: 194156, ChEBI; [67.6]; {CCMSLIB00005467723}



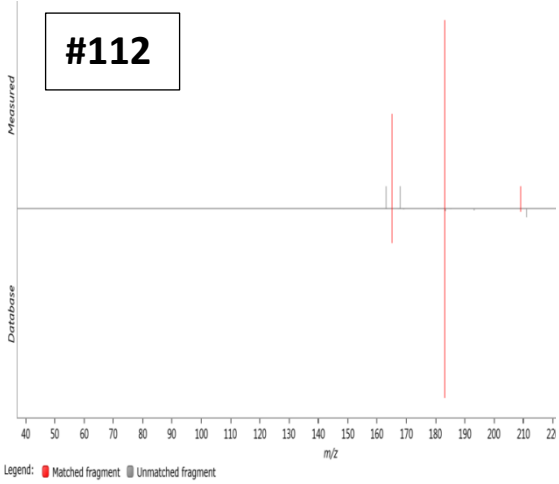
Swertiamarin, CID: 442435, ChEBI; [91.9]; {CCMSLIB00005467762}



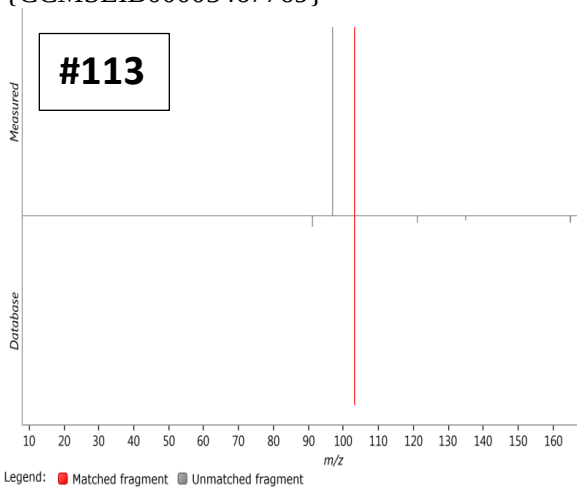
Tetradecanedioic acid, CID: 13185, METLIN; [71.7]; {CCMSLIB00005467764}



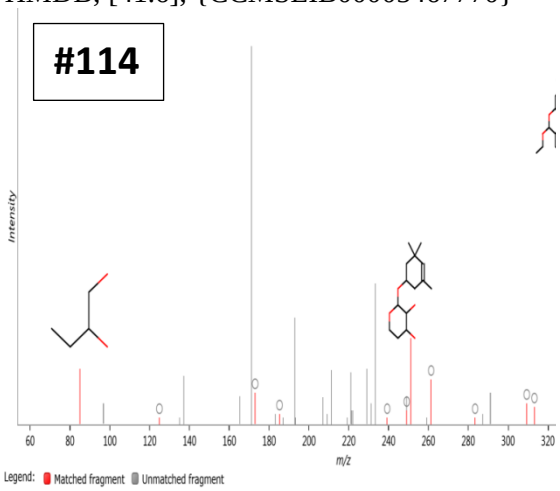
Traumatic Acid, CID: 5283028, METLIN; [77.4]; {CCMSLIB00005467765}



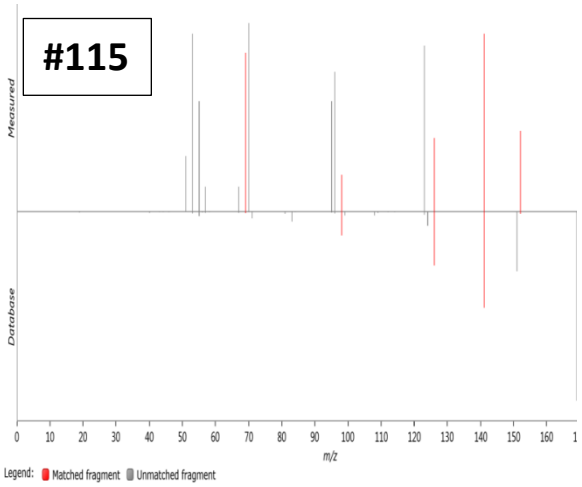
Tropic acid, CID: 10726, METLIN; [69.0]; {CCMSLIB00005467769}



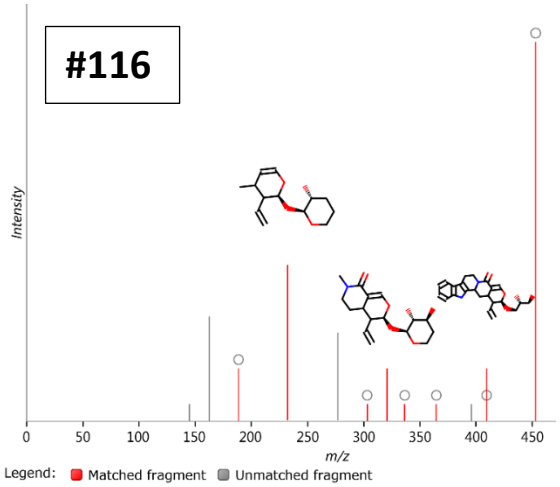
Tsangane L 3-glucoside, CID: 73981648, HMDB; [41.6]; {CCMSLIB00005467770}



Uric acid, CID: 1175, METLIN; [56.4]; {CCMSLIB00005467724}



Vincosamide, CID: 10163855, ChEBI; [79.9]; {CCMSLIB00005467725}



Xanthurenic acid, CID: 5699, METLIN; [76.6]; {CCMSLIB00005467726}

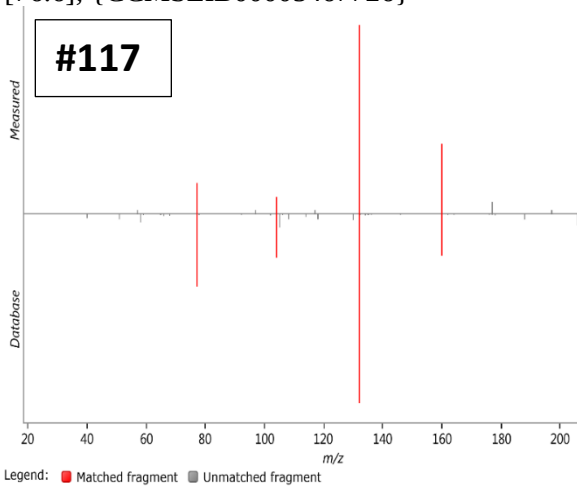
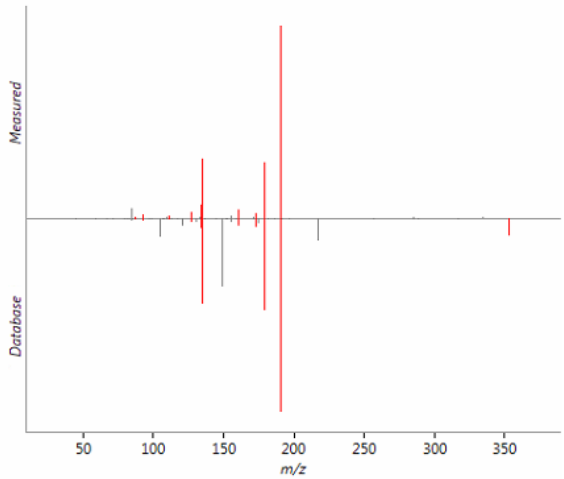
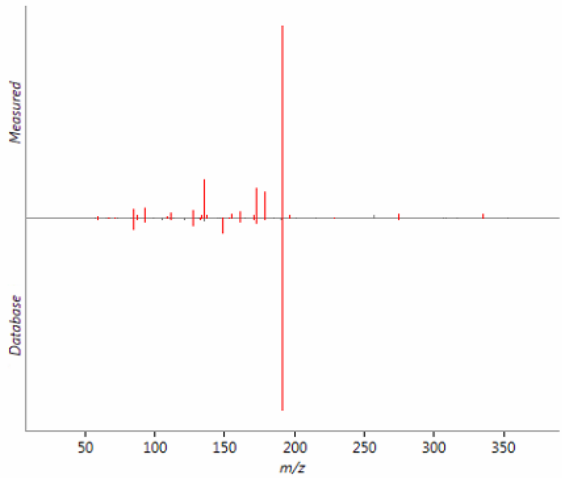


Figure AS8: MS/MS spectra for compounds present in CA water extracts that were identified using authentic standards (L1 annotations). MS/MS score is indicated in [ ].

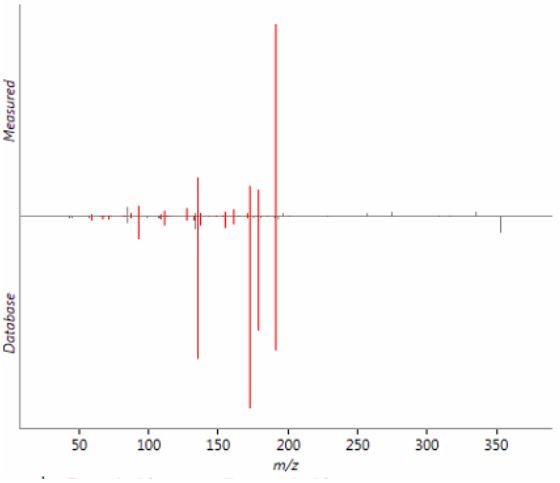
5-O-Caffeoylquinic acid; [81.8]



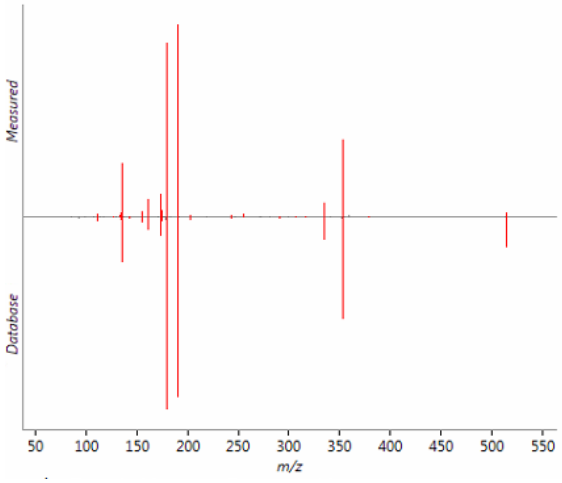
3-O-Caffeoylquinic acid; [87.9]



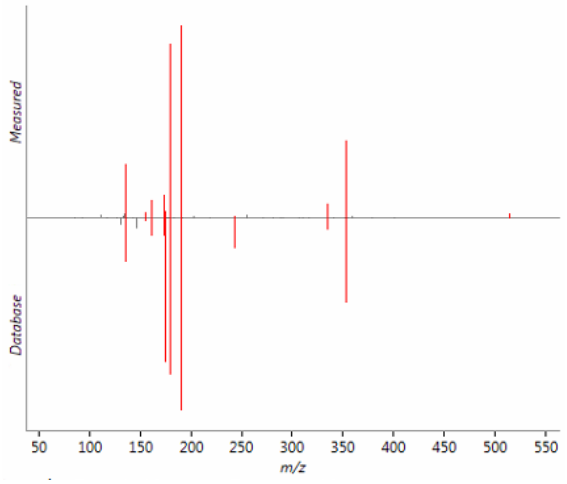
4-O-Caffeoylquinic acid; [79.9]



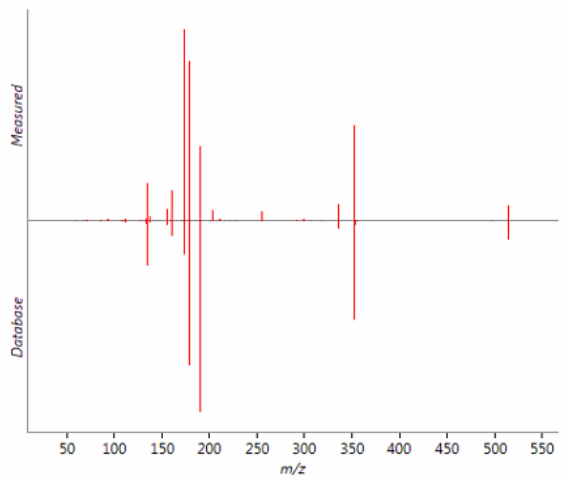
1,5-Dicaffeoylquinic acid; [92.8]



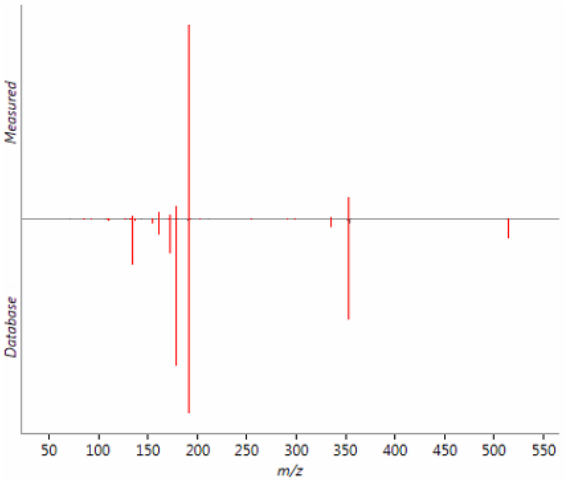
1,3-Dicaffeoylquinic acid; [85.7]



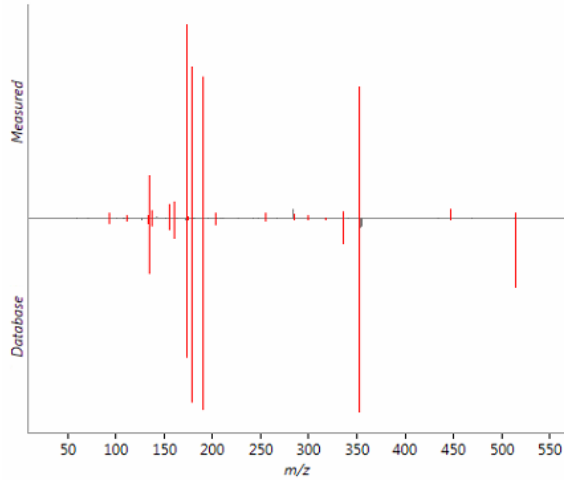
3,4-Dicaffeoylquinic acid; [91.9]



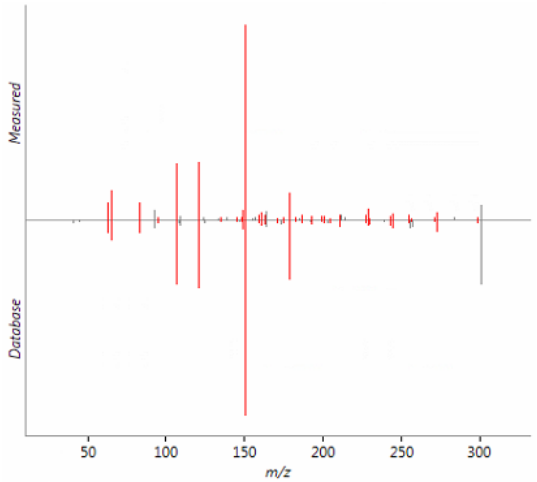
3,5-Dicaffeoylquinic acid; [93.2]



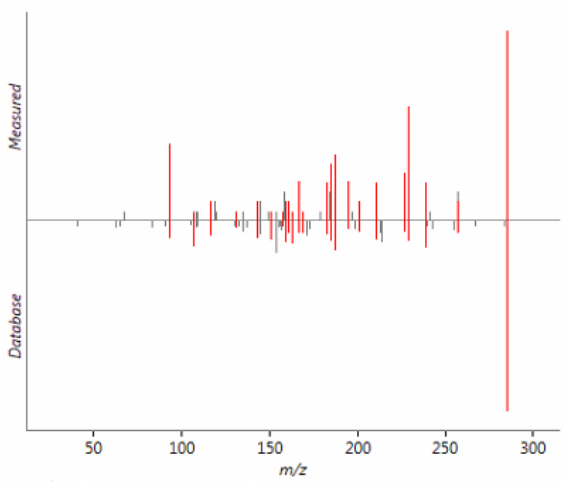
4,5-Dicaffeoylquinic acid; [91.2]



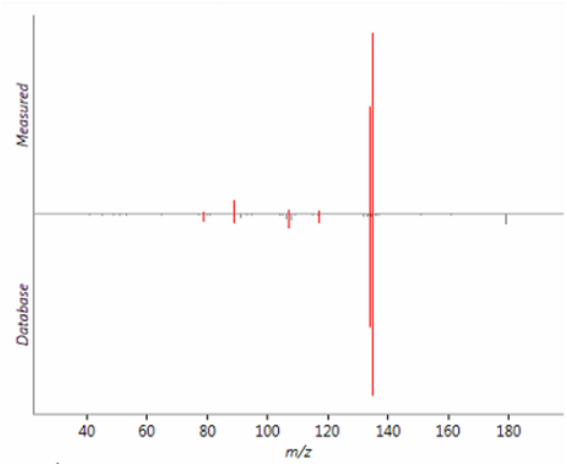
Quercetin; [76.1]



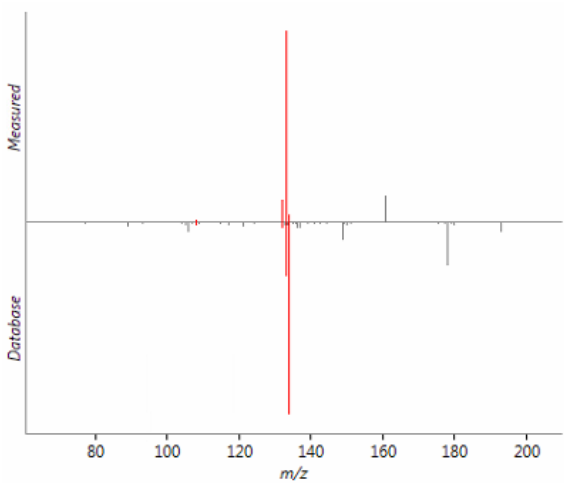
Kaempferol; [92.6]



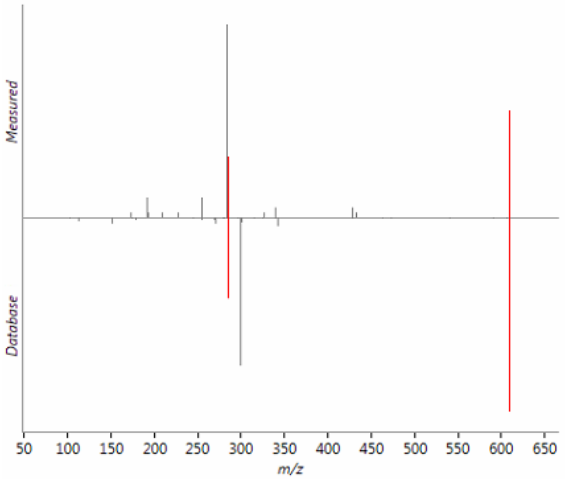
Caffeic acid; [98.2]



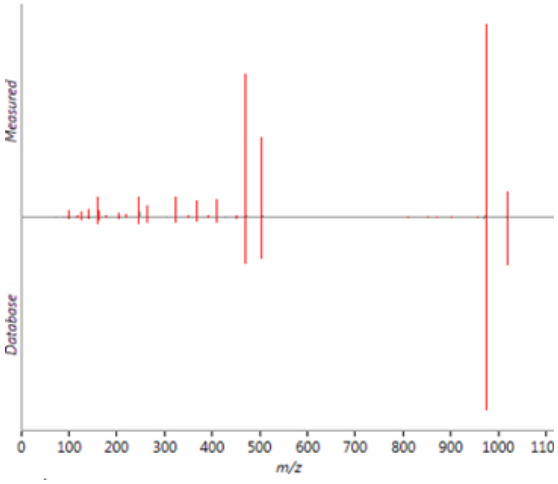
Ferulic acid; [48.4]



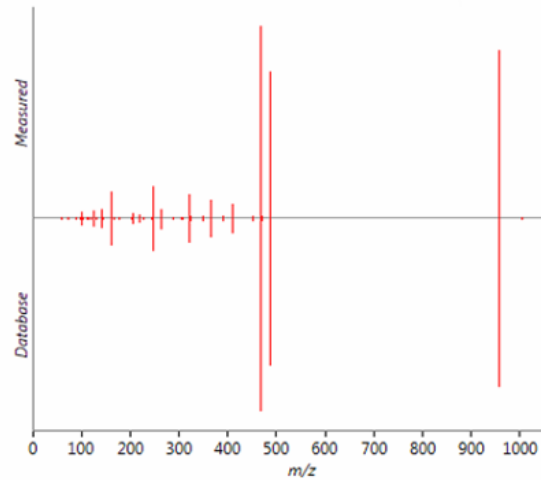
Rutin; [45.4]



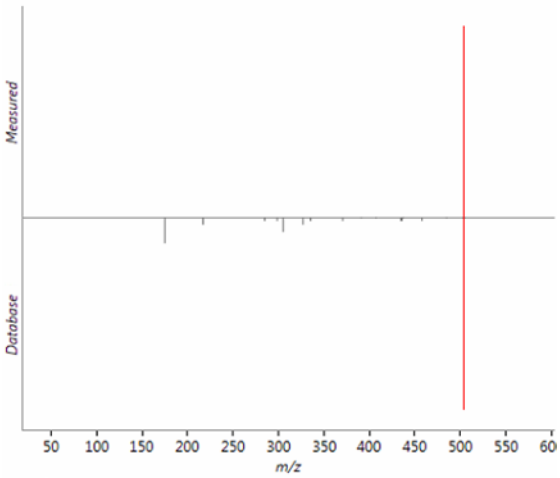
Madecassoside; [97.7]



Asiaticoside; [86.7]



Madecassic Acid; [47.3]



Asiatic acid; [54.7]

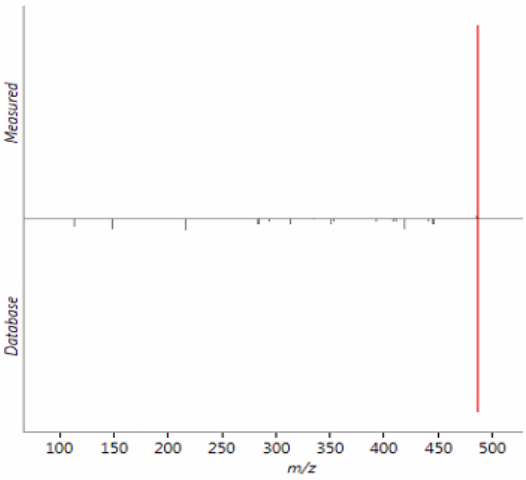


Table AS1 Additional parameters for 117 identified or tentatively identified compounds detected in Centella asiatica aqueous extracts using positive and negative ion mode. Compounds confirmed using authentic standards are shown in bold. For tentatively assigned compounds (L2 annotations) the MS/MS spectral matches that supported the annotation are compiled in Figure AS7.

#[a]	Accepted Description	m/z	RT (min)	Detected Adducts	Δ ppm	Neutral Formula[b]	PubChem CID
1	**1,3-Dicaffeoylquinic acid[c]**	515.1184	17.49	$[M-H]^-$	2.3	$C_{25}H_{24}O_{12}$	6474640

2	1,4-Dicaffeoylquinic acid	515.1184	19.71	$[M-H]^-$	2.3	$C_{25}H_{24}O_{12}$	12358846
3	**1,5-Dicaffeoylquinic acid**	515.1184	17.49	$[M-H]^-$	2.3	$C_{25}H_{24}O_{12}$	5281769
4	12-Oxodihydrophytodienoic acid	317.2090	25.11	$[M+H-H_2O]^+$, $[M+Na]^+$	1.6	$C_{18}H_{30}O_3$	5716902
5	16-hydroxypalmitic acid	271.2271	24.81	$[M-H]^-$	3.1	$C_{16}H_{32}O_3$	10466
6	1-beta-D-Glucopyranosyl-L-tryptophan	367.1501	12.29	$[M+H]^+$, $[M+Na]^+$	1.1	$C_{17}H_{22}N_2O_7$	11772967
7	1-Caffeoyl-5-feruloylquinic acid	553.1296	19.16	$[M+Na]^+$	3.3	$C_{26}H_{26}O_{12}$	121225501
8	2,6-Piperidinedicarboxylic acid	174.0754	5.94	$[M+H]^+$	2.3	$C_7H_{11}NO_4$	557515
9	26-(2-Glucosyl-6-acetylglucosyl]-1,3,11,22-tetrahydroxyergosta-5,24-dien-26-oate	863.4065	23.57	$[M+Na-2H]^-$	1.8	$C_{42}H_{66}O_{17}$	131752817
10	2'-O-Methyladenosine	282.1195	13.45	$[M+H]^+$, $[M+Na]^+$	0.4	$C_{11}H_{15}N_5O_4$	102213
11	2-Pyrrolidone-5-carboxylic acid	130.0491	7.54	$[M+H]^+$	3.8	$C_5H_7NO_3$	499
12	3 Hydroxycoumarin	163.0380	15.27	$[M+H]^+$	4.2	$C_9H_6O_3$	13650
13	**3,4-Dicaffeoylquinic acid**	515.1189	19.11	$[M-H]^-$, $[M+Na-2H]^-$	1.3	$C_{25}H_{24}O_{12}$	5281780
14	3,4-Dihydroxybenzaldehyde	139.0382	14.29	$[M+H]^+$	3.5	$C_7H_6O_3$	8768
15	**3,5-Dicaffeoylquinic acid**	515.1189	19.45	$[M-H]^-$, $[M+Na-2H]^-$	1.3	$C_{25}H_{24}O_{12}$	6474310
16	3,5-Dihydroxy-2-methylphenyl beta-D-glucopyranoside	325.0896	9.73	$[M+Na]^+$	1.3	$C_{13}H_{18}O_8$	46184089
17	3,5-Dihydroxyphenyl 1-O-(6-O-galloyl-beta-D-glucopyranoside)	463.0845	13.58	$[M+H-H_2O]^+$, $[M+Na]^+$	0.1	$C_{19}H_{20}O_{12}$	131752603
18	3-Hydroxy-2-oxo-3-phenylpropanoic acid	163.0375	19.95	$[M+H-H_2O]^+$	4.6	$C_9H_8O_4$	71581094
19	**3-O-Caffeoylquinic acid**	353.0872	15.28	$[M-H]^-$	1.9	$C_{16}H_{18}O_9$	1794427
20	**4,5-Dicaffeoylquinic acid**	515.1184	19.91	$[M-H]^-$	2.3	$C_{25}H_{24}O_{12}$	6474309
21	4-Guanidinobutanoic acid	146.0915	5.45	$[M+H]^+$	4.3	$C_5H_{11}N_3O_2$	25200642
22	4-Hydroxybenzaldehyde	121.0297	17.04	$[M-H]^-$	1.0	$C_7H_6O_2$	126
23	**4-O-Caffeoylquinic acid**	353.0874	11.86	$[M-H]^-$, $[M+Na-2H]^-$	1.4	$C_{16}H_{18}O_{10}$	9798666
24	5'-Deoxy-5'-(methylsulfinyl)adenosine	314.0917	11.71	$[M+H]^+$, $[M+Na]^+$	0.7	$C_{11}H_{15}N_5O_4S$	165114
25	5-Methoxy-L-tryptophan	217.0970	14.58	$[M+H-H_2O]^+$	2.6	$C_{12}H_{14}N_2O_3$	151018
26	5-Methoxysalicylic acid	167.0351	17.38	$[M-H]^-$	0.3	$C_8H_8O_4$	75787
27	**5-O-Caffeoylquinic acid**	353.0874	11.86	$[M-H]^-$, $[M+Na-2H]^-$	1.4	$C_{16}H_{18}O_9$	5280633
28	6-Amino-9H-purine-9-propanoic acid	208.0837	9.09	$[M+H]^+$	5.2	$C_8H_9N_5O_2$	255450
29	6-C-(alpha-L-arabinosyl)-8-C-(beta-L-arabinosyl)apigenin	557.1259	17.40	$[M+Na]^+$	0.8	$C_{25}H_{26}O_{13}$	122391238
30	6-Docosenamide	360.3234	26.15	$[M+H]^+$, $[M+Na]^+$	0.1	$C_{22}H_{43}NO$	44584605
31	6-Oxo-2-piperidinecarboxylic acid	126.0541	12.80	$[M+H-H_2O]^+$	1.1	$C_6H_9NO_3$	3014237

32	8-Acetoxy-4'-methoxypinoresinol 4-glucoside	593.2188	19.59	$[M+H]^+$	6.4	$C_{29}H_{36}O_{13}$	73830447
33	Adenine	136.0612	7.80	$[M+H]^+$	2.2	$C_5H_5N_5$	190
34	Adenosine	312.0945	9.09	$[M-H]^-$, $[M+FA-H]^-$	0.8	$C_{10}H_{13}N_5O_4$	60961
35	Aesculin	363.0688	14.37	$[M+H]^+$, $[M+Na]^+$	1.0	$C_{15}H_{16}O_9$	5281417
36	Apimaysin	541.1355	4.67	$[M-H_2O-H]^-$	1.0	$C_{27}H_{28}O_{13}$	101920411
37	**Asiatic acid**	487.3421	21.91	$[M-H]^-$	1.8	$C_{30}H_{48}O_5$	119034
38	**Asiaticoside**	957.5101	20.50	$[M-H]^-$	3.8	$C_{48}H_{78}O_{19}$	24721205
39	Astragalin	447.0916	20.07	$[M-H]^-$	3.9	$C_{21}H_{20}O_{11}$	5282102
40	b-Chlorogenin 3-[4'-(2'-glucosyl-3'-xylosylglucosyl)galactoside]	1049.5217	21.01	$[M-H]^-$	4.0	$C_{50}H_{82}O_{23}$	74193143
41	Betaine	118.0853	4.00	$[M+H]^+$	5.7	$C_5H_{11}NO_2$	247
42	**Caffeic acid**	179.0350	15.86	$[M-H]^-$	0.3	$C_9H_8O_4$	689043
43	cAMP	328.0448	7.85	$[M-H]^-$	1.6	$C_{10}H_{12}N_5O_6P$	6076
44	Caprylic acid	125.0977	21.76	$[M-H_2O-H]^-$	5.4	$C_8H_{16}O_2$	379
45	Carlosic acid methyl ester	223.0615	17.32	$[M-H_2O-H]^-$	2.1	$C_{11}H_{14}O_6$	122391261
46	**Catechin**	289.0710	14.48	$[M-H]^-$	2.9	$C_{15}H_{14}O_6$	9064
47	Choline	104.1066	6.46	$[M+H]^+$	1.0	$C_5H_{13}NO$	305
48	Choline O-Sulfate	184.0627	6.73	$[M+H]^+$	4.5	$C_5H_{13}NO_4S$	486
49	Citric acid	191.0193	5.82	$[M-H]^-$	2.6	$C_6H_8O_7$	19782904
50	Coumarin	147.0432	18.38	$[M+H]^+$	3.9	$C_9H_6O_2$	323
51	Cytosine	112.0498	5.18	$[M+H]^+$	4.1	$C_4H_5N_3O$	597
52	D-1-[(3-Carboxypropyl)amino]-1-deoxyfructose	266.1233	4.85	$[M+H]^+$	0.6	$C_{10}H_{19}NO_7$	131752417
53	Daucic acid	203.0189	6.08	$[M-H]^-$, $[M+Na-2H]^-$	4.4	$C_7H_8O_7$	5316316
54	Deoxyfructosazine	305.1343	5.34	$[M+H]^+$	0.8	$C_{12}H_{20}N_2O_7$	73452
55	Digalacturonate	369.0673	4.51	$[M-H]^-$	0.6	$C_{12}H_{18}O_{13}$	439694
56	Dihydroactinidiolide	181.1219	23.39	$[M+H]^+$, $[M+Na]^+$	0.7	$C_{11}H_{16}O_2$	27209
57	**Dihydrocaffeic acid**	181.0501	17.72	$[M-H]^-$	3.3	$C_9H_{10}O_4$	348154
58	**Dihydroferulic acid**	195.0665	19.02	$[M-H]^-$	0.7	$C_{10}H_{12}O_4$	14340
59	Dysolenticin B	451.3197	23.72	$[M+H-H_2O]^+$, $[M+H]^+$	1.5	$C_{30}H_{42}O_3$	56601655
60	Enicoflavine	210.0771	17.99	$[M-H]^-$	0.7	$C_{10}H_{13}NO_4$	5281564
61	**Epicatechin**	289.0715	16.80	$[M-H]^-$	1.2	$C_{15}H_{14}O_6$	72276
62	**Epigallocatechin**	305.0666	13.48	$[M-H]^-$	0.5	$C_{15}H_{14}O_7$	72277
63	**Ferulic acid**	193.0500	19.54	$[M-H]^-$	3.7	$C_{10}H_{10}O_4$	445858
64	Folinic acid	474.1733	17.01	$[M+H]^+$	0.9	$C_{20}H_{23}N_7O_7$	6006
65	Furaneol 4-(6-malonylglucoside)	397.0745	12.70	$[M+Na-2H]^-$	2.7	$C_{15}H_{20}O_{11}$	131750900
66	Gentiopicroside	379.0999	18.91	$[M+Na]^+$	0.5	$C_{16}H_{20}O_9$	88708

67	Ginkgoic acid	345.2435	26.26	$[M-H]^-$	0.2	$C_{22}H_{34}O_3$	5281858
68	Ginsenoyne K	299.1613	24.98	$[M+Na]^+$	0.8	$C_{17}H_{24}O_3$	15736266
69	Glabraoside A	597.1607	22.37	$[M-H]^-$, $[M+Na-2H]^-$	1.2	$C_{30}H_{30}O_{13}$	102393599
70	Guanosine	282.0841	7.83	$[M-H]^-$	1.3	$C_{10}H_{13}N_5O_5$	6802
71	**Isoferulic acid**	193.0497	20.52	$[M-H]^-$	5.2	$C_{10}H_{10}O_4$	736186
72	Isovalerylglucuronide	301.0896	10.01	$[M+Na]^+$	1.5	$C_{11}H_{18}O_8$	137383
73	**Kaempferol**	285.0404	21.96	$[M-H]^-$	0.5	$C_{15}H_{10}O_6$	5280863
74	Kuwanon Y	605.1794	7.24	$[M+Na]^+$	2.4	$C_{34}H_{30}O_9$	14334307
75	Kynurenic acid	190.0491	17.62	$[M+H]^+$	2.6	$C_{10}H_7NO_3$	3845
76	L-Arginine	175.1184	4.34	$[M+H]^+$	1.5	$C_6H_{14}N_4O_2$	28782
77	Linustatin	410.1663	8.22	$[M+H]^+$	2.2	$C_{16}H_{27}NO_{11}$	119301
78	Longicamphenylone	207.1742	24.81	$[M+H-H_2O]^+$, $[M+H]^+$, $[M+Na]^+$	0.7	$C_{14}H_{22}O$	91747202
79	Longifolenaldehyde	221.1897	23.87	$[M+H]^+$	0.0	$C_{15}H_{24}O$	565584
80	L-Ribulose	149.0451	5.19	$[M-H]^-$, $[M+FA-H]^-$	3.5	$C_5H_{10}O_5$	644111
81	**Madecassic Acid**	503.3367	23.79	$[M-H]^-$	2.3	$C_{30}H_{48}O_6$	73412
82	**Madecassoside**	973.5017	21.35	$[M-H]^-$	0.3	$C_{48}H_{78}O_{20}$	91885295
83	Malate	133.0143	5.32	$[M-H_2O-H]^-$, $[M-H]^-$	0.2	$C_4H_6O_5$	20130941
84	Mangiferin	421.0759	16.80	$[M-H]^-$	4.3	$C_{19}H_{18}O_{11}$	5281647
85	Muramic acid	234.0967	4.65	$[M+H-H_2O]^+$, $[M+H]^+$	0.9	$C_9H_{17}NO_7$	433580
86	N-(1-Deoxy-1-fructosyl)phenylalanine	328.1393	10.27	$[M+H-H_2O]^+$, $[M+H]^+$, $[M+Na]^+$	1.5	$C_{15}H_{21}NO_7$	101039148
87	N1,N5,N10,N14-Tetra-trans-p-coumaroylspermine	809.3504	23.17	$[M+H]^+$, $[M+Na]^+$	1.8	$C_{46}H_{50}N_4O_8$	9810941
88	N-Acetyl-D-glucosamine	222.0961	5.14	$[M+H]^+$	3.8	$C_8H_{15}NO_6$	899
89	N-Acetyl-L-glutamic acid	212.0525	7.54	$[M+H-H_2O]^+$, $[M+H]^+$, $[M+Na]^+$	1.0	$C_7H_{11}NO_5$	70914
90	**Naringin**	579.1710	20.12	$[M-H]^-$	1.7	$C_{27}H_{32}O_{14}$	442428
91	Niacin (Nicotinic acid)	124.0383	5.32	$[M+H]^+$	5.9	$C_6H_5NO_2$	938
92	Nomilinic acid 17-glucoside	695.2913	20.15	$[M+H-H_2O]^+$	1.6	$C_{34}H_{48}O_{16}$	444212
93	Palmitic acid	255.2323	25.44	$[M-H]^-$	2.8	$C_{16}H_{32}O_2$	985
94	Pantothenic Acid	220.1176	10.51	$[M+H]^+$, $[M+Na]^+$	0.3	$C_9H_{17}NO_5$	6613
95	Pelargonidin 3-O-glucoside	431.0967	22.47	$[M-H]^-$	4.0	$C_{21}H_{20}O_{10}$	443648
96	Phlorin	311.0739	14.27	$[M+H]^+$, $[M+Na]^+$	1.2	$C_{12}H_{16}O_8$	476785
97	Phosphocholine	184.0727	5.27	$[M+H]^+$	1.9	$C_5H_{14}NO_4P$	1014
98	Purgic acid B,(-)-Purgic acid B	1179.5675	22.81	$[M-H]^-$	2.0	$C_{52}H_{92}O_{29}$	16091605
99	**Quercetin**	301.0349	21.10	$[M-H]^-$	1.8	$C_{15}H_{10}O_7$	5280343
100	Quercetin 3-(6'-acetylglucoside)	505.0996	20.30	$[M-H]^-$	1.5	$C_{23}H_{22}O_{13}$	44259187

101	Quercetin 3-O-glucoside	463.0884	18.77	$[M-H]^-$	0.3	$C_{21}H_{20}O_{12}$	5280804
102	**Rutin**	609.1465	18.88	$[M-H]^-$	0.5	$C_{27}H_{30}O_{16}$	5280805
103	Sambacin	563.2089	19.43	$[M+Na]^+$	1.4	$C_{26}H_{36}O_{12}$	131752486
104	shanzhiside	391.1215	14.13	$[M-H]^-$	8.1	$C_{16}H_{24}O_{11}$	11948668
105	Soyacerebroside I	758.5428	26.17	$[M-H]^-$, $[M+FA-H]^-$	0.9	$C_{40}H_{75}NO_9$	131751281
106	Stachyose	689.2104	4.58	$[M+Na]^+$	0.7	$C_{24}H_{42}O_{21}$	439531
107	Succinate	117.0195	7.39	$[M-H]^-$	0.8	$C_4H_6O_4$	1110
108	Succinoadenosine	384.1156	11.33	$[M+H]^+$	2.3	$C_{14}H_{17}N_5O_8$	126969142
109	Succinyl-L-proline	238.0706	7.31	$[M+Na]^+$	7.3	$C_9H_{13}NO_5$	194156
110	Swertiamarin	395.0978	17.90	$[M+Na-2H]^-$	3.8	$C_{16}H_{22}O_{10}$	442435
111	Tetradecanedioic acid	279.1623	23.65	$[M-H]^-$, $[M+Na-2H]^-$	8.3	$C_{14}H_{26}O_4$	13185
112	Traumatic Acid	227.1285	22.67	$[M-H]^-$	2.0	$C_{12}H_{20}O_4$	5283028
113	Tropic acid	165.0551	17.88	$[M-H]^-$	4.2	$C_9H_{10}O_3$	10726
114	Tsangane L 3-glucoside	395.2039	22.28	$[M+Na-2H]^-$	3.9	$C_{19}H_{34}O_7$	73981648
115	Uric acid	169.0344	5.66	$[M+H]^+$	5.6	$C_5H_4N_4O_3$	1175
116	Vincosamide	499.2067	20.70	$[M+H]^+$	1.0	$C_{26}H_{30}N_2O_8$	10163855
117	Xanthurenic acid	206.0439	14.77	$[M+H]^+$	2.9	$C_{10}H_7NO_4$	5699

[a] Matched with Table A.T3.7
[b] Neutral formula obtained for annotated compounds
[c] **Bold -** Compounds confirmed using authentic standards

Table AS2*:* Linear regression functions calculated for 24 precursor ions. Calibration curves where stablished in negative ion mode with correlation factor r > 0.990. Calibration curves were prepared from mixed standard solutions containing 0.0, 0.005, 0.01, 0.05, 0.10, 0.50, 1.00, 5.0 and 10.0 mg/L. Compounds are sorted by retention time.

Compound	[M-H]- (exact mass), XIC	Regression function	r
5-O-Caffeoylquinic acid	353.0875	A^a=(4.082 C^b + 0.0502)*10^6	0.9999
Epigallocatechin	305.0666	A=(2.745 C - 0.0063)*10^6	0.9999
Catechin	289.0710	A=(3.514 C + 0.0954)*10^6	0.9990
Dihydrocaffeic acid	181.0511	A=(2.661 C + 0.0249)*10^6	0.9998
4-O-Caffeoylquinic acid	353.0872	A=(3.312 C + 0.0331)*10^6	0.9998
3-O-Caffeoylquinic acid	353.0869	A=(4.072 C + 0.0682)*10^6	0.9994
Caffeic acid	179.0349	A=(4.516 C + 0.0664)*10^6	0.9996
Epicatechin	289.0715	A=(3.524 C + 0.0731)*10^6	0.9994
1,5-Dicaffeoylquinic acid	515.1189	A=(9.217 C + 0.0927)*10^6	0.9999
1,3-Dicaffeoylquinic acid	515.1189	A=(9.217 C + 0.0927)*10^6	0.9999

Rutin	609.1452	A=(5.673 C + 0.0103)*10^6	0.9999
Dihydroferulic acid	195.0665	A=(3.234 C + 0.0581)*10^6	0.9984
3,4-Dicaffeoylquinic acid	515.1190	A=(3.451 C + 0.0377)*10^6	1.0000
3,5-Dicaffeoylquinic acid	515.1190	A=(3.045 C + 0.0300)*10^6	1.0000
Ferulic acid	193.0510	A=(1.551 C + 0.0241)*10^6	0.9992
4,5-Dicaffeoylquinic acid	515.1192	A=(5.546 C + 0.0570)*10^6	0.9999
Naringin	579.1718	A=(4.730 C + 0.0510)*10^6	0.9999
Iso Ferulic acid	193.0510	A=(0.417 C + 0.0032)*10^6	0.9999
Quercetin	301.0340	A=(7.879 C + 0.2305)*10^6	0.9975
Madecassoside	973.5018	A=(0.647 C + 0.0401)*10^6	0.9997
Asiaticoside	957.5087	A=(0.506 C + 0.0107)*10^6	0.9996
Kaempferol	285.0404	A=(12.150 C + 0.3562)*10^6	0.9916
Madecassic Acid	503.3370	A=(1.232 C + 0.3609)*10^6	0.9895
Asiatic acid	487.3428	A=(1.937 C + 0.1587)*10^6	0.9866

[a]A: peak area for the quantifier
[b]c: concentration, mg l^{-1}.

Appendix B

Supplemental Material for Chapter 4.

Figure BS1: Volcano plot comparing the differences in the lipid profile across 306 lipids in heart tissues due to the storage conditions of OCT and liquid nitrogen. The adjusted p-value and fold change were obtained using false discovery rate (FDR) and the mean ratio of lipid levels respectively across OCT and liquid nitrogen storage conditions. 14 lipids with FDR ≤ 0.1 (statistically significant) are represented in red of which only PE(42:2) had FDR ≤ 0.05 whereas the lipids with FDR > 0.1 are represented in blue. The two storage

conditions accounted for 4.6% less than 5%) significant differences in the lipidomic profile comprising of 306 lipids found in the heart tissues.

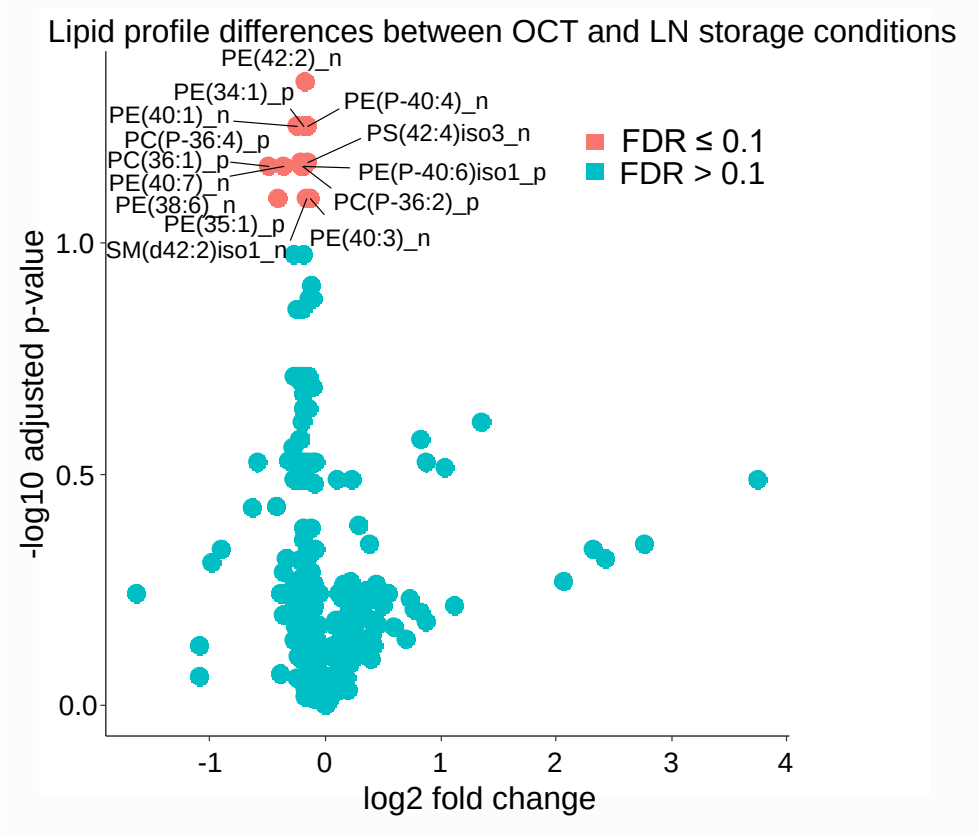


Table BS1: Summary of fold change, p-value, false discovery rate adjusted p-value for 306 lipid species identified across heart tissue samples for OCT and liquid nitrogen storage conditions using reversed phase LC-MS/MS in both ESI+ (_p) and ESI- (_n) mode.

Compound	p-value	FDR	FC	log2FC	log10p
10-HETE_n	0.52	0.78	1.29	0.36	0.11
11,12-Epoxyeicosatrienoic acid_n	0.10	0.45	1.30	0.38	0.35
13E-Docosenamide_p	0.85	0.94	0.95	-0.08	0.03
13-Hexadecenoic acid_n	0.23	0.59	1.66	0.73	0.23
13-HODE_n	0.10	0.46	5.02	2.33	0.34
14-methyl palmitic acid_n	0.16	0.55	1.12	0.16	0.26

15,16-Dihydroxyoctadecanoic acid_n	0.03	0.28	0.82	-0.28	0.56
19-HETE_n	0.12	0.48	5.38	2.43	0.32
20-HETE ethanolamide_p	0.07	0.37	0.64	-0.63	0.43
2-hydroxydocosanoic acid_n	0.05	0.30	0.67	-0.59	0.52
3-O-Sulfogalactosylceramide (d36:1)_p	0.04	0.30	1.83	0.87	0.52
3-O-Sulfogalactosylceramide (d42:1)_p	0.03	0.27	1.78	0.83	0.57
3-O-Sulfogalactosylceramide (d42:2)_p	0.05	0.31	2.07	1.05	0.51
4,8,12,15,19-Docosapentaenoic acid_n	0.28	0.63	1.77	0.83	0.20
C16 Sphinganine_p	0.15	0.54	0.83	-0.27	0.27
CE(17:0)_n	0.47	0.75	1.12	0.17	0.13
Cer(d32:0)_p	0.46	0.74	0.47	-1.09	0.13
Cer(d36:1)_p	0.69	0.89	0.93	-0.10	0.05
Cer(d38:0)_p	0.79	0.93	0.96	-0.07	0.03
Cer(d38:1)_n	0.94	0.98	1.01	0.01	0.01
Cer(d40:0)iso1_n	0.55	0.80	0.88	-0.18	0.09
Cer(d40:2)iso2_n	0.82	0.93	1.03	0.05	0.03
Cer(d41:1)_n	0.31	0.66	1.07	0.10	0.18
Cer(d44:1)_n	0.75	0.92	1.06	0.09	0.03
Ceramide (d34:1)iso1_n	0.60	0.84	0.95	-0.08	0.07
Ceramide (d34:1)iso2_n	0.85	0.94	0.98	-0.03	0.03
Ceramide (d36:1)iso1_n	0.50	0.77	0.91	-0.13	0.11
Ceramide (d36:1)iso2_n	0.82	0.93	0.98	-0.03	0.03
Ceramide (d40:1)_n	0.32	0.66	1.09	0.12	0.18
Ceramide (d42:1)_n	0.75	0.92	0.94	-0.09	0.03
Coenzyme Q10_p	0.93	0.97	0.99	-0.01	0.01
Coenzyme Q9_p	0.39	0.71	0.88	-0.19	0.15
Dehydrophytosphingosine_p	0.26	0.61	1.16	0.22	0.21
DG(40:8)_n	0.54	0.80	1.32	0.40	0.10

DG(38:5)iso1_n	0.47	0.75	1.34	0.43	0.13
DG(33:2)_n	0.26	0.61	2.18	1.13	0.21
DG(36:2)iso1_p	0.67	0.88	1.06	0.09	0.06
DG(36:2)iso2_p	0.92	0.97	0.95	-0.07	0.01
DG(36:4)_p	0.81	0.93	1.09	0.13	0.03
DG(37:5)_n	0.41	0.72	1.62	0.70	0.14
DG(38:4)iso1_p	0.99	1.00	1.00	0.00	0.00
DG(38:4)iso2_p	0.86	0.95	1.03	0.04	0.02
DG(38:4)iso3_p	0.06	0.33	1.07	0.10	0.49
DG(38:5)iso2_p	0.70	0.90	1.12	0.16	0.05
Docosahexaenoic acid_n	0.36	0.68	1.51	0.59	0.17
Galabiosylceramide (d36:1)_n	0.80	0.93	0.98	-0.02	0.03
Glucosylceramide (d43:1)_n	0.01	0.14	0.84	-0.25	0.86
GPSer(40:5)_n	0.42	0.73	0.96	-0.07	0.14
Hexadecadienoic acid_n	0.26	0.62	1.70	0.77	0.21
Lactosylceramide (d34:1)_n	0.52	0.78	0.85	-0.23	0.11
Linoleyl carnitine_n	0.19	0.57	0.32	-1.64	0.24
LysoPA(16:0)_p	0.66	0.88	1.14	0.19	0.06
LysoPA(18:0)_p	0.65	0.87	1.06	0.09	0.06
LysoPC(15:0)_n	0.42	0.73	1.21	0.28	0.14
LysoPC(16:0)iso1_p	0.22	0.59	1.11	0.15	0.23
LysoPC(16:0)iso2_p	0.20	0.57	0.77	-0.37	0.24
LysoPC(17:0)iso1_n	0.50	0.77	1.15	0.21	0.11
LysoPC(17:0)iso2_n	0.17	0.56	0.87	-0.21	0.26
LysoPC(18:0)iso1_p	0.56	0.81	1.09	0.12	0.09
LysoPC(18:0)iso2_p	0.76	0.92	0.97	-0.05	0.03
LysoPC(18:2)_p	0.19	0.57	1.46	0.55	0.24
LysoPC(20:4)_p	0.27	0.62	1.24	0.31	0.21
LysoPC(22:6)_p	0.58	0.83	0.91	-0.14	0.08
LysoPC(P-16:0)_p	0.43	0.73	1.26	0.33	0.14
LysoPC(P-18:0)_p	0.42	0.73	1.23	0.30	0.14

LysoPC(P-18:1)_p	0.34	0.67	1.37	0.46	0.17
LysoPE(16:0)iso1_n	0.98	1.00	1.00	0.01	0.00
LysoPE(16:0)iso2_n	0.58	0.83	0.92	-0.12	0.08
LysoPE(18:0)iso1_n	0.45	0.74	1.21	0.28	0.13
LysoPE(18:0)iso2_n	0.85	0.94	1.05	0.07	0.03
LysoPE(18:0)iso3_p	0.77	0.92	0.94	-0.08	0.03
LysoPE(18:0)iso4_p	0.90	0.97	1.03	0.04	0.01
LysoPE(18:1)_p	0.19	0.57	1.38	0.46	0.24
LysoPE(18:1)iso1_n	0.77	0.92	1.08	0.11	0.03
LysoPE(18:1)iso2_n	0.22	0.59	0.81	-0.30	0.23
LysoPE(18:2)iso1_n	0.31	0.65	1.29	0.37	0.19
LysoPE(18:2)iso2_n	0.77	0.92	1.09	0.12	0.03
LysoPE(20:0)_n	0.65	0.87	1.04	0.06	0.06
LysoPE(20:3)_p	0.28	0.63	1.25	0.32	0.20
LysoPE(20:4)_n	0.24	0.61	1.42	0.50	0.21
LysoPE(20:5)_n	0.17	0.57	1.27	0.35	0.25
LysoPE(22:0)iso1_n	0.19	0.57	0.89	-0.18	0.24
LysoPE(22:0)iso2_n	0.60	0.84	0.96	-0.06	0.08
LysoPE(22:1)iso1_n	0.37	0.70	1.11	0.15	0.16
LysoPE(22:1)iso2_n	0.55	0.81	1.16	0.22	0.09
LysoPE(22:2)_n	0.23	0.59	1.33	0.42	0.23
LysoPE(22:5)_n	0.25	0.61	0.88	-0.18	0.21
LysoPE(22:6)_n	0.30	0.64	1.25	0.32	0.19
Methyltestosterone_n	0.08	0.41	1.23	0.30	0.39
MG(13:0)_n	0.15	0.54	1.17	0.22	0.27
MG(15:0)_n	0.06	0.33	1.17	0.23	0.49
MG(16:0)_n	0.84	0.94	0.98	-0.03	0.03
MG(18:0)_n	0.67	0.88	1.03	0.05	0.06
MG(i-17:0)iso1_n	0.47	0.75	1.05	0.07	0.13
MG(i-17:0)iso2_n	0.21	0.57	0.95	-0.08	0.24
N-hexadecanoylsphinganine-1-phosphocholine_p	0.02	0.19	0.87	-0.20	0.71

N-Stearoyl GABA_p	0.78	0.93	1.15	0.20	0.03
Octadecanamideiso1_p	0.65	0.87	0.47	-1.09	0.06
Octadecanamideiso2_p	0.40	0.72	1.23	0.30	0.14
Oleoylcarnitine_p	0.76	0.92	1.09	0.12	0.03
PA(34:1)_p	0.84	0.94	1.04	0.06	0.03
PA(36:2)_n	0.09	0.41	0.88	-0.19	0.38
PA(36:3)_p	0.32	0.66	1.34	0.42	0.18
PA(38:4)_n	0.14	0.53	0.87	-0.20	0.28
PA(40:2)_n	0.62	0.86	0.76	-0.39	0.07
PA(44:10)_n	0.29	0.64	0.85	-0.24	0.19
PA(48:1)_p	0.99	1.00	1.00	0.00	0.00
Palmitic amide_p	0.10	0.45	6.83	2.77	0.35
Palmitoleoyl Ethanolamide_p	0.32	0.66	1.83	0.88	0.18
Palmitoylcarnitine_p	0.95	0.98	1.02	0.03	0.01
PC(30:0)_p	0.29	0.63	0.82	-0.29	0.20
PC(31:0)_p	0.12	0.48	0.80	-0.33	0.32
PC(32:0)_p	0.05	0.30	0.85	-0.23	0.52
PC(34:1)_p	0.12	0.48	0.79	-0.34	0.32
PC(34:2)iso1_n	0.16	0.54	4.19	2.07	0.27
PC(34:2)iso2_p	0.90	0.97	0.99	-0.02	0.02
PC(34:3)_p	0.34	0.67	0.92	-0.12	0.17
PC(36:1)_p	0.00	0.07	0.71	-0.49	1.17
PC(36:2)_p	0.11	0.46	0.95	-0.08	0.34
PC(36:3)_p	0.02	0.20	0.86	-0.22	0.70
PC(36:4)iso1_p	0.26	0.61	0.91	-0.13	0.21
PC(36:4)iso2_p	0.70	0.90	0.94	-0.09	0.05
PC(36:5)_p	0.54	0.80	0.87	-0.21	0.10
PC(38:2)_n	0.22	0.59	0.87	-0.21	0.23
PC(38:6)_p	0.19	0.57	0.88	-0.18	0.24
PC(40:5)_n	0.21	0.57	0.77	-0.38	0.24
PC(44:9)_n	0.02	0.19	0.85	-0.23	0.71

PC(DiMe(20,8))_n	1.00	1.00	1.00	0.00	0.00
PC(o-44:4)_p	0.72	0.91	0.95	-0.07	0.04
PC(P-30:0)_p	0.39	0.71	0.91	-0.13	0.15
PC(P-34:1)_p	0.78	0.93	0.95	-0.07	0.03
PC(P-34:2)iso1_p	0.39	0.71	0.92	-0.12	0.15
PC(P-34:2)iso2_n	0.20	0.57	1.08	0.12	0.24
PC(P-36:2)_p	0.00	0.07	0.87	-0.20	1.17
PC(P-36:3)_n	0.59	0.84	0.95	-0.07	0.08
PC(P-36:4)iso1_n	0.77	0.92	0.98	-0.02	0.03
PC(P-36:4)iso2_n	0.19	0.57	0.87	-0.21	0.24
PC(P-36:4)iso3_p	0.00	0.07	0.86	-0.21	1.17
PC(P-36:5)_n	0.06	0.33	0.83	-0.27	0.49
PC(P-38:4)_n	0.15	0.54	0.88	-0.19	0.27
PC(P-38:5)_p	0.03	0.24	0.87	-0.19	0.61
PC(P-38:6)_p	0.16	0.55	0.90	-0.15	0.26
PC(P-42:6)_n	0.30	0.64	0.78	-0.36	0.19
PE(33:2)_n	0.20	0.57	1.17	0.23	0.24
PE(34:1)_p	0.00	0.06	0.88	-0.19	1.25
PE(34:2)iso1_n	0.06	0.33	0.90	-0.16	0.49
PE(34:2)iso2_p	0.82	0.93	1.06	0.09	0.03
PE(35:0)_p	0.36	0.68	0.89	-0.16	0.17
PE(35:1)_p	0.00	0.08	0.76	-0.40	1.10
PE(36:0)_n	0.02	0.21	0.88	-0.18	0.67
PE(36:1)_p	0.68	0.88	1.05	0.07	0.06
PE(36:2)iso1_n	0.02	0.23	0.88	-0.18	0.64
PE(36:2)iso2_n	0.21	0.57	0.96	-0.07	0.24
PE(36:2)iso3_n	0.02	0.23	0.91	-0.14	0.64
PE(36:3)_n	0.01	0.13	0.90	-0.14	0.88
PE(36:4)_p	0.81	0.93	1.02	0.03	0.03
PE(36:4)iso1_n	0.13	0.51	0.77	-0.37	0.29
PE(36:4)iso2_n	0.02	0.21	0.89	-0.17	0.69

PE(36:5)_n	0.20	0.57	0.97	-0.05	0.24
PE(37:2)_n	0.18	0.57	0.82	-0.28	0.24
PE(38:0)_n	0.04	0.30	0.90	-0.16	0.52
PE(38:1)iso1_n	0.62	0.86	0.96	-0.05	0.07
PE(38:1)iso2_n	0.23	0.59	0.87	-0.20	0.23
PE(38:2)iso1_n	0.82	0.93	1.01	0.02	0.03
PE(38:2)iso2_n	0.47	0.75	0.92	-0.11	0.13
PE(38:3)iso1_n	0.87	0.95	1.02	0.03	0.02
PE(38:3)iso2_n	0.20	0.57	0.91	-0.13	0.24
PE(38:4)iso1_n	0.05	0.33	0.83	-0.27	0.49
PE(38:4)iso2_n	0.76	0.92	0.98	-0.03	0.03
PE(38:4)iso3_n	0.26	0.61	0.84	-0.24	0.21
PE(38:5)iso1_n	0.01	0.11	0.88	-0.18	0.98
PE(38:5)iso2_n	0.45	0.74	1.09	0.13	0.13
PE(38:6)_n	0.00	0.07	0.86	-0.21	1.17
PE(40:1)_n	0.00	0.06	0.84	-0.24	1.25
PE(40:2)iso1_n	0.06	0.33	0.90	-0.15	0.49
PE(40:2)iso2_n	0.04	0.30	0.87	-0.19	0.52
PE(40:2)iso3_n	0.21	0.57	0.85	-0.23	0.24
PE(40:3)_n	0.00	0.08	0.91	-0.13	1.10
PE(40:4)iso1_n	0.10	0.46	0.90	-0.15	0.34
PE(40:4)iso2_n	0.45	0.74	0.91	-0.13	0.13
PE(40:5)iso1_n	0.80	0.93	0.95	-0.07	0.03
PE(40:5)iso2_n	0.43	0.73	0.92	-0.12	0.14
PE(40:6)iso1_p	0.14	0.52	0.93	-0.11	0.29
PE(40:6)iso2_n	0.06	0.33	0.94	-0.09	0.48
PE(40:6)iso3_n	0.01	0.19	0.89	-0.17	0.71
PE(40:7)_n	0.00	0.07	0.78	-0.36	1.17
PE(42:2)_n	0.00	0.04	0.89	-0.17	1.35
PE(42:3)_n	0.01	0.11	0.83	-0.26	0.98
PE(42:4)_n	0.09	0.44	0.88	-0.19	0.36

PE(42:5)_n	0.03	0.27	0.86	-0.22	0.57
PE(42:8)_n	0.78	0.93	0.90	-0.16	0.03
PE(44:2)_n	0.79	0.93	0.90	-0.16	0.03
PE(44:5)_n	0.33	0.66	0.85	-0.23	0.18
PE(44:6)_n	0.40	0.72	0.92	-0.13	0.14
PE(48:2)_p	0.98	1.00	1.00	0.00	0.00
PE(DiMe(20,8))_n	0.76	0.92	1.10	0.14	0.03
PE(P-16:0e)_n	0.27	0.62	0.89	-0.16	0.21
PE(P-32:0)_p	0.49	0.76	1.08	0.11	0.12
PE(P-34:1)_n	0.49	0.77	0.95	-0.07	0.12
PE(P-34:2)_n	0.56	0.81	0.92	-0.12	0.09
PE(P-34:3)_n	0.40	0.72	0.91	-0.14	0.14
PE(P-36:0)_n	0.17	0.56	0.88	-0.18	0.26
PE(P-36:2)iso1_n	0.39	0.71	0.93	-0.11	0.15
PE(P-36:2)iso2_n	0.16	0.55	0.94	-0.09	0.26
PE(P-36:3)iso1_n	0.41	0.72	0.87	-0.20	0.14
PE(P-36:3)iso2_n	0.92	0.97	0.98	-0.03	0.01
PE(P-36:4)_n	0.15	0.54	0.92	-0.12	0.27
PE(P-36:5)_n	0.13	0.51	0.90	-0.15	0.30
PE(P-38:0)_p	0.02	0.19	0.83	-0.27	0.71
PE(P-38:1)iso1_n	0.36	0.68	0.84	-0.26	0.17
PE(P-38:1)iso2_n	0.07	0.37	0.75	-0.42	0.43
PE(P-38:2)iso1_n	0.95	0.98	1.00	0.00	0.01
PE(P-38:2)iso2_n	0.06	0.33	0.88	-0.19	0.49
PE(P-38:3)_n	0.09	0.41	0.92	-0.12	0.38
PE(P-38:4)_n	0.01	0.12	0.92	-0.12	0.91
PE(P-38:5)_n	0.04	0.30	0.93	-0.10	0.52
PE(P-38:6)_n	0.34	0.67	0.94	-0.10	0.17
PE(P-40:2)_n	0.01	0.13	0.93	-0.10	0.88
PE(P-40:4)_n	0.00	0.06	0.90	-0.15	1.25
PE(P-40:5)_n	0.35	0.67	0.97	-0.04	0.17

PE(P-40:6)iso1_p	0.00	0.07	0.88	-0.18	1.17
PE(P-40:6)iso2_p	0.64	0.87	0.97	-0.05	0.06
PE(P-40:6)iso3_p	0.98	1.00	1.00	0.00	0.00
PE(P-40:7)_n	0.30	0.64	1.15	0.20	0.19
PE-NMe(38:2)_n	0.25	0.61	0.94	-0.09	0.21
PE-NMe(38:3)_n	0.32	0.66	0.89	-0.16	0.18
PE-NMe(38:4)_n	0.89	0.96	0.97	-0.05	0.02
PE-NMe(42:5)_n	0.87	0.95	0.97	-0.04	0.02
PE-NMe2(36:2)_n	0.58	0.83	0.95	-0.08	0.08
PE-NMe2(40:5)_p	0.60	0.84	0.90	-0.15	0.07
PE-NMe2(42:8)_n	0.80	0.93	1.09	0.13	0.03
Pentadecanoic acid_n	0.53	0.79	1.05	0.07	0.10
Pentadecanoylcarnitine_n	0.12	0.49	0.51	-0.98	0.31
PG(34:1)_n	0.02	0.19	0.91	-0.14	0.71
PG(34:2)_n	0.42	0.73	0.83	-0.27	0.14
PG(34:3)iso1_n	0.74	0.92	0.88	-0.18	0.03
PG(34:3)iso2_n	0.06	0.33	0.85	-0.24	0.49
PG(36:4)_n	0.87	0.95	0.95	-0.07	0.02
PG(38:4)_p	0.34	0.67	1.13	0.18	0.17
PI(36:3)_n	0.91	0.97	1.02	0.02	0.01
PI(38:4)_n	0.94	0.98	0.98	-0.03	0.01
PI(40:4)_n	0.35	0.67	1.06	0.08	0.17
PI(40:6)_n	0.64	0.87	1.03	0.04	0.06
PIP(34:1)_p	0.21	0.58	1.24	0.31	0.24
PIP(38:2)_p	0.44	0.73	1.13	0.18	0.13
PS(28:0)iso1_p	0.63	0.86	0.97	-0.05	0.07
PS(28:0)iso2_p	0.89	0.96	0.88	-0.18	0.02
PS(28:0)iso3_p	0.05	0.33	13.44	3.75	0.49
PS(28:0)iso4_p	0.93	0.97	0.94	-0.09	0.01
PS(29:0)_p	0.30	0.65	1.15	0.20	0.19
PS(30:0)_p	0.89	0.96	1.01	0.02	0.02

PS(32:3)_p	0.38	0.71	1.33	0.41	0.15
PS(35:1)_n	0.91	0.97	0.97	-0.04	0.01
PS(36:0)_n	0.64	0.87	1.03	0.04	0.06
PS(36:2)_p	0.42	0.73	0.84	-0.25	0.14
PS(36:3)_n	0.49	0.77	0.90	-0.15	0.12
PS(37:0)_n	0.52	0.78	1.03	0.04	0.11
PS(38:0)_n	0.15	0.54	0.90	-0.15	0.27
PS(38:1)_n	0.72	0.91	1.04	0.06	0.04
PS(38:2)_n	0.04	0.30	0.80	-0.32	0.53
PS(38:3)_p	0.51	0.78	1.03	0.04	0.11
PS(38:5)iso1_n	0.25	0.61	0.89	-0.17	0.21
PS(38:5)iso2_n	0.75	0.92	1.03	0.05	0.03
PS(39:0)_n	0.75	0.92	0.97	-0.05	0.03
PS(40:1)_n	0.05	0.30	0.95	-0.08	0.52
PS(40:3)_n	0.01	0.14	0.87	-0.20	0.86
PS(40:6)_n	0.51	0.78	0.96	-0.06	0.11
PS(42:1)_n	0.11	0.47	0.91	-0.13	0.32
PS(42:2)_n	0.40	0.72	0.91	-0.14	0.14
PS(42:3)_n	0.32	0.66	0.92	-0.12	0.18
PS(42:4)iso1_n	0.27	0.62	0.92	-0.12	0.21
PS(42:4)iso2_n	0.05	0.30	0.84	-0.26	0.52
PS(42:4)iso3_n	0.00	0.07	0.90	-0.15	1.17
PS(42:5)iso1_n	0.28	0.63	0.87	-0.20	0.20
PS(42:5)iso2_n	0.69	0.89	0.94	-0.10	0.05
PS(42:5)iso3_n	0.06	0.33	0.86	-0.22	0.49
PS(44:2)_n	0.02	0.21	0.93	-0.10	0.69
PS(44:3)_n	0.56	0.81	0.92	-0.11	0.09
PS(44:4)iso1_n	0.67	0.88	0.94	-0.09	0.06
PS(44:4)iso2_n	0.29	0.64	0.90	-0.15	0.19
PS(44:5)_n	0.27	0.62	0.93	-0.11	0.21
PS(44:6)_n	0.25	0.61	0.93	-0.11	0.21

PS(46:1)_n	0.66	0.87	0.85	-0.24	0.06
PS(DiMe(24,8))_n	0.82	0.93	0.98	-0.03	0.03
SM(d34:1)_p	0.13	0.51	0.89	-0.16	0.29
SM(d36:1)_p	0.04	0.30	0.87	-0.19	0.52
SM(d38:1)_p	0.25	0.61	0.90	-0.15	0.21
SM(d40:1)_p	0.03	0.27	0.86	-0.21	0.57
SM(d40:2)_p	0.12	0.49	0.86	-0.22	0.31
SM(d41:1)_p	0.19	0.57	0.89	-0.16	0.24
SM(d42:1)_p	0.46	0.74	0.88	-0.19	0.13
SM(d42:2)iso1_n	0.00	0.08	0.89	-0.17	1.10
SM(d42:2)iso2_n	0.82	0.93	1.02	0.03	0.03
SM(d44:2)iso1_n	0.26	0.62	0.89	-0.16	0.21
SM(d44:2)iso2_n	0.44	0.73	0.91	-0.14	0.13
Stearic Acid ethyl ester_n	0.98	1.00	1.00	0.00	0.00
Tetracosanoic acid_n	0.50	0.77	0.91	-0.14	0.11
TG(52:4)_n	0.96	0.99	1.00	-0.01	0.01
TG(i-47:0)_p	0.21	0.57	0.83	-0.27	0.24
TG(i-48:0)_p	0.33	0.67	1.20	0.26	0.17
Trihexosylceramide (d34:1)_p	0.16	0.55	1.36	0.44	0.26
Trihexosylceramide (d42:1)_n	0.77	0.92	1.07	0.09	0.03
Trihexosylceramide (d42:2)_n	0.93	0.97	1.02	0.03	0.01
Trihexosylceramide (d43:1)_n	0.65	0.87	1.09	0.12	0.06
Undecanedioic acid_n	0.10	0.46	0.54	-0.90	0.34
Vaccenyl carnitine_p	0.03	0.24	2.56	1.35	0.61

Abbreviation index

PC: phosphatidylcholines

PE: phosphatidylethanolamines

SM: sphingomyelins

MG: monoacylglycerols

DG: diacylglycerols

TG: triacylglycerols

PC(P), PE(P), LysoPC/PE(P): plasmalogens

Appendix C

Supplemental Material for Chapter 5.

Figure CS1: Metabolic pathways identified by pathway analysis using KEGG identifiers based on 254 metabolites detected in MDA-MB-231 cells using Metaboanalyst v5.0. a. Tricarboxylic acid cycle b. glutamine-glutamate pathway c. alanine-aspartate-glutamate metabolic pathway d. glutathione metabolic pathway.

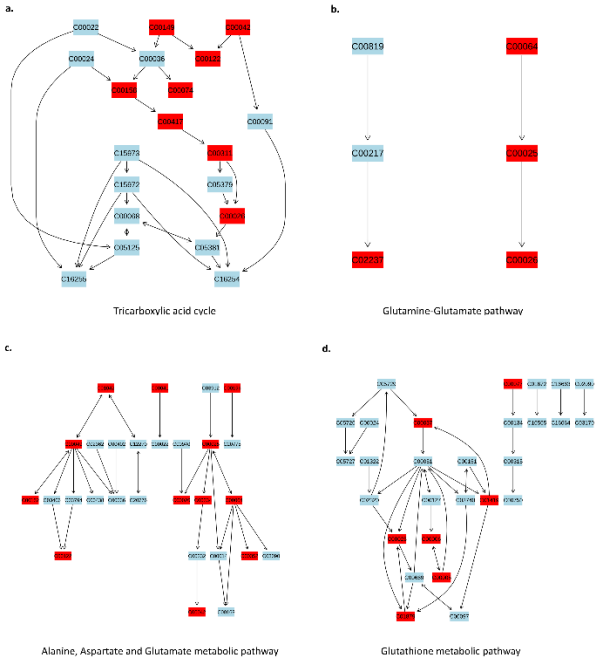


Table CS1: Metabolite name, theoretical mass, observed mass, Δ ppm and retention time (RT) in minutes for 166 metabolites identified using ESI positive ion $[M+H]^+$ in MDA-MB 231 cell extracts.

Metabolite	Theoretical mass	Observed mass	Δ ppm	RT (Min)

NAD	663.10912	663.1091	0.060322	9.47
Glutamine	146.06914	146.0691	0.273843	8.75
Inosine 5'-Phosphate	348.0471	348.0471	0.114927	9.95
Taurine	125.01467	125.0146	0.319962	7.78
Inosine	268.08077	268.0807	0.149209	1.9
Isoleucine	131.09463	131.0946	0.305123	8.1
Glutamic Acid	147.05316	147.0531	0.27201	9.49
Beta-Alanine	89.04768	89.04764	0.449198	8.11
Dihydroorotate	158.03276	158.0327	0.253112	2.2
Alanine	89.04768	89.04764	0.449198	8.11
Proline	115.06333	115.0633	0.347635	6.89
Thymine	126.04293	126.0429	0.317352	1.11
Carnosine	226.10659	226.1066	0.176908	11.09
Nicotinamide	122.04801	122.048	0.32774	1.35
Phenylalanine	165.07898	165.0789	0.242308	4.52
Uracil	112.02728	112.0272	0.357056	2.23
Aspartate	133.03751	133.0375	0.300667	9.99
Creatine	131.06948	131.0694	0.305182	8.1
Guanosine	283.09167	283.0916	0.141297	5.15
Lysine	146.10553	146.1055	0.273775	15.94
Tyrosine	181.07389	181.0739	0.220904	6.7
Asparagine	132.05349	132.0535	0.302908	8.1
Valine	117.07898	117.0789	0.34165	6.11
Guanine	151.04941	151.0494	0.264814	5.15
Deoxyadenosine Monophosphate	331.06817	331.0681	0.120821	8.27
Nicotinamide Mononucleotide	334.05661	334.0566	0.11974	9.62
Thiourea	76.00952	76.00948	0.52625	1.1
Cys-Gly	178.04122	178.0412	0.224667	28.83
Guanidinoacetate	117.05383	117.0538	0.341723	6.11
Creatinine	113.05891	113.0589	0.353798	1.21
Aconitate	174.01644	174.0164	0.229863	28.82
3-Dehydroshikimate	172.03717	172.0371	0.232508	28.82
Leucine	131.09463	131.0946	0.305123	8.1
Betaine	117.07898	117.0789	0.34165	6.11
Tryptophan	204.08988	204.0898	0.195992	7.25

5-Oxo-D-Proline	129.04259	129.0426	0.309975	9.49
Methylguandine	73.064	73.06396	0.547465	4.64
Caffeine	194.08038	194.0803	0.2061	1.13
Methyladenosine	281.11241	281.1124	0.142292	1.1
4-Hydroxyproline	131.05824	131.0582	0.305208	8.1
Urocanate	138.04293	138.0429	0.289765	6.27
5-Oxo-L-Proline	129.04259	129.0426	0.309975	9.49
Lysine	146.10553	146.1055	0.273775	15.94
Cytidine 5'-Diphosphocholine	489.11459	489.1146	0.08178	10.63
Phosphonoacetate	139.98746	139.9874	0.28574	28.97
Arginine	174.11168	174.1116	0.229738	15.36
4-Hydroxy-L-Proline	131.05824	131.0582	0.305208	8.1
Xanthosine	284.07569	284.0757	0.140808	6.83
Thyrotropin Releasing Hormone	362.17025	362.1702	0.110445	1.14
2'-Deoxyguanosine 5'-Monophosphate	347.06309	347.0631	0.115253	9.43
Nicotinamide Hypoxanthine Dinucleotide	665.1004	665.1004	0.060141	9.47
Adenosine 5'-Monophosphate	347.06309	347.0631	0.115253	9.43
Sodium D-Gluconate	178.04774	178.0477	0.224659	28.83
Deoxycarnitine	145.11028	145.1102	0.275652	10.18
Mevalolactone	130.063	130.063	0.307543	1.18
Uridine 5'-Diphosphoglucose	566.05503	566.055	0.070665	9.69
Thiamine	265.11231	265.1123	0.150879	8.88

2'-Deoxyguanosine 5'-Diphosphate	427.02941	427.0294	0.09367	10.47
5-Methylcytosine	125.05891	125.0589	0.319849	1.18
Cytidine 2',3'-Cyclic Monophosphate	305.04129	305.0413	0.13113	7.5
Guanosine 5'-Diphosphoglucose	443.02432	443.0243	0.090288	11.46
A-D-Galactose 1-Phosphate Dipotassium Salt Pentahydrate	260.02971	260.0297	0.153829	10.82
Mannose 6-Phosphate	260.02972	260.0297	0.153829	10.82
Carnitine	161.10519	161.1052	0.248285	8.94
Thymidine 5'-Monophosphate	322.05661	322.0566	0.124202	10.38
Adenosine-5'-Diphosphoglucose	427.02942	427.0294	0.09367	10.48
Octopamine	153.07897	153.0789	0.261303	8.11
Uridine 5'-Diphosphogalactose	566.05503	566.055	0.070665	9.69
5-Aminolevulinic Acid	131.05824	131.0582	0.305208	8.1
2'-Deoxyuridine 5'-Monophosphate	308.04096	308.0409	0.129853	7.7
Xanthosine 5'-Monophosphate	364.04202	364.042	0.109877	10.56
Flavin Adenine Dinucleotide	785.15714	785.1571	0.050945	8.21
Methylnicotinamide	137.07094	137.0709	0.29182	6.91

	338.06275	338.0627	0.118321	8.52
5-Aminoimidazole-4-Carboxamide-1-?-D-Ribofuranosyl 5'-Monophosphate				
	607.08154	607.0815	0.065889	9.37
Uridine 5'-Diphospho-N-Acetylgalactosamine				
(5'-Adenosyl)-L-Homocysteine	384.12159	384.1216	0.104134	8.27
Adenine Hydrochloride	135.0545	135.0545	0.296177	1.89
Phosphocreatine	211.03581	211.0358	0.189541	10.67
Uridine 5'-Diphosphoglucuronic Acid	580.0343	580.0343	0.068961	9.03
Guanosine 5'-Monophosphate	363.058	363.058	0.110175	10.46
N-Acetyl-Alanine	131.05824	131.0582	0.305208	8.1
4-Guanidino-Butanoate	145.08513	145.0851	0.2757	10.18
Methyl Beta-D-Galactoside	194.07904	194.079	0.206102	2.19
CMP	323.05186	323.0518	0.123819	10.36
6-Hydroxynicotinate	139.02694	139.0269	0.287714	6.8
Inosine 5'-Monophosphate	348.0471	348.0471	0.114927	9.95
Pantothenic Acid	219.11067	219.1106	0.182556	4.79
DTMP	322.05661	322.0566	0.124202	10.37
3-Ureidopropionic Acid	132.05349	132.0535	0.302908	8.1
Norleucine	131.09463	131.0946	0.305123	8.1
Adenosine	267.09676	267.0967	0.149758	1.9
Raffinose	594.22187	594.2218	0.067315	19.89
4-Hydroxy-D-Proline	131.05824	131.0582	0.305208	8.09

Choline	104.10754	104.1075	0.384218	6.51
4-Aminobutanoic Acid	103.06333	103.0633	0.388111	9.26
Formyl-Methionyl	177.04597	177.0459	0.22593	28.82
Alanine	89.04768	89.04764	0.449198	8.11
Histidine	155.06948	155.0694	0.257949	16.56
Creatine Phosphate Dibasic Tetrahydrate	211.03581	211.0358	0.189541	9.04
Adenosine 5'-Diphosphoribose	559.07168	559.0716	0.071547	9.31
Trigonelline	137.04768	137.0476	0.291869	6.91
Sn-Glycerol-3-Phosphocholine	258.11011	258.1101	0.154973	9.66
N-Acetyl-Serine	147.05316	147.0531	0.27201	9.49
Uridine 5'-Diphosphate	404.0022	404.0022	0.099009	9.36
Alpha-Glucose 1-Phosphate	260.02972	260.0297	0.153829	10.82
Glucosamine 6-Sulfate	259.03619	259.0362	0.154419	9.55
Tyramine	137.08406	137.084	0.291792	6.91
Cortisol	362.20933	362.2093	0.110433	1.13
Melatonin	232.12118	232.1211	0.172324	1.1
3-Hydroxybenzyl Alcohol	124.05243	124.0524	0.322444	26.33
Lumichrome	242.08038	242.0803	0.165234	1.22
Glycocholate	465.30904	465.309	0.085964	2.33
3-Amino-5-Hydroxybenzoic Acid	153.0426	153.0426	0.261365	8.11
Guaiacol	124.05243	124.0524	0.322444	26.33
Methyl Vanillate	182.05791	182.0579	0.21971	6.67
3-(4-Hydroxyphenyl)Pyruvate	180.04226	180.0422	0.22217	5.1
Cortisol 21-Acetate	404.21989	404.2199	0.098956	1.11

3(2-Hydroxyphenyl)Propanoate	166.063	166.063	0.240872	4.6
Dihydroxymandelic Acid	184.03718	184.0371	0.217347	7.1
Dethiobiotin	214.13174	214.1317	0.186801	1.2
Azelaic Acid	188.10486	188.1048	0.212647	6.11
3Alpha,11Beta,17Alpha,21-Tretrahydroxy-5Alpha-Pregnan-20-One	366.24062	366.2406	0.109218	1.13
Biotin	244.08817	244.0881	0.163875	2.2
Mercaptopyruvate	119.98812	119.9881	0.333366	28.82
Lithium Potassium Acetyl Phosphate	139.9875	139.9875	0.28574	8.91
Dehydroascorbate	174.01644	174.0164	0.229863	28.82
Sucrose	342.11622	342.1162	0.116919	8.55
Mannitol	182.07904	182.079	0.219685	6.68
Melibiose	342.11622	342.1162	0.116919	8.55
Maltose	342.11622	342.1162	0.116919	8.55
Gulonic Acid	178.04774	178.0477	0.224659	28.83
Cellobiose	342.11622	342.1162	0.116919	8.55
Palatinose	342.11621	342.1162	0.116919	8.55
Vitamin D2	396.33922	396.3392	0.100924	1.08
4-Coumarate	164.04735	164.0473	0.243832	6.73
Cholesteryl Acetate	428.36543	428.3654	0.093378	1.09
Petroselinic Acid	282.25588	282.2558	0.141715	1.14
1,2-Dipalmitoyl-Sn-Glycerol	568.50668	568.5066	0.07036	1.03
Deoxycholic Acid	392.29266	392.2926	0.101965	1.1
3Alpha-Hydroxy-5-Beta-Cholanate	376.29775	376.2977	0.106299	1.08
Protoporphyrin	562.25801	562.258	0.071142	1.08
Menaquinone	444.30283	444.3028	0.090029	1.13

Estradiol-17Alpha	272.17763	272.1776	0.146963	1.3
Myristic Acid	228.20893	228.2089	0.175278	1.13
17A,21-Dihydroxy-4-Pregnene-3,20-Dione	346.21441	346.2144	0.115535	1.12
Sodium Taurolithocholate	483.30185	483.3018	0.082764	1.13
Sphinganine	301.29808	301.298	0.132759	1.41
Erucic Acid	338.31848	338.3184	0.118232	1.09
Deoxycholic Acid	392.29266	392.2926	0.101965	1.1
Leukotriene B4	336.23006	336.23	0.118966	1.1
Cholestra-5,7-Dien-3Beta-Ol	384.33922	384.3392	0.104075	1.11
Bis(2-Ethylhexyl)Phthalate	390.27701	390.277	0.102491	1.06
Cholic Acid	408.28758	408.2875	0.09797	1.1
Phylloquinone	450.34978	450.3497	0.08882	1.12
Diethyl-2-Methyl-3-Oxosuccinate	202.08413	202.0841	0.197937	1.08
Sodium Phenylpyruvate	164.04735	164.0473	0.243832	6.72
Oleate	282.25588	282.2558	0.141715	1.14
Stearate	284.27153	284.2715	0.140711	1.08
Beta-Carotene	536.4382	536.4382	0.074566	1.12
25-Hydroxycholesterol	402.34978	402.3497	0.099416	1.11
Nervonic Acid	366.34978	366.3497	0.109185	1.14
Deoxycorticosterone	372.23006	372.23	0.10746	1.17
Oleoyl-Rac-Glycerol	356.29266	356.2926	0.112267	1.11
Tocopherol	430.38108	430.381	0.092941	1.09
Cortisone	360.19368	360.1936	0.111051	1.15
Corticosterone	346.21441	346.2144	0.115535	1.12

Table CS2: Metabolite name, theoretical mass, observed mass, Δ ppm and retention time (RT) in minutes for 153 metabolites identified using ESI negative ion $[M-H]^+$ in MDA-MB 231 cell extracts.

Metabolite	Theoretical mass	Experimental mass	ppm	RT (min)
NAD	663.10912	663.10916	0.0603219	10.02
Glutamine	146.06914	146.06918	0.27384292	9
Inosine 5'-Phosphate	348.0471	348.04714	0.11492697	10.53
Citrate	192.02701	192.02705	0.20830403	14.26
Threonine	119.05824	119.05828	0.33597003	8.31
N-Acetylneuraminate	309.10598	309.10602	0.12940546	8.58
Aspartate	133.03751	133.03755	0.30066708	11.13
Serine	105.04259	105.04263	0.38079792	9.36
Taurine	125.01467	125.01471	0.31996245	7.9
Nicotinate	123.03203	123.03207	0.32511859	1.83
Inosine	268.08077	268.08081	0.14920876	3.23
4-Aminobutanoate	103.06333	103.06337	0.38811088	9.99
Isoleucine	131.09463	131.09467	0.3051231	4.98
Pyrazole	67.02962	67.02966	0.59675111	19.54
Glutamic Acid	147.05316	147.0532	0.27201048	10.01
Beta-Alanine	89.04768	89.04772	0.44919755	8.23
Sarcosine	89.04768	89.04772	0.44919755	8.23
Gluconic Acid	196.0583	196.05834	0.20402095	8.8
Adenine	135.0545	135.05454	0.29617673	1.88
Thymidine	242.09027	242.09031	0.16522762	1.21
Alanine	89.04768	89.04772	0.44919755	8.23
Tryptophan	204.08988	204.08992	0.19599208	5.15
Uridine-5-Monophosphate	324.03587	324.03591	0.12344312	10.33
Proline	115.06333	115.06337	0.34763465	6.83
Uridine	244.06954	244.06958	0.16388772	2.22
Carnosine	226.10659	226.10663	0.17690771	11.63
Shikimate	174.05283	174.05287	0.22981528	7.3
Succinate	118.02661	118.02665	0.33890662	10.25
Phenylalanine	165.07898	165.07902	0.24230826	4.26
Malate	134.02153	134.02157	0.29845951	11.15
Aspartate	133.03751	133.03755	0.30066708	10.52
2'-Deoxycytidine 5-Monophosphate	307.05694	307.05698	0.130269	9.13
Hypoxanthine	136.03851	136.03855	0.29403439	2.15
3,4-Dihydroxy-1-Phenylalanine	197.06881	197.06885	0.20297479	11.16

Guanosine	283.09167	283.09171	0.14129699	5.14
Malate	134.02153	134.02157	0.29845951	11.16
Lysine	146.10553	146.10557	0.27377472	15.71
Tyrosine	181.07389	181.07393	0.22090429	6.68
Valine	117.07898	117.07902	0.34164971	6.16
Homoserine	119.05824	119.05828	0.33597003	8.31
Folic Acid	441.13968	441.13972	0.09067423	10.42
A-Galactose 1-Phosphate Dipotassium Salt Pentahydrate	260.02971	260.02975	0.15382858	11.38
Mannose 6-Phosphate	260.02972	260.02976	0.15382857	11.38
O-Acetyl-L-Serine	147.05316	147.0532	0.27201048	10.01
Thymidine 5'-Monophosphate	322.05661	322.05665	0.12420177	11.43
Fructose 6-Phosphate	260.02972	260.02976	0.15382857	11.38
Nitro-Tyrosine	226.05897	226.05901	0.17694498	9.12
Uridine 5'-Diphosphogalactose	566.05503	566.05507	0.07066451	10.37
5-Aminolevulinic Acid	131.05824	131.05828	0.30520782	8.25
2'-Deoxyuridine 5'-Monophosphate	308.04096	308.041	0.12985286	9.13
Xanthosine 5'-Monophosphate	364.04202	364.04206	0.10987743	11.27
Flavin Adenine Dinucleotide	785.15714	785.15718	0.05094522	8.56
Orotic Acid	174.02767	174.02771	0.22984851	7.14
5-Aminoimidazole-4-Carboxamide-1-Ribofuranosyl 5'-Monophosphate	338.06275	338.06279	0.11832123	10.97
Uridine 5'-Diphospho-N-Acetylgalactosamine	607.08154	607.08158	0.06588901	10.01
S-(5'-Adenosyl)-Homocysteine	384.12159	384.12163	0.10413369	8.29
Uridine 5'-Diphospho-N-Acetylglucosamine	607.08158	607.08162	0.065889	10.01
Uridine 5'-Diphosphoglucuronic Acid	580.0343	580.03434	0.06896144	11.08
Thiamine Monophosphate	344.07081	344.07085	0.11625514	10.35
Guanosine 5'-Monophosphate	363.058	363.05804	0.11017523	11.07
CMP	323.05186	323.0519	0.12381913	10.38
Guanosine 5'-Diphospho-D-Mannose	605.07716	605.0772	0.06610727	11.11
N-Acetyl-Dl-Glutamic Acid	189.06372	189.06376	0.21156888	9.63

2,4-Dihydroxypteridine	164.03343	164.03347	0.24385273	1.19
Inosine 5'-Monophosphate	348.0471	348.04714	0.11492697	10.53
Pantothenic Acid	219.11067	219.11071	0.18255615	4
2-Amino-2-Methyl-Propanoate	103.06333	103.06337	0.38811088	9.99
DTMP	322.05661	322.05665	0.12420177	11.43
Norleucine	131.09463	131.09467	0.3051231	4.67
Adenosine	267.09676	267.0968	0.14975846	1.87
Saccharic Acid	210.03757	210.03761	0.19044212	9.14
4-Imidazoleacetic Acid	126.04293	126.04297	0.31735219	9.34
Methionine Sulfoximine	180.05687	180.05691	0.22215203	8.79
4-Hydroxy-Proline	131.05824	131.05828	0.30520782	8.25
4-Aminobutanoic Acid	103.06333	103.06337	0.38811088	10
Allothreonine	119.05824	119.05828	0.33597003	8.31
Adenosine 5'-Diphosphoribose	559.07168	559.07172	0.07154718	9.95
2'-Deoxycytidine 5'-Diphosphate	387.02327	387.02331	0.10335296	11.07
Ribose 5-Phosphate	230.01916	230.0192	0.17389856	10.69
Alpha-Glucose 1-Phosphate	260.02972	260.02976	0.15382857	11.38
Glucosamine 6-Sulfate	259.03619	259.03623	0.15441858	10.29
Cortisol	362.20933	362.20937	0.11043338	1.09
3-Hydroxybenzaldehyde	122.03678	122.03682	0.32777004	1.85
Maleic Acid	116.01096	116.011	0.34479501	1.33
3-Hydroxybenzyl Alcohol	124.05243	124.05247	0.32244431	1.88
Mandelic Acid	152.04734	152.04738	0.26307596	4.16
Benzoate	122.03678	122.03682	0.32777004	1.85
Lumichrome	242.08038	242.08042	0.16523437	1.19
Oxaloacetate	132.00588	132.00592	0.3030168	8.3
2,3-Dihydroxybenzoate	154.02661	154.02665	0.25969539	4.28
Thiopurine S-Methylester	166.03132	166.03136	0.2409184	1.14
10-Hydroxydecanoate	188.14125	188.14129	0.21260622	1.38
Ethylmalonic Acid	132.04226	132.0423	0.30293332	1.21
Fumarate	116.01096	116.011	0.34479501	11.11
4-Hydroxybenzaldehyde	122.03678	122.03682	0.32777004	1.85
3(2-Hydroxyphenyl)Propanoate	166.063	166.06304	0.24087244	4.11
3-Methoxytyramine	167.09464	167.09468	0.2393853	4.06
Pterine	163.04941	163.04945	0.24532441	1.19
Indole-3-Pyruvic Acid	203.05824	203.05828	0.19698782	6.66
Anthranilate	137.04768	137.04772	0.29186922	2.15

2-Methylmaleate	130.02661	130.02665	0.30762934	1.39
Mono-Methyl Glutarate	146.05791	146.05795	0.27386398	9.02
2-Methylglutaric Acid	146.05791	146.05795	0.27386398	9.02
Adipic Acid	146.05791	146.05795	0.27386398	9.02
Homovanillate	182.05791	182.05795	0.21971031	5.83
Salicylamide	137.04768	137.04772	0.29186922	2.15
3-Hydroxybenzoate	138.0317	138.03174	0.2897885	2.19
3Alpha,11Beta,17Alpha,21-Tretrahydroxy-5Alpha-Pregnan-20-One	366.24062	366.24066	0.10921781	1.14
Indole-3-Acetaldehyde	159.06841	159.06845	0.25146413	2.18
3-Hydroxyphenylacetate	152.04735	152.04739	0.26307594	4.3
Salicylic Acid	138.0317	138.03174	0.2897885	2.19
3(4-Hydroxyphenyl)Lactate	182.05791	182.05795	0.21971031	5.83
Biotin	244.08817	244.08821	0.16387521	2.18
Dehydroascorbate	174.01644	174.01648	0.22986334	7.14
Sorbose	180.06339	180.06343	0.22214399	8.79
Xylitol	152.06848	152.06852	0.26303939	4.19
Ribitol	152.06848	152.06852	0.26303939	4.19
Myo-Inositol	180.06339	180.06343	0.22214399	8.79
Mannose	180.06339	180.06343	0.22214399	8.79
Galactose	180.06339	180.06343	0.22214399	8.79
Glucose	180.06339	180.06343	0.22214399	8.79
Allose	180.06339	180.06343	0.22214399	8.79
Mannitol	182.07904	182.07908	0.21968481	5.82
Melibiose	342.11622	342.11626	0.11691933	7.93
Maltose	342.11622	342.11626	0.11691933	7.93
Cellobiose	342.11622	342.11626	0.11691933	7.93
Arabitol	152.06848	152.06852	0.26303939	4.3
Palatinose	342.11621	342.11625	0.11691934	7.93
4-Coumarate	164.04735	164.04739	0.24383204	1.19
Nonanoate	158.13068	158.13072	0.25295534	1.24
Chenodeoxycholate	392.29266	392.2927	0.10196469	1.1
Caprylic Acid	144.11503	144.11507	0.27755606	1.26
Petroselinic Acid	282.25588	282.25592	0.14171538	1.16
Deoxycholic Acid	392.29266	392.2927	0.10196469	1.1
3Alpha-Hydroxy-5-Beta-Cholanate	376.29775	376.29779	0.1062988	1.1
Menaquinone	444.30283	444.30287	0.09002869	1.16
Elaidic Acid	282.25588	282.25592	0.14171538	1.15
Myristic Acid	228.20893	228.20897	0.17527798	1.19
Palmitoleic Acid	254.22458	254.22462	0.1573412	1.16
Palmitate	256.24023	256.24027	0.15610351	1.17

Triiodothyronine	650.79007	650.79011	0.06146375	1.33
Lauric Acid	200.17763	200.17767	0.19982253	1.2
Arachidic Acid	312.30283	312.30287	0.12808081	1.13
Deoxycholic Acid	392.29266	392.2927	0.10196469	1.1
Linoleate	280.24023	280.24027	0.14273468	1.16
Bis(2-Ethylhexyl)Phthalate	390.27701	390.27705	0.1024913	1.1
Retinoate	300.20893	300.20897	0.13324054	27.93
Cholic Acid	408.28758	408.28762	0.09797016	1.35
Phylloquinone	450.34978	450.34982	0.08881985	1.18
Docosahexaenoic Acid	328.24023	328.24027	0.12186197	28.34
Sodium Phenylpyruvate	164.04735	164.04739	0.24383204	1.2
Oleate	282.25588	282.25592	0.14171538	1.16
Decanoate	172.14633	172.14637	0.23236046	28.05

Table CS3: List of metabolic pathways impacted by hits (metabolites identified in the pathways) and the significance (p-value of hits), along with total compounds present in the metabolic pathway and false discovery rate (FDR) for the hits adjusted for total number of compounds present in the metabolic pathway for metabolites identified using ESI positive ion [M+H]+ and ESI negative ion mode [M-H]- in MDA-MB 231 cell extracts. Number of metabolite annotations used for this analysis were 254 and the metabolic pathways listed are grouped into four super metabolic pathways namely central carbon metabolism, one carbon metabolism, methionine metabolism and glutathione metabolism.

Pathway	**Total**	**Expected**	**Hits**	**p-value**	**FDR**	**Impact**
Purine metabolism	65	8.30	23	1.4E-06	7.5E-05	0.40
Aminoacyl-tRNA biosynthesis	48	6.13	19	1.8E-06	7.5E-05	0.17
Arginine biosynthesis	14	1.79	9	8.6E-06	2.4E-04	0.48
Alanine, aspartate and glutamate metabolism	28	3.58	12	6.4E-05	1.4E-03	0.76
Pantothenate and CoA biosynthesis	19	2.43	8	1.3E-03	2.2E-02	0.26
Citrate cycle (TCA cycle)	20	2.55	8	2.0E-03	2.7E-02	0.35
Pyrimidine metabolism	39	4.98	12	2.3E-03	2.7E-02	0.32
D-Glutamine and D-glutamate metabolism	6	0.77	4	3.1E-03	3.3E-02	0.50
Valine, leucine and isoleucine biosynthesis	8	1.02	4	1.2E-02	1.1E-01	0.00
Amino sugar and nucleotide sugar metabolism	37	4.73	10	1.4E-02	1.2E-01	0.25
Arginine and proline metabolism	38	4.85	10	1.7E-02	1.3E-01	0.45
Glycine, serine and threonine metabolism	33	4.22	9	1.9E-02	1.3E-01	0.63
Lysine degradation	25	3.19	7	3.2E-02	2.0E-01	0.29
Nicotinate and nicotinamide metabolism	15	1.92	5	3.3E-02	2.0E-01	0.46

beta-Alanine metabolism	21	2.68	6	4.2E-02	2.1E-01	0.56
Glyoxylate and dicarboxylate metabolism	32	4.09	8	4.2E-02	2.1E-01	0.20
Histidine metabolism	16	2.04	5	4.3E-02	2.1E-01	0.43
Glutathione metabolism	28	3.58	7	5.6E-02	2.6E-01	0.20
Starch and sucrose metabolism	18	2.30	5	6.9E-02	2.8E-01	0.41
Ascorbate and aldarate metabolism	8	1.02	3	7.1E-02	2.8E-01	0.50
Taurine and hypotaurine metabolism	8	1.02	3	7.1E-02	2.8E-01	0.71
Phenylalanine, tyrosine and tryptophan biosynthesis	4	0.51	2	8.2E-02	3.1E-01	1.00
Vitamin B6 metabolism	9	1.15	3	9.6E-02	3.5E-01	0.57
Butanoate metabolism	15	1.92	4	1.1E-01	4.0E-01	0.03
Galactose metabolism	27	3.45	6	1.2E-01	4.0E-01	0.13
Nitrogen metabolism	6	0.77	2	1.7E-01	5.6E-01	0.00
Neomycin, kanamycin and gentamicin biosynthesis	2	0.26	1	2.4E-01	7.4E-01	0.00
Pyruvate metabolism	22	2.81	4	3.1E-01	9.2E-01	0.03
Phenylalanine metabolism	10	1.28	2	3.7E-01	1.0E+00	0.36
Riboflavin metabolism	4	0.51	1	4.2E-01	1.0E+00	0.00
Glycolysis / Gluconeogenesis	26	3.32	4	4.3E-01	1.0E+00	0.11
Fructose and mannose metabolism	20	2.55	3	4.8E-01	1.0E+00	0.19
Phosphonate and phosphinate metabolism	6	0.77	1	5.6E-01	1.0E+00	0.00
Thiamine metabolism	7	0.89	1	6.2E-01	1.0E+00	0.00
Cysteine and methionine metabolism	33	4.22	4	6.3E-01	1.0E+00	0.15
Tyrosine metabolism	42	5.37	5	6.4E-01	1.0E+00	0.20
Pentose and glucuronate interconversions	18	2.30	2	6.9E-01	1.0E+00	0.13
Biosynthesis of unsaturated fatty acids	36	4.60	4	6.9E-01	1.0E+00	0.00
Glycerophospholipid metabolism	36	4.60	4	6.9E-01	1.0E+00	0.07
Ubiquinone and other terpenoid-quinone biosynthesis	9	1.15	1	7.1E-01	1.0E+00	0.00
Fatty acid biosynthesis	47	6.00	5	7.4E-01	1.0E+00	0.01
Biotin metabolism	10	1.28	1	7.5E-01	1.0E+00	0.00
Porphyrin and chlorophyll metabolism	30	3.83	3	7.6E-01	1.0E+00	0.03
Inositol phosphate metabolism	30	3.83	3	7.6E-01	1.0E+00	0.13
Sphingolipid metabolism	21	2.68	2	7.7E-01	1.0E+00	0.01
Valine, leucine and isoleucine degradation	40	5.11	4	7.7E-01	1.0E+00	0.02
Pentose phosphate pathway	22	2.81	2	7.9E-01	1.0E+00	0.22
Propanoate metabolism	23	2.94	2	8.1E-01	1.0E+00	0.00
Mannose type O-glycan biosynthesis	17	2.17	1	9.0E-01	1.0E+00	0.00
Tryptophan metabolism	41	5.24	3	9.1E-01	1.0E+00	0.15
Selenocompound metabolism	20	2.55	1	9.4E-01	1.0E+00	0.00
Fatty acid degradation	39	4.98	2	9.7E-01	1.0E+00	0.12
Phosphatidylinositol signaling system	28	3.58	1	9.8E-01	1.0E+00	0.04
Primary bile acid biosynthesis	46	5.88	2	9.9E-01	1.0E+00	0.02
Fatty acid elongation	39	4.98	1	1.0E+00	1.0E+00	0.00

Table CS4: Fold Change (FC), standard error (StdError), false discovery rate (FDR), coefficient of determination (R2), retention time (RT) in minutes, m/z for 54 metabolites identified using ESI positive ion $[M+H]^+$ and ESI negative ion mode [M-H]- in MDA-MB 231 cell extracts. Number of metabolite annotations in mass spectrometry were 254, substantially high but the number of metabolites identified using NMR were only 29, therefore a subset of 54 metabolites were selected from HILIC-MS in ESI positive mode, negative mode and NMR to identify the impact of doxycycline on four super metabolic pathways namely central carbon metabolism, one carbon metabolism, methionine metabolism and glutathione metabolism.

Metabolite	**FoldChange**	**StdError**	**FDR**	**R^2**	**RT**	**[M+H]+**	**[M-H]-**
5'-MethylThioAdenosine_n[1]	2.08	0.29	0.04		1.47		296.082
Acetylcarnitine_p[2]	1.42	0.17	0.16		7.61	204.123	
Adenine_p	1.40	0.33	0.76		2.01	136.062	
Adenosine_p	1.24	0.31	0.95		1.9	268.105	
ADP-Ribose_p	0.87	0.23	0.86		9.55	560.080	
Alanine_n	1.62	0.19	0.15	0.66	8.65		88.040
Arginine_p	0.26	0.10	0.06		15.98	175.119	
Asparagine_p	0.43	0.12	0.13		9.66	133.061	
Aspartate_nmr[3]	0.45	0.05	0.03	0.47	NA		
Beta-Alanine_nmr	1.75	0.40	0.23		NA		
Betaine_p	1.39	0.11	0.38		6.51	118.087	
CAMP_p	0.67	0.11	0.15		9.52		
Choline_p	1.75	0.11	0.02	0.52	6.51	105.115	
Citrate_n	1.43	0.45	0.19		14.26		191.019
CMP_p	1.92	0.46	0.38		10.13	324.060	
Creatine phosphate_nmr	1.63	0.12	0.01		NA		
Creatine_p	1.62	0.08	0.06		8.58	132.077	
Creatinine_p	0.62	0.23	0.61		1.21	114.067	
Deoxycarnitine_p	2.42	0.36	0.02		10.14	146.117	
Fumarate_n	0.82	0.22	0.75		11.11		115.003
Fumarate_nmr	1.19	0.10	0.19		NA		

							146.045
Glutamate_n	2.16	0.23	0.02	0.95	10.01		
Glutamine_p	0.32	0.17	0.13	-	8.75	147.077	
Glutathione_nmr	1.12	0.04	0.34		NA		
Glutathione_p	1.01	0.13	0.42		9.68	308.091	
Glycine_nmr	1.38	0.11	0.02		NA		
Histidine_p	1.18	0.06	0.76		16.56	156.077	
							118.050
Homoserine_n	1.79	0.21	0.13		8.31		
Hypotaurine_p	1.12	0.18	0.76		9.34	110.028	
							135.031
Hypoxanthine_n	0.03	0.01	0.13		2.15		
							347.039
Inosine 5'-Phosphate_n	0.69	0.10	0.29		10.53		
							267.073
Inosine_n	0.73	0.11	0.38		3.23		
Isoleucine_p	1.03	0.10	0.39	0.37	8.1	132.102	
				0.313			
Lactate_n	0.19	0.38	0.38		4.57		89.024
Leucine_nmr	1.06	0.06	0.61		8.1		
							145.098
Lysine_n	1.45	0.19	0.33		15.71		
							133.014
Malate_n	0.85	0.24	0.76		11.15		
Methionine_p	1.06	0.08	0.38		6.34	150.059	
Myo-Inositol_nmr	1.28	0.07	0.19		NA		
							662.101
NAD_n	1.84	0.16	0.02		10.02		
NADP+_p	0.91	0.21	0.86		9.47	744.080	
							164.071
Phenylalanine_n	1.76	0.22	0.14	0.73	4.26		
							116.071
Proline_p	1.53	0.12	0.23		6.89		
Propionylcarnitine_p	1.40	0.15	0.16		6.72	218.138	
Pyruvate_nmr	1.05	0.15	0.80		NA		
S-Adenosylmethionine_p	0.64	0.10	0.17		13.79	399.140	
Sarcosine_p	1.37	0.09	0.38		9.68	90.055	
Serine_p	0.39	0.11	0.07		9.36	106.049	
Succinate_NMR	0.61	0.10	0.09		NA		
Taurine_p	1.57	0.08	0.02	0.8	7.78	126.022	
Threonine_p	1.12	0.12	0.75	0.69	9.34	120.066	
Tryptophan_p	0.82	0.15	0.38		7.25	205.098	
Tyrosine_p	0.96	0.11	0.38	0.83	6.7	182.082	
Valine_p	1.40	0.11	0.38	0.63	6.11	118.087	

[1]ESI negative, [2] ESI positive, [3] NMR

Table CS5: Fold Change (FC), standard error (StdError), false discovery rate (FDR) for metabolites identified in HS578T cell extracts using NMR.

Metabolite	**FC**	**StdError**	**FDR**
Alanine	0.87	0.14	0.30
Creatine phosphate	0.50	0.05	0.03
Glutamate	0.67	0.14	0.12
Glycine	0.78	0.23	0.31
Isoleucine	0.76	0.14	0.12
Lactate	0.28	0.12	0.04
Lysine	0.91	0.14	0.30
Methionine	0.94	0.23	0.76
NAD+	0.80	0.20	0.31
Propylene glycol	0.96	0.38	0.89
Pyruvate	0.63	0.08	0.30
Threonine	0.66	0.15	0.11
Tyrosine	1.37	0.34	0.41
Valine	0.94	0.15	0.68
myo-Inositol	0.42	0.08	0.03
sn-Glycero-3-phosphocholine	0.97	0.08	0.80

Table CS6: Fold Change (FC), standard error (StdError), false discovery rate (FDR), retention time (RT) in minutes, m/z for 32 lipids in MDA-MB 231 cell extracts using ESI positive ion $[M+H]^+$ and ESI negative ion mode $[M-H]^-$.

Lipid	**FC**	**StdError**	**FDR**	**RT**	**$[M+H]^+$**	**$[M-H]^-$**
CL(51:0)_n	2.99	0.39	2.4E-03	7.98		1191.74
CL(71:2)_p	0.32	0.02	2.9E-02	7.82	1412.04	
PC(30:2)_n	1.65	0.07	6.9E-04	8.13		746.49
PC(34:3)_n	1.60	0.11	1.5E-02	8.18		736.53
PC(36:1)_p	1.35	0.17	2.9E-01	8.21	752.59	
PC(36:2)_n	1.25	0.07	3.3E-02	8.88		806.57
PC(36:4)_n	1.80	0.12	2.4E-03	8.34		762.54
PC(38:3)_n	1.33	0.07	1.1E-02	9.24		792.59
PC(38:5)_n	1.31	0.09	3.3E-02	8.2		828.55
PC(40:4)_n	1.48	0.00	2.4E-03	9		858.60
PC(40:5)_n	1.19	0.01	1.5E-01	8.74		856.58
PC(42:7)_n	0.63	0.01	6.3E-02	8.58		880.59
PC(42:9)_n	1.15	0.10	5.4E-01	7.92		876.55

PE(34:1)_p	1.21	0.07	1.7E-01	8.63	718.53	
PE(37:2)_n	1.23	0.07	3.3E-02	8.33		778.54
PE(37:5)_n	1.30	0.08	2.9E-02	7.75		750.51
PE(37:6)_n	1.30	0.08	2.9E-02	7.33		748.49
PE(38:3)_p	1.17	0.06	2.6E-01	8.92	770.57	
PE(38:7)_n	1.22	0.07	5.7E-02	7.84		760.49
PE(40:7)_n	1.29	0.07	2.0E-02	8.38		788.52
PE(42:5)_p	1.33	0.07	3.3E-02	8.58	786.58	
PE(46:3)_p	0.96	0.04	8.0E-01	10.38	846.68	
PE(46:5)_p	1.13	0.06	4.2E-01	9.57	842.65	
PC(P-36:5)_p	1.40	0.11	6.6E-02	9.26	750.55	
PC(P-38:3)_n	1.36	0.10	4.2E-02	9.24		778.57
PC(P-40:2)_n	1.49	0.07	2.4E-03	9.92		846.64
PE(P-36:3)_p	1.20	0.06	2.1E-01	8.16	690.52	
PE(P-38:5)_p	1.42	0.10	4.0E-02	9.26	750.55	
PE(P-40:3)_p	1.43	0.07	1.4E-02	8.88	746.59	
PE(P-40:6)_p	1.32	0.24	2.8E-01	7.19	740.53	
PE(P-42:2)_n	1.50	0.12	5.3E-02	9.72		832.62
SM(d42:1)_p	0.84	0.08	6.9E-01	9.58	793.66	

Table CS7: Fold Change (FC), standard error (StdError), false discovery rate (FDR), retention time (RT) in minutes, m/z for 32 lipids in HS578T cell extracts using ESI positive ion $[M+H]^+$ and ESI negative ion mode $[M-H]^-$.

Lipid	**FC**	**StdError**	**FDR**	**RT**	**$[M+H]^+$**	**$[M-H]^-$**
PC(30:1)_p	0.93	0.03	2.9E-01	5.63	726.51	
PC(32:1)_n	0.83	0.02	5.8E-03	7.19		730.54
PC(32:2)_n	0.79	0.04	3.9E-02	6.55		728.52
PC(32:3)_n	1.30	0.04	2.8E-03	9.75		726.5
PC(34:1)_n	1.12	0.03	5.2E-02	6.76		804.58
PC(35:1)_n	1.04	0.02	2.9E-01	6.76		818.59
PC(36:0)_p	0.83	0.02	2.3E-03	7.39	812.62	
PC(36:1)_p	0.67	0.02	8.0E-04	7.03	810.6	
PC(36:2)_n	1.10	0.03	1.1E-01	6.93		830.59
PC(36:3)_n	1.01	0.05	8.6E-01	6.45		828.57
PC(36:4)_n	1.05	0.03	3.0E-01	7.51		826.56
PC(36:5)_n	0.85	0.03	2.2E-02	6.69		760.53
PC(38:2)_n	1.14	0.04	5.1E-02	7.7		858.62
PC(38:3)_p	0.65	0.02	2.9E-05	6.84	834.6	

PC(38:4)_p	0.91	0.02	5.1E-02	6.18	832.58	
PC(40:1)_p	0.89	0.05	2.9E-01	9.35	844.68	
PE(34:1)_n	0.97	0.03	4.7E-01	6.82		716.52
PE(36:1)_n	0.92	0.02	1.0E-01	6.87		766.54
PE(36:2)_n	1.43	0.03	2.9E-05	7.06		724.53
PE(38:1)_n	0.95	0.02	2.9E-01	6.9		792.56
PE(38:2)_n	0.93	0.04	1.4E-01	7.68		770.57
PE(40:2)_p	0.86	0.02	5.5E-03	7.48	800.62	
PE(40:3)_n	0.89	0.02	5.3E-02	7.75		778.58
PC(P-34:1)_p	1.12	0.02	5.2E-02	7.48	744.59	
PE(P-34:1)_n	1.04	0.03	2.9E-01	6.85		724.53
PE(P-36:0)_n	1.41	0.04	1.8E-04	7.76		752.56
PE(P-36:1)_n	1.15	0.03	3.5E-02	7.93		728.56
PE(P-38:4)_p	0.89	0.04	2.0E-01	7.43	752.56	
PE(P-38:5)_n	1.09	0.03	3.5E-01	6.79		748.53
PE(P-40:2)_p	0.86	0.03	5.5E-02	8.75	784.62	
PE(P-40:6)_n	0.95	0.03	2.9E-01	6.63		774.55
SM(d34:1)_p	0.83	0.02	1.8E-03	6.03	703.57	

Table CS8: Lipid name, m/z and retention time (RT) in minutes for 69 lipids identified using ESI positive ion [M+H]+ and ESI negative ion [M-H] – mode in MDA-MB 231 cell extracts.

Lipid	**RT (min)**	**[M+H]+**	**[M-H]-**
3-O-Sulfogalactosylceramide (d36:1)_n	8.04		852.55
3-O-Sulfogalactosylceramide (d38:1)_n	8.88		816.57
3-O-Sulfogalactosylceramide (d42:1)_n	9.75		872.63
Cerebroside B_p	8.23	692.54	
CL(51:0)_n	7.98		1191.74
CL(71:2)_p	7.82	1412.04	
DG(38:2n6)_p	0.60	599.54	
DG(40:5)_n	10.28		669.55
DG(42:6)_p	12.17	714.61	
Lactosylceramide (d36:1)_n	9.36		870.63
N-Palmitoylsphingosine_p	10.86	555.54	
PA(36:2)_n	7.82		721.48
PC(30:2)_n	8.13		746.49
PC(34:3)_n	8.18		736.53
PC(34:4)_n	7.94		774.51

PC(36:1)_p	8.21	752.59	
PC(36:2)_n	8.88		806.57
PC(36:4)_n	8.34		762.54
PC(38:3)_n	9.24		792.59
PC(38:5)_n	8.20		828.55
PC(38:7)_n	8.09		824.52
PC(40:4)_n	9.00		858.60
PC(40:5)_n	8.74		856.58
PC(40:8)_n	8.09		850.54
PC(42:6)_n	9.44		860.62
PC(42:7)_n	8.58		880.59
PC(42:9)_n	7.92		876.55
PE(30:0)_n	8.14		708.48
PE(34:0)_n	7.62		740.53
PE(34:1)_n	8.73		762.53
PE(35:5)_n	7.09		722.48
PE(36:2)_p	8.70	744.55	
PE(37:2)_n	8.33		778.54
PE(37:4)_n	8.11		774.51
PE(37:4)_n	8.08		734.51
PE(37:4)_n	8.00		734.51
PE(37:4)_p	9.26	736.53	
PE(37:5)_n	7.75		750.51
PE(37:6)_n	7.33		748.49
PE(38:7)_n	7.84		760.49
PE(40:3)_n	8.70		818.57
PE(40:4)_p	9.25	778.57	
PE(40:6)_n	8.16		812.52
PE(40:7)_n	8.38		788.52
PE(42:10)_n	7.97		810.51
PE(42:5)_p	8.58	786.58	
PE(42:6)_n	8.70		840.56
PE(42:7)_n	8.25		816.55
PE(44:11)_n	8.02		836.52
PE(44:12)_n	7.85		834.51
PE(44:6)_p	8.82	812.60	
PE(44:7)_p	8.35	810.58	
PE(46:3)_p	10.38	846.68	
PE(46:5)_n	9.59		858.64
PE(46:6)_p	9.33	840.63	
PC(O-34:2)_n	8.73		764.56
PC(o-44:6)_n	9.91		856.67
PC(P-36:5)_p	9.26	764.56	

PC(P-38:3)_n	9.24		816.59
PC(P-40:2)_n	9.92		846.64
PE(P-36:3)_p	8.16	690.52	
PE(P-40:3)_p	8.88	746.59	
PE(P-40:6)_p	7.19	740.53	
PE(P-42:2)_n	9.72		832.62
PG(36:1)_p	8.46	794.60	
PS(32:4)_n	5.91		772.44
PS(36:0)_n	8.34		812.54
PS(38:0)_p	8.87	820.60	
SM(d42:1)_p	9.58	793.66	
Tetradecanoylcarnitine_p	6.06	394.29	
TG(53:7)_p	11.73	827.69	
TG(59:9)_p	12.10	907.75	

Table CS9: Lipid name, m/z and retention time (RT) in minutes for 78 lipids identified using ESI positive ion [M+H]+ and ESI negative ion [M-H] – mode in HS578T cell extracts.

Metabolite	**RT (min)**	**[M+H]+**	**[M-H]-**
Cer(d40:0)_p	8.91	624.63	
Ceramide (d34:1)_p	6.23	520.51	
Ceramide (d34:1)_p	6.22	538.52	
Ceramide (d42:2)_p	9.09	648.63	
Cerebroside B_p	7.00	1478.12	
DG(36:1)_p	8.87	623.56	
DG(36:1n5)_p	8.88	605.55	
DG(36:2)_p	7.07	603.54	
DG(38:3)_p	6.27	629.55	
DG(38:4)_p	8.30	667.53	
DG(i-36:0)_p	9.38	647.56	
DG(i-40:0)_p	9.91	663.63	
Glucosylceramide (d34:1)_p	6.23	700.57	
Glucosylceramide (d40:1)_p	8.26	784.66	
Glucosylceramide (d42:1)_p	8.86	812.70	
Glucosylceramide (d42:2)_p	8.24	810.68	
GPCho(32:1)_p	6.32	1464.10	
GPEtn(34:1)_p	7.00	718.54	
GPEtn(38:5)_p	6.44	766.54	

GPEtn(40:5)_p	7.10	794.57	
LysoPC(18:1)_p	2.85	544.34	
LysoSM(d18:0)_p	1.09	449.35	
MG(16:0)_p	3.46	313.27	
MG(18:0)_p	9.40	359.32	
MG(18:0)_p	4.07	381.30	
MG(18:0)_p	8.82	341.31	
MG(18:0)_p	4.07	739.61	
N-[3-hydroxytetracos-15-enoyl]sphingosine-1-phosphocholine_p	7.67	811.66	
N-myristoylsphingosine-1-phosphocholine_p	5.38	697.53	
PC(14:1(9Z)/16:0)_p	5.63	726.51	
PC(30:1)_p	5.67	704.52	
PC(32:1)_n	7.19		730.54
PC(32:2)_n	6.55		728.52
PC(32:3)_n	9.75		726.50
PC(34:1)_n	6.76		804.58
PC(35:1)_n	6.76		818.59
PC(36:0)_p	7.39	812.62	
PC(36:1)_p	7.03	810.60	
PC(36:2)_n	6.93		830.59
PC(36:3)_n	6.45		828.57
PC(36:4)_n	7.51		826.56
PC(36:5)_n	6.69		760.53
PC(38:2)_n	7.70		858.62
PC(38:3)_p	6.84	834.60	
PC(38:4)_p	6.18	832.58	
PC(40:1)_p	9.35	844.68	
PC(o-34:0)_p	8.37	770.60	
PC(P-34:1)_p	7.48	744.59	
PE(34:1)_n	6.82		716.52
PE(35:0)_n	6.71		778.56
PE(36:1)_n	6.87		766.54
PE(36:2)_n	7.06		724.53
PE(38:1)_n	8.19		772.59
PE(38:2)_n	7.68		770.57
PE(38:3)_n	7.05		750.54
PE(40:2)_p	7.48	800.62	
PE(40:3)_n	7.75		778.58
PE(40:5)_n_i1	7.05		774.54
PE(40:5)_n_i2	6.78		774.55
PE(P-34:1)_n	6.85		724.53
PE(P-36:0)_n	7.76		752.56
PE(P-36:1)_n	7.93		728.56

PE(P-38:5)_n	6.79		748.53
PE(P-38:6)_n	6.12		746.51
PE(P-40:2)_p	8.75	784.62	
PE(P-40:6)_n	6.63		774.55
PE-NMe(35:3)_p	5.98	764.52	
PE-NMe(36:3)_n	6.90		736.53
PI(36:0)_p	6.49	889.58	
PS(36:1)_n	6.28		810.53
PS(36:2)_p	6.27	788.54	
PS(37:0)_n	6.12		826.56
PS(38:3)_p	6.49	814.56	
PS(40:5)_n	6.67		858.53
SM(d34:1)_p	6.03	703.57	
sphinganine_p	3.95	284.30	
TG(64:12)_p	3.46	515.40	
TG(i-37:0)_p	8.21	653.57	
Trihexosylceramide (d42:1)_p	8.32	1136.80	

Publications by Ashish Vaswani

1. Vaswani, A., Alcazar Magana, A., Zimmermann, E., Hasan, W., Raman, J., & Maier, C. S. (2021). Comparative LC–MS/MS lipidomic analysis of macaque heart tissue flash frozen or embedded in optimal cutting temperature polymer (OCT): Practical considerations. Rapid Communications in Mass Spectrometry, e9155.
2. Le, D. E., García-Jaramillo, M., Bobe, G., Alcazar Magana, A., Vaswani, A., Minnier, J., ... & Kaul, S. (2021). Plasma oxylipins: a potential risk assessment tool in atherosclerotic coronary artery disease. Frontiers in Cardiovascular Medicine, 8, 257.
3. Esther Yap Shiau Ping, Apiradee Uthairatanakij, Natta Laohakunjit, Pongphen Jitareerat, Ashish Vaswani, Armando Alcazar Magana, Jeffrey Morre, Claudia S. Maier (2021). Plant growth and metabolic changes in “Super Hot” Chili Fruit (Capsicum annuum) exposed with supplemental LED Lights. Plant Science, 110826.
4. Alcazar Magana, A, Wright, K, Vaswani, A, et al. Integration of mass spectral fingerprinting analysis with precursor ion (MS1) quantification for the characterisation of botanical extracts: application to extracts of Centella asiatica (L.) Urban. Phytochemical Analysis. 2020; 1– 17.