SERUM METABOLITES CHARACTERISTICS IN PATIENTS WITH NON-ALCOHOLIC FATTY LIVER DISEASE AND RELATED CORONARY ARTERY DISEASE

by

Fang Chiu

A Thesis
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Master of Science
Nutrition

Committee:

_____	Margaret Slavin, Ph.D., Co-Thesis Director
_____	Robin Couch, Ph.D., Co-Thesis Director
_____	Sina Gallo, Ph.D., Committee Member
_____	Lawrence J. Cheskin, Ph.D., Department Chairperson
_____	Robert M. Weiler, Ph.D., Interim Associate Dean for Academic Affairs, College of Health and Human Services
_____	Germaine M. Louis, Ph.D., Dean, College of Health and Human Services

Date: _____ Summer Semester 2019
George Mason University
Fairfax, VA

Serum Metabolites Characteristics in Patients with Non-Alcoholic Fatty Liver Disease and Related Coronary Artery Disease

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at George Mason University

by

Fang Chiu
Bachelor of Science
Taipei Medical University, 2017

Co-Director: Margaret Slavin, Associate Professor
Department of Nutrition and Food Studies
Co-Director: Robin Couch, Associate Professor
Department of Chemistry and Biochemistry

Summer Semester 2019
George Mason University
Fairfax, VA

DEDICATION

To my parents, for always encouraging me to pursue my goals and providing every kinds of support. To my grandmother, for the endless care and love (and food). To Yun and Ryan, for giving me enough courage and laughter to overcome the challenges.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to Dr. Couch and Dr. Slavin, for the guidance and the constant encouragement. To the members in the Couch Lab and Dr. Dailey, thank you for your advice and assistance. I would also like to give special thanks to the metabolomics group, including Fatima, Grey, and Hameed, for providing instructions and suggestions in this project. And lastly, I really appreciate the University Writing Center, my tutor Funmi and Tanya, for the weekly thesis consultation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

Atmospheric pressure chemical ionization ..... APCI
Automated data analysis pipeline ..... ADAP
Body mass index ..... BMI
Branched-chain amino acids ..... BCAA
Cardiovascular disease ..... CVD
Carotid intima-media thickness ..... CIMT
Centers for Disease Control and Prevention ..... CDC
Collision-induced dissociation ..... CID
Competitive Fragmentation Modeling for Metabolite Identification ..... CFM ID
Computed tomography ..... CT
Coronary artery calcification ..... CAC
Coronary artery disease ..... CAD
Coronary heart disease ..... CHD
c-reactive protein ..... CRP
Cyclooxygenase-1 ..... COX-1
Dalton ..... Da
Docosahexaenoic acid ..... DHA
Eicosapentaenoic acid ..... EPA
Electrospray ionization ..... ESI
Extract ion chromatograms ..... EIC
False discovery rate ..... FDR
High-density lipoprotein cholesterol ..... HDL-c
High-performance liquid chromatography ..... HPLC
Human Metabolome Database ..... HMDB
Interleukin-8 ..... IL-8
Institutional/Independent Review Board ..... IRB
Liquid chromatography–mass spectrometry ..... LC-MS
Low-density lipoprotein cholesterol ..... LDL-c
Mass-of-charge ..... m/z
Metabolic syndrome ..... MetS
Mitogen-activated protein kinase ..... MAPK
Multiple reactions monitoring ..... MRM
Myocardial infarction ..... MI
National Heart, Lung, and Blood Institute ..... NHLBI
National Institute of Diabetes and Digestive and Kidney Diseases ..... NIDDK
Non-alcoholic fatty liver disease ..... NAFLD

Non-alcoholic steatohepatitis.....NASH
Partial least square-discriminant analysis.....PLS DA
Phosphotidylcholine.....PC
Principle component analysis.....PCA
Quadrupole time of flight mass analyzer.....QToF
Reactive oxygen species.....ROS
Signal-to-noise ratio.....s/n
Sphingomyelin.....SM
Tumor necrosis factor alpha.....TNF-α
Toll-like receptor 2.....TLR2
Total ion count.....TIC
Triacylglyceride.....TAG
Triglycerides.....TG
Trimethylamine N-oxide.....TMAO
Type 2 diabetes mellitus.....T2DM
Ultra-performance liquid chromatography.....UPLC
Vascular cell adhesion molecule 1.....VCAM1
Very low-density lipoprotein.....VLDL

ABSTRACT

SERUM METABOLITES CHARACTERISTICS IN PATIENTS WITH NON-ALCOHOLIC FATTY LIVER DISEASE AND RELATED CORONARY ARTERY DISEASE

Fang Chiu, Master of Science Candidate

George Mason University, 2019

Thesis Co-Directors: Dr. Margaret Slavin

Dr. Robin Couch

Coronary artery disease (CAD) is more likely to develop among individuals with non-alcoholic fatty liver disease (NAFLD), as compared to healthy individuals. However, there are currently limited biomarkers to detect heart disease before cardiac symptoms occur. Applying metabolomic fingerprinting and profiling to these patients might improve prediction and diagnosis of CAD. Thus, the aims of this study were to identify metabolomic pattern differences among various CAD severities and to address metabolite signatures that can become potential biomarkers.

Liquid chromatography-mass spectrometry was performed on 83 serum sample extractions from participants diagnosed with NAFLD and CAD severity levels ranging from no CAD through CAD level 4. The data underwent multivariate analysis to distinguish metabolomic patterns variation, and results showed that there is a separation between CAD stages, indicating the difference in their metabolomic fingerprints by stage. Then, Student's t-test with FDR correction, frequency calculation, and log2 fold change

were applied to select metabolites associated with CAD. Eight candidate metabolites were chosen and were further analyzed by tandem mass spectrometry to yield metabolomic profiles. Compound identification was attempted by comparing these sample profiles with the NIST database. According to the results, one of the targeted profiles matched a drug derived metabolite, 4-hydroxyatorvastatin lactone, and it is CAD 0 stage specific. Based on patient medical history report, the use of atorvastatin was found to be higher in this stage, which confirmed the result. In conclusion, the identified metabolite can be considered as an internal standard; it verified the data processing and method of selecting metabolites in the present research. The other 7 CAD sensitive metabolites were unable to be identified using current databases of known compounds and may be investigated in future projects.

BACKGROUND

Coronary artery disease (CAD) is a type of cardiovascular disease (CVD) featuring narrowed coronary arteries; it is also known as coronary heart disease (CHD). Usually caused by the buildup of plaques in the arteries, this heart disease can decrease or block the oxygen-rich blood flow to the heart, which leads to ischemic heart disease and heart failure. CAD has become one of the most common types of heart disease in the United States, according to the Centers for Disease Control and Prevention (CDC).[1,2] However, patients with CAD often do not notice their conditions until they experience several complications, such as angina pectoris or chest pain.

Diet, lifestyle, ethnicity, and genetics are factors that affect the development of plaques in the coronary arteries. In addition, non-alcoholic fatty liver disease (NAFLD) is one of the important risk factors that is associated with CAD.[3] Since plaques in coronary arteries build up due to inflamed cholesterol deposits, and dyslipidemia is highly prevalent among NAFLD patients, the relationship between CAD and NAFLD is predictable. People with NAFLD are at higher risk of developing CAD and other kinds of cardiovascular disease than those without NAFLD.[4]

Considering the association between NAFLD and CAD, and the symptomless progress of CAD among patients, an easily-measured indicator of CAD that reflects coronary artery conditions before an actual heart attack would be helpful. The metabolic

status in a human body changes when disease occurs, meaning that there is an alteration in quality and quantity of metabolites in the diseased person. Connecting this metabolomics concept with clinical demands, if specific metabolites that are related to CAD can be successfully identified in NAFLD patients, they can be defined as new biomarkers for predicting CAD in this population. Therefore, we would be able to identify the heart disease at an early stage by measuring specific metabolites in the body, making the diagnosis of CAD more efficient, and patients could receive treatments before their heart disease worsens.

The following study focuses on the metabolomic profiles of NAFLD-related CAD patients and the identification of CAD specific metabolites as potential biomarkers.

Disease Overview

Non-Alcoholic Fatty Liver Disease (NAFLD)

According to the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), NAFLD is a series of conditions occurring when excess lipids build up in the liver causing steatosis and/or inflammation but not due to heavy use of alcohol.[5] NAFLD is considered as the most common form of chronic liver disease in the United States and its estimated prevalence rate had increased from 18% in 1991 to 31% in 2012.[6]

A wide variety of risk factors can result in NAFLD, including overweight or obesity, metabolic syndrome (MetS), and dyslipidemia. Overall, insulin resistance is considered as the primary metabolic defect that leads to this fatty liver disease.[7] Elevated insulin levels and impaired insulin sensitivity result in a positive energy balance, which

expands the storage of triglycerides in adipocytes. This also induces low-grade adipose tissue inflammation, which increases lipolysis and intrinsic fatty acids release. Free fatty acids produced by adipose tissue travel in vessels until they are taken by the liver; this uptake process is a hepatic response to the high level of lipolysis and insulin resistance in a body. As the lipolysis level in the peripheral adipocytes elevates, hepatic cells react to the influx of free fatty acids and tend to generate triglycerides (TG) in hepatocytes. The function of liver in very low density lipoprotein (VLDL) generation is also interrupted due to the decrease of apolipoprotein B-100, and this impacts the packaging and shipping of triglyceride in the liver. Instead of distribution to the body, the TG remains within the liver under this condition.[8]

As lipid accumulation increases, the level of β-oxidation in the hepatocytes also elevated, which leads to mitochondrial dysfunction. During this up-regulated fatty acid metabolism process, more reactive oxygen species, including H_2O_2 and O_2• (-), are produced at the electron transport chain, and they increase intracellular oxidative stress and lipid peroxidation. This oxidation reaction, therefore, connects to cell apoptosis and the release of malondialdehyde and 4-hydroxynonenal; both of the secreted compounds promote cell inflammation and collagen synthesis, which can induce hepatic fibrosis.[9,10]

Summing up the facts, the presence of lipid droplets in hepatocytes, accompanied by mitochondrial dysfunction and lipid peroxidation, introduce cell apoptosis and higher oxidative stress. At this stage, liver inflammation and fibrosis can occur, leading to severe malfunction and other complications.[8]

Within the etiology of NAFLD, two separate sub-classification, simple steatosis and non-alcoholic steatohepatitis (NASH), have been recognized. The former condition is defined by a greater than 5% liver triglyceride accumulation with no or little inflammation and hepatocyte damage. Simple steatosis is relatively benign and can be reversible. About 10-20% of the NAFLD patients will progress to more severe complications that eventually develops into NASH; symptoms such as hepatic inflammation and injuries are included. NASH has characteristics of hepatitis in addition to accumulation of excess TG within the liver. The inflammation in NASH is due to the exacerbation of insulin resistance and the abnormal lipogenesis in the hepatocytes. In this condition, the liver has a higher oxidation and cell apoptosis rate. If the hepatitis continues, it will eventually cause fibrosis or scarring, and can further lead to liver carcinoma and cirrhosis[11].

Several comorbidities are associated with NAFLD, including type 2 diabetes mellitus (T2DM), metabolic syndrome (MetS), and cardiovascular disease (CVD). CVD is the general term for all types of diseases that affect the heart or blood vessels, and CAD typically refer to the heart disease caused by blockage of the coronary arteries. Hence, CAD is considered as a subtype of CVD, while CVD covers a wide range of vessel diseases and cardiac symptoms. Research showed that NAFLD serves as an independent risk factor for CVD.[12] NAFLD patients are more likely to develop CVD than those without NAFLD; the significance still persists after adjusting confounding variables, such as glycemic control and overweight. Atherosclerosis and coronary calcification, which results in CAD, was also found to be connected with NAFLD and has a higher

prevalence among NAFLD patients.[13] Several factors are suggested to be involved in the development of CAD among NAFLD patients. These include a higher oxidative stress level and systemic inflammation caused by liver steatosis, and the lipotoxicity due to adipose tissue lipolysis and its resulting free fatty acids. Studies also found a higher rate of endothelial dysfunction among people with NAFLD.[14,15] The mechanisms that connect NAFLD and CAD will be explained more thoroughly in the following section.

Coronary Artery Disease (CAD)

According to the National Heart, Lung, and Blood Institute (NHLBI), CAD is the most common type of heart disease in the United States. Initially, plaque builds up in the coronary artery and narrows the blood vessels, and this process is also referred to as atherosclerosis. The waxy clot hardens overtime and might ultimately partially or completely block the blood flow to the heart, leading to angina or myocardial infarction (MI).[16] Obstructive CAD is defined as an obstruction of greater than 50% in the left main coronary artery or a 70% or greater obstruction of a major coronary vessel.[17]

The risk factors of CAD largely overlap with those of NAFLD, including MetS, hypercholesterolemia, obesity, and T2DM.[13] Former studies suggested that elevated levels of low-density lipoprotein cholesterol (LDL-c) and lower high-density lipoprotein cholesterol (HDL-c) levels are closely related to atherosclerosis and CAD.[18] Cholesterol oxidation enhances atherosclerosis and plaque formation. When the LDL-c in blood vessels enters through the endothelium, it can be oxidized by the free radicals and is turned into pro-inflammatory lipids. After detecting the oxidized components, monocytes and macrophages migrate to the inflammatory spot and devour the oxidized LDL-c by

endocytosis. Through the consumption of oxidized LDL-c, these white blood cells gradually transform into lipid containing foam cells. Foam cells adhere to the intima when cell apoptosis happens, and the buildup of lipid droplets-contained foam cells at the endothelial site is considered as plaque formation or atherosclerosis. The existence of plaques can trigger the endothelial cells to release multiple cytokines and pro-inflammatory compounds, such as interleukin-8 (IL-8) and toll-like receptor 2 (TLR2). Plaques also encroach the lumen of the arteries, causing narrowing and reduction of blood flow. Once a plaque ruptures, the released pro-coagulant factors will cause an acute blood coagulation, at which it generates blood clots that exist in the circulatory system and increase the risk of obstruction or stroke.[19,20]

Recent findings also associate insulin resistance with CAD among non-diabetic patients, which might be due to a lower secretion of anti-inflammatory components.[21] Former research also introduced several pathways that connect insulin resistance with atherosclerosis. In a state of insulin resistance, nitric oxide is down-regulated and its anti-inflammatory and vasodilation effect is suppressed. Whereas the, mitogen-activated protein kinase (MAPK) pathway, which has a pro-atherogenic effect, may be enhanced due to higher inflammation levels in the body. Also, insulin receptors are present on the surface of circulating monocytes and macrophages. The body's status of insulin resistance thus impacts macrophage apoptosis; foam cells are more likely to accumulate and grow into plaques when a patient has insulin resistance.[21,22]

Combining the biological mechanisms described above, the combination of an elevated pro-inflammatory response with depressed anti-oxidative functions in the body

plays an important role in both NAFLD and CAD progression. Meanwhile, insulin resistance is a possible factor that can lead to the two diseases. The following paragraph describes observational studies to highlight the relationship between and severity of CAD among NAFLD patients.

Association between NAFLD and CAD

The association between CAD and NAFLD was mentioned in previous research highlighting an increase of carotid plaques and CVD prevalence among ultrasound-diagnosed NAFLD patients.[3,23] An increased risk of nonfatal CVD events was shown to be associated with NAFLD, and the relationship was independent of common risk factors such as body weight, lifestyle, and blood lipid profile[24]. A smaller study found that the prevalence of coronary calcification and plaques obstruction are significantly higher in NAFLD patients.[14] These cross-sectional studies considered NAFLD as a predictor of CVD and heart disease.

Results from prospective studies further confirmed the relation of NAFLD and CAD. Zeb et al. conducted a cohort study with 4119 participants and an average follow-up interval of 7.6 years. Results showed that NAFLD was associated with non-fatal CHD incidence, and was independent from other CHD risks factors such as T2DM and dyslipidemia.[25,26] Another large population based study tested the coronary artery calcification (CAC) score to measure calcium deposit level and stiffness of an artery. Calcification of vessel intima or media is a feature of atherosclerosis, it can reflect the narrowed ratio of arteries and indicate plaque accumulation levels.[27] Results showed that increased CAC score is associated with NAFLD, in both unadjusted and risk variable

adjusted models.[28] As most of the studies focused on the incidence of coronary artery events and the one-time measure indicators, Park *et al.* looked into the development and progression of CAC.[29] This retrospective cohort study collected both initial and follow-up CAC scores of the participants. Those with NAFLD at baseline had higher CAC development rate than those without NAFLD. A larger prospective study by Sinn et al. also yielded similar results, which showed a higher increase in carotid intima-media thickness (CIMT) score and coronary plaques in persistent NAFLD patients throughout the 3.3 years follow-up duration[30]. Combining these lines of evidence, it can be concluded that NAFLD has an independent association with coronary atherosclerosis and can be considered as one of the risk factors of CAD.

The biological pathway between NAFLD and CAD is still unclear, but it is recognized to be multifactorial. The progress involves multiple mechanisms, including dyslipidemia, lipotoxicity, oxidative stress (free radicals), and inflammatory response.[13] Overall, however, excessive calorie intake and unbalanced energy expenditure are believed to be the major causes of NAFLD. The over consumed energy turns into adipose tissue and stores in the body. When an individual is obese, enlarged adipose cells enhance lipolysis and induce lipotoxicity; this further increases inflammatory reactions and signaling. The accumulated fat also increases the secretion of adipokines, which is a series of pro-inflammatory hormones that can affect vascular and immune functions.[31] In short, lipotoxicity and insulin resistance are linked to liver fat accumulation, while the widely promoted inflammation status by adipokines and cytokines is showed to be a trigger of free radicals and reactive oxygen species (ROS) synthesis.[32] A body that

contains a larger amount of free radicals and ROS is more likely to develop CAD, since these unstable atoms can attack serum cholesterol and lipoproteins. Several enzymes in liver, such as superoxide dismutase and glutathione peroxidase, are responsible for processing ROS and reducing oxidative damage. However, the presence of excess lipid in the liver of an individual with NAFLD, will cause elevated hepatic cytokines (hepatokines) levels, which accompanied by adipokines and cytokines, exacerbating insulin resistance and inflammation in the body.[33,34] The ability of the liver reducing free radicals might also be limited; lipoproteins are more likely to be oxidized and produce plaques.[35,36] At this point, the reduced function of the liver becomes a determining factor of CAD in NAFLD patients. Thus, based on epidemiology studies, clinical reviews, and supported by mechanistic connections, NAFLD patients are more likely to develop CAD than normal people.

Nutrition Overview: NAFLD and CAD

The etiology of NAFLD and CAD involves glucose homeostasis, fat metabolism, and inflammation, which are factors closely related to diet. A previous review investigated possible dietary patterns that are linked to the disease, and overnutrition was considered as a cause of the imbalance.[37] Overnutrition refers to a malnutrition status, in which the body is consuming more energy or nutrients than necessary. Excessive carbohydrate, fat, or energy intakes are some of the factors that can influence body metabolism and induce NAFLD and CAD.[37] In a case-control study, over consumption of total calories was found to be associated with the incidence of NAFLD.[38] A more recent

article also compared the daily total energy intake using a case-control study design, and overall, NAFLD patients consumed more calories than healthy controls.[39] The mechanism behind overconsumption is that, when energy conditioned are adequate, excessive calories are transformed into adipose tissues. Accumulated body or visceral fat might lead to obesity and other NAFLD triggering pathways, such as inflammation and decreased insulin sensitivity. As previously described in the overview of NAFLD, unrestricted storage of fat results in breakdown of triglycerides and lipotoxicity; these free fatty acids released into systemic circulation are more likely to be taken up by the liver, as compared to other organs and tissues. In contrast, Zelber-Sagi et al. demonstrated that calorie intake was not different between groups of NAFLD patients and healthy controls.[40] However, the body mass indices (BMIs) of the NAFLD patients were significantly higher, which indicates that though they did not consume excess calories during the study, their increased body weight had become a risk factor for NAFLD.

In terms of the nutritional contents of the diet, previous research has observed a high-fat diet with increased total fat intake was observed among NAFLD patients. It was believed to be related to the abnormal distribution of fat in the liver.[38,41] Researchers further investigated the composition of dietary fat and found that the types of fatty acid in participant's diets can be linked to insulin resistance and fatty liver.[40,42,43] In a review article by McCarthy et al., the adverse impact of saturated and trans fatty acids on lipid and glucose metabolism were linked with worsened NAFLD and atherosclerosis.[42] A similar article by Zivkovic et al. supported this theory and suggested that dietary

consumption of saturated fat is associated with insulin resistance.[37] In a cross sectional study, the consumption of saturated fat was found to be higher in participants with NASH, the hepatitis caused by fat accumulation and the more severe form of NAFLD.[44] These patients also ingested less polyunsaturated fat than the healthy group. Besides fatty liver disease, saturated fat also has impact on CVD; by replacing dietary saturated fat with polyunsaturated fat, the incidence of CVD was significantly reduced in a previous study.[45]

However, even in studies which did not detect a difference in saturated fat intake in NAFLD populations, differences were detected in consumption of other types of fatty acids. Cortez-Pinto et al. detected a difference in the forms of unsaturated fatty acids that were consumed. NAFLD patients consumed higher amounts of ω-6 fatty acid and they also had a higher ω-6 to ω-3 ratio in their diet, compared to controls.[43] The adverse impact of n-6 unsaturated fat might originate from the pro-inflammatory compound, arachidonic acid, which is generated during lipid metabolism. Arachidonic acid is the intermediate compound of multiple inflammatory pathways and can be the precursor of several eicosanoids that induce inflammation and vasoconstriction, such as prostaglandins and thromboxanes. On the other hand, ω-3 unsaturated fat does not promote synthesis of arachidonic acids and produces anti-inflammatory chemical mediators, which is beneficial in reducing body inflammation. Studies found that increasing ω-3 fatty acids consumption decreases serum triglyceride and improves liver fat contents in NAFLD patients. Eicosapentaenoic acid (EPA) and docosahexaenoic acid

(DHA) are common forms of ω-3 fatty acid supplements, and were also found to reduce insulin resistance in NAFLD.[46]

In conclusion, dietary factors play an important role in body metabolism, and its relation with diseases is also worth investigating. The concept of comparing and identifying human metabolic changes will be explained in the following paragraphs.

Current Problems and Diagnostic Difficulties

CAD has long been called the silent killer, due to its symptomless development and the acute attack. The artery disease has characteristics of endothelial dysfunction and vessel calcification. Thus, the prediction and diagnosis of CAD can be approached with both blood tests and imaging. Results from blood test can be a primary prediction of heart conditions; it is usually based on the inflammation levels in a body. Some proteins that are sensitive to oxidative stress and ROS are used as CAD biomarkers in current clinical practice. At the same time, these biomarkers, such as c-reactive protein (CRP) and vascular cell adhesion molecule 1 (VCAM1), cannot directly reflect atherosclerosis and plaque status. The underlying problem is due to the various cause of body inflammation; multiple conditions can lead to an increased VCAM1 level, and while CAD is one of them, carcinoma and metabolic syndrome are also factors that affect VCAM1 expression. Thus, existing biomarkers are unable to accurately predict CAD or evaluate its severity before any cardiac symptom appears.[47]

Direct diagnosis of CAD features artery imaging and scanning strategies. Computed tomography (CT) scan is an imaging method that assesses CAC levels by looking for calcium deposits and narrowed arteries. Similarly, ultrasound measures vessel

thickness or obstruction levels of vessels with a 3-12 Hz linear array transducer.[48] These imaging methods are non-invasive and have a high sensitivity to plaques, which is useful in diagnosing CAD and detecting the risk of acute heart events[49]. However, these imaging services require professionals to interpret the results and are usually unavailable in general practice clinics. Individuals who have cardiac scans are typically patients with known higher severity levels of CAD and had experienced certain prognostic symptoms, including chest pain and myocardial infarction. The low approachability and prevalence of cardiac imaging make this precise diagnosis method hard to generalize to larger populations. These limitations result in a delayed detection of CAD among symptomless patients and ultimately a higher rate of mortality when cardiac events occur.

In conclusion, CAD is more likely to occur among NAFLD patients, but it is difficult to address before cardiac symptoms develop. Current obstacles in CAD detection come in two domains. The accurate CAD diagnosis method, cardiac imaging, is time consuming and cannot be widely performed as a screening tool for CAD; while the blood test, though is more common in clinical practice and can aid in early detection, relies on non-specific biomarkers that overlap with other metabolic diseases. Therefore, blood testing methods are currently conducted as a risk assessment rather than a CAD predicting standard. If new biomarkers specific to NAFLD related CAD can be recognized, it will greatly improve the early detection of this disease and decrease the CAD mortality rate by enabling the application of treatments earlier in the disease process.

Metabolomics Overview

The exploration of human biological systems has expanded throughout the past decade. The field of molecular biology addresses the four "omics", including genomics, transcriptomics, proteomics, and metabolomics. The "omics" is a cascade of biochemical and biological mechanisms that leads to different phenotypes presented in human bodies. Genomics and transcriptomics are involved in gene expression and regulation. The former category studies gene, their functions, and the factors that can affect gene expression; the later one focuses on the transcriptomes (mRNA) that are the results of actively expressed gene. Both of them are at the upper level of phenotype modulation, indicating the genetic responses and expressions from microarray changes. On the other hand, proteomics and metabolomics are downstream of the resulting phenotype. They investigate proteins and metabolites which are dynamic reflections of gene and the influencing factors. Slightly different from proteins, metabolites are the final products of the gene transcription cascade, which makes them the closest compounds to the studied phenotypes. The changes in the metabolome are also amplified relative to changes in the transcriptome and the proteome in a biological system. Besides, metabolomics are more diverse that other "omics" when reflecting present body conditions.[50] By understanding the influence of the "omics", we will be able to integrate different levels of information and clarify the human biological system.[51,52]

Metabolites are low-molecular-weight chemicals that are usually less than 1000 Da, and these compounds are generally considered as results of the interaction between system's genome and its environment. At the same time, metabolites are not merely the

end product of gene expression; they can also be reactants and intermediates of biochemical reactions that participate in the human regulatory system. Due to the important role of metabolites in body homeostasis, the impact of environmental factors, lifestyles, diseases, and changes in genotypes that influence biological status can be largely shown in metabolite profiles. Therefore, metabolites-related studies look into biological functions at a cellular level and try to recognize the cause-and-effect behind a phenotype.[52,53] Based on this aspect, the identification and quantification of metabolites is referred to as metabolomics. It provides an instantaneous snapshot of the cell physiology, while improving our understanding of pathetical processes in human bodies.

The approaches of metabolomics can be divided into two main categories, targeted and untargeted analyses. The former measures a limited number of known metabolites; it quantifies the concentration of targeted molecules and defines those molecules precisely. The later can be further approached by two methods. One of them is metabolomic fingerprinting, which analyze intact metabolic patterns to identify similarities or differences between fingerprints. The analysis can be approached by using multivariate methods to visualize biological discrimination of different groups or phenotypes.[42] The second untargeted approach is metabolic profiling, which is the identification and quantification of a large number of metabolites, including known and unknown ones that generally related to the metabolic pathways of interest. Profiling usually narrows down to compound identification, in which the method selects metabolites that might contribute to the phenotype and analyzes its structure and composition. Comparing metabolomic fingerprints is useful in building biological patterns and predicting models, while

profiling finds specific metabolites that vary in certain populations. Together, they aid in differentiation among cohorts and discover of new biomarker for phenotypes detection and disease diagnosis[41].

Diet, Nutrients, and Metabolomics

Nutrition-related metabolomics is a relatively new field, making large strides in the past decade. The connection between diet and metabolites starts from an insight of body digestion and metabolism.[56] The food sources that are consumed by a biological body undergo digestion and absorption. According to different status and necessity, the substances from the food sources participate in various biological mechanisms to maintain the interior homeostasis, which includes degradation of the original molecules, chemical modifications to allow for transport within the body to distant sites, involvement of metabolic pathways, and synthesis of required compounds. This series of metabolization generates different intermediates or products that are considered as metabolites. These small biological remnants are distributed throughout the body, and present in serum, urine, and feces. Thus, the substances that are found in the body or in the microbiome can largely reflect what had been consumed in the diet. Tracing back these metabolites might provide information about the diet and the current condition of the body.

A metabolomics review done by Jin et al. provides evidence on this concept.[57] People consuming the Mediterranean diet, which has features including high unsaturated fats and vegetable oils (olive oil) intake and limited saturated fats, sweets, and red or processed meat consumption, showed different metabolic patterns than those following

the regular Western diet. Specific metabolites in their serum were at a significantly different level than the control group. Bondia-Pons et al.[58] suggested that plasma phosphotidylcholine levels were lower after participants adopted the Mediterranean diet pattern. The diet intervention placed a protective impact on CVD events, and was reflected in the metabolites concentration presented in the blood. This fact implied that the CVD pathways are linked to certain diets, and metabolomics can potentially interact with them. The Mediterranean diet served as a factor that impacted metabolite levels in this study. Similarly, harmful factors can also be reflected in the metabolomic profiles. Thus, metabolites specific to diets, diseases, or risk factors can be identified as potential biomarkers. In this case, NAFLD and CAD are diseases that are highly related to diet and lifestyles, such as a high-saturated fat diet and excessive energy consumption, which may reflect in certain metabolites levels in the patients. Detecting the metabolic difference and identifying the key metabolites are the critical steps to achieve risk clarification and early diagnosis.

In short, metabolites are snapshots that reflect the instant status of a body; they represent the biochemical reactions that are happening or have happened. Risk factors, such as lifestyle, environment, and diet, influence metabolite distributions and concentrations. This further implies that metabolites can become potential indicators of diseases. Thus, the identification of metabolites specific to certain phenotypes can improve current procedures in disease diagnosis that better account for variations in disease phenotypes and medication effectiveness. It can also aid in tracing back possible causes of a disease, such as diet. The current study is an untargeted analysis that includes

both fingerprinting and profiling. It focuses on identifying the metabolic differences among NAFLD patients with CAD; aim to recognize metabolites that are specific to different severity levels of CAD and define them as potential biomarkers.

Liquid chromatography–mass spectrometry (LC-MS)

LC-MS is a combination of liquid chromatography (LC) and mass spectrometry (MS). It attempts to separate the chemical content in a mixture and define the molecules by their masses. LC-MS, is one of the analytical techniques that is commonly used in metabolomics studies. It has advantages of having a high resolving power, a high sensitivity, and a wide applicability to sample types. According to the data yielded from LC-MS, the metabolites in a sample mixture can be identified by comparing their metabolic patterns and conducting statistical analysis based on their quantified data.

Liquid chromatography

The first part of LC-MS is the liquid chromatography. Metabolites in the samples are carried by solvents that are known as the mobile phase; together, the mixture travels through the solid phase, which is a column packed with adsorbent materials, and are separated due to the affinity difference between the metabolites and the column. There are two kinds of mobile phase applied in a LC-MS run, one with higher polarity and is more hydrophilic, the other is at a lower level of polarity and is primarily an organic solvent. In usual runs, the initial mobile phase is highly hydrophilic; the proportion of the hydrophobic mobile phase gradually increases over the course. The chromatography run ends when the mobile phase is entirely hydrophobic; thus, the eluted compounds come in

a gradient from mostly hydrophilic (or polar) to mostly hydrophobic (or non-polar). Initially, the aqueous mobile phase goes through the column; organic and hydrophobic compounds in the serum are trapped in the stationary phase column due to a higher affinity to carbon. Later, as the concentration of the organic mobile phase gradually increases, the affinity of the metabolites for the mobile phase also slowly rises. The adhered molecules in the column are eluted by the organic mobile phase eventually. These substances arrive at the end detector later than the hydrophilic molecules. This separation based on their affinity to the stationary phase is the main goal of LC. It is an important step for composition analysis and component identification.

As a sample mixture loads the column in LC, the ability of separating the compounds is associated to the resolution power of the column. Decreasing particle sizes of the stationary phase is a way to increase resolution, since under a certain column volume, smaller particles have more surface area than larger particles, which brings better resolution and reduces coeluting analytes.[59] Traditional LC relies on gravity to force the mobile phase flow through the column. In order to yield a better separation and decrease the column particle size, though the power of separation might be improved, the flow rate will be reduced due to the strong resistance force (back-pressure) from the column. Thus, high performance LC (HPLC) and ultra-performance LC (UPLC) are introduced to resolve this problem. Both of these methods results in a higher resolution by pressurizing the mobile phase. But the particle size of the solid phase usually needs to compromise with the consistent flow in the HPLC system; columns packed with smaller particles have higher resolution but a delayed retention time, while larger particle size columns do not

separate metabolites effectively. This leads to limitations in column options and retention efficiency. UPLC, as an update version of HPLC, allows the column to be at a smaller packing size (less than 500 Da) and has special pumping techniques that can withstand the high back-pressure.[60] The introduction of samples in UPLC requires lower volume than HPLC and the injection cycle is also faster; these further allow the column to be smaller and shorter. Together, the characteristics of UPLC as compared to other types of analytical chromatography include high speed, good resolution, and minimal required sample volume.[60,61]

Mass-spectrometry and Tandem Mass-spectrometry

Mass spectrometry (MS) is an analytical method that ionizes the chemical compounds and sorts them by their mass-to-charge (m/z) ratio. Once separated, qualification or quantification analysis can be conducted based on the m/z ratio and the abundance of the compound.

The m/z ratio is created by ionization of analytes; different compounds generate unique ionic species after they pass through the ionization source and the m/z ratio can be considered to be representative of each species. The ionized compounds travel through the mass spectrometer, in which various mass separation methods are applied; common process includes passing through electric fields for mass selection and accelerated in a field-free space to distinguish travel distance between masses. The metabolites eventually arrive at the bottom detector, where they are counted and documented to generate MS data. This result with information of the abundance of ionic species is known as the intensity of the respective m/z ratio. Signals and spectra detected at the end of the MS are

presented as peaks in a two-dimensional plot known as a mass spectrum; m/z ratios are on the x axis and signal intensities are on the y axis. These spectra are used to determine the elemental or isotopic signature of a sample, the masses of particles and molecules, and to elucidate the chemical structures of molecules and other chemical compounds. Usually, ion peaks that have higher intensities than the background noise are considered as the intact ionized molecules. However, the analytes or molecules in a sample might not be fully ionized when they elute from the ion source; it creates neutral species which are eliminated in the mass spectrometer when selecting ions. This process is defined as the neutral loss. Since the neutral species are filtered out at the beginning of the run, and the detector in mass spectrometry can only sense charged analytes, the distance between peaks shown on the mass spectrum indicates that it contains neutral loss of an ionized compound. The unique patterns that metabolites generate due to its ionic and neutral species can be presented in the mass spectrum; this information can be used for both metabolomic fingerprinting and profiling.[62]

MS/MS, or tandem mass spectrometry, is an analysis with multiple mass spectrometers involved. Fragmentations are induced between sections to breakdown the intact molecules. It creates fragmented species that can be analyzed in the following spectrometer. Thus, the process of MS/MS starts with the same UPLC for compound separation. Then, the samples are ionized and travel through the first mass spectrometer to select the metabolites with the targeted m/z ratio; these compounds are later fragmented and go through the second mass spectrometer. The m/z ratios of the fragments are determined at this stage and yield the MS/MS spectrum. MS/MS is an

analytical method that aids in mass determination and composition assessment; the coped mass spectrometers provide fragmentation and metabolites structure identification, which are analysis that MS alone cannot approach.

MS and MS/MS Instrumentation

The composition of a typical mass spectrometer can be divided into three parts, the ion source, the mass analyzer, and the detector.

To fully conduct a LC-MS, it requires an ion source to produce gas or condensed phase ions for further mass analysis; thus, transferring the molecules eluted from UPLC into ionic species is critical. Electrospray ionization (ESI) is a common ionizer used in LC-MS, while atmospheric pressure chemical ionization (APCI) is also frequently seen. Both of the ionization sources are conducted under atmospheric environments and are part of the atmospheric pressure ionization. APCI nebulizes and vaporizes the solvent eluted from the LC; the gaseous sample and solvent are ionized by a corona discharge and injected into a mass analyzer after acceleration and desolvation. APCI is able to ionize non-polar compounds, but in view of the high temperature at the vaporization unit, the analytes must be thermally stable and volatile.[63] In contrast, ESI does not require high heat vaporization. The electrospray that connects to the LC can provide high voltage that ionizes the liquid phase analytes and solvent. The ending tip of the electrospray forms ionic droplets that later are treated with nitrogen gas and an elevated temperature in order to remove the solvent and further reduce the droplet size.[64,65] As the droplet reaches a point at which the surface ions are able to be ejected into the gaseous phase, the emitted

ions enter the skimmer cone and are then accelerated into the mass analyzer for analysis of molecular mass and measurement of ion intensity[66,67].

The mass analyzer is the second part of a mass spectrometer, it accepts the gas phase ionic products from the ion source, and its purpose is to separate or filter the ionized analytes according to their m/z ratio. The combination of multiple mass analyzers in sequence is defined as MS/MS. A common configuration of mass analyzers is the Quadrupole Time of Flight mass analyzer (QToF), where a quadrupole is followed by a time of flight analyzer for MS/MS.

A quadrupole is an ion path that is made up of four parallel rods; these rods are charged during the mass analysis and can select or filter ions. To do so, a certain range of m/z ratio is given before the run, and the ion that meets this standard will travel through the quadrupole, hitting the detector at the end. Other ions with unstable trajectories will collide with the rod and be eliminated. By applying different voltage to the rods, the quadrupole can perform different function when conducting mass analysis. An alternating current voltage is used when MS is running; both stable molecules and oscillating ions, which are unstable, can pass through the quadrupole, yielding a broad range of ions from the injected sample. On the other hand, when direct current voltage is combined to the quadrupole, it creates a mass filter function. Only ions with certain masses can pass through the rods under this status, and those that do not have the right mass will discharge on the rods. This function is applied when conducting MS/MS, in which the quadrupole selects the targeted molecule and the second mass spectrometer performs mass analysis.[68,69]

Within a QToF system, after traveling through the quadrupole, the selected ions arrive at the collision cell. This hexapole collision cell in the QToF is where collision-induced dissociation (CID) happens. Ionic species are applied with an electrical energy that increases the ion kinetics in the cell. The ionic compound collides and increases inner energy; eventually, the accumulated force breaks down the chemical bonds of the ionic molecules.[70,71] This process can also be considered as fragmentation; it generates both ionic and neutral fragments from the parent ion species, breaking down the structure for further mass analysis. MS only analyzes the original molecules and fragmentation is unnecessary. Thus, both the quadrupole and the collision cell function as an ion guide during MS, which avoids the breakdown of intact molecules. While in MS/MS, the first mass spectrometer filters out undesired masses and maintains molecules with an optimal m/z, followed by the CID in the collision cell. The second mass analyzer detects the fragmented molecules and performs mass analysis. During the MS/MS run, the collision cell applies different voltage to collide the original species. Ions with higher m/z ratio usually require higher energy; commonly used voltage levels are 10, 20, and 40 eV.

The orthogonal acceleration-ToF spectrometer, which contains an ion source, a flight tube, and a detector, is the second mass analyzer in the instrument. This method involves pulsed ion extraction, which extracts short packets of ions from the source periodically and is a technique to connect continuous ionization and ToF analysis.[72] Ionic fragments that are released from the hexapole collision cell move into the orthogonal accelerator as an ion beam. The ions are accumulated and later released into the ToF tube. Initially, the voltage of the accelerator remains 0V, creating a field-free region that allows

the ions to move in their original direction. Then an injection pulse voltage applies to the plate, resulting in an electric field and pushes the ions in an orthogonal direction. As the high voltage releases the ions, the orthogonal accelerator is refilled with new ion species for the next impulse. Flight cycles end when the ion with the highest m/z ratio hits the detector; a new cycle begins at this point with the plate generating voltage and pushing the ions forward again.[72]

The purpose of the orthogonal accelerator is to correct the initial energy difference in the ionic species and gives the ions an equal electrical pulse to travel through the mass analyzer. The velocities and time of flight of the ion fragments are determined by their m/z ratio, and this results in mass separations.[66,73] The detector at the end of the QToF identifies the m/z ratio and the abundance of an ion species. The electrons that reach the detector go through three sections, the microchannel plate, the scintillator, and the photomultiplier tube. The signals from an ion species are first multiplied and freed to strike the scintillator; photons are then emitted as a response to the strike. The number of photons is determined by the intensity of each ionic compound. Finally, the photons are amplified and transfer back into electrical signals for m/z identification and quantification. This conversion of ion signals can alter the electrical voltage of the species, in order to adapt the fields in the flight tube and the photomultiplier tube. The detected signals are shown in the MS or MS/MS spectrum as peaks that represent certain m/z raios.[74]

Mass Spectrums and Chromatograms

Since MS only measures the m/z of the compound in the sample and MS/MS measures both the compounds (precursor ions) and their fragments (product ions), they give different levels of information in their output spectrums. The detector in the QToF scans almost every second, creating multiple spectrums that includes the eluted metabolites throughout the mass spectrometry process. The output of LC/MS is a chromatogram with mass spectrums. It can be explained as a 3D plot containing facts about the compounds. The chromatogram is created by the UPLC; it includes the intensity and retention time of all the ionic species in a sample. And since each scan from the QToF creates a spectrum that includes the eluted compounds, these 2D spectrums can be considered as a snapshot of the eluted content at the time point. Thus, MS spectrums contain the m/z ratio and the intensity level of the intact metabolites. While in MS/MS, in addition to the chromatogram and the MS spectrums, another series of data containing the spectrums of the product ions is collected. The MS/MS spectrums present peaks that represent the fragments from the precursor ions. Different compounds break differently in the collision cell and yield unique spectra patterns. As a result, the MS/MS spectrums can be considered as the profiles of the targeted precursor ions, and compound identification is conducted according to these identical spectrums by comparing the break-down pattern of known molecules in the library databases against the breakdown pattern of unknown molecules in a sample.

SPECIFIC AIMS AND PROJECT OVERVIEW

CAD among NAFLD patients often develops in a gradual and silent process, and different severity levels of CAD might present unique metabolic patterns. Analyzing these patterns will aid in addressing the disease. Moreover, recognizing the metabolomic profile and metabolites that are unique in certain CAD stages can benefit in early diagnosis and disease prevention. Thus, the goal of the current research is to identify specific metabolites that reflect certain CAD severity levels. To approach this goal, there are two sections of investigation contained in this research project, including a comparison of metabolic finger prints and an untargeted analysis with compound identification. The aims are listed below:

Specific Aim 1: Use metabolomics to compare and contrast serum metabolites among NAFLD patients with varying severity levels of CAD

Serum samples from NAFLD patients are extracted with organic solvent and analyzed by LC-QToF. The chromatograms are statistically processed and the samples are compared to identify molecular features that differentiate each stage.

Specific Aim 2: Use MS/MS to identify stage specific CAD-associated metabolites in NAFLD patients with CAD

Targeted MS/MS analysis of the molecular features identified in Specific Aim 1 are used for feature identification.

INTRODUCTION

The most common form of heart disease, CAD, , starts with the deposition of plaques in the coronary arteries. The development of atherosclerosis narrows the blood vessels, which further leads to a blockage of cardiac blood flow and cardiac events, such as MI. The progression of this disease, however, is usually symptomless, making CAD hard to detect until cardiac events occur. The risk factors of CAD include diet, obesity, dyslipidemia, and insulin resistance. NAFLD is also associated with atherosclerosis and heart attacks due to the shared biological pathways.[13,34] Though research has shown that people with NAFLD are more likely to develop CAD, the patients are detected only after CAD has established, and the treatment options at this stage have limited effectiveness, which highlighted the importance of early diagnosis of CAD.[3]

Metabolomics is a field that involves qualification and quantification of metabolites. It can be applied in studies that aim for biomarker identification and metabolic pathway discovery; this approach is beneficial in detecting disease process at an earlier stage. Metabolites, which are low molecular weight chemicals, interact and participate in biochemical reactions in the body system; they are also products from biological mechanisms, which make them a good reflection of the body status. Thus, the presence of individual metabolites or particular combinations of metabolites can imply risks of certain diseases or conditions. Metabolic fingerprinting compares panels of metabolites based on their characteristics, while metabolomic profiling investigates the

composition of chemical compounds. Tracing unknown metabolites and identifying their structure is considered a facet of untargeted metabolomics, which can be approached by analytical techniques such as liquid chromatography-mass spectrometry (LC-MS) and tandem mass spectrometry (MS/MS). These methods aid in disease detection and biomarker development when combined with quantitative methods and statistical analysis.[51]

Metabolomic approaches could be useful in identification of metabolomic fingerprints and certain metabolite abnormalities specific to NAFLD-related-CAD. However, there has been limited research in metabolomics that creates methods to identify the risk and severity of CAD; most of the previous studies focused on discovering biomarkers in the field of proteomics instead. One cross-sectional study examined the serum inflammatory components of NAFLD patients, in order to detect biomarkers of coronary atherosclerosis. Results showed that circulating interleukin-6 (IL-6), an immune-glycoprotein, is independently associated with subclinical atherosclerosis prevalence and severity.[75] Another study investigated potential biomarkers of endothelial function, and found that serum endocan, a protein released by damaged endothelial cells, is positively related to atherosclerosis severity in patients with NAFLD and CAD.[76] Proteomics is one approach to identification of the body's condition; it is the study of measuring and investigating proteins. However, the presence of certain proteins does not indicate that they have been activated and are participating in biochemical pathways. On the other hand, metabolomics collects and analyzes end products of biochemical reactions, which shows that metabolites are evidence of the occurrence of certain body mechanisms.

In the past decade, several studies focused on the relation between metabolomics and heart disease. Yannell *et al.* compared the serum metabolites patterns between CAD patients and normal controls, followed by profiling the metabolic biomarkers using LC-QToF and multiple reactions monitoring (MRM).This is a classic metabolomics untargeted analysis which set to discover disease specific metabolites.[77] The levels of choline, carnitine, p-cresol sulfate, triacylglycerides (TAG), trimethylamine N-oxide (TMAO), sphingomyelins (SMs) and phosphotidylcholine (PC) were found to be significantly different in CAD patients. Li and others also did an untargeted metabolomics study for biomarker identification in coronary heart disease (CHD), using LC-MS.[78] They suggested that naphthol, methylitaconate, N-acetyl-D-glucosamine 6-phosphate and carnitine might be the potential biomarkers for CHD diagnosis. Similarly, dicarboxylacylcarnitines was associated with CAD and predicted the mortality rate of MI in metabolomics research by Shah *et al.*[79] Branched-chain amino acids (BCAA) and urea cycle related metabolites were also found to be related to CAD. Though these studies provided a series of biomarkers that are sensitive to CAD and some of the determined compounds shared the same category, the consistency of these results might still be low due to the difference in data analysis methods and MS/MS instruments. Besides, the serum processing method also varied, which can lead to difference in collected metabolite contents. Some common sample preparations included methanol extraction, ethyl acetate extraction, and the dilute-and-shoot method.

Recently, an untargeted metabolomics study involved a large population and a longer follow up duration that documented the CHD incidence. [80] According to the

results, 19 metabolites were found to be positively or negatively associated with coronary heart disease. Seven of the compounds were reported in previous research to be related to cardiac disease or its risk factors; these metabolites include a smoking-related biomarker, 4-vinylphenol sulfate, and renal dysfunction indicators, N-acetylthreonine and N-acetyl-1-methylhistidine. Also, N-acetylalanine, p-cresol sulfate, 1-arachidonoylglycerophosphocholine, and 2-methylbutyrylcarnitine were reported as CHD biomarkers in previous studies. Their presence suggested the importance of amino acid metabolism in heart disease occurrence. The novel dietary-related metabolites found in this study were erythritol and α -linoleic acid, which reflect the risks to which patients with CHD had been exposed to higher artificial sweetened beverage intake and lower unsaturated fat consumption. The two metabolites that are specific to certain heart conditions are influenced by dietary factors which showed difference between CHD patients and the healthy controls. The success in the metabolic analysis in these prior studies confirmed the applicability of the fingerprinting and profiling methods. Overall, the identified biomarkers were related to choline, carnitine, and amino acids, especially BCAA and alanine.

The goal of the current study is to explore CAD-related metabolites in a different population; NAFLD patients. Already at a higher risk of developing CAD, these people with the fatty liver disease have different biological conditions. And as the CAD severity increase, their metabolomic fingerprints might also alter, making noticeable difference between various CAD stages. Comparing the metabolites collected from distinct severity levels can bring better identification of CAD stages. Thus, the objective of this study is to

identify the metabolic patterns and determine CAD stage-specific metabolites. This approach can provide advance disease detection and prevention of CAD in NAFLD patients, while tracing the possible causes of NAFLD and CAD.

MATERIALS AND METHODS

Sample collection

Eighty-three human serum samples were provided by INOVA, with 63 male donors and 20 female donors. The participant recruitment and sample collection was approved by the INOVA IRB. All the participants who were included in this study were diagnosed with NAFLD by ultrasound. Further, they received an ultrasound scan in the carotid artery to determine the stenosis level of the vessels. Based on the number of obstructed arteries and the stenosis percentage, the samples were divided into five groups during serum collection. The diagnoses and classifications were performed at INOVA; then, the samples were sent to our facility for analysis. The classification criteria are listed below:

Table 1 Group classification criteria

Group	Carotid stenosis level description
NAFLD (n=21)	Normal carotid vessels without stenosis
CAD 0 (n=15)	Vessel stenosis level between 10 and 69%
CAD 1 (n=29)	One vessel greater than 70%
CAD 2 (n=13)	Two vessels greater than 70%
CAD 3 (n=5)	Three vessels greater that 70%

NAFLD group (N group) contains NAFLD participants without vessel stenosis. The other CAD groups are participants having different CAD severity levels in addition to NAFLD. CAD 0 is the group with one vessel that was 10 to 69% carotid stenosis. While CAD 1, CAD 2, CAD 3 refers to one, two, and three coronary arteries having greater than 70% stenosis, respectively.

Serum extraction

Serum metabolites were extracted by following existing protocols.[81] Initially, 450μL of serum sample, an equal amount of ethyl acetate, and 1.5μL of reserpine were placed in a 2 mL microcentrifuge tube. After the completion of aliquoting, the mixture was vortexed for 30 seconds followed by a 10 minutes centrifuge at 17,000 x G (Fisher Scientific accuSpin Micro 17) and the upper organic layer was collected. This extraction procedure was repeated once, and the combined serum extract was vacuum-dried and stored at -20° C. Before LC-MS analysis, a resuspension was performed by adding 60μL of acetonitrile to the serum extract. The solution was filtered by a 3mL syringe and a 4mm 0.45μm nylon filter into a 2mL autosampler vial and was set for analysis by LC-MS.

Aim 1 Methods

LC-MS analysis

Samples were analyzed via an Agilent 1260 Infinity HPLC coupled with an Agilent 6530 Accurate Mass Q-ToF running in ESI positive mode. The HPLC was equipped with a C18 column (Zorbax Eclipse XDB-C18, 2.1 x 100 mm, 1.8μm), and the column temperature was maintained at 50°C. The total collecting time was 25 minutes, and the

injection was followed by a needle wash. Each injection volume was 5μL. For the solvent used in LC, mobile phase A (aqueous) was water with 0.1% formic acid, while mobile phase B (organic) was acetonitrile with 0.1% formic acid. During the 25 minutes run, the flow rate was 0.4 mL / minute, and the dispensed gradient started from 80% A solvent and 20% B solvent at 0.5min to 3% A solvent and 97% B solvent at 15 minutes. The LC eventually ended at 25 minutes, when the solvent composition going back to 80% A solvent and 20% B solvent to re-equilibrate. The ionization mode in the Q-ToF was ESI positive. Nebulizer pressure was 20psi, with a drying gas temperature at 350℃ and a sheath gas temperature at 300 ℃. For the gas-flow rate, the drying gas-flow rate and the sheath gas-flow rate were 8 L/min and 6 L/min, respectively. In the LC-MS analysis, the collision cell was not turned on, leaving the voltage at 0 without any fragmentation. Mass range was set between 100 to 1500 m/z. The MS scan rate was 1 spectra/s. Reference ions were 121.0507 and 922.0098.

MS Data processing

Pre-processing

The chromatograms and the MS spectra were viewed using the Agilent Mass Hunter Qualitative Analysis B.06.00 software. For data analysis and calculation, the MS data collected from the LC-QToF was converted into XML files by MSConvert software (ProteoWizard Tool). Then, they were pre-processed using mzMine 2 (version 2.37); the steps included mass detection, automated data analysis pipeline (ADAP) chromatogram deconvolution, alignment, and gap filling. To detect peaks and create extracted chromatograms of eachfeature, the noise level was determined by the first scanned

spectrum and was set at 1500. The retention time window was between 1 and 24 minutes. The minimum group size in number of scans was 12, and the minimum peak intensity was set at 5000. The m/z tolerance was 0.01m/z. During deconvolution, the signal-to-noise ratio (s/n) threshold was 15. Considering the peaks in the spectra, minimum feature height was set at 5000, with the coefficient/area threshold set at 100. For sample-wise alignment and gap filling, the m/z tolerance and weight were set at 0.015 and 0.17 m/z, respectively. Retention time tolerance was 0.5 minute and its weight was set at 5. The processed dataset was exported into an Excel spread sheet with data including m/z values of the metabolites and their retention time and intensity. Each metabolite can be seen as a feature; its identity is the m/z value and the retention time, and its abundance is the peak intensity value (area) of all the samples. Any missing value was presented as 0.

MS FLO

Duplicates and isotopes detection was done by the online resource, MS FLO. The m/z tolerance was set at 0.01 m/z in both duplicate and isotope detection, while the retention time tolerance was 0.5 and 0.02 minutes respectively. For removal of duplicates, the peak height tolerance was 100 m/z, and the minimum peak match ratio was 0.85. The tolerance settings for the adduct joiner were 0.01 for m/z and 0.02 minutes for retention time. The initial adducts were set at M+H, while all of the final adduct options were selected, including M-H2O+H, M+NH4, M+Na, M+CH3OH+H, M+K, and M+CAN+H. In the MS FLO processed file, duplicates and isotopes that were 100% matched were removed automatically. Other potential duplicates and isotopes were flagged and marked with their percentage match. Then, the data was examined manually to remove these flagged

subjects. For two metabolites that were flagged as duplicates, the one with a lower average intensity was eliminated. However, for isotopes, only the one that meets both a higher m/z value and a lower average intensity was deleted.

Frequency Calculation

The total frequency (f_t) is the occurrence of a detected metabolite (feature) among the total samples. Likewise, the cohort frequency (f_{co}) is defined as the occurrence of the metabolite (feature) in each CAD group or the N group. It was calculated by the following equation:

Equation 1 Frequency:

$$f_t = \frac{number\ of\ occurences\ of\ a\ metabolite}{total\ sample\ number} \times 100\%$$

$$f_{co} = \frac{number\ of\ occurences\ of\ a\ metabolite\ in\ a\ cohort\ (group)}{cohort\ (group)\ sample\ number} \times 100\%$$

Then, the dataset went through a frequency cutoff, including a total cutoff and a cohort-wise cutoff. In the total cutoff, metabolites with a f_t less than 4.8% were eliminated, and in the cohort-wise cutoff, those with a f_{co} less than 80% were deleted.

Normalization

Total ion count (TIC) normalization was performed to the data matrix, and the value was timed by a scale factor (=10^6).[82] The TIC equation is listed below:

Equation 2 TIC normalization:

$$y = \frac{I_i}{\sum_i Ii} \times 10^6$$

In which, *y* is the normalized intensity value and *I* is the original intensity value. $\sum_i Ii$ is the sum of all the metabolite intensities in a sample. After this normalization, outliers among samples of the same metabolite were detected by the interquartile method[83,84] and replaced by the median of this metabolite category. Missing values were also addressed and imputed by the median.

Multivariate Analysis

Principle component analysis (PCA) and partial least square-discriminant analysis (PLS DA) were conducted for dataset dimension-reduction.[85–87] The analysis was done by R-statistics and the samples were presented in a 3D plot containing three major principle components.

Metabolite Selection

Fold Change

The cohort absolute $\log_2$ fold change (F)[88,89] of each detected metabolite was calculated as the formula below:

Equation 3 Fold change:

$$F = \left|\log_2 \frac{CAD\ group\ median}{N\ median}\right|$$

If the $\frac{CAD\ group\ median}{N\ median}$ was less than 1, an inverse value ($\frac{N\ median}{CAD\ group\ median}$) was used instead.

Statistical Analysis

To identify stage specific features, each CAD group (0-3) was compared with the NAFLD Group (N) using a univariate Student's t-test. At the same time, given the number of features in a metabolomics data set, a false discovery rate (FDR) correction was performed as the multiple testing correction method. All statistical analysis was conducted via XLSTAT and R-statistics; the resulting values (p) were considered statistically significant when $p < 0.05$.

Weight Score

When comparing a CAD group with the N group, a novel equation was applied to the data matrix to yield a weight score (W) of each feature. The formula is shown below:

Equation 4 Weight score (1):

$$W = F \times (|f_N - f_C| + 1) \times |f_N - f_C| \times |\log p|$$

The absolute value of the frequency difference between cohorts was presented as $|f_N - f_C|$; in which f_N is the frequency of the N group and f_C is that of a CAD group. If $|f_N - f_C|$ was found to be $\leq 0.5 \times f_N$, an alternative equation was applied; in which the formula was multiplied by f_C instead of $|f_N - f_C|$:

Equation 5 Weight score (2):

$$W = F \times (|f_N - f_C| + 1) \times f_C \times |\log p|$$

Since each CAD group (CAD0, CAD1, CAD2, and CAD3) was compared with the N group, the features in each comparison pair were ranked by their *W* values. Following this ranking, the top 20 features with the highest *W* values in each of the 4 CAD groups were selected, collecting a total of 80 potential stage-specific features.

Manual evaluation

The selected features then underwent a manual evaluation, in which features with f_{CO} values less than 90% (in both CAD groups and N group; see equation 1) and those with $\frac{CAD\ group\ median}{N\ median}$ values less than 2 were eliminated. Additionally, if the original $\frac{CAD\ group\ median}{N\ median}$ value was less than 1, the inverse value was used instead.

Metabolite evaluations

Extract ion chromatograms (EIC) and Box-and-Whisker plots were used as an evaluation method for the selected features. The EIC was obtained by mzMine 2, and the mass

window was set at ± 0.01 to extract the optimal mass. Box-and-whisker plots compared the distribution of samples in a CAD group and the N group based on their intensity of a metabolite. This approach was done by using R-statistics.

Aim 2 Methods

MS/MS analysis

Selected features were analyzed by the MS/MS method to pursue component and structure identification of the targeted masses. The LC-QToF instrument was used with a fragmentation step in the collision cell. The collision energy levels were 10, 20, and 40 eV. For scanning the metabolites, the mass range was 100 to 3000 m/z; the MS scan rate was still 1 spectra / sec, but the MS/MS scan rate was increased to 2 spectra / sec. Each injection volume was 3 μL. The m/z ratios of the targeted metabolites were input into the LC-QToF computer with a window of 0.01 m/z for MS/MS. The range for retention time tolerance was 0.5 minute. Other settings in the LC-QToF remained the same as the LC-MS analysis.

MS/MS data processing

MSConvert software (ProteoWizard Tool) was used to convert the MS/MS data into XML files for further data analysis and calculation. The MS/MS spectra, during a retention time window in which the targeted metabolites were eluted, were viewed by mzMine 2, and the spectrum number that contained the profile of the metabolite was noted. Due to the three collision energy levels applied, there are three spectra files (10, 20, 40 eV) for each analyzed feature.

Metabolite Identification

NIST Database

Targeted feature MS/MS spectra were compared to MS/MS spectra contained in the National Institute of Standard and Technology (NIST) database. Spectra in the NIST library that matched the targeted unknown feature spectra were documented. All three spectra (different collision energy levels) were compared to the NIST database to identify a feature.

Online Databases

Each feature's MS/MS spectra was also used in online database matching. Four online sources were used, including METLIN MS/MS Spectrum Search, The Human Metabolome Database (HMDB) LC-MS/MS Search, Competitive Fragmentation Modeling for Metabolite Identification (CFM-ID) version 3.0 (compound identification), and MetFrag (Web Tool). The precursor (parent/candidate) mass tolerance was set at 25 ppm, and the MS/MS m/z tolerance was at 0.01 Da.

RESULTS AND DISCUSSION

Sample Description

As detailed in Materials and Methods, the total study participant number was 83, with 63 male and 20 female participants. Basic information and anthropometric data were collected and are listed below (**Table 2**) One of the participants had missing values in the weight, height, and body mass index (BMI) categories, while two participants had missing blood pressure values. To complete the statistical analysis, the missing data was eliminated during the calculation of that category. But other data of the participants that had missing values was still applied and underwent the calculation. For the CAD severity stages, the sample numbers in each group are presented in the following table, with the detected feature number in each cohort (selected after a 80% cohort frequency cutoff). (**Table 3**)

Table 2 Participant information and anthropometric data

	N mean	CAD0 mean	CAD1 mean	CAD2 mean	CAD3 mean	Total mean	Total standard deviation	Total range
Age (year)	56	66.20	61.36	66.38	63	61.61	10.07	31-84
Weight (kg)	201.64	202.65	194.99	208	208.16	91.42	16.26	48.64-163.64
Height (cm)	67.97	68.93	68.34	69.86	71.40	174.81	9.63	152.40-200.66

BMI (kg/m^2)	30.26	30.06	29.49	30.05	28.77	29.85	4.81	15.43-48.82
Underweight	0%	0%	3.45%	0%	0%	1.2%		
Normal weight	9.52%	18.33%	13.79%	7.69%	0%	10.84%		
Overweight	52.38%	40%	34.48%	46.15%	80%	44.58%		
Obese	38.10%	41.67%	48.28%	46.16%	20%	43.37%		
Systolic blood pressure (mmHg)	128.71	124.73	129.14	136.38	145.40	130.31	19.25	89-185
Diastolic blood pressure (mmHg)	70.71	68.80	69.39	68.85	78	70.17	9.23	44-89
Elevated blood pressure	4.76%	13.33%	24.14%	7.69%	0%	13.25%		
Hypertension (Stage 1)	23.8%	26.67%	13.79%	53.85%	0%	24.10%		
Hypertension (Stage 2 or higher)	33.33%	26.67%	31.03%	30.77%	80%	33.73%		
Gender	Male: 75.9% Female: 24.1%							
Ethnicity	White (Non-Hispanic): 86.75% African American: 2.41% Asian: 8.43% Others: 2.41%							

All data were collected during the initial visit, except one participant. The range of a category indicates the minimum and maximum value among the available samples.

Table 3 CAD group artery stenosis level and sample number

Artery stenosis level (Group/Cohort abbreviation)	Sample number	Detected feature number
NAFLD (N)	21	4361

10 to 69% (CAD 0)	15	4395
One vessel > 70% (CAD 1)	29	4345
Two vessel > 70% (CAD 2)	13	4345
Three vessel > 70% (CAD 3)	5	4413

Only features that have a cohort frequency ≥80% were included, the number refers to the counted features in that group.

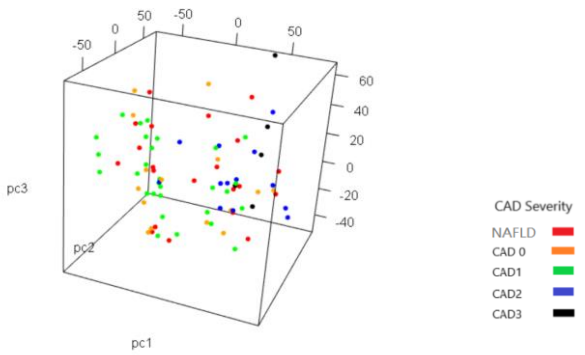
Multivariable Analysis

To address the difference of metabolomic fingerprints among the CAD groups, multivariable analyses, including PCA and PLS DA analysis, were performed. PCA analysis transforms the input data, which are individual variables measured as features and corresponding intensities, into principle axes based on the maximum variation. This operation preserves as much original variation as possible and reveals separations according to the original data matrix. If the within-group variability is sufficiently less than between-group variability, the samples will cluster together by their CAD severity level.

In contrast, unlike PCA which only considers features and their intensities, PLS DA also considers the categorical CAD stages. This discriminated analysis supervises the data by assigning the group categories prior to the data transformation. The major difference between PCA and PLS DA is that the principal components in PCA are based on the direction of maximum variation in the raw data matrix, and those in PLS DA are factors that can maximize the correlation between the data matrix and the CAD scores.[87,90]

In the PCA plot (**Figure 1a**), samples in the same CAD stage group are assigned a representative color and, as seen in **Figure 1**, the entire examined population is separated into two distinct cohorts, but not by the CAD severity. However, the accuracy of the stage classification was unable to estimate. There are possibilities of errors occurring during ultrasound reading and grouping. According to the PCA plot, the total variance explained from the first three principal components was 44.79% (pc1: 24.42%, pc2: 11.55%, pc3: 8.82%). On the other hand, there is a mild separation in the PLS DA plot with samples gathered in cohorts based on the CAD severity levels (**Figure 1b**). The two variables included in the PLS DA analysis, x- and y-variables, were the metabolites data matrix and the assigned CAD stages respectively. The three major PLS components in a PLS DA plot can explain 24.92% of the variance in the x-variables (pls1: 13.10%, pls2: 9.31%, pls3: 2.50%). While in the y-variables, 41.46% of the variance was explained (pls1: 13.72%, pls2: 9.51%, pls3: 18.23%). Error rate in the PLS DA plot was 7.23%.

a.



b.

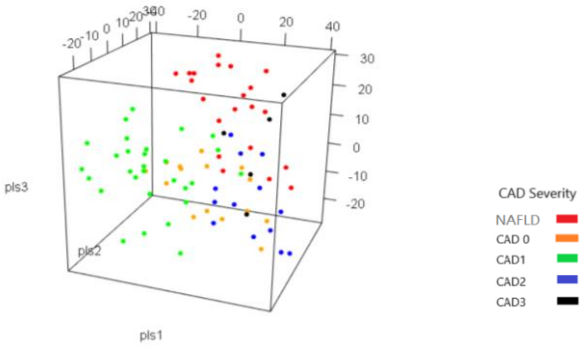


Figure 1 Difference between PCA and PLS DA multivariable analysis

(a) The axes of the 3D PCA plot represent the first 3 principal components (pc1, pc2, pc3). The first principal component (pc1) explains the maximum variation in the data matrix; then, the second one (pc2) continues to explain the remaining variance, and the third one (pc3) follows this pattern. In this PCA plot, the samples in the same CAD group did not cluster together.

(b) The axes of the 3D PLS DA plot are the first 3 partial-leas square component (pls1, pls2, pls3). Their function is equivalent to principal components in PCA plots. In this PLS DA plot, there is a mild separation by the CAD severity levels.

PCA and PLS DA plotting are applied in various "omics" studies and component analysis to detect the difference in baseline patterns between phenotypes.[91,92] Previous research also confirmed the use of PCA and PLS DA analysis for variable and dimension reduction in metabolomics data.[93] Raamsdonk *et al*., who aimed to identify the metabolic pattern difference between gene mutation phenotypes, applied the two analyses to a series of mass spectrums and observed clusters based on their metabolite content.[94] The PLS DA plot in the current study also had better separation than the PCA plot due to the algorithm that maximized between-group variance. For the metabolic patterns unique in heart diseases, a recent study by Gao *et al*. performed PLS DA to visualize the separation between patients with coronary atherosclerosis and non-patients; they observed a separation between groups and further confirmed the metabolic difference of the two

cohorts.[95] Another non-targeted metabolomics study performed PLS DA analysis to a dataset containing serum samples with and without coronary heart disease (CHD).[96] It depicted a clear difference of the metabolite content, which enabled the study moving forward to the selection and identification of important metabolites that distinguish the CAD stages. The cohort separation trend in previous research is consistent with our findings, indicating that heart disease patients present a distinguishable metabolic pattern than normal controls. This confirmed the metabolites variation in the CAD groups can cause the cohort separation, and it is worth investigating when it comes to stage specific biomarker detection.

Metabolite Selection

After calculating the weight score (*W*), the top 20 features with the highest *W* score in each CAD group were selected. This indicates that the selected features were possible factors that differentiate CAD groups from the N group, due to their higher *W* scores than other features. A total of 80 features were then filtered by the cohort frequency (f_{co}) and the fold change score (*F*). Those features which had low f_{co} in both the CAD group and N group were deleted. An ideal biomarker would have high f_{co} in one group and low f_{co} in the other. However, in the present study, most of the selected features have high f_{co} (>90%) in both groups; thus, these features were kept in the selection list and went further to the *F* score filtering.

According to the method, features with a fold change ($\frac{CAD\ group\ median}{N\ median}$) less than 2 were eliminated. The process cut down a significant number of features, and only 11

features met the criteria. All these features had higher intensity levels in the CAD groups than the N group, except features 165.0712 m/z and 166.0879 m/z. And while some of the features were single stage specific, others were distinctive in more than one group. The selected features, their m/z ratio, retention time, and represented stage are listed in **Table 4**

Table 4 Selected features

Features (m/z)	Retention time (min)	Specific CAD Stage
406.1767	9.435	CAD 0
499.2359	9.400	CAD 0
422.2082	8.577	CAD 0 / CAD 1
380.1971	9.435	CAD 0 / CAD 2
165.0712	4.620	CAD 1
166.0879	3.878	CAD 1
563.2302	8.407	CAD 1 / CAD 2
743.2325	0.564	CAD 1 / CAD 3
579.2193	8.134	CAD 2 / CAD 3
423.1977	7.335	CAD 3
624.3059	9.500	CAD 3

Several features were found to represent more than one group. They are factors that might be specific to two CAD stages. All features will be evaluated by EIC and the Box-and-Whisker plot, to confirm their stage differentiating ability.

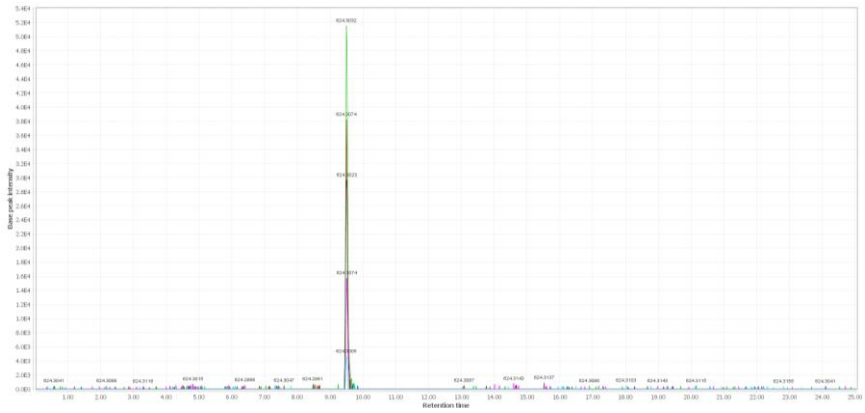
EIC Evaluation

The EICs of the selected metabolites in each CAD stage were visualized using mzMine 2 and are presented in **Appendix 1**. If a metabolite was specific to two groups,

both of the EIC were extracted and compared. **Figure 2** shows a comparison of a good EIC and a bad one. An ideal feature EIC has a symmetric Gaussian distribution peak, with a low noise level. A single peak throughout the entire chromatogram is expected, though there is still a possibility of a peak at several time points due to the same m/z ratio the features have. However, in that case, the peaks should be clearly separated by the retention time. As long as the peaks are not high in baseline, with asymmetrical peak shape and a long tail, having a sudden intensity drop near the apex position, or with a zigzagging shape, it can be considered as a good EIC.[97]

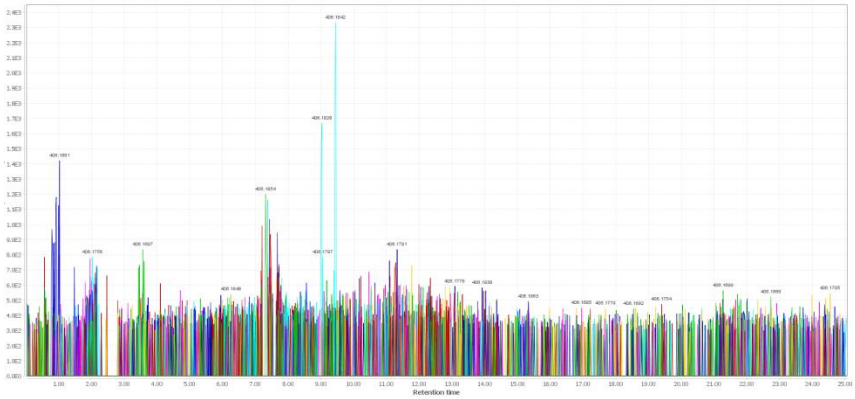
Figure 2 Good and bad EIC comparison
EIC is a chromatogram showing only the extracted peaks; a window of m/z can be defined previously and the EIC is generated by extracting signals that only fit in the m/z range.

a.



(a) A good EIC features a high-intensity and symmetrical peak with a quiet baseline.

b.



(b) This is an example of a bad EIC, in which there are numerous signals which appear as a noisy baseline. It will be difficult to distinguish whether the targeted peak is a real metabolite or background noise.

In the results of the present study, some of the EIC appeared to have 2 peaks at a similar retention time, but one of them was lower in intensity and has a slightly different m/z ratio then the targeted m/z (**Figure 3**). Thus, the higher signal that is closer to the targeted retention time was regarded as the major peak (targeted features). Overall, features 165.0712 m/z and 166.0879 m/z did not yield a good EIC and were eliminated from the selected metabolite list. Another deleted feature, 743.2325 m/z, has a retention time less than 1 minute and a noisy baseline in the EIC. After the removal, 8 features remained as CAD stage specific features (**Table 5**).

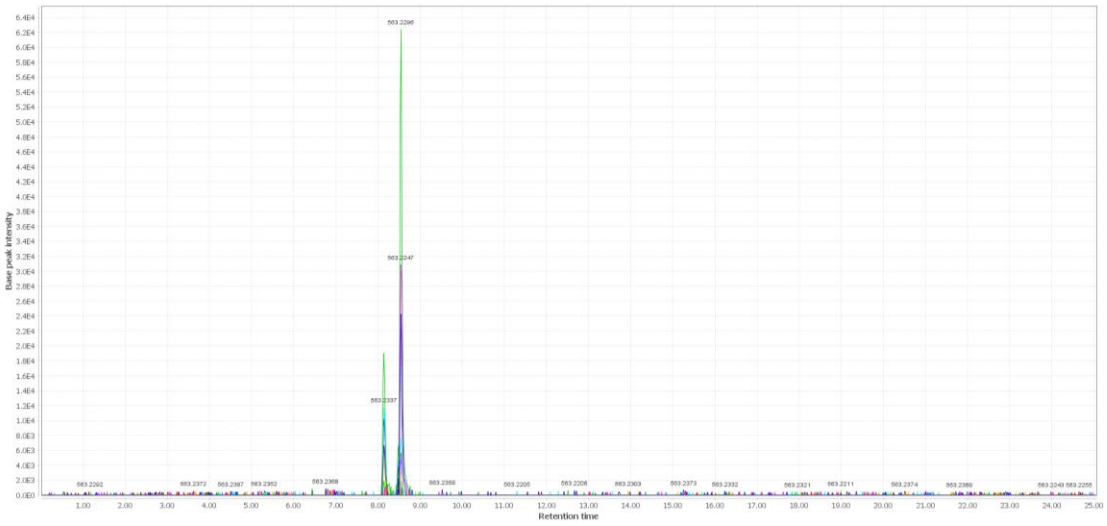


Figure 3 An example of an EIC with 2 peaks
The targeted metabolite is 563.2302 m/z and its retention time is 8.407 minutes. A smaller peak appeared at around 8 minute and its detected m/z is 563.2337, while the higher peak is 563.2296 m/z at 8.4 minute. The latter is considered as the major signal for the targeted feature, due to the higher intensity and the more accurate retention time.

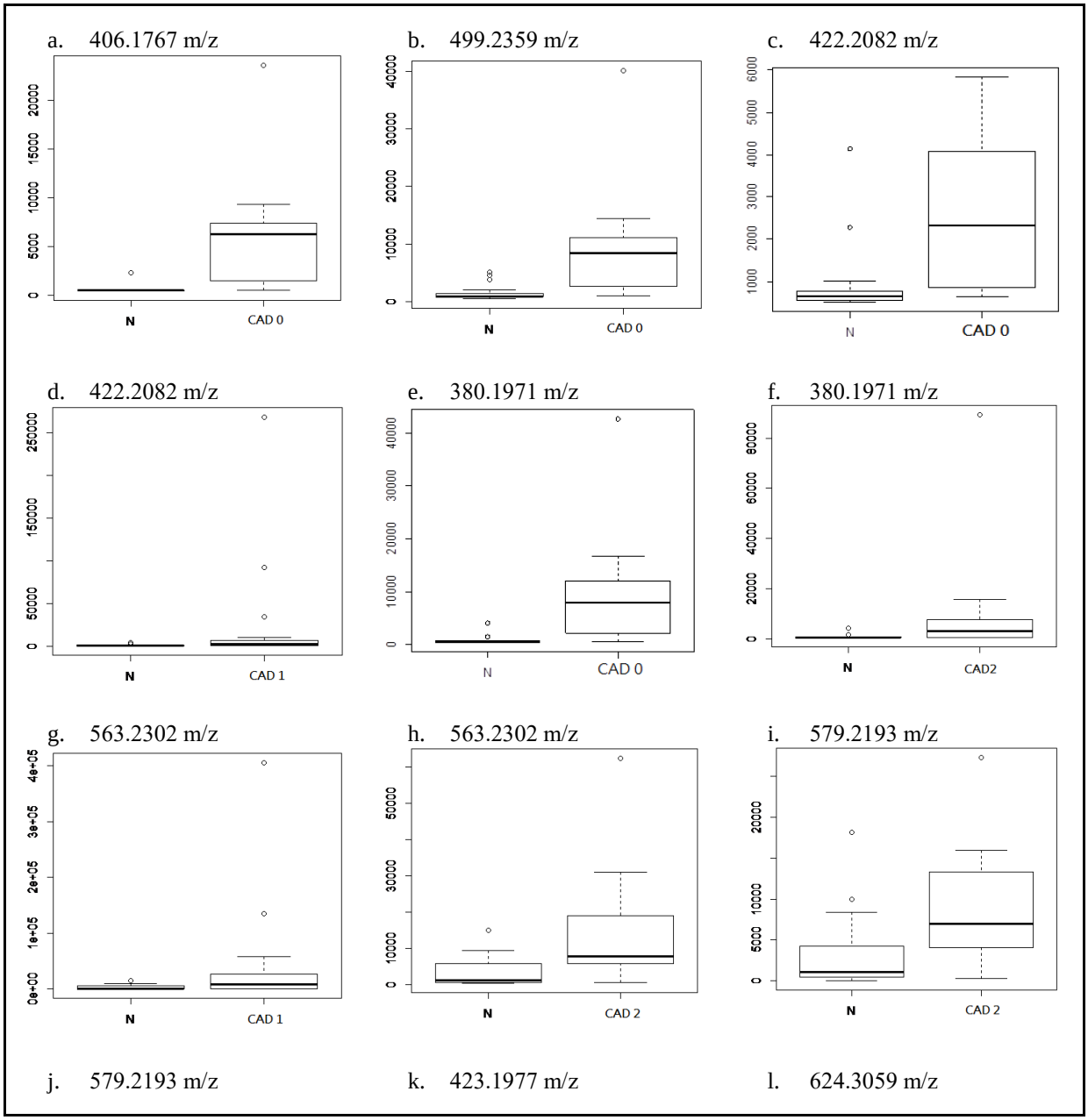
Table 5 Selected features

ID	Metabolite (m/z)	Retention time (min)	Specific CAD Stage
1	406.1767	9.435	CAD 0
2	499.2359	9.400	CAD 0
3	422.2082	8.577	CAD 0 / CAD 1
4	380.1971	9.435	CAD 0 / CAD 2
5	563.2302	8.407	CAD 1 / CAD 2
6	579.2193	8.134	CAD 2 / CAD 3
7	423.1977	7.335	CAD 3
8	624.3059	9.500	CAD 3

Feature No. 3, No. 4, No. 5, and No. 6 were found unique in more than one group.

Box-and-Whisker Plot Evaluation

The second step of metabolite evaluation used Box-and-Whisker plots to visualize the difference between N group and CAD group. All the samples in the groups were included when generating the plots, and the results are included in **Figure 4.** The y-axis in the plot is the metabolite intensity level, and the bold line in the interquartile box represents the median of the group. While evaluating the distribution of samples, the median and the interquartile values are important determinants. An ideal plot consists of a separation of median by intensities, little or non-overlapped boxes, and few outliers. On the other hand, non-significant metabolites will have a Box-and-Whisker plot with characteristics including equivalent median levels and similar interquartile range distribution.


a. 406.1767 m/z
b. 499.2359 m/z
c. 422.2082 m/z
d. 422.2082 m/z
e. 380.1971 m/z
f. 380.1971 m/z
g. 563.2302 m/z
h. 563.2302 m/z
i. 579.2193 m/z
j. 579.2193 m/z
k. 423.1977 m/z
l. 624.3059 m/z
N
CAD 0
CAD 1
CAD2
CAD 2

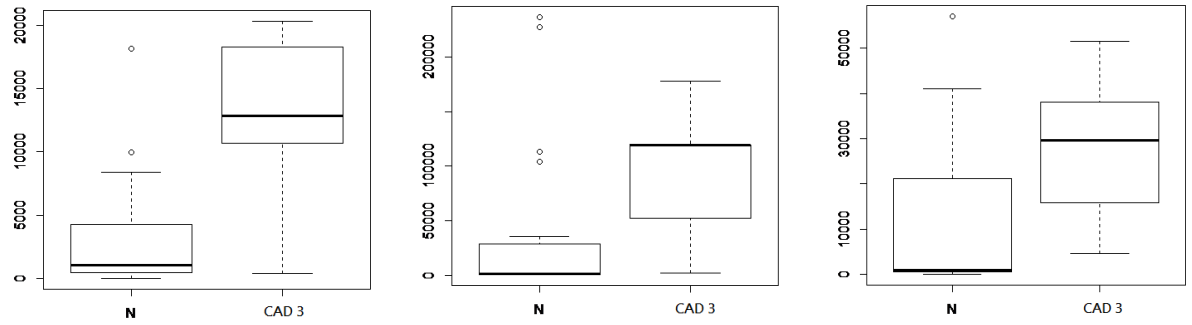


Figure 4 Box-and-whisker plot
In general, the intensity levels in the CAD groups are usually higher than the N group, based on the results. The purpose of a Box-and-Whisker plot is to confirm the ability of a metabolite to identify CAD stages from the N group and serve as a biomarker. Thus, a distinguishable difference in the intensity level between two groups indicates a good CAD stage specific metabolite. (d) 422.2082 m/z is an example of a bad plot for the purposes of identifying biomarkers, with narrow boxes in both groups that have little separation in intensity levels. (j) 579.2193 m/z is a good Box-and-Whisker plot, due to its clear distinction between groups. The two medians almost had a 1000 times difference.

With a separation of N group and CAD group in the box plots, features 406.1767 m/z, 499.2359 m/z, 422.2082 m/z, and 380.1971 m/z were distinctive in CAD 0. Features 563.2302 m/z and 579.2193 m/z are CAD 2 specific features, whereas, the plot of 579.2193 m/z shows some overlaid area. But the median of the two cohorts showed almost a 1000 times difference, so this feature was still included for CAD 2 characteristic. CAD 3 stage has three specific features, including 579.2193 m/z, 423.1977 m/z, and 624.3059 m/z. The first two features have good Box-and-Whisker plot, while metabolite 624.3059 m/z was considered acceptable because of the partially overlapped intensity levels and the separated median. Metabolite 422.2082 m/z and 563.2302 m/z in CAD 1 and 380.1971 m/z in CAD 2 showed a compressed interquartile range in both N group and CAD group; their median levels also presented poor separation, which leads to the elimination of the features.

Finalized Metabolites

After the EIC and Box-and-Whisker plot evaluation, the 8 features remained the same; however, some of them were found to no longer be distinctive in certain CAD stages. The final feature list and corresponding CAD stages is shown in **Table 6**.

Table 6 Finalized features

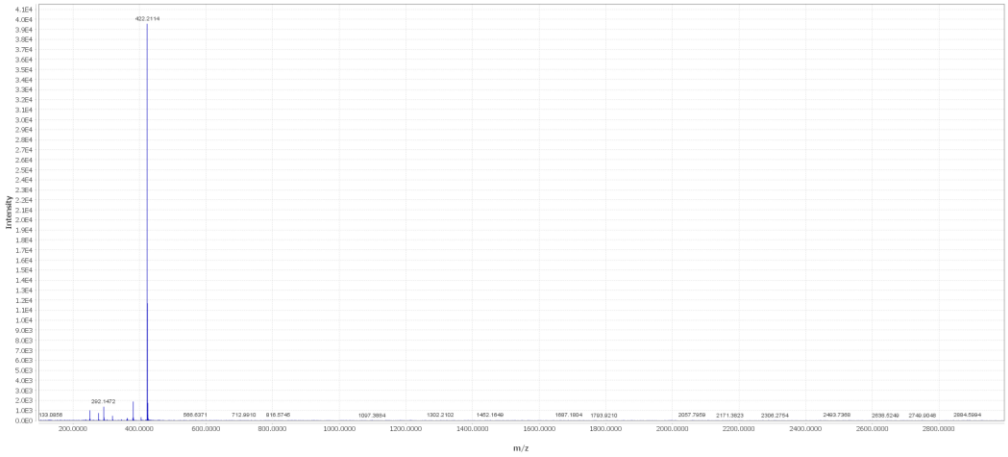
ID	Features (m/z)	Retention time (min)	Specific CAD Stage
1	406.1767	9.435	CAD 0
2	499.2359	9.4	CAD 0
3	422.2082	8.577	CAD 0
4	380.1971	9.435	CAD 0
5	563.2302	8.407	CAD 2
6	579.2193	8.134	CAD 2 / CAD 3
7	423.1977	7.335	CAD 3
8	624.3059	9.5	CAD 3

Among the 8 metabolites, 4 of them are CAD 0 specific, 2 of them are CAD 2 specific, and 3 of them are CAD 3 specific. CAD 3 and CAD 2 also have one metabolite in common. In the finalized list, none of the metabolites can distinguish CAD 1.

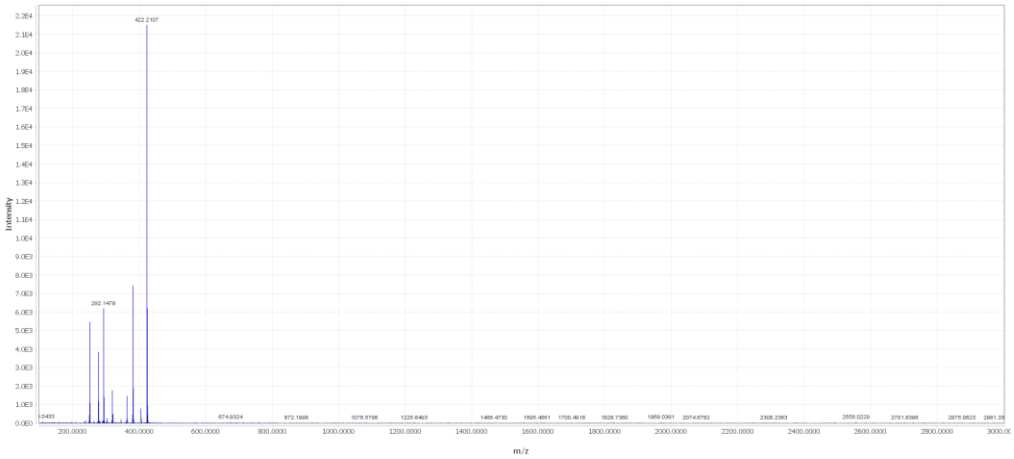
MS/MS Spectra

Results from the MS/MS came in three collision energy levels; each feature yielded three spectra that contained the fragments collected after collision induced fragmentation. **Figure 5** includes the MS/MS spectra of one feature (422.2082 m/z); the results of the seven other features are included in **Appendix 2**:

a. Collision energy: 10 eV



b. Collision energy: 20 eV



c. Collision energy: 40 eV

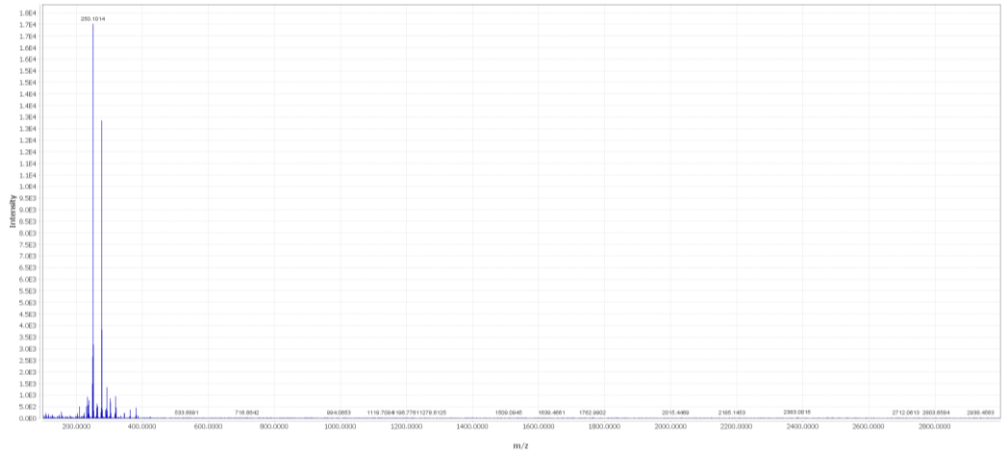


Figure 5 MS/MS spectrums
The parent ion (422.2082 m/z) was targeted and fragmented by the induced energy in the collision cell. (a.) At 10 eV, the ion did not generate many fragments, which indicated that the energy was not enough to break the parent ion structure. (b.) As the collision energy increased, there are several signals appearing on the left of the parent ion peak. These signals are the fragmented ions with lower m/z ratio than the original molecule. (c.) The parent ion was fully broken down at 40 eV, leaving the spectrum without the parent ion peak.

During MS/MS spectrum selection, results showed that three features (563.2302 m/z, 423.1977 m/z, and 624.3059 m/z) were unable to fragment completely even under the highest collision energy level used. Thus, the three metabolites were analyzed again with an alternative collision energies; 5 eV and 60 eV. Their resulting spectra are presented in **Appendix 2**. The 5 eV collision energy was not strong enough to induce fragmentations among the 3 features; the parent ion signals still have high intensity levels, and there are no signals of fragments. While the 60 eV spectra showed high noise levels, indicating that the collision energy was too high that the fragments of the targeted features were lost in the background noises. Overall, 40 eV yields better signals of the parent ions than the other energy levels.

NIST Database Matching

The matching in the NIST database was done by manually evaluating the peaks of the targeted unknown spectra to those of compounds in the database. **Figure 6** shows an example of a good and bad match. In the comparison, the unknown sample spectra is colored in red, while the NIST library spectra is shown in a reversed orientation (mirrored, relative to the unknown spectra) and colored in blue. The x-axis is the m/z ratio and the y-axis is the intensity level (in percentage). An ideal match of metabolite

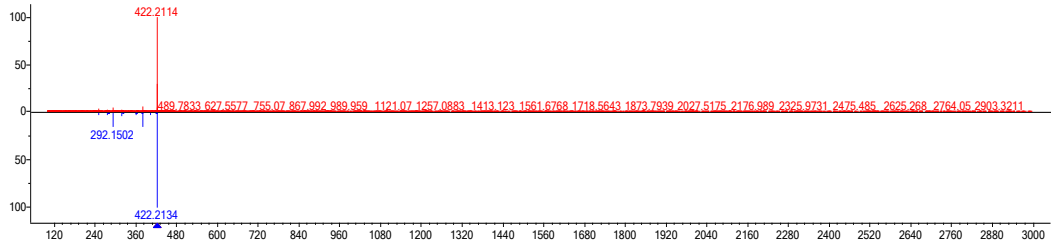
profiles in NIST requires similar m/z ratios and intensity levels, which indicates that the major peaks in the spectrums should mirror with each other. For peaks in the sample spectrums and the library references that are considered as a match, the variance of the matching signals is within 0.1 m/z, and the low intensity signals, as well as the baseline noise, are ignored. In addition, spectrums created from the three collision energy levels were compared at the same time during matching. When yielding a good match, the spectrums of the other 2 energy levels were also extracted and compared with the library spectrum. If the match is correct, the library metabolite might have the same fragmentation trend, which breaks down into fragments with similar m/z, as the unknown metabolite.

Figure 6 Good and bad match (in NIST database) comparison
(a) A good match of the database showed matching in the major peaks (238.1017 m/z in the example) and most of the signals above the baseline. (b) The signals in the library spectrum did not match the unknown peaks, which pointed out that this is not an ideal match.

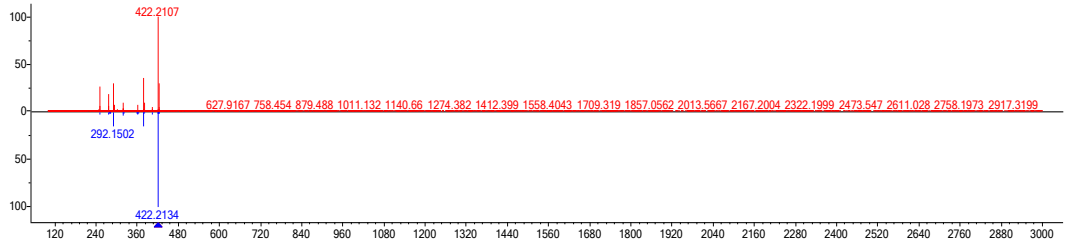
The matches of the 8 metabolites collected in NIST are included in the **Figure 7**. Metabolite 422.2082 m/z had a good match at all three collision energy levels, and the substance is identified as 4-hydroxyatorvastatin lactone. Another match was found in 380.1971 m/z at 40 eV, with identified metabolite 1-[[4,5-Bis (4-methoxyphenyl) -2-thiazolyl]carbonyl]-4- methylpiperazine, a cyclooxygenase-1 (COX-1) inhibitor (FR122047).

a. 422.2082 m/z match: 4-hydroxyatorvastatin lactone

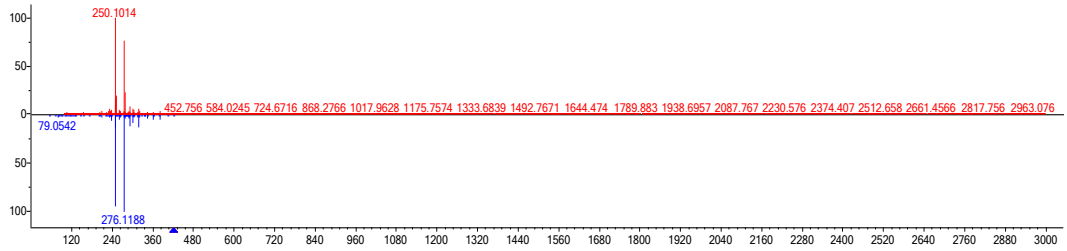
10 eV



20 eV



40 eV



b. 380.1971 m/z match: 1-[[4,5-Bis(4-methoxyphenyl)-2-thiazolyl]carbonyl]-4-methylpiperazine

40 eV

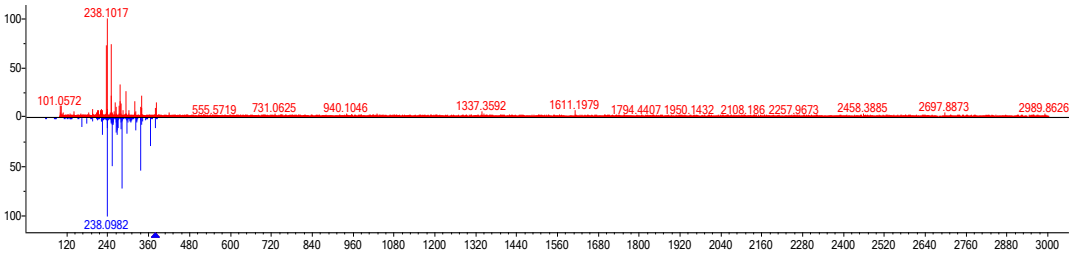


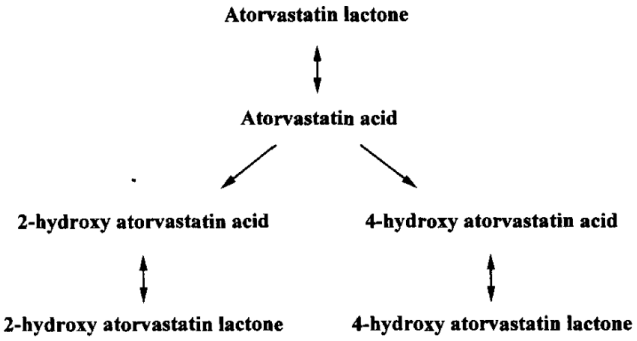
Figure 7 NIST database matches

(a) A match of 422.2082 m/z at all three collision energy levels; the metabolite was identified as 4-hydroxyatorvastatin lactone. (b) The metabolite 380.1971 m/z has a match at 40 eV, and the compound was a COX-1 inhibitor.

Based on the identified metabolites from the NIST database, 380.1971 m/z is considered as a COX-1 inhibitor (FR122047). However, it is unlikely to be presented in human body, since FR122047 is a material for in vitro studies to investigate prostaglandin related pathways; it was also applied for reducing muscle nerve response in human induced heart failure events in animal experiments.[98] Previous research also applied the substance as an analgesic in rats.[11] Despite its wide use in mechanism identification and pain reduction, this inhibitor has not been established as drug nor applied to humans. As a result, though the spectrum matching of FR122047 was good, the identification and verification of this metabolite failed, leaving this substance (380.1971 m/z) still unknown.

Among the NIST database results, 4-hydroxyatorvastatin lactone matched with metabolite 422.2082 m/z at all three energy levels. The known metabolite has a precursor ion of 422.2126 m/z, and the exact mass is 556.23735 Da. According to previous studies, 4-hydroxyatorvastatin lactone is the lactone form of the atorvastatin (Lipitor®), an HMG-CoA (3-hydroxy-3-methyl-glutaryl-coenzyme A) reductase inhibitor. The drug belongs to the second generation of statins and has been widely used to delay atherosclerosis progression and reduce plasma cholesterol levels.[100,101] During the biological metabolism of atorvastatin, lactonization was conducted to generate the lactone form statin, which has better affinity to cytochrome P450 and can be deposited easier.[102] As presented in **Figure 8**., the active form of the drug is the atorvastatin acid. As the body metabolism operates and the liver detoxifies the statin drug, lactonization occurs to transform atorvastatin

acids into atorvastatin lactones.[103] As a result, 4-hydroxyatorvastatin lactone is one of the metabolic products of the atorvastatin drug clearance pathways.



The pathway was adapted from Kantola et al. (1998)

Figure 8 Metabolism of atorvastatin
The identified metabolite, 4-hydroxyatorvastatin lactone, is one of the metabolic products in the atorvastatin elimination pathway.

In the finalized metabolites list (as seen in **Table 6**), the intensity level of metabolite 422.2082 m/z, which was identified as 4-hydroxyatorvastatin lactone, showed a difference between the N group and CAD 0 group. This indicates that 4-hydroxyatorvastatin lactone is a metabolite that can distinguish CAD 0 group from the N group. Based on this assumption and the fact that this metabolite originates from exogenous sources, tracing back the use of atorvastatin in the participants is an important step for determining this stage specific metabolite. The following table shows the proportion of participants using atorvastatin in each group, as reported in the medical history record from INOVA (**Table 7**). In the N group, 9.52% of the people were

assigned this drug, while 80% of the patients in CAD 0 were using atorvastatin. The use of statin drug in the other three CAD groups was 65.62%, 53.84%, and 60% respectively.

Table 7 Use of atorvastatin in the sample population

Group/Cohort	Proportion of use (%)
NAFLD (N)	9.52
CAD 0	80.00
CAD 1	65.62
CAD 2	53.84
CAD 3	60

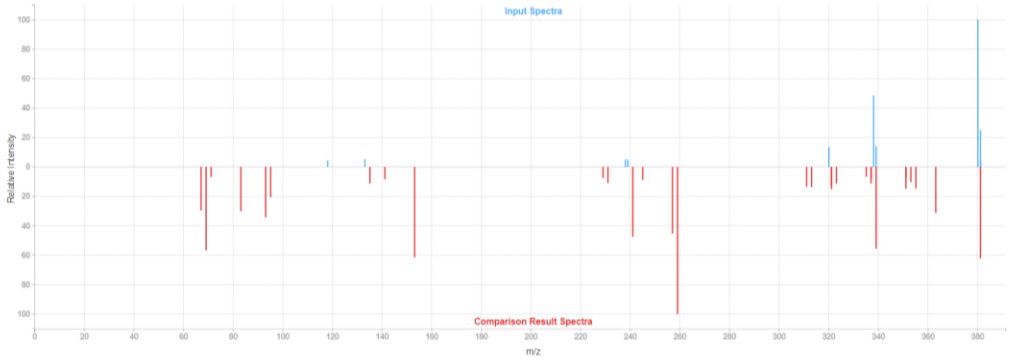
Since CAD 0 has the highest percentage of atorvastatin use and N group has the lowest level, the difference between the two cohorts might contribute to the calculation result and further impact the selection of stage specific metabolites. The possible reason that 4-hydroxyatorvastatin lactone was not detected in as a stage-specific metabolite CAD 1, CAD 2, and CAD 3 can be attributed to the frequency calculation. Those who were not taking atorvastatin might have low intensity levels or missing values of 4-hydroxyatorvastatin lactone; consequently, the frequency of this feature will be low if fewer participants received the drug treatment. Therefore, in CAD 1, CAD 2, and CAD 3, the metabolite might have been eliminated as a candidate stage-specific metabolite during the frequency cutoff and metabolite evaluation steps. On the contrary, the occurrence of 4-hydroxyatorvastatin lactone in CAD 0 was relatively high, and this group showed a difference with N group; due to that, the metabolite was able to remain in the finalized list and detected as a stage determining factor. Thus, the targeted metabolite (422.2082

m/z) was identified as 4-hydroxyatorvastatin lactone by the NIST database, and it can be considered as a CAD 0 specific metabolite in the sample population.

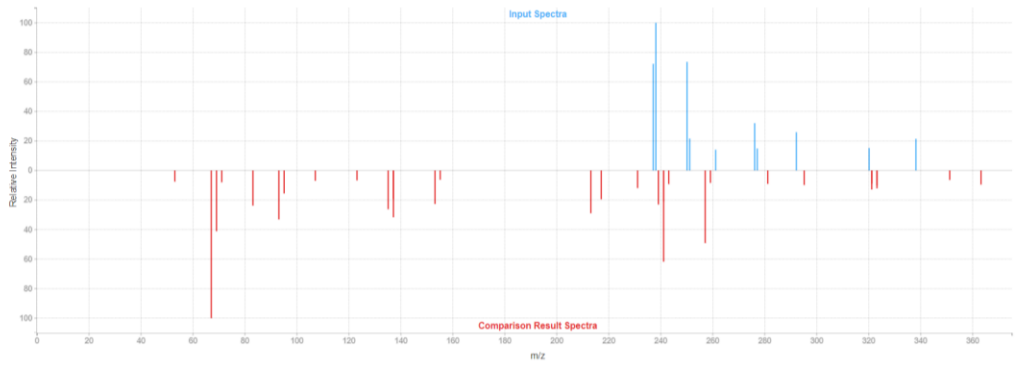
Online Databases Matching

The NCBI database matching results are listed in **Figure 9**. Similar to the NIST database, the results were presented as a symmetrical graph, with blue spectrums indicating the sample peaks and the red ones are the library spectrums. There were several matches for 380.1971 m/z at 20 and 40 eV, and a match for 624.3059 m/z at 60 eV. But overall, the results collected from NCBI are not good matches that predicted the targeted metabolites. Thus, the matches were considered not reliable.

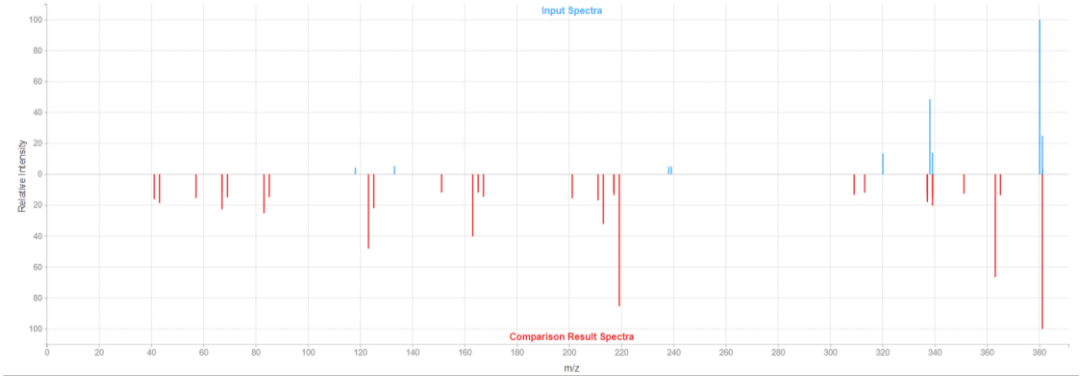
a. 380.1971 m/z match: Kinocoumarin
20 eV



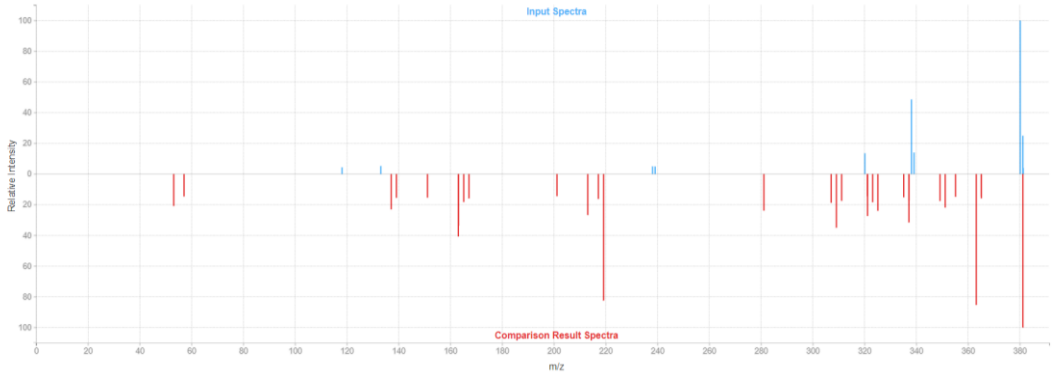
40 eV



b. 380.1971 m/z match: (3'x,5'a,9'x,10'b)-O-(6-Oxo-7-drimen-11-yl)umbelliferone
20 eV



c. 380.1971 m/z match: Conferone
20 eV



d. 624.3059 m/z match: Neferine
60 eV

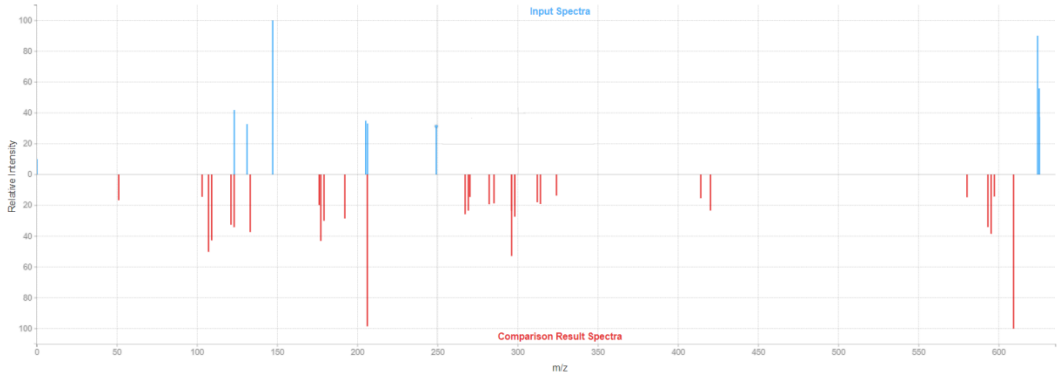
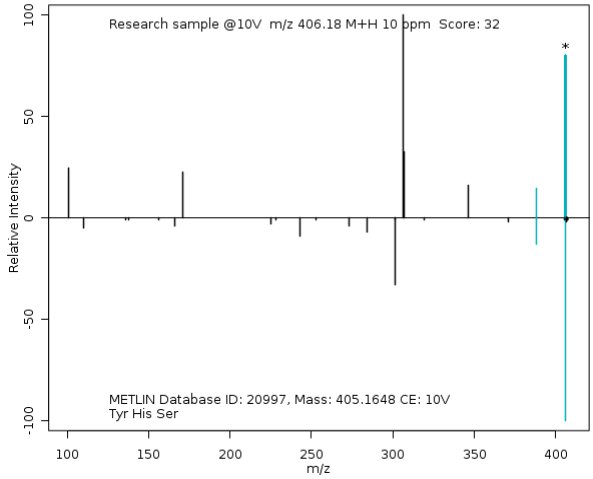


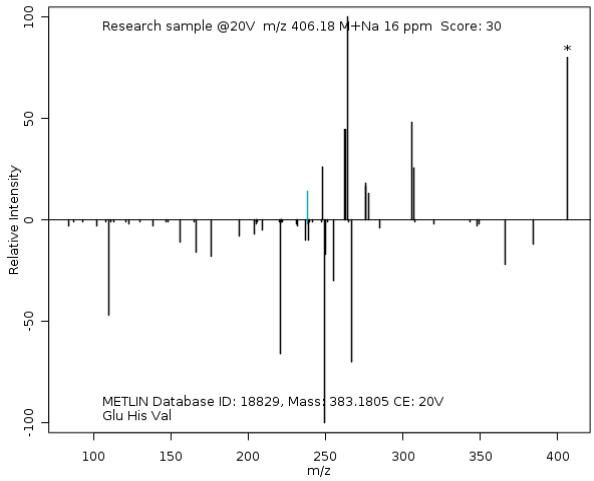
Figure 9 NCBI database matching results
(a) Spectrums of 20 and 40 eV matched with Kinocoumarin, a constituent of Citrus. (b) The match was not a good one, and the identified metabolite is a food compound found in spices. (c) The matched metabolite is isolated from Ferula species. (d) The collision level was 60 eV, but the library spectrum did not match with the sample peaks. Neferine is an alkaloid found in coffee products.

Metlin is another online database that is commonly applied in metabolomics global analyses. In a matching graph, the upper spectrums were the sample signals and the lower reversed peaks were from the library. The matched peaks were colored in blue, while those did not match remained black. According to the figure showing below (as seen in **Figure 10**), most of the matches are peptides containing 3 amino acids. During Metlin database search, each sample spectrum was compared with the library spectrums at 10, 20, and 40 eV. Thus, there are matches that include a pair of spectrums with different collision energy level. However, none of the results were considered as good matches for the targeted 8 metabolites.

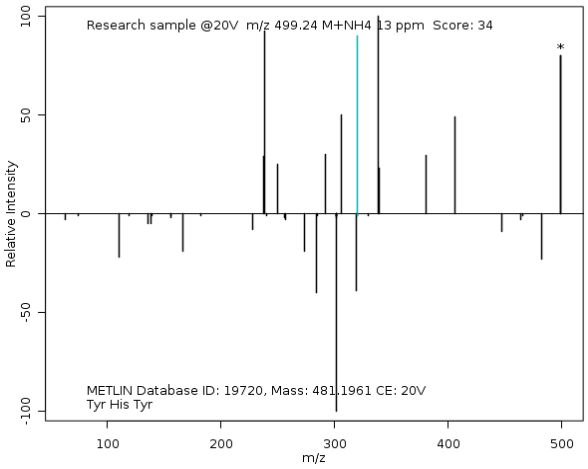
a. 406.1767 m/z match:
20 eV (matched with 10 eV spectrums): Try His Ser



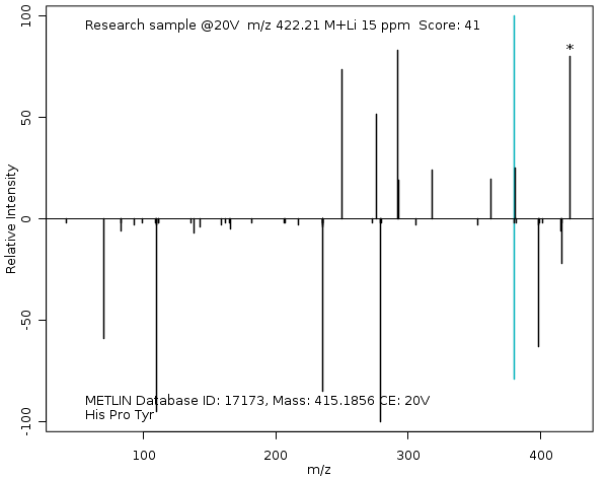
40 eV (matched with 20eV spectrums): Glu His Val



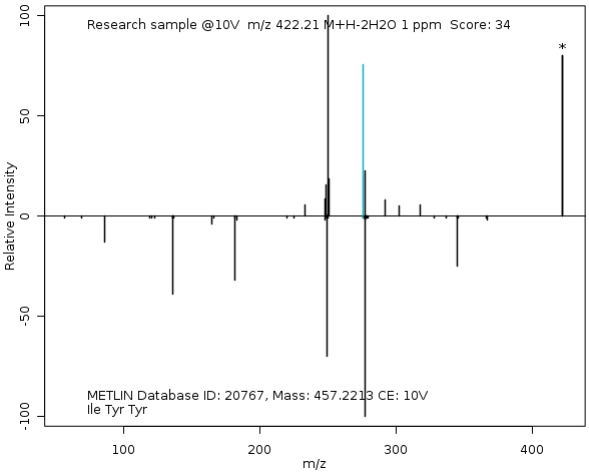
b. 499.2359 m/z match:
40 eV (matched with 20eV spectrums): Try His Try



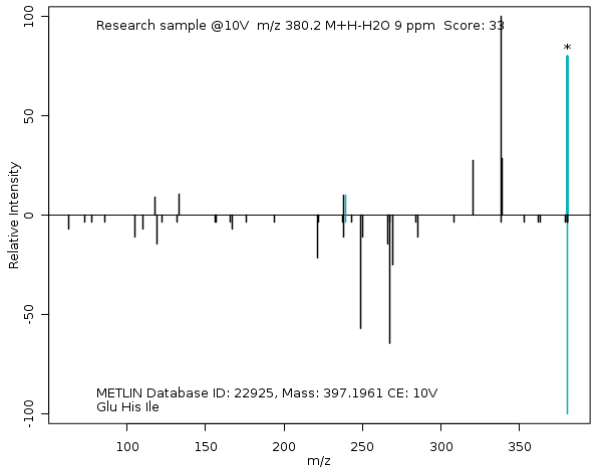
c. 422.2082 m/z match:
20 eV: Hi Pro Try



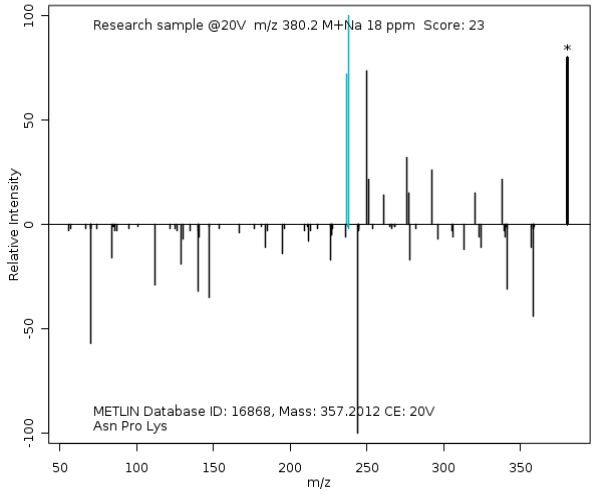
40 eV (matched with 10 eV spectrums): Ile Try Try



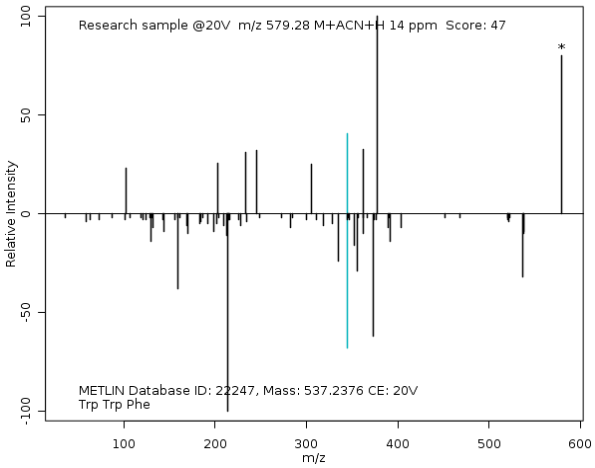
d. 380.1971 m/z match:
10 eV: Glu His Ile



40 eV (matched with 20 eV spectrums): Asn Pro Lys



e. 579.2193 m/z match:
40 eV (matched with 20 eV spectrums): Trp Trp Phe



f. 423.1977 m/z match:
60 eV (matched with 10 eV spectrums): Ile Cys Phe

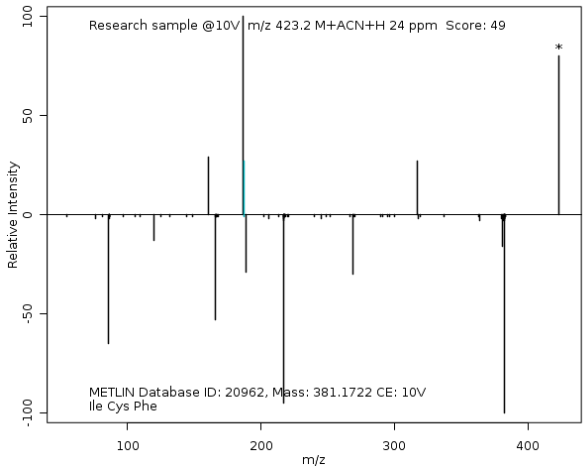
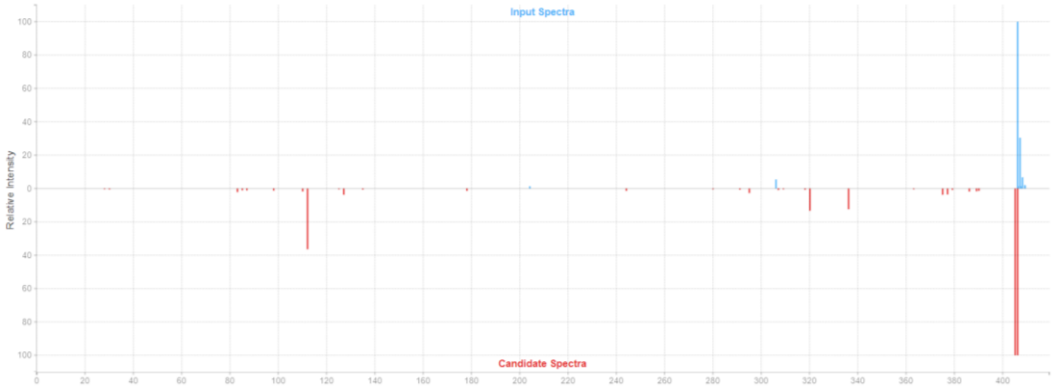


Figure 10 Metlin database matching results
(a) There is a match of 20eV sample spectrums with 10 eV library spectrums, Try His Ser, in which the parent ions are at the same m/z. But there is no good match at 40 eV. (b) Only one of the signals matched. (c) One signal with 380 m/z matched the library peak at 20 eV. The 40 eV spectrums do not yield good matching results. (d) The parent ion matched at 10 eV, but as the collision energy increases, the fragments do not have a good match. (e) The sample spectrum at 40 eV has a match with a 20 eV spectrum, Trp Trp Phe. But the major peak at 378 m/z in the sample does not have a match in the library spectrum. (f) This is a match of 60 eV and 10 eV, the difference between the collision energy and the less similar patterns refer that it is not an ideal match.

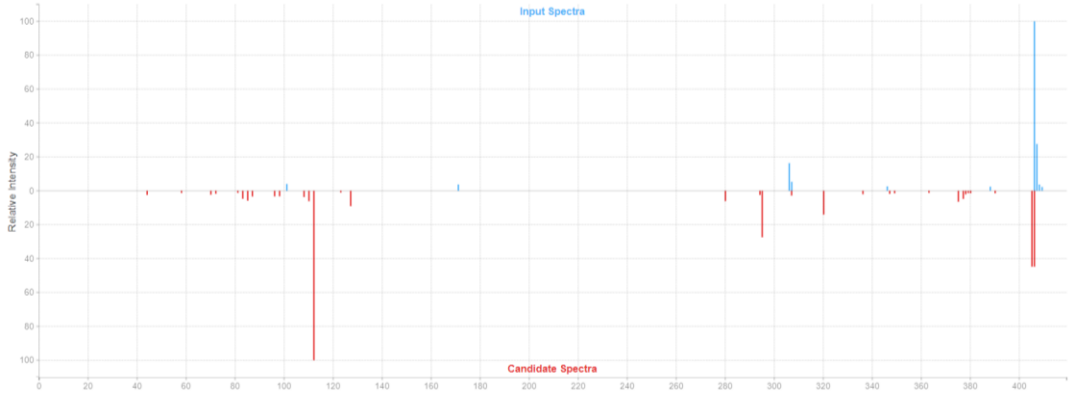
CFM ID is another online matching resource that compares three energy levels at the same time. The output of a matched metabolite contains 3 graphs, with the input spectrums colored in blue and the database spectrums colored in red. The spectrums were also symmetrically arranged to compare their similarities. Results collected in CFM ID are presented in **Figure 11**, but none of the matches were considered good:

a. 406.1767 m/z match: SGI-1776

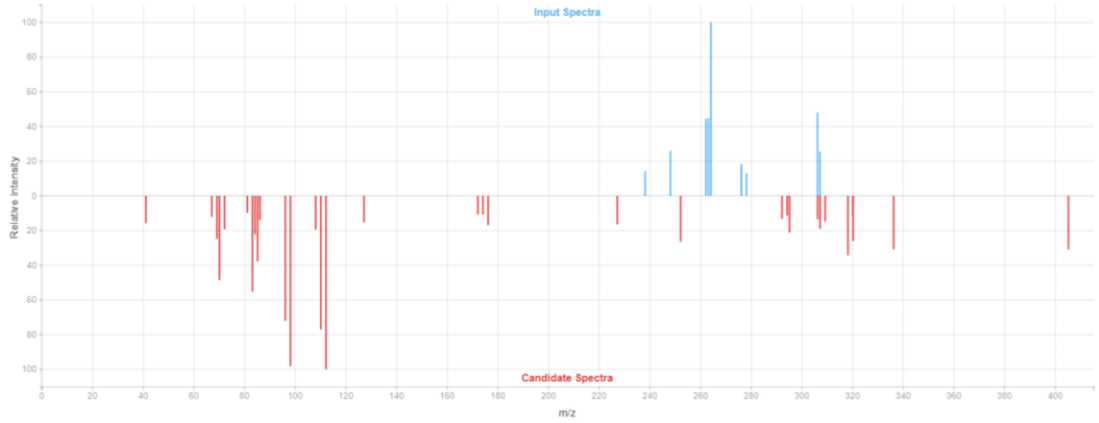
10 eV



20 eV

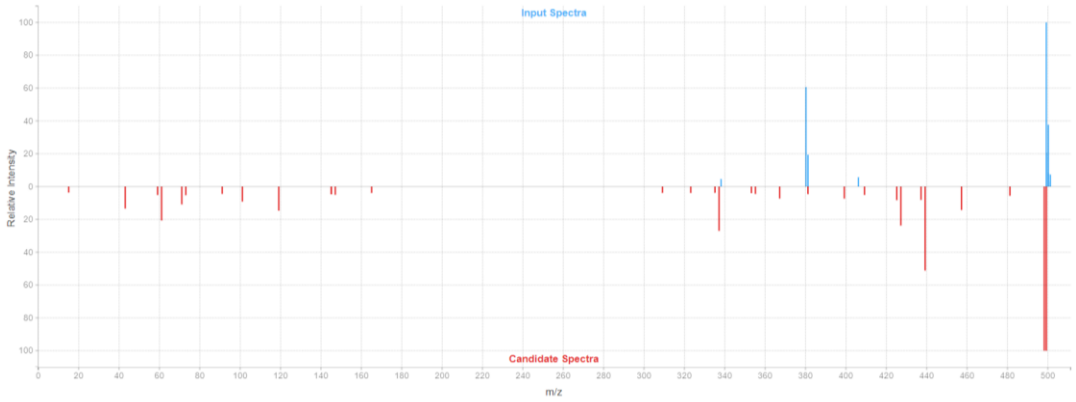


40 eV

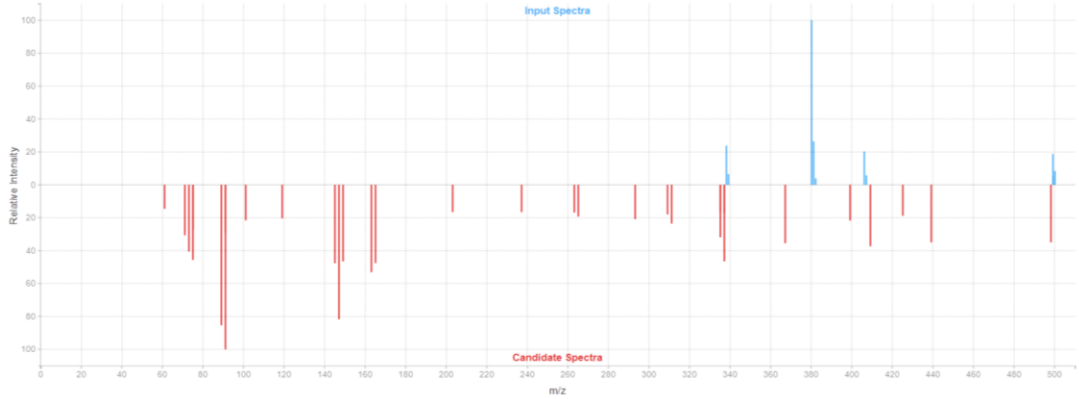


b. 499.2359 m/z match: Siloxane

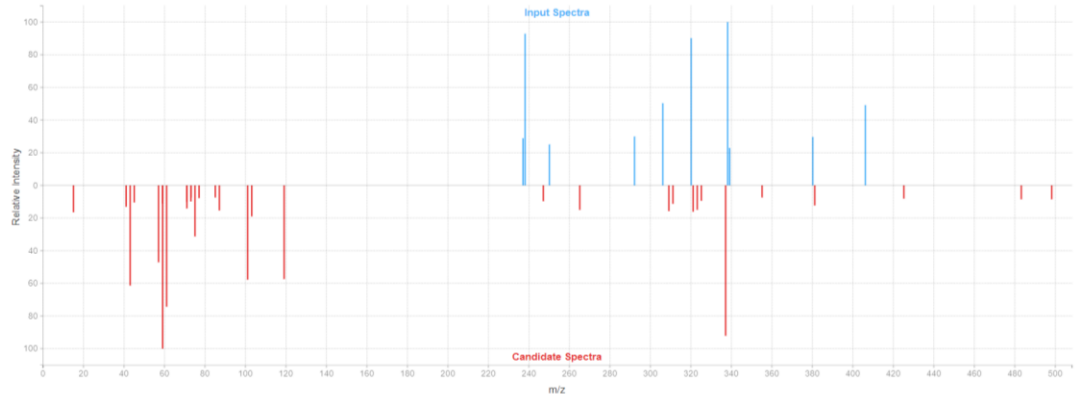
10 eV



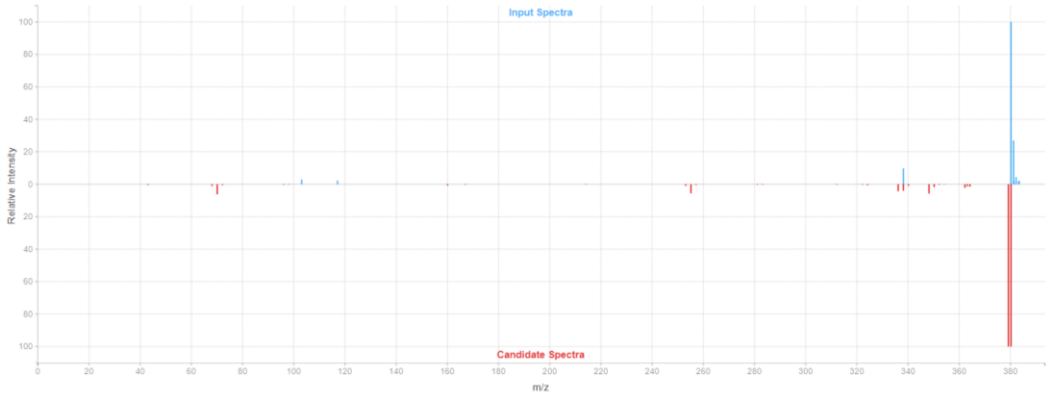
20 eV



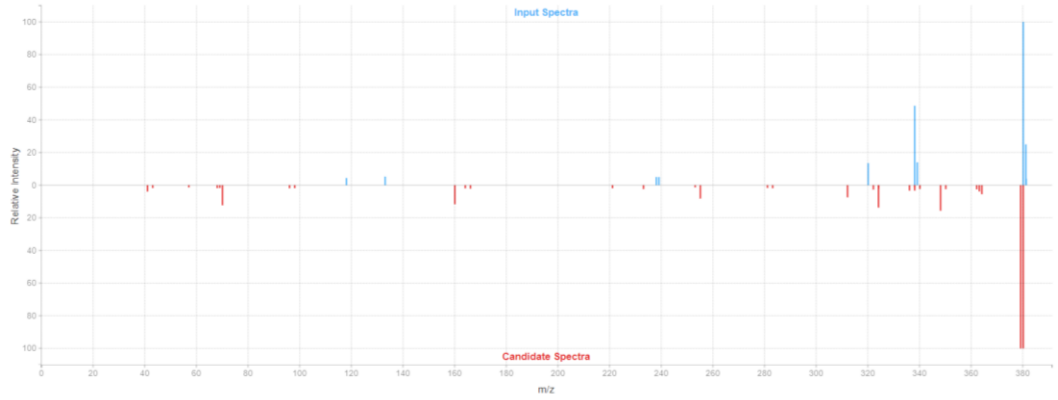
40 eV



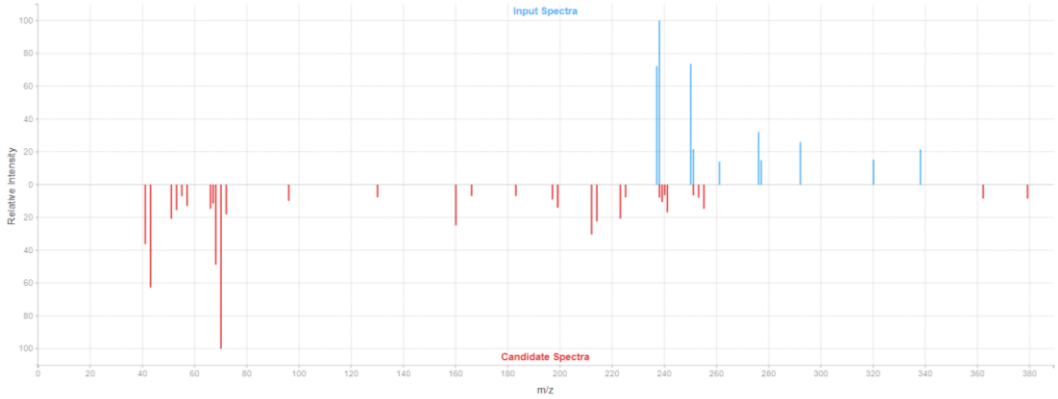
c. 380.1971 m/z match: Fumitremorgin C
10 eV



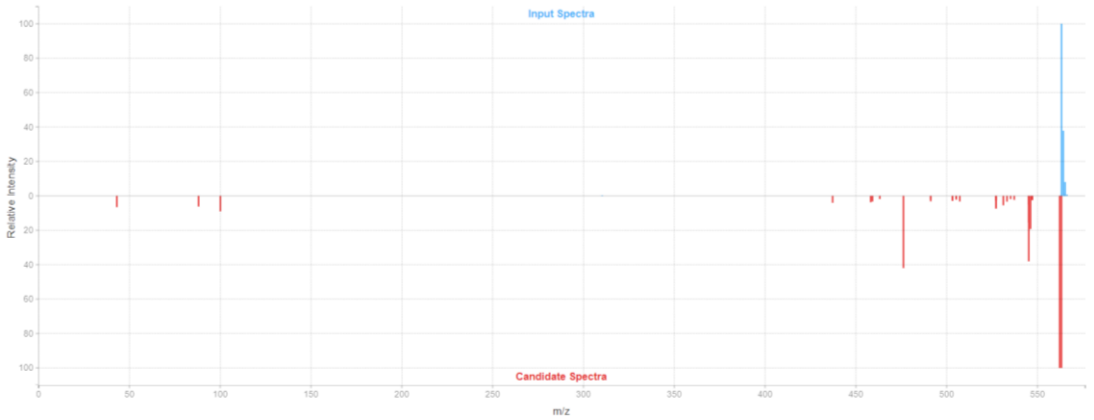
20 eV



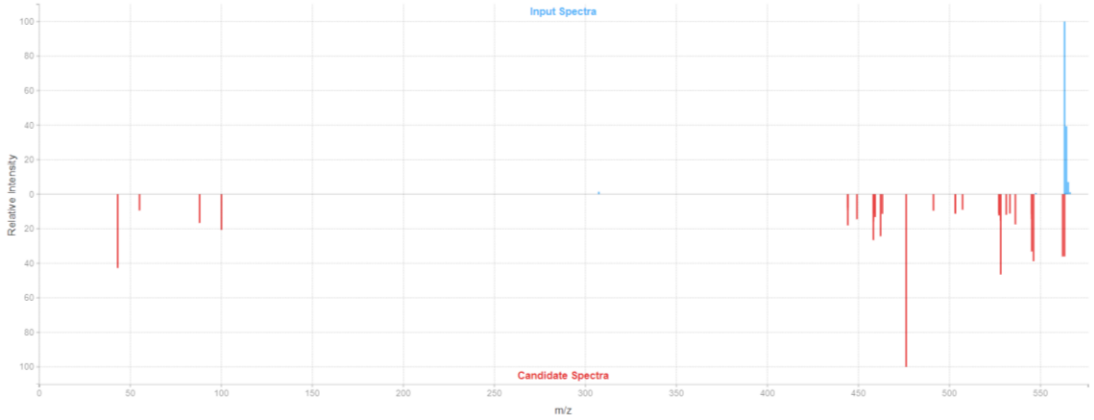
40 eV



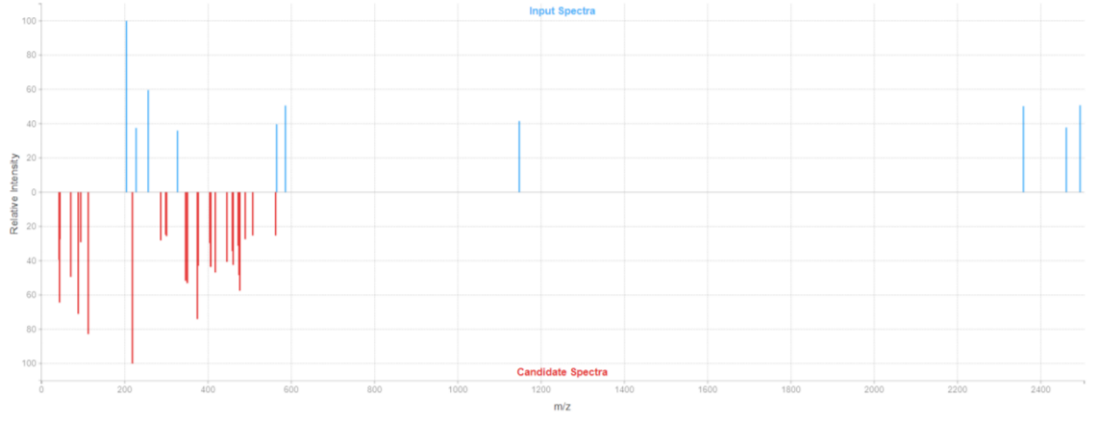
d. 563.2302 m/z match: Saframycin A
10 eV



20 eV



60 eV



e. 423.1977 m/z match: 8-hydroxycarvedilol
10 eV

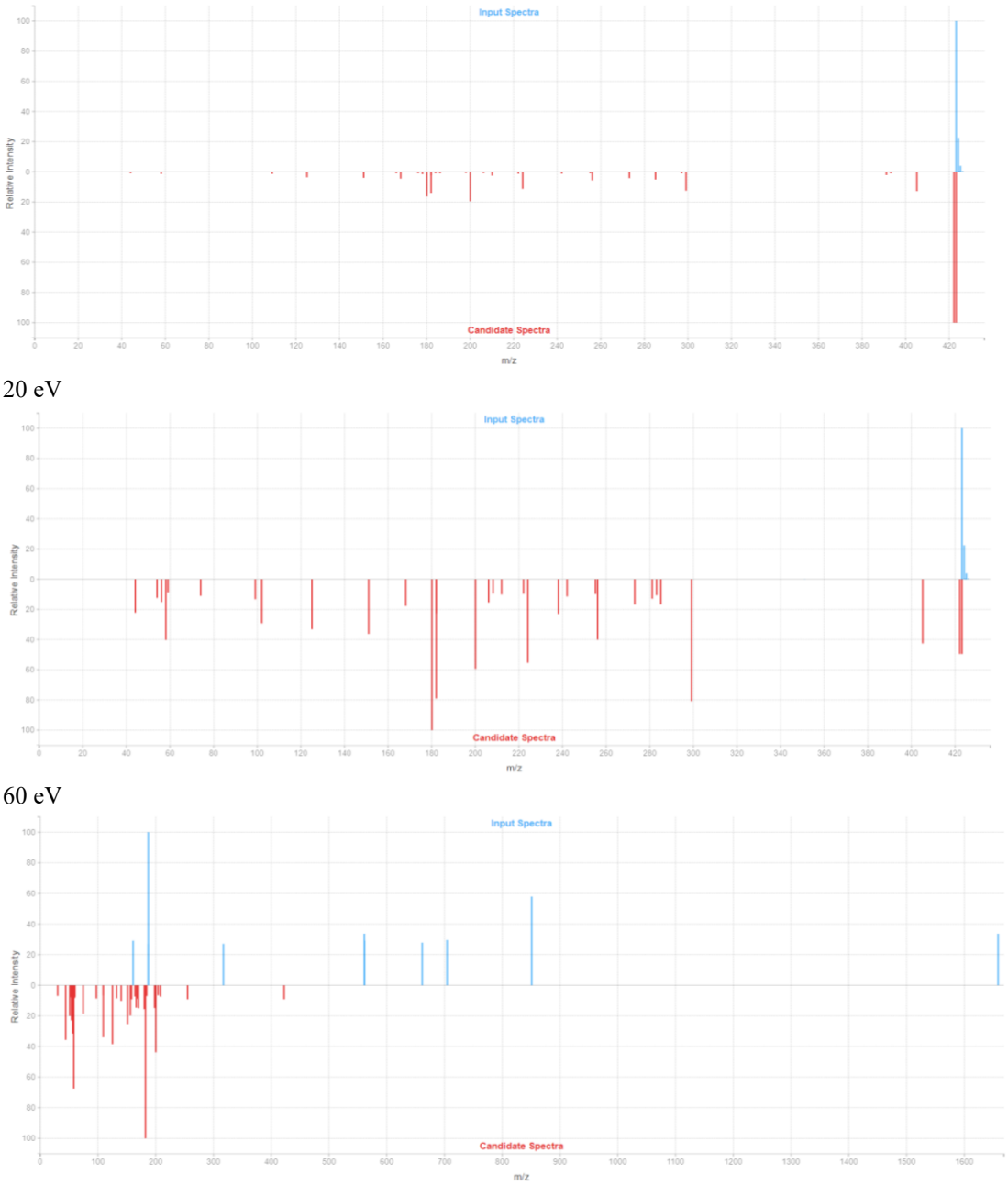


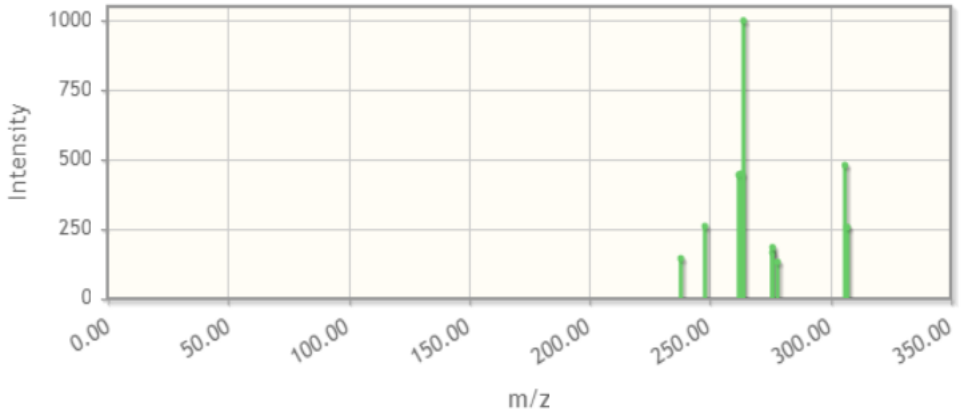
Figure 11 CFM ID database matching results
(a) The metabolite matched with SGI-1776, a pim-kinase inhibitor used as anti-cancer drug. At 40 eV, the lower m/z fragments in the library did not match the sample spectrum. (b) Siloxane is a human-synthesized compound that is commonly applied as industrial additive. Only the parent ion at 10 eV matched the library spectrum. (c) Metabolite 380.1971 m/z matched with an indole alkaloid produced by several funguses that serve as a protein inhibitor in cell experiments; it is unlikely to present in

human serum. The match at 40 eV was also not ideal. (d) Saframycin A is an antibiotic drug, but most of the spectrums at the highest energy (60 eV) did not match with each other. (e) The precursor ion matched at the lower energy levels but not at 60 eV. 8-hydroxycarvedilol is a metabolite produced from Carvedilol, a drug used to treat high blood pressure and heart failure.

The online database, MetFrag, provides structure identification and spectrum matching. However, unlike other databases, MetFrag does not visualize a comparison of sample spectrums and the library spectrums. In the resulting figures, the green peaks are the signals in the sample spectrums that matched the library data, while the blue peaks are those that do not meet the correct m/z range. Gray peaks, if present, are the signals that have lower intensity levels and were excluded. Results from MetFrag are limited since examiners are unable to assess the actual spectrum comparison, but it can be used to distinguish possible structures. (**Figure 12**)

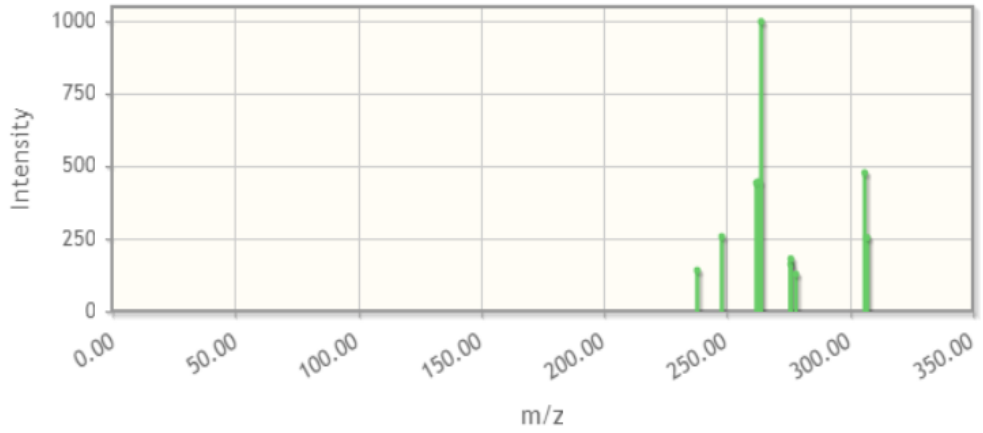
a. 406.1767 m/z match: 2-[(7-fluoro-2-oxo-1, 3-benzoxazin-4-yl)amino]-4, 4-dimethyl-N-(2-methyl-4-oxo-tetrahydrofuran-3-yl) pentanamide

40 eV



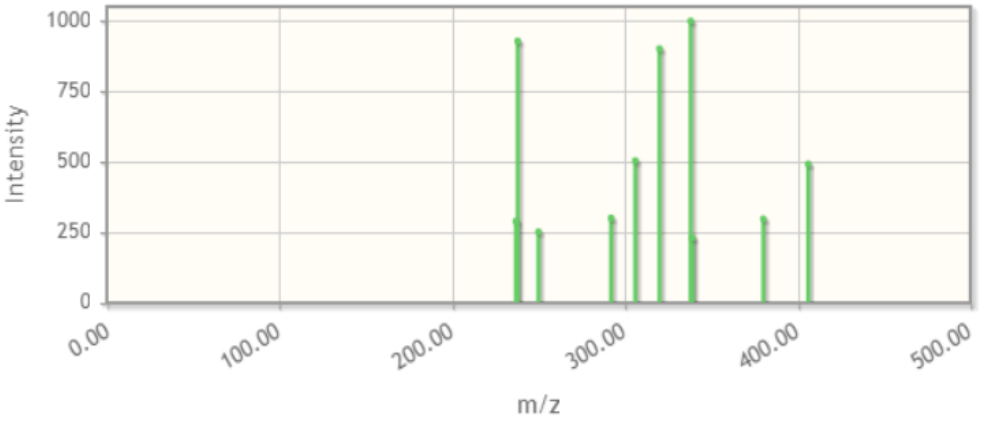
b. 406.1767 m/z match: cyclopropylmethyl 2-[[3-(2-methoxyphenyl) -1-phenyl-pyrazole -4-carbonyl]amino]acetate

40 eV



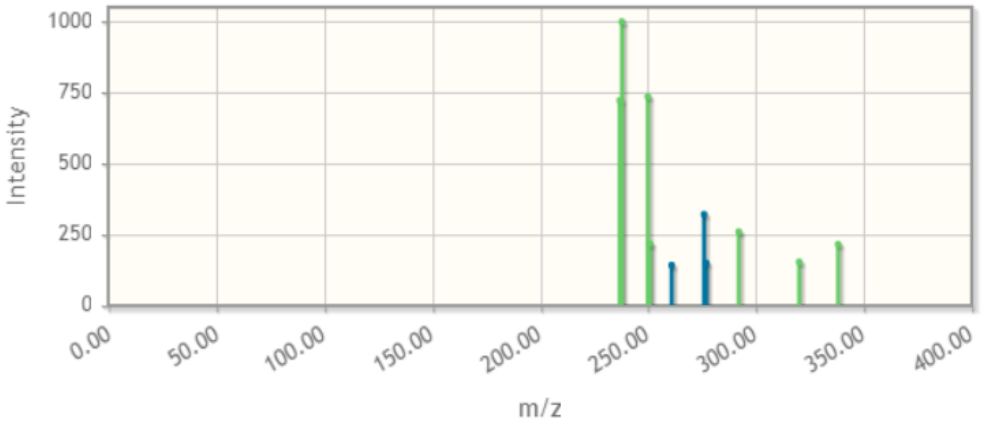
c. 499.2359 m/z match: 2-(4-methyl-1-oxo-phthalazin-2-yl) -N-(3-pyridylmethyl) -N-[[4-(tetrahydrofuran-2-ylmethoxy)phenyl]methyl]acetamide

40 eV



d. 380.1971 m/z match: Fumitremorgin C

40 eV



e. 579.2193 m/z match: tert-butyl 4- [5-oxo-3-[[[2-[4-(trifluoromethoxy) phenyl] acetyl]amino]carbamoyl]-1H-pyrazolo[1,5-a]pyrimidin-7-yl]piperidine-1-carboxylate 40 eV

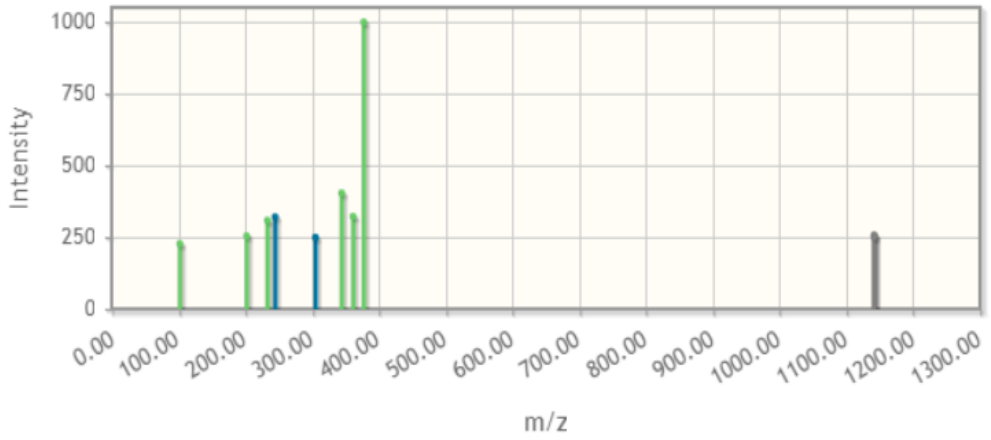


Figure 12 MetFrag database matching results
(a) (b) The two metabolites both matched the targeted metabolite 406.1767 m/z, and they shared the same amide structure. (c) According to the match, the unknown metabolite might contain a pyridine and an amide. (d) The matched metabolite, Fumitremorgin C, is consistent with the results from CFM ID. However, it is a protein inhibitor that has not been used in human. (e) The two blue peaks were those that did not match, and the gray one was eliminated. Possible structure is a benzene ring and a pyrimidin.

In-Source Fragmentation

High throughput metabolomics such as LC MS enables metabolite separation and identification by their retention time and m/z ratios. However, in-source fragmentation might occur even when using the most gentle ionization method, electrospray ionization.[104] This happens between the ion source and the mass spectrometer, in which the metabolites are accelerated by applying voltages, prompting collisions with surrounding species that can produce sufficient energy to yield fragment ions.[105] However, this fragmentation is not desired since the process of ionic species breakdown is only expected in the collision cell. When in-source fragmentation happens, the peaks in the LC-MS spectrum that correspond to the prematurely broken down parts of the intact ionic species may be falsely considered as different compounds. This results in false detection when selecting candidate metabolites and errors in future MS/MS matching. One of the possible solutions is to check the finalized metabolite list for features having the same retention time, because the fragments generated during an in-source fragmentation are originated from the same metabolite with a fixed retention time collected from the liquid chromatography. Thus, combining the MS/MS spectrums of the possible fragments and considering them as one metabolite when doing database matching are an effective approach to correct these errors from in-source fragmentation. According to the finalized metabolites in the present study, metabolites 406.1767 m/z and 380.1971 m/z have the same retention time at 9.435 minute, which highlights the possibility of them being the products of in-source fragmentation. To address that, the spectrums of the two metabolites were combined and searched together in the online

databases. Only the Metlin database yielded a match at 20 eV with a parent ion of 406 m/z. **Figure 13** summarized the results, showing that although there is a match for the combined spectrum, it was not considered as a good spectrum, and the occurrence of in-source fragmentation was unable to be confirmed.

a. Combination of 406.1767 m/z and 380.1971 m/z match: Cefaclor

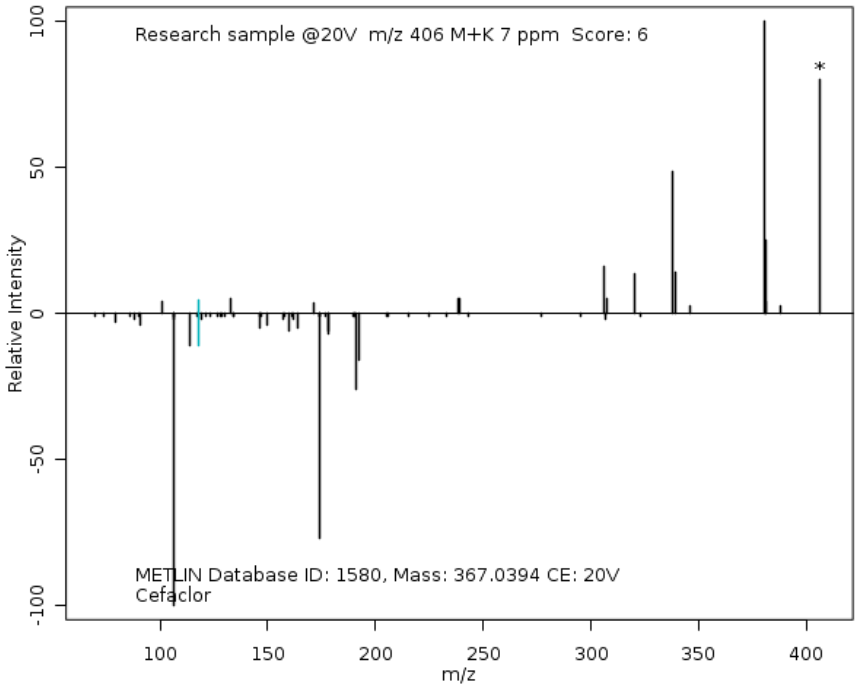


Figure 13 In-source fragmentation Metlin database matching results
(a) The combined spectrum showed a result of Cefaclor, a respiratory tract infection drug, in the Metlin database. But the sample peaks did not match with the library ones, indicating it as a bad match.

CONCLUSION AND SUMMARY

The non-targeted analysis yielded eight unknown metabolites that are potential biomarkers for the CAD stages. After performing MS/MS and database matching, one of the targeted metabolites was identified and confirmed to be 4-hydroxyatorvastatin lactone. The statin drug-derived metabolite was specific to CAD 0 among the study participants. Although it is not a factor that can be applied as disease biomarkers and be generalized to the population, 4-hydroxyatorvastatin lactone can be regarded in the present study as being similar to an internal standard. It was a stage-determining metabolite caused by medication, which was later confirmed during the untargeted analysis. This suggests that the mass spectrometry and the multivariate analysis (Figure 1.b) was correctly performed and was able to address the metabolic difference of CAD in NAFLD patients by metabolomic fingerprinting. The success in metabolite profiling also validated the metabolomics pipeline, the weight score equation, and the manual evaluation applied to the data. Together, it provides support that our algorithm of variables detecting is reliable in selecting CAD relevant metabolites.

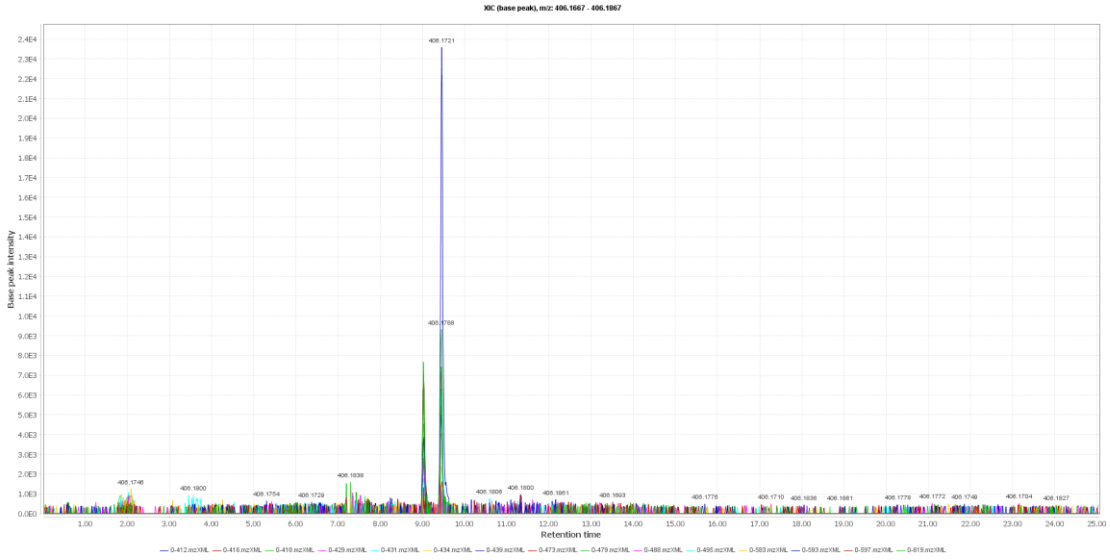
Several limitations, however, appeared during the research. One of them is the uneven sample number in each CAD stage. Since the analysis required normally distributed data, errors and over explanation might occur when the sample size is too small. Another limitation is the low identification rate in MS/MS spectrum matching,

especially among the online databases. Lastly, almost half of the participants were diagnosed with diabetes or pre-diabetes state; which might become a confounding variable when investigating metabolic patterns.

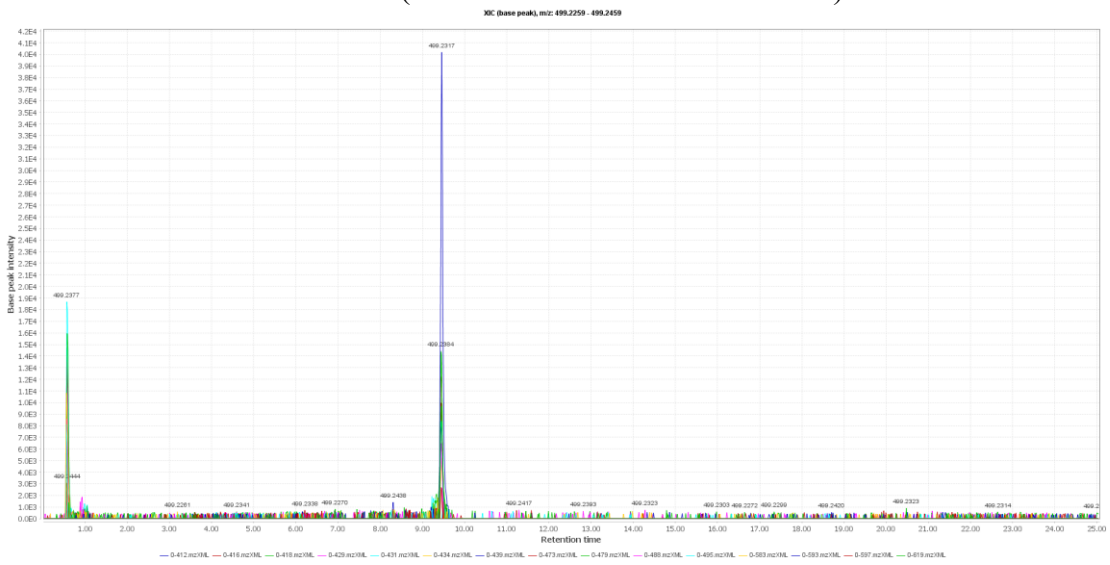
Future research may include expanding the sample number and improving preliminary steps in participant recruitments. Improving MS/MS spectrum matching methods will also aid in identifying the remained unknown CAD sensitive metabolites.

APPENDIX 1

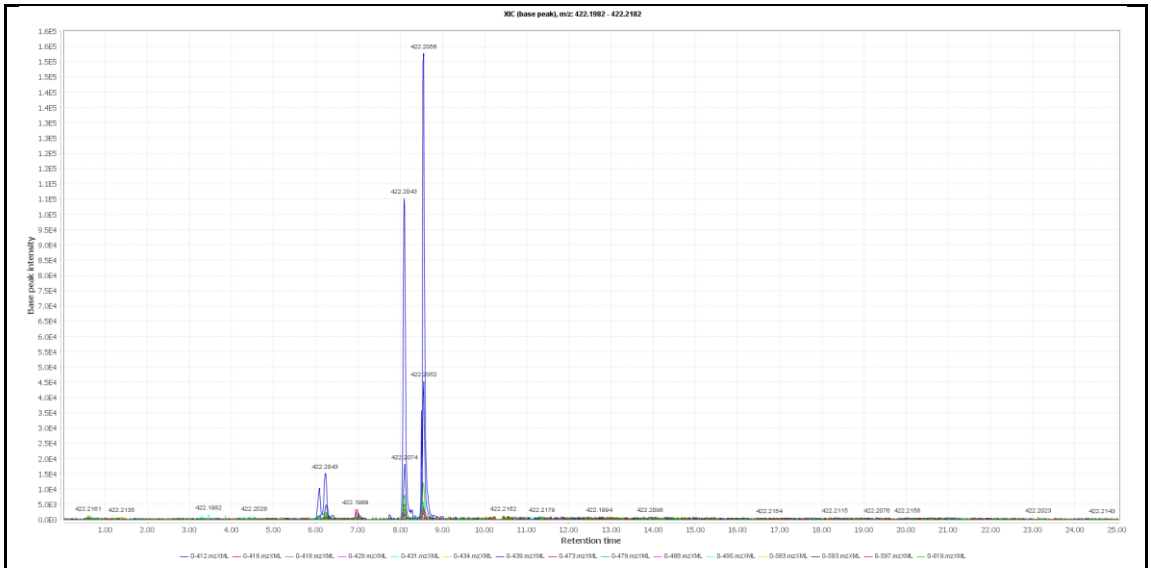
a. 406.1767 m/z EIC: CAD 0



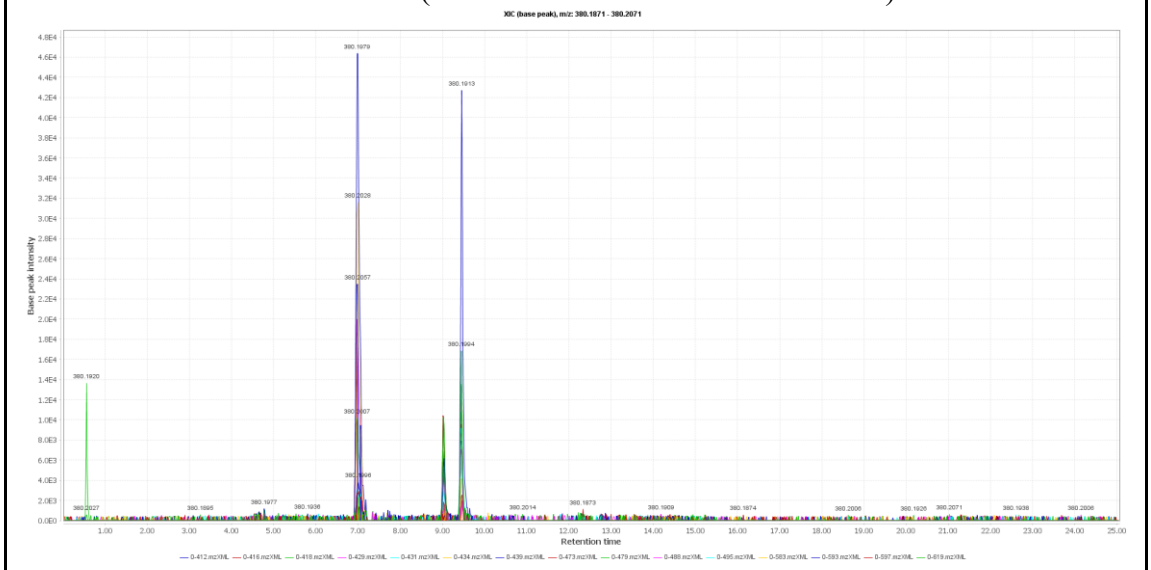
b. 499.2359 m/z EIC: CAD 0 (The retention time is 9.4 minutes)



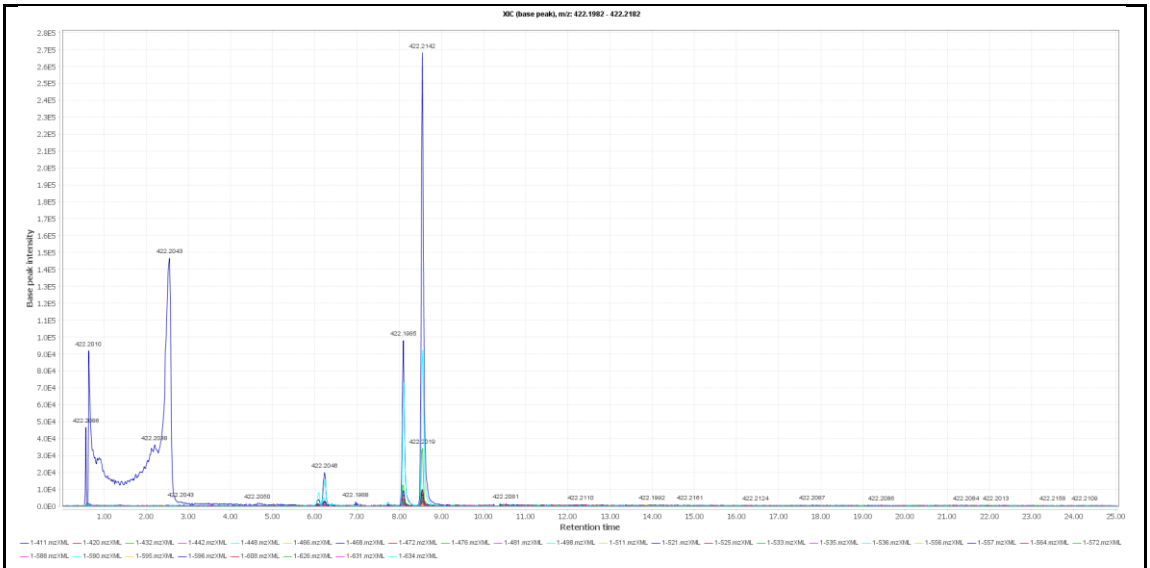
c. 422.2082 m/z EIC: CAD 0



d. 380.1971 m/z EIC: CAD 0 (The retention time is 9.435 minutes)

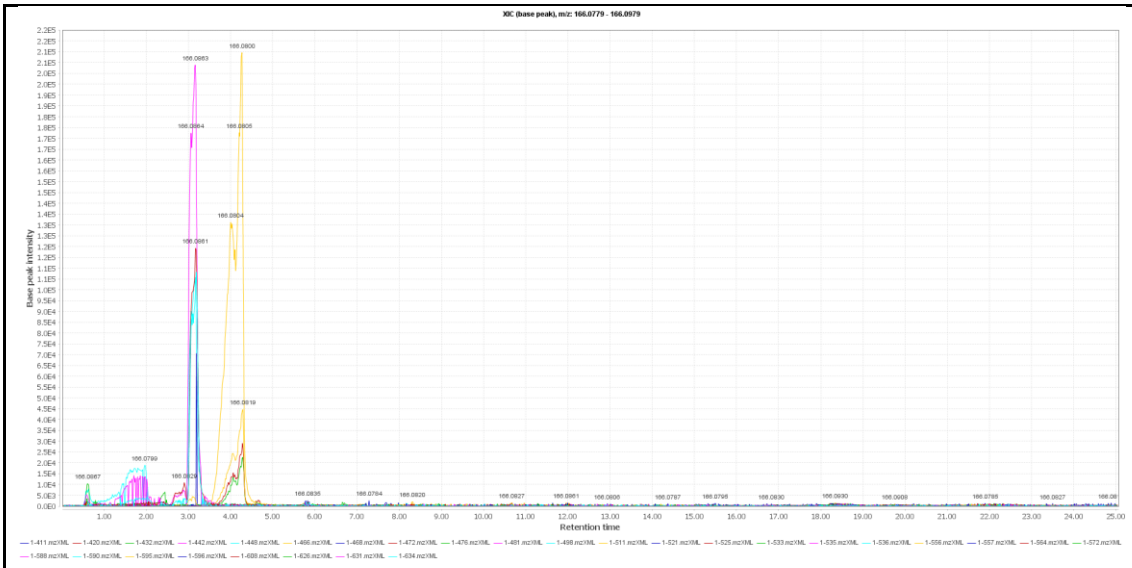


e. 422.2082 m/z EIC: CAD 1 (The retention time is 8.577 minutes)

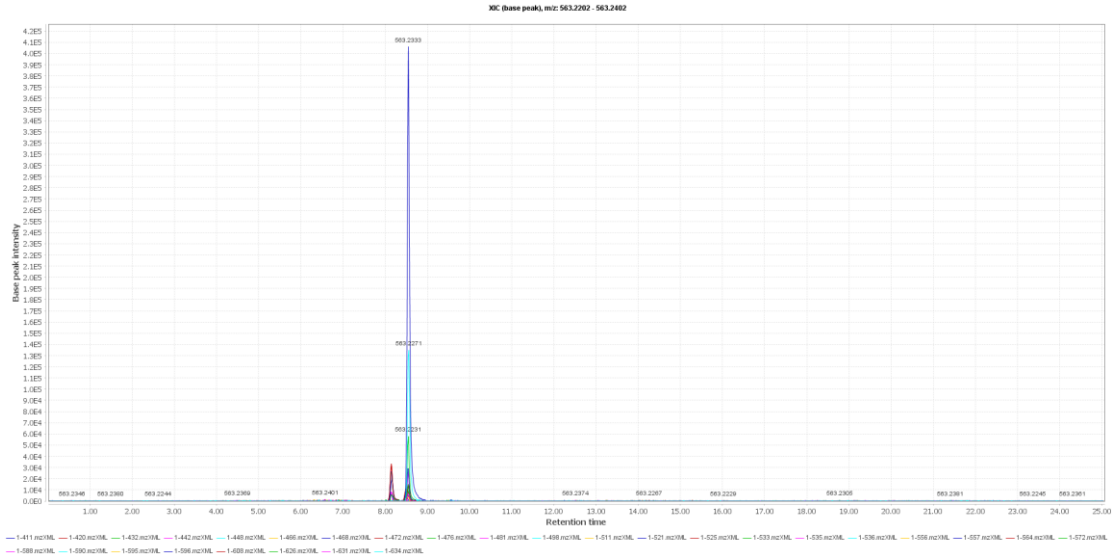


f. 165.0712 m/z EIC: CAD 1 (The retention time is 4.62 minutes)
There is no clear signal of the metabolite at 4.62 minute, and the yellow peak has a zigzag shape indicating this is not a good EIC.

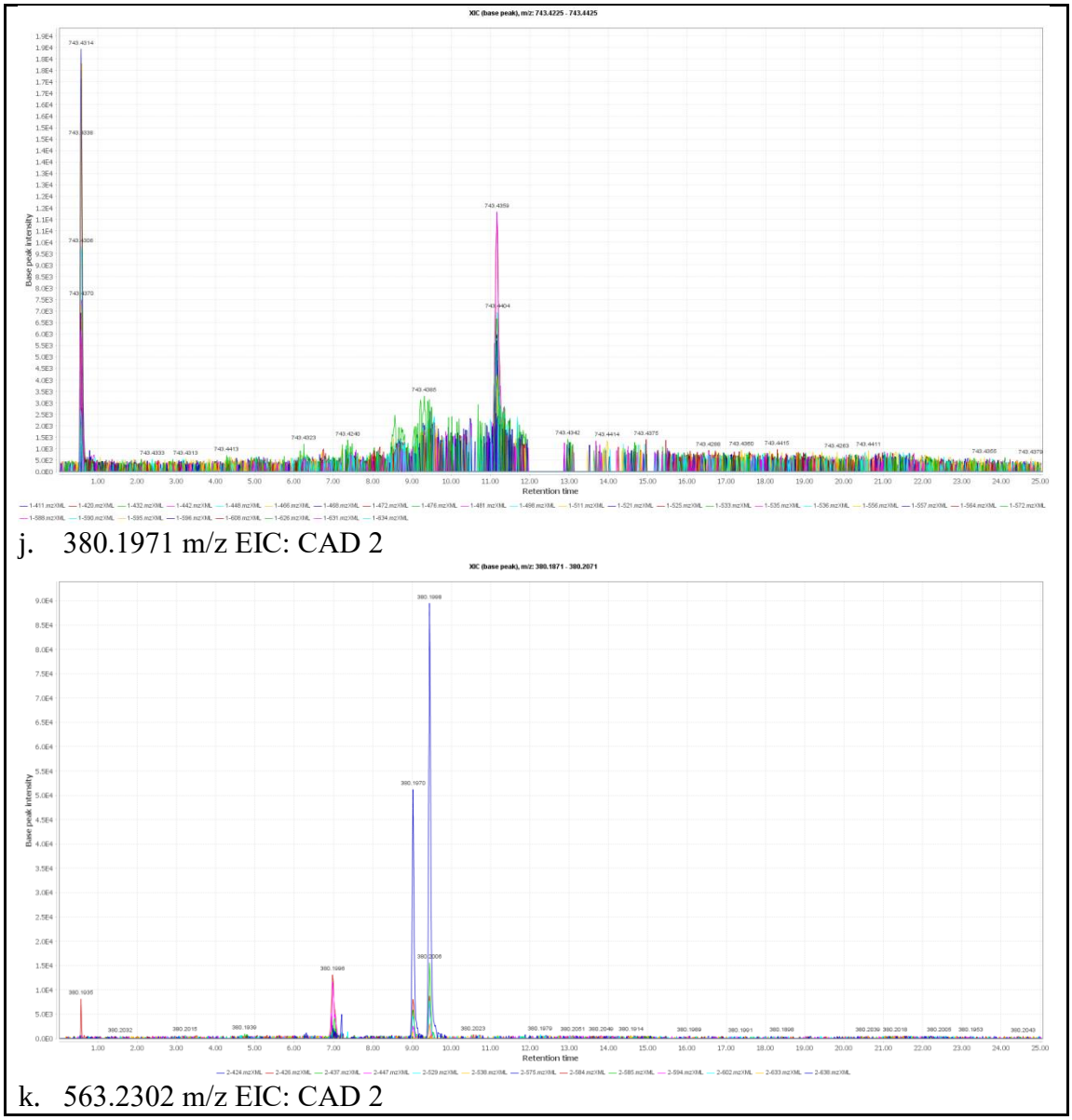
g. 166.0879 m/z EIC: CAD 1 (The retention time is 3.878 minutes)
The peaks at around 3.8 minute are wider and contain multiple signals that cannot be addressed. This is not an ideal EIC.



h. 563.2302 m/z EIC: CAD 1

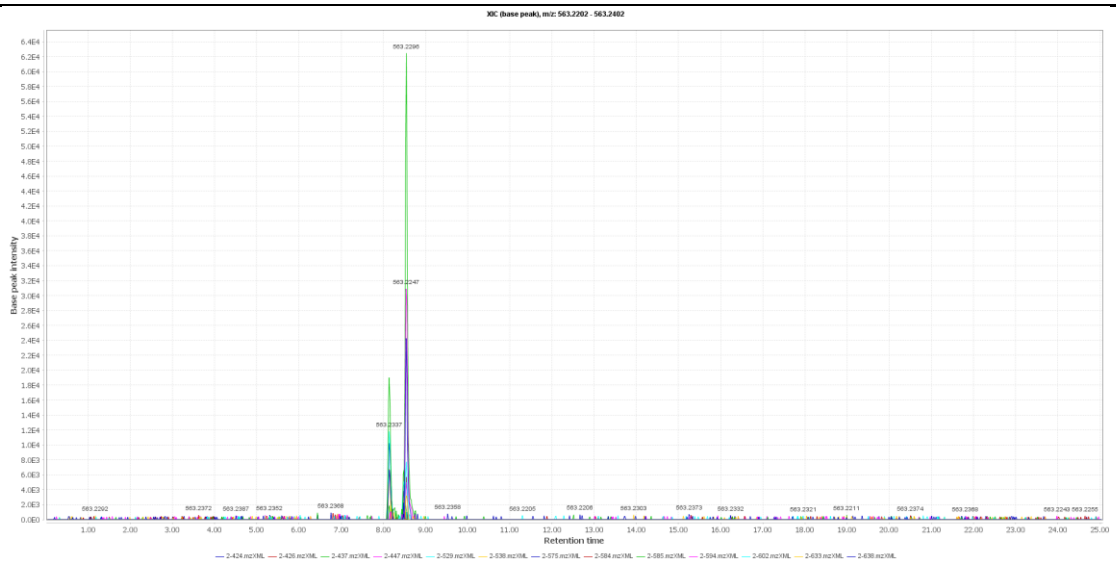


i. 743.2325 m/z EIC: CAD 1 (The retention time is 0.564 minutes)
Metabolites eluted before 1 minute and after 24 minute are considered less reliable due to the commonly observed ion suppression, and they should not be selected. Also, the baseline is too noisy.

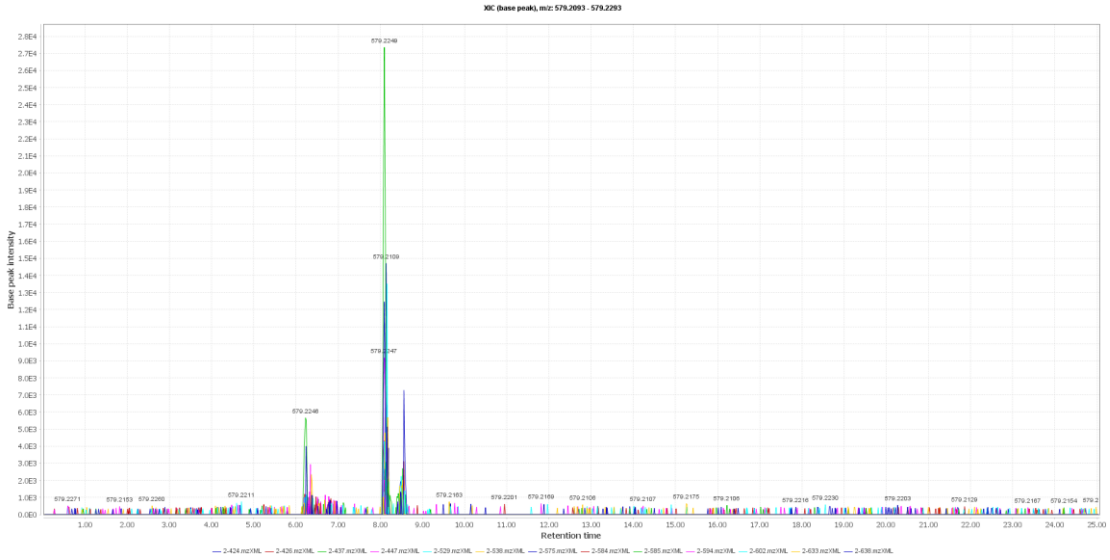


j. 380.1971 m/z EIC: CAD 2

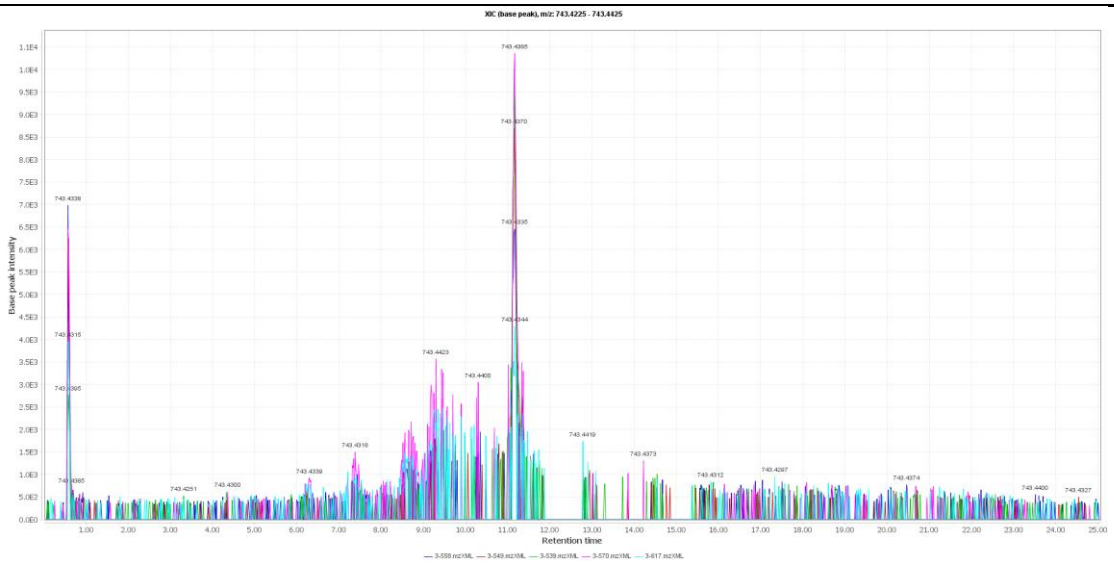
k. 563.2302 m/z EIC: CAD 2



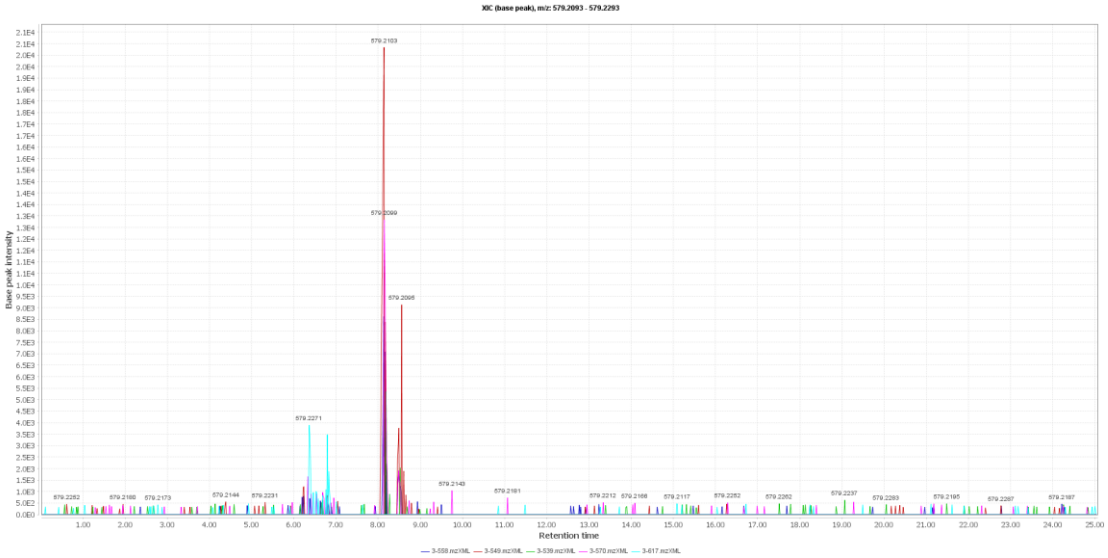
l. 579.2193 m/z EIC: CAD 2



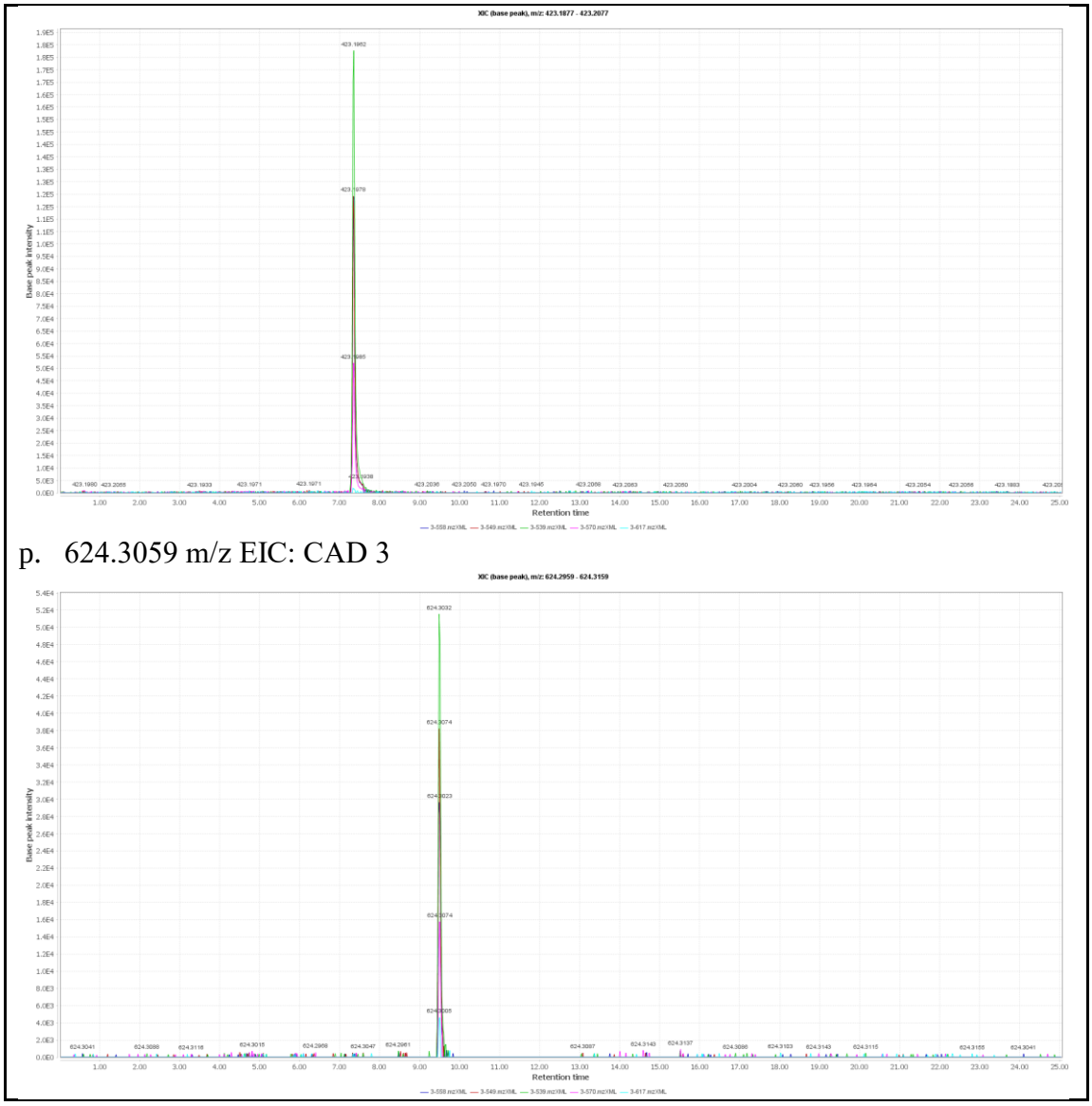
m. 743.2325 m/z EIC: CAD 3 (The retention time is 0.564 minutes)
The EIC shows the same trend as CAD 1, so this metabolite is eliminated.



n. 579.2193 m/z EIC: CAD 3



o. 423.1977 m/z EIC: CAD 3

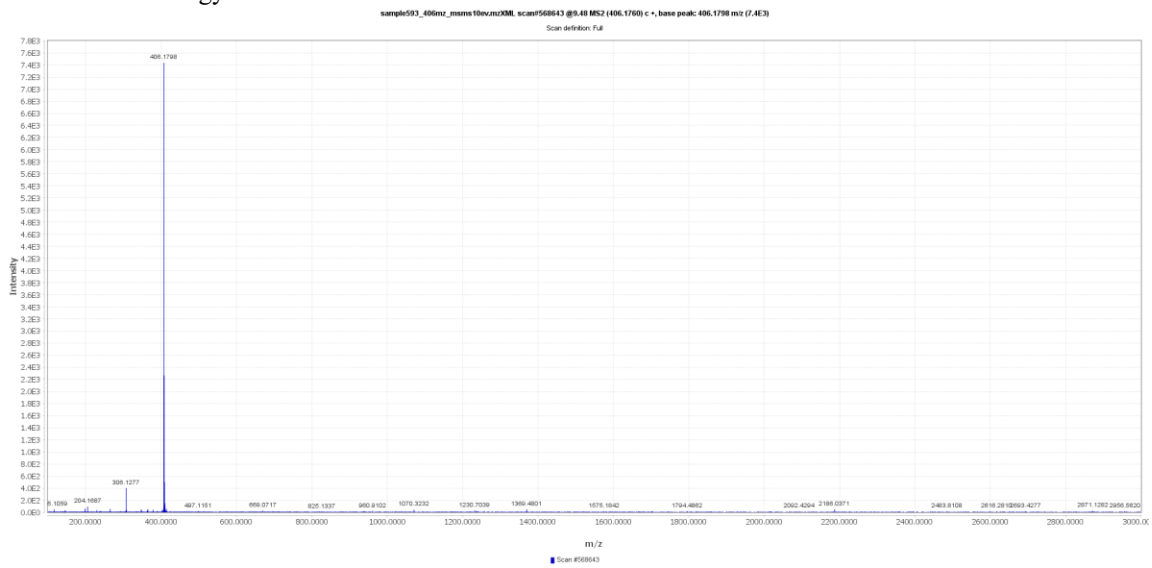


p. 624.3059 m/z EIC: CAD 3

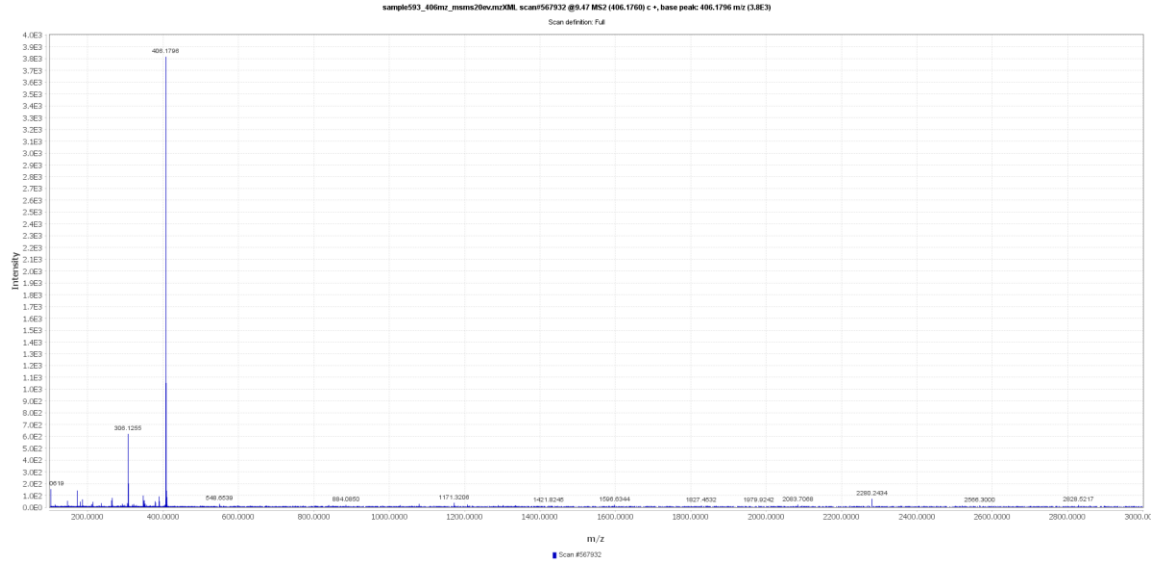
APPENDIX 2

a. 406.1767 m/z MS/MS spectrums:

Collision energy: 10 eV



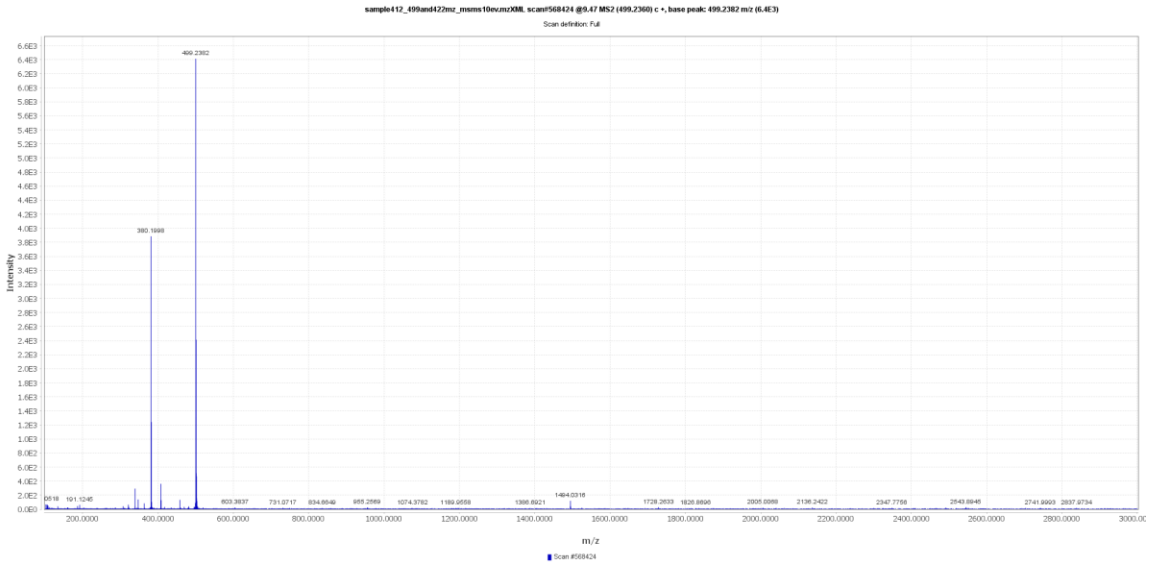
Collision energy: 20 eV



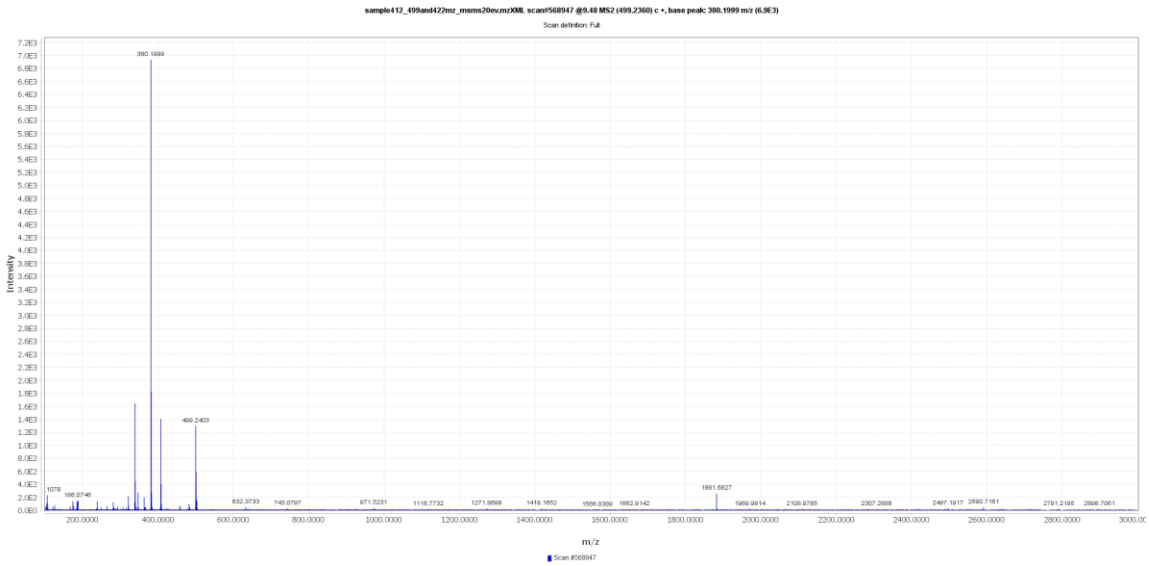
Collision energy: 40 eV

b. 499.2359 m/z MS/MS spectrums:

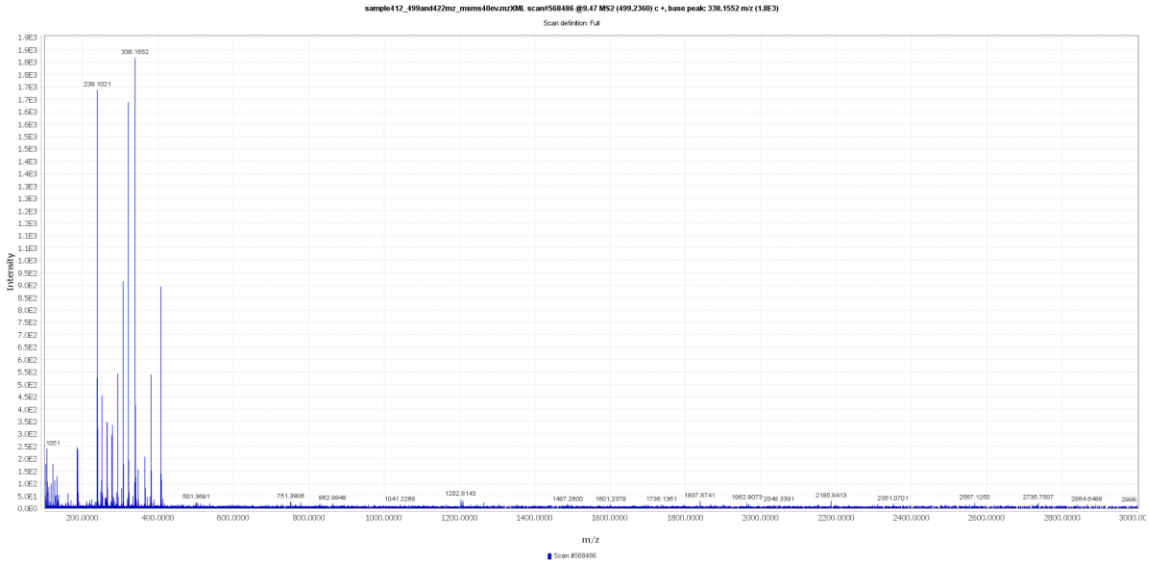
Collision energy: 10 eV



Collision energy: 20 eV

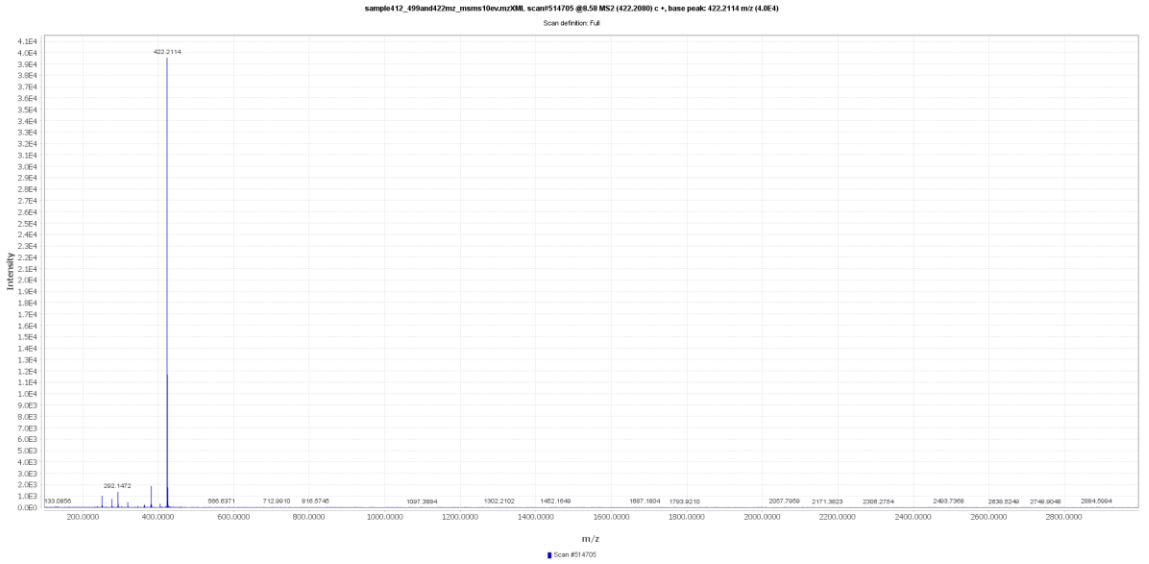


Collision energy: 40 eV

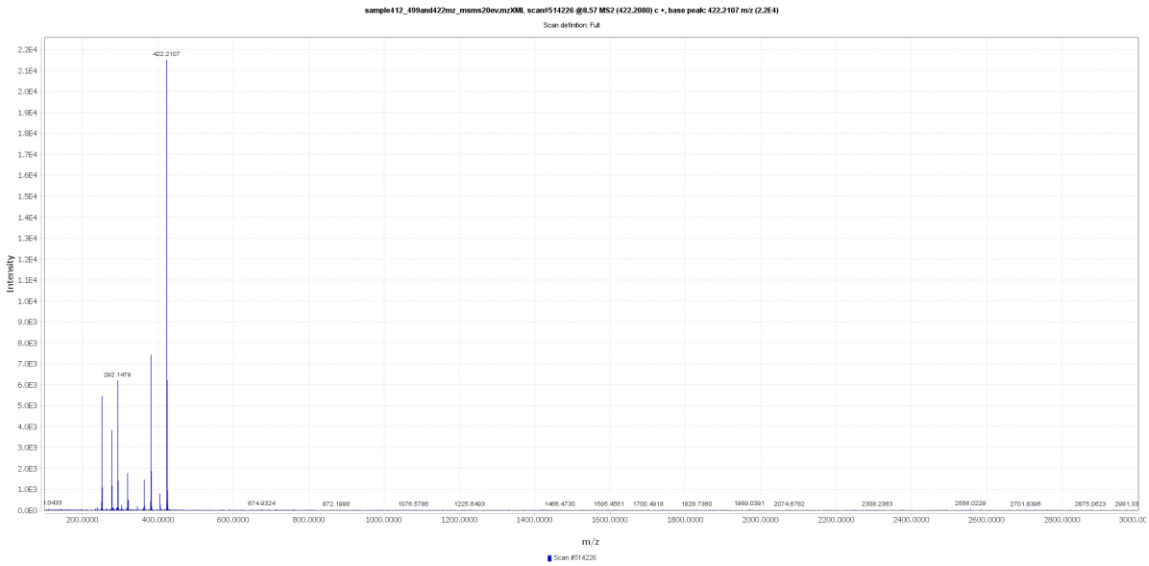


c. 422.2082 m/z MS/MS spectrums:

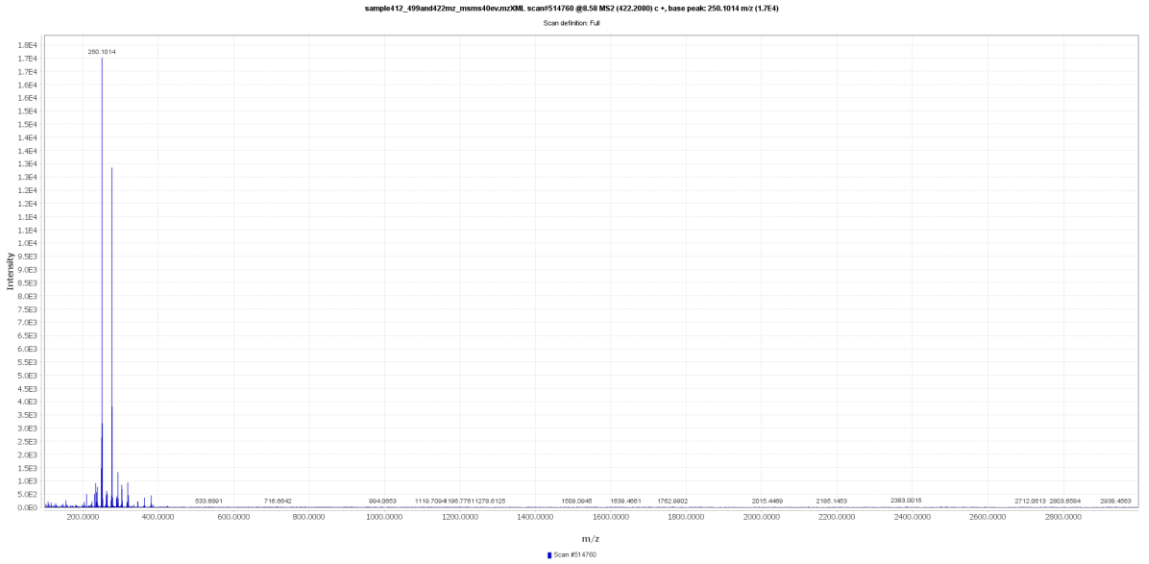
Collision energy: 10 eV



Collision energy: 20 eV

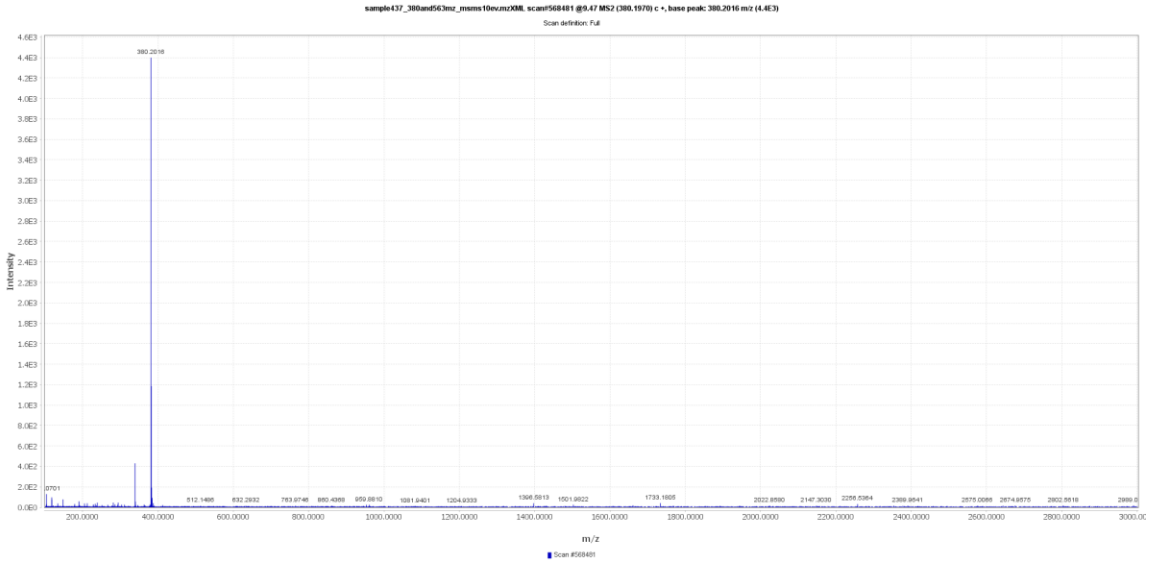


Collision energy: 40 eV

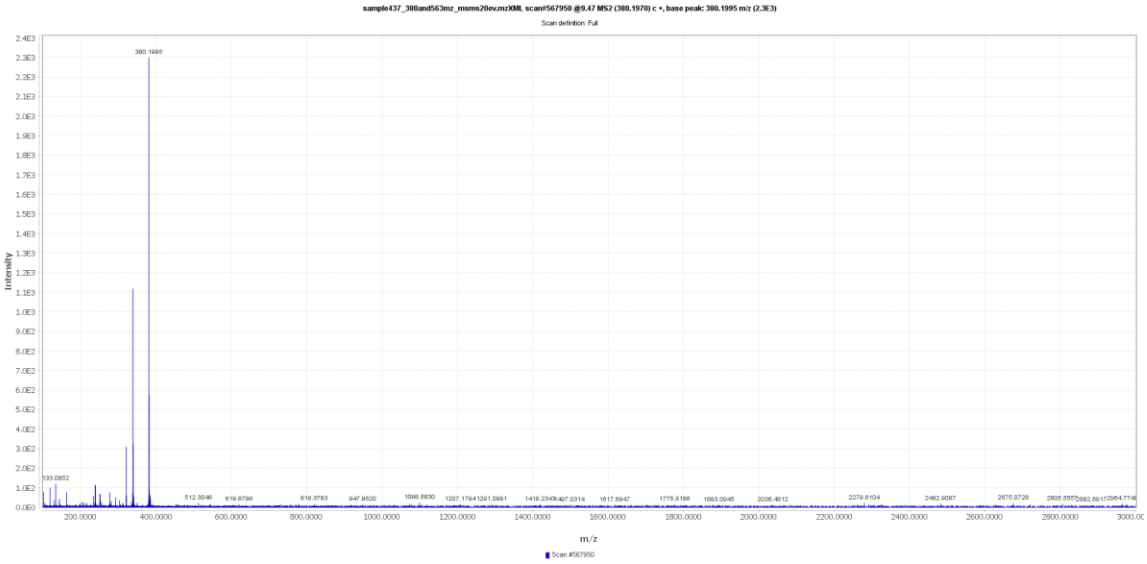


d. 380.1971 m/z MS/MS spectrums:

Collision energy: 10 eV



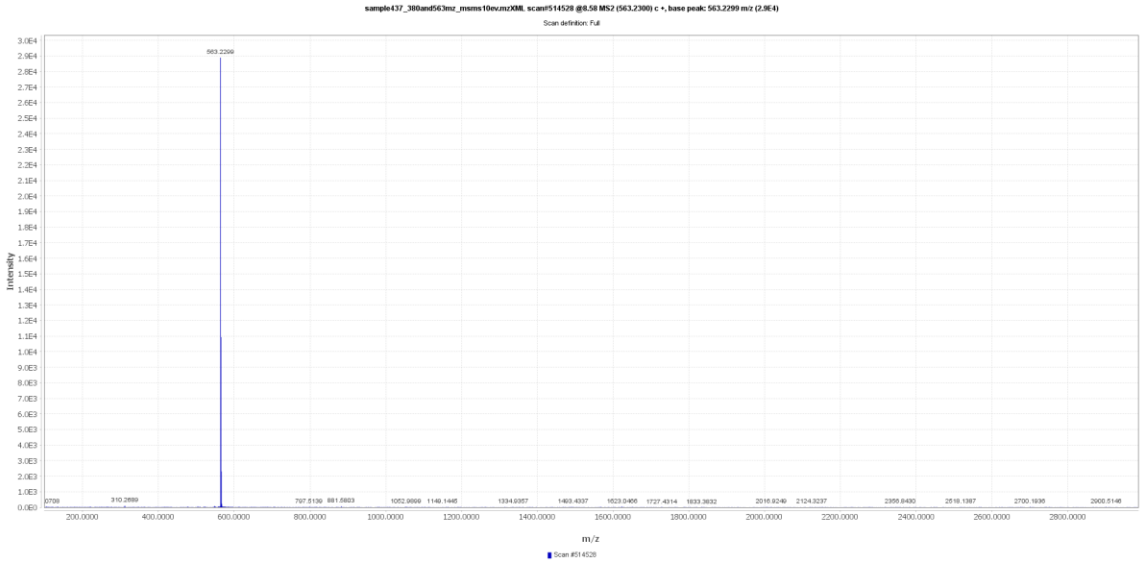
Collision energy: 20 eV



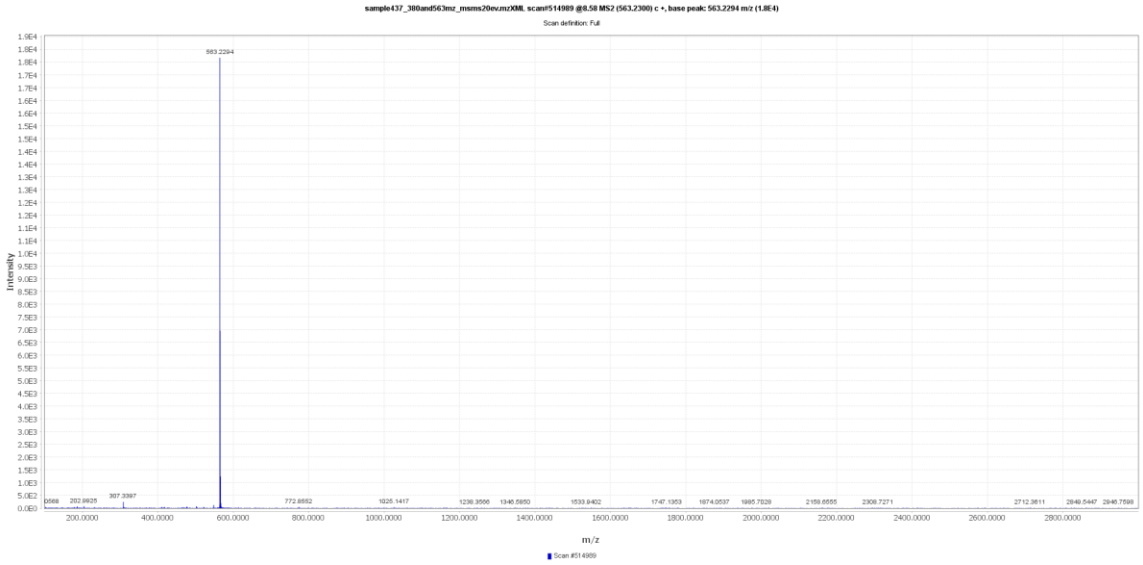
Collision energy: 40 eV

e. 563.2302 m/z MS/MS spectrums:

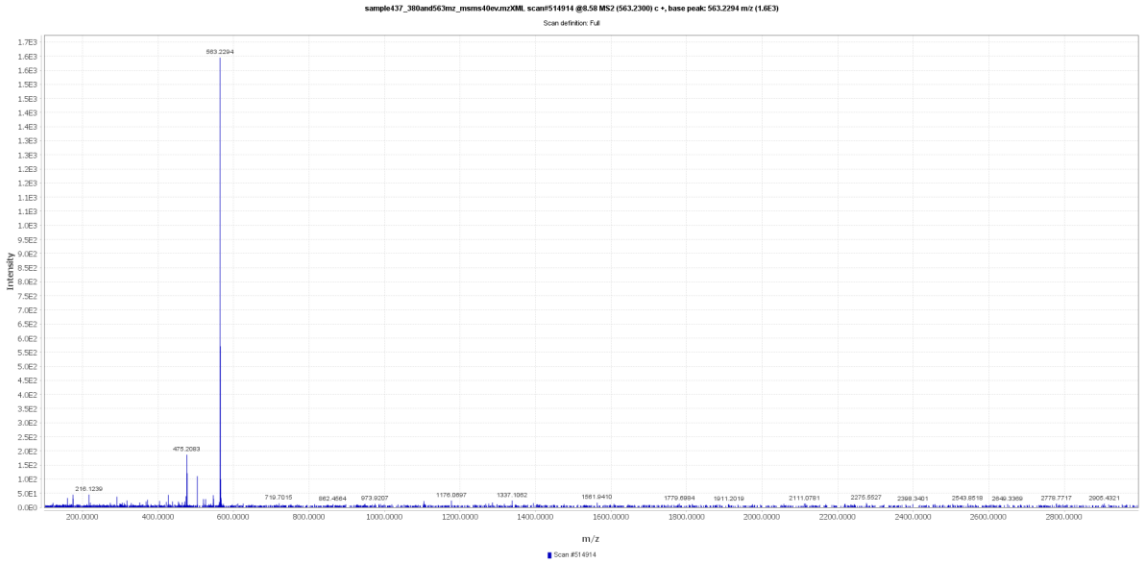
Collision energy: 10 eV



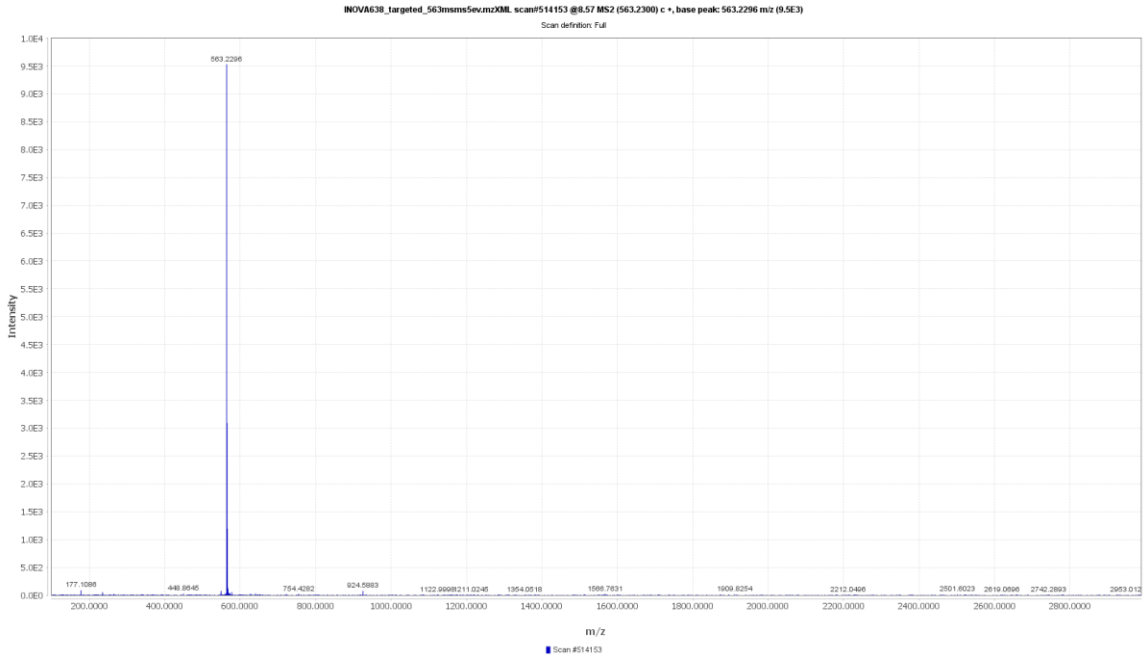
Collision energy: 20 eV



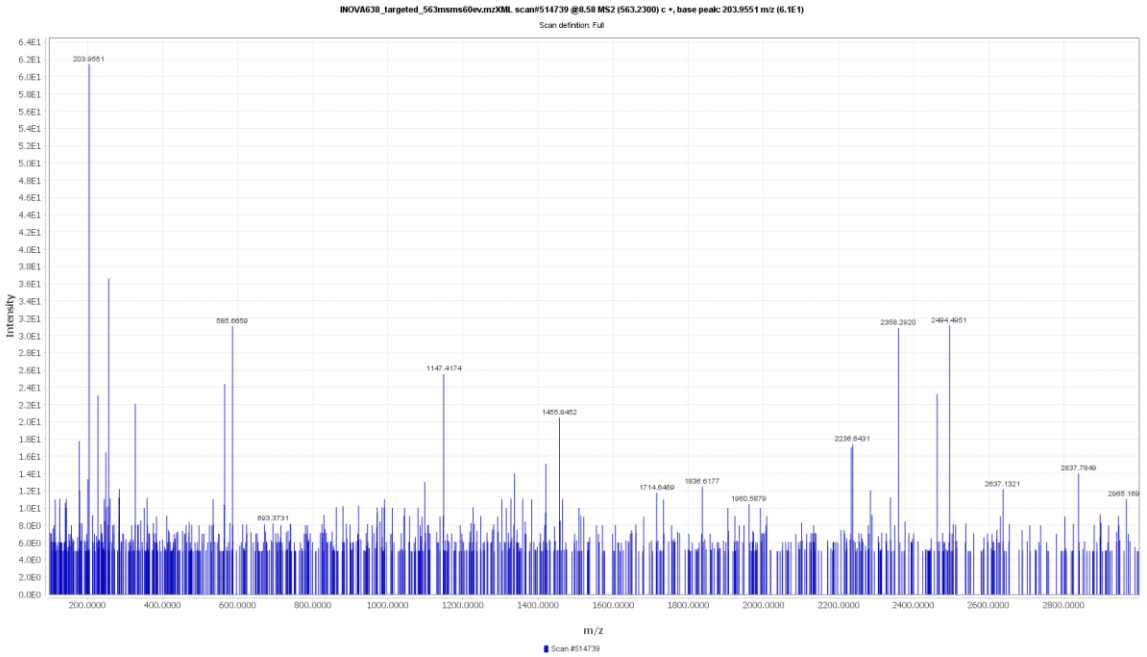
Collision energy: 40 eV



Collision energy: 5 eV

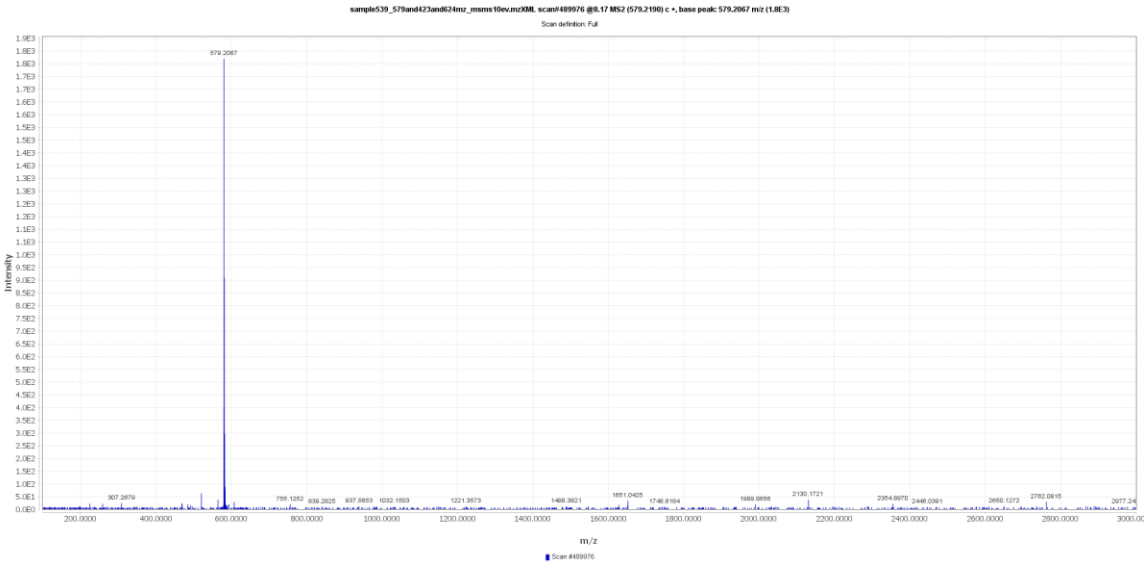


Collision energy: 60 eV

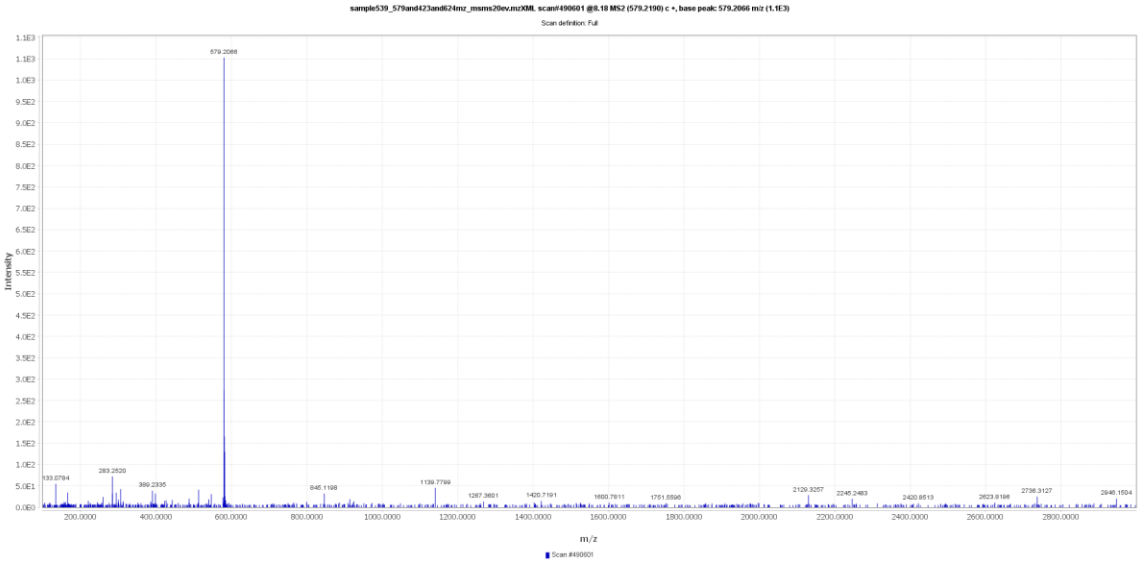


f. 579.2193 m/z MS/MS spectrums:

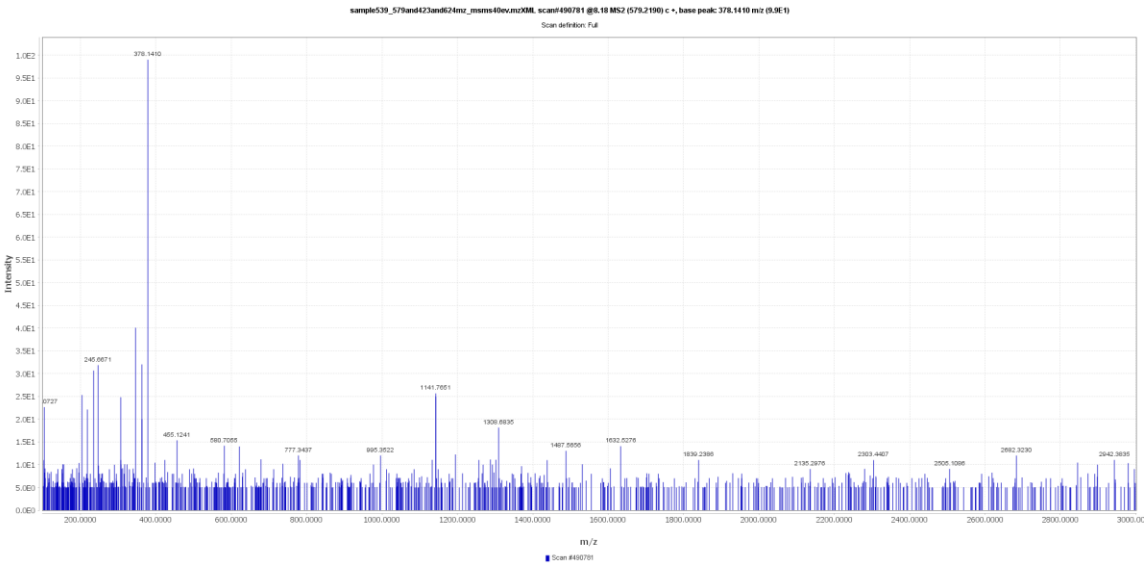
Collision energy: 10 eV



Collision energy: 20 eV

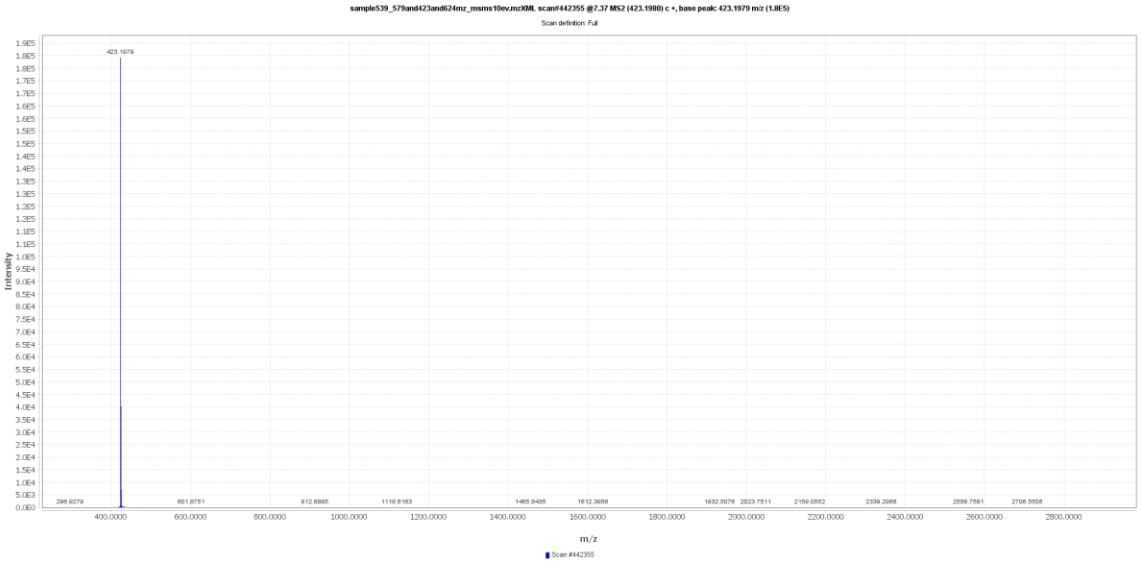


Collision energy: 40 eV

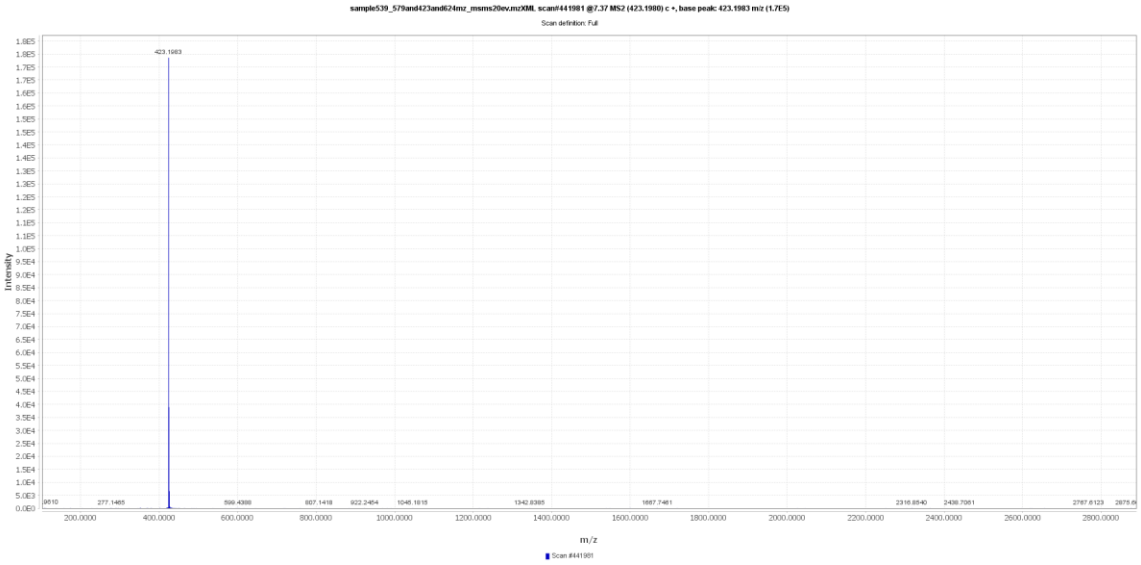


g. 423.1977 m/z MS/MS spectrums:

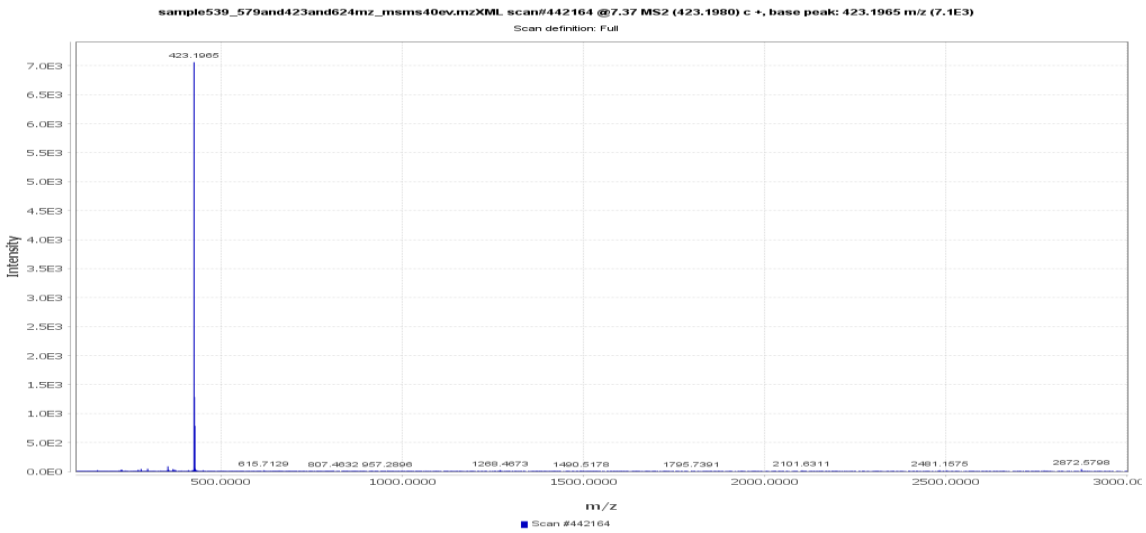
Collision energy: 10 eV



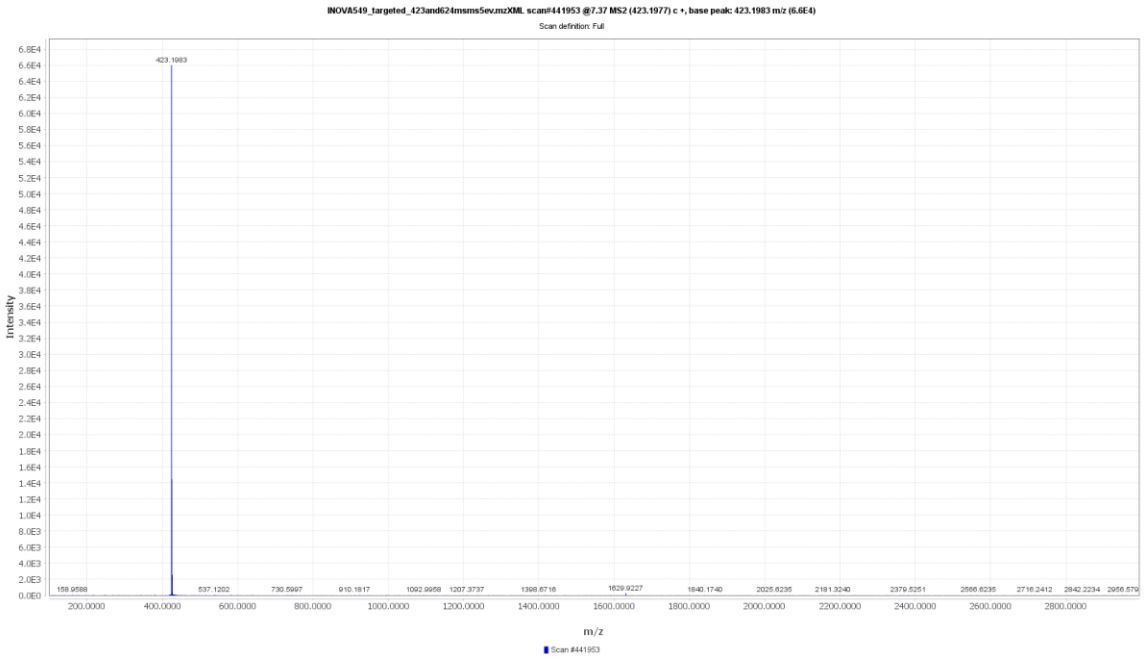
Collision energy: 20 eV



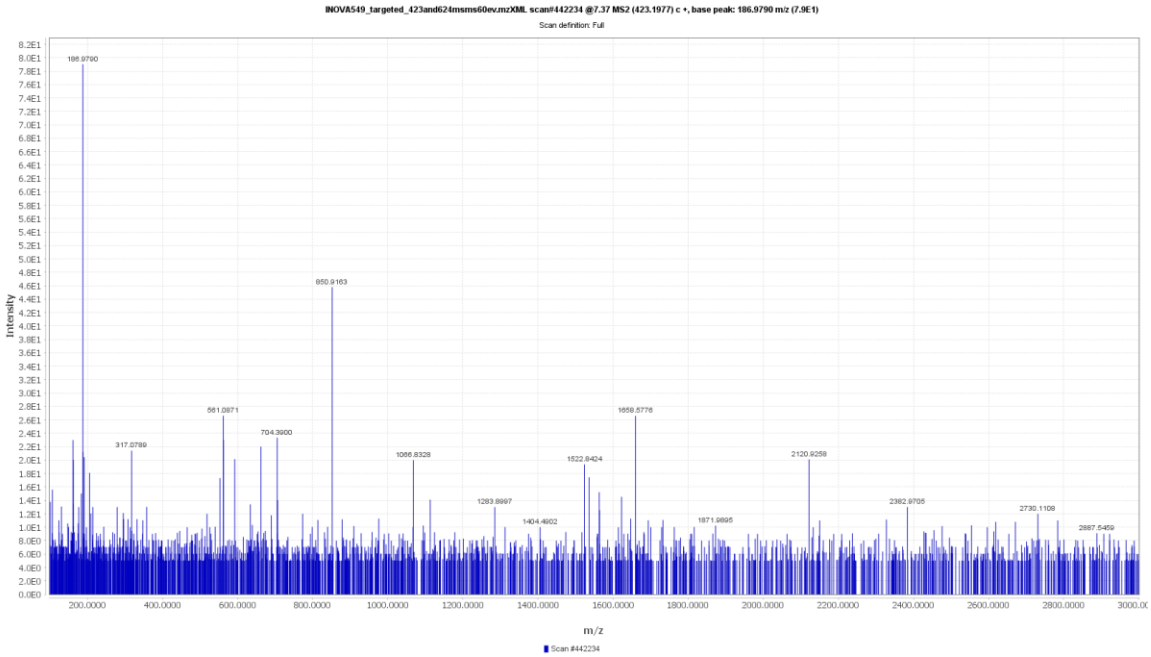
Collision energy: 40 eV



Collision energy: 5 eV

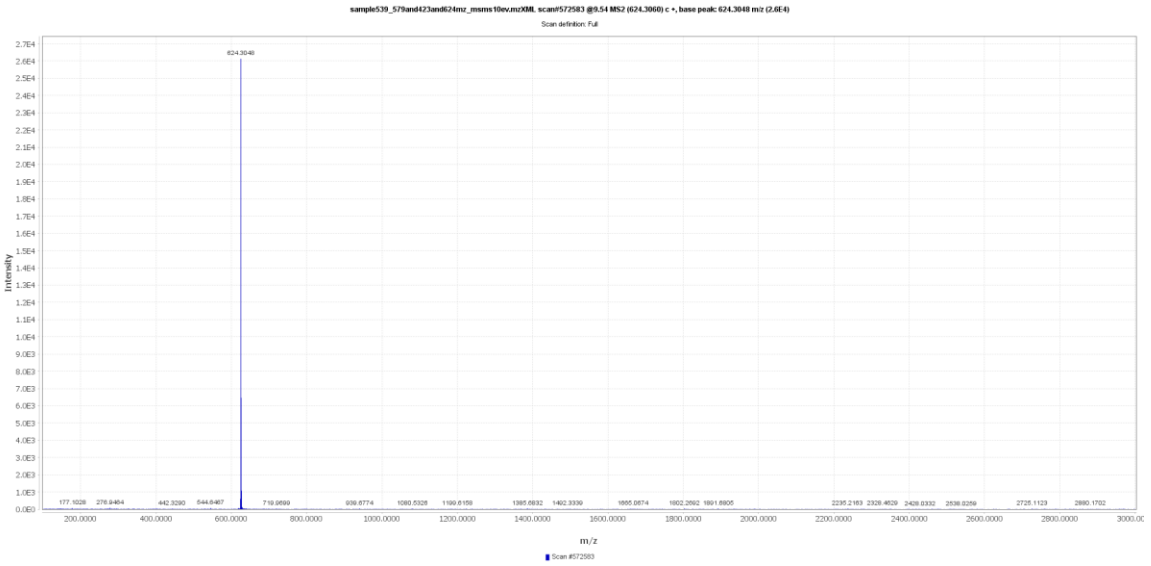


Collision energy: 60 eV

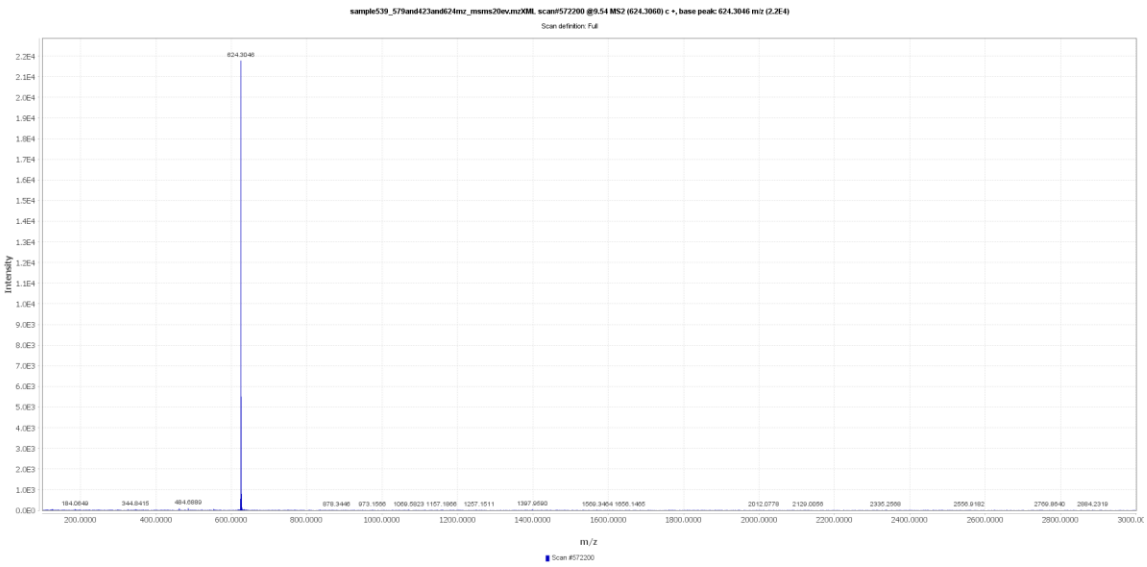


h. 624.3059 m/z MS/MS spectrums:

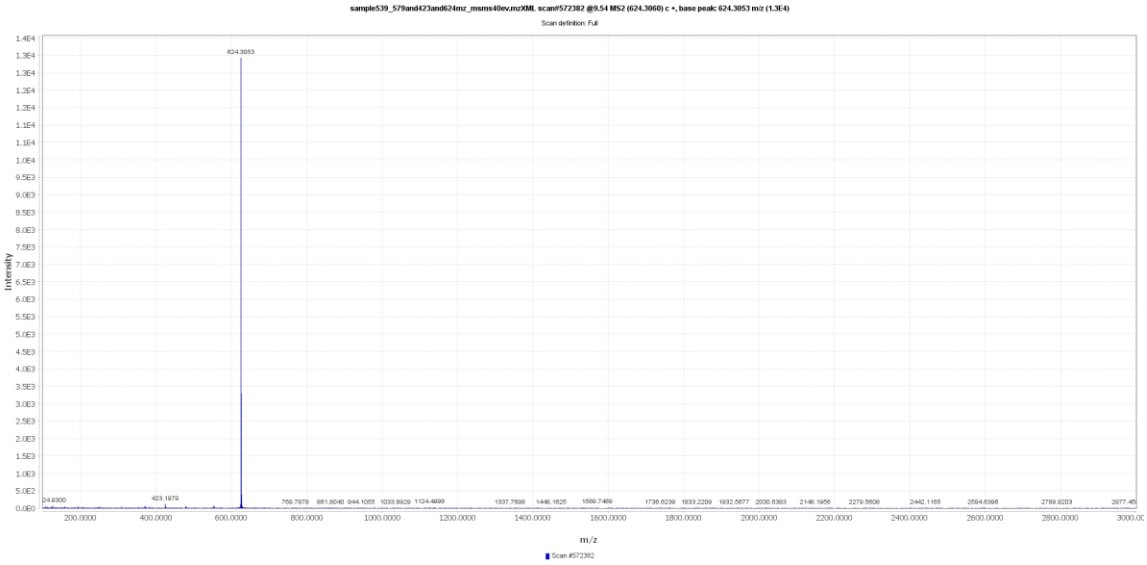
Collision energy: 10 eV



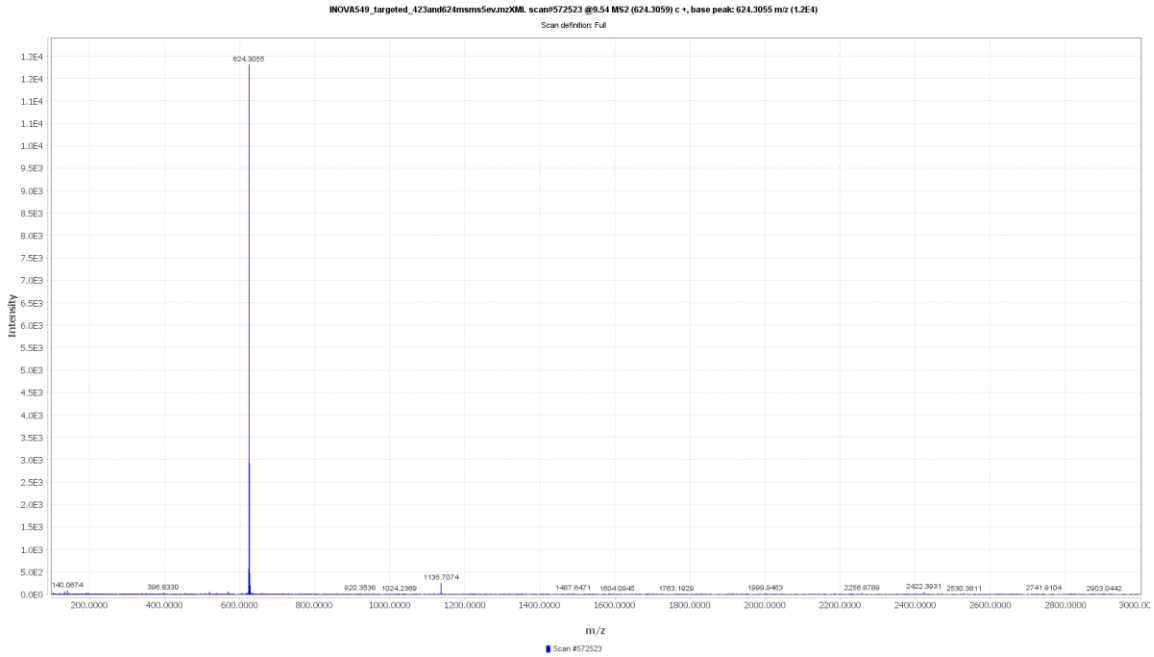
Collision energy: 20 eV



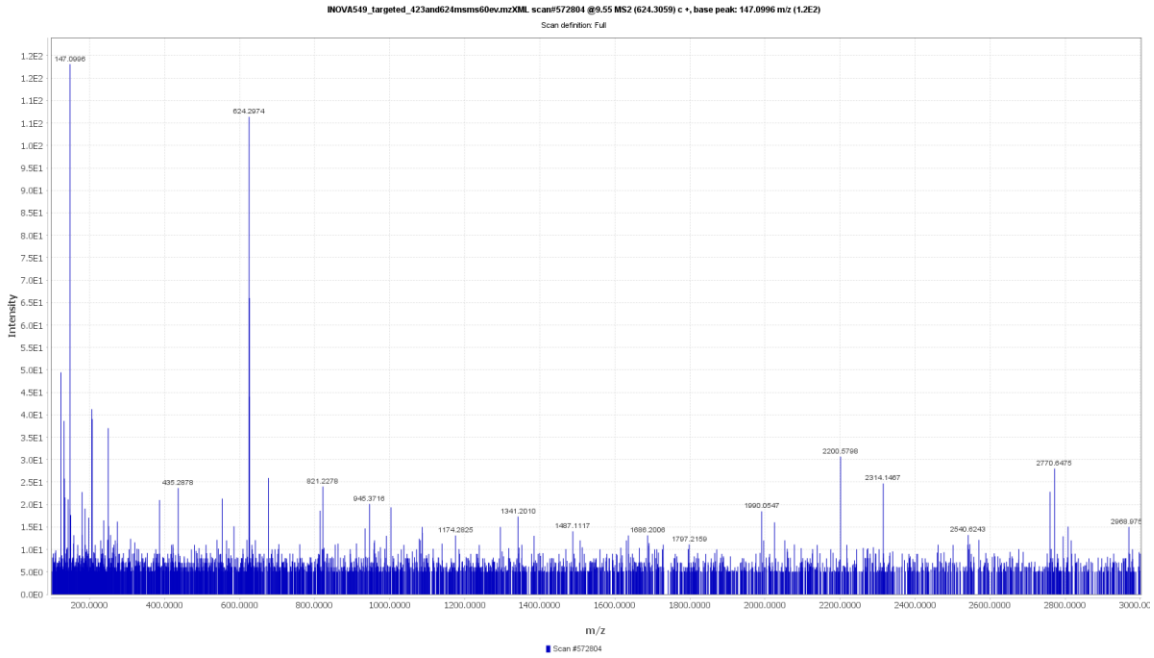
Collision energy: 40 eV



Collision energy: 5 eV



Collision energy: 60 eV



REFERENCES

1. Ma J, Ward EM, Siegel RL, Jemal A. Temporal Trends in Mortality in the United States, 1969-2013. *JAMA*. 2015;314(16):1731-1739. doi:10.1001/jama.2015.12319

2. Heart Disease Facts & Statistics | cdc.gov. https://www.cdc.gov/heartdisease/facts.htm. Published October 9, 2018. Accessed May 6, 2019.

3. Targher G, Day CP, Bonora E. Risk of Cardiovascular Disease in Patients with Nonalcoholic Fatty Liver Disease. *New England Journal of Medicine*. 2010;363(14):1341-1350. doi:10.1056/NEJMra0912063

4. Patil R, Sood GK. Non-alcoholic fatty liver disease and cardiovascular risk. *World J Gastrointest Pathophysiol*. 2017;8(2):51-58. doi:10.4291/wjgp.v8.i2.51

5. Definition & Facts of NAFLD & NASH | NIDDK. National Institute of Diabetes and Digestive and Kidney Diseases. https://www.niddk.nih.gov/health-information/liver-disease/nafld-nash/definition-facts. Accessed November 16, 2018.

6. Ruhl CE, Everhart JE. Fatty liver indices in the multiethnic United States National Health and Nutrition Examination Survey. *Alimentary Pharmacology & Therapeutics*. 2015;41(1):65-76. doi:10.1111/apt.13012

7. Schuppan D, Schattenberg JM. Non-alcoholic steatohepatitis: Pathogenesis and novel therapeutic approaches. *Journal of Gastroenterology and Hepatology*. 28(S1):68-76. doi:10.1111/jgh.12212

8. Cohen DE, Anania FA. Nonalcoholic Fatty Liver Disease. In: Greenberger NJ, Blumberg RS, Burakoff R, eds. *CURRENT Diagnosis & Treatment: Gastroenterology, Hepatology, & Endoscopy*. 3rd ed. New York, NY: McGraw-Hill Education; 2016. accessmedicine.mhmedical.com/content.aspx?aid=1119990644. Accessed November 20, 2018.

9. Masarone M, Rosato V, Dallio M, et al. Role of Oxidative Stress in Pathophysiology of Nonalcoholic Fatty Liver Disease. *Oxid Med Cell Longev*. 2018;2018. doi:10.1155/2018/9547613

10. Angulo P. Nonalcoholic Fatty Liver Disease. *New England Journal of Medicine*. 2002;346(16):1221-1231. doi:10.1056/NEJMra011775

11. Koch LK, Yeh MM. Nonalcoholic fatty liver disease (NAFLD): Diagnosis, pitfalls, and staging. *Annals of Diagnostic Pathology*. 2018;37:83-90. doi:10.1016/j.anndiagpath.2018.09.009

12. Bang KB, Cho YK. Comorbidities and Metabolic Derangement of NAFLD. *J Lifestyle Med*. 2015;5(1):7-13. doi:10.15280/jlm.2015.5.1.7

13. Azzam H, Malnick S. Non-alcoholic fatty liver diseaseâ€ " The heart of the matter. *World J Hepatol*. 2015;7(10):1369-1376. doi:10.4254/wjh.v7.i10.1369

14. Assy N, Djibre A, Farah R, Grosovski M, Marmor A. Presence of Coronary Plaques in Patients with Nonalcoholic Fatty Liver Disease. *Radiology*. 2010;254(2):393-400. doi:10.1148/radiol.09090769

15. Chen C-H, Nien C-K, Yang C-C, Yeh Y-H. Association Between Nonalcoholic Fatty Liver Disease and Coronary Artery Calcification. *Dig Dis Sci*. 2010;55(6):1752-1760. doi:10.1007/s10620-009-0935-9

16. Coronary Heart Disease | National Heart, Lung, and Blood Institute (NHLBI). https://www.nhlbi.nih.gov/health-topics/coronary-heart-disease. Accessed November 26, 2018.

17. Patel MR, Peterson ED, Dai D, et al. Low Diagnostic Yield of Elective Coronary Angiography. *New England Journal of Medicine*. 2010;362(10):886-895. doi:10.1056/NEJMoa0907272

18. Bandara EMS, Ekanayake S, Wanigatunge CA, Kapuruge A. Lipoprotein(a) and lipid profiles of patients awaiting coronary artery bypass graft; a cross sectional study. *BMC Cardiovascular Disorders*. 2016;16(1):213-213.

19. Libby P. Inflammation in atherosclerosis. *Nature*. 2002;420(6917):868. doi:10.1038/nature01323

20. Linton MF, Yancey PG, Davies SS, et al. The Role of Lipids and Lipoproteins in Atherosclerosis. In: Feingold KR, Anawalt B, Boyce A, et al., eds. *Endotext*. South Dartmouth (MA): MDText.com, Inc.; 2000. http://www.ncbi.nlm.nih.gov/books/NBK343489/. Accessed May 7, 2019.

21. Vafaeimanesh J, Parham M, Norouzi S, Hamednasimi P, Bagherzadeh M. Insulin resistance and coronary artery disease in non-diabetic patients: Is there any correlation? *Caspian J Intern Med*. 2018;9(2):121-126. doi:10.22088/cjim.9.2.121

22. Karrowni W, Li Y, Jones PG, et al. Insulin Resistance is Associated with Significant Clinical Atherosclerosis in Non-Diabetic Patients with Acute Myocardial Infarction. *Arterioscler Thromb Vasc Biol*. 2013;33(9). doi:10.1161/ATVBAHA.113.301585

23. Sookoian S, Pirola CJ. Non-alcoholic fatty liver disease is strongly associated with carotid atherosclerosis: A systematic review. *Journal of Hepatology*. 2008;49(4):600-607. doi:10.1016/j.jhep.2008.06.012

24. Hamaguchi M, Kojima T, Takeda N, et al. Nonalcoholic fatty liver disease is a novel predictor of cardiovascular disease. *World Journal of Gastroenterology*. 2007;13(10):1579-1584. doi:10.3748/wjg.v13.i10.1579

25. Zeb I, Li D, Budoff MJ, et al. Nonalcoholic Fatty Liver Disease and Incident Cardiac Events: The Multi-Ethnic Study of Atherosclerosis. *Journal of the American College of Cardiology*. 2016;67(16):1965-1966. doi:10.1016/j.jacc.2016.01.070

26. Zeb I, Li D, Nasir K, Katz R, Larijani VN, Budoff MJ. Computed Tomography Scans in the Evaluation of Fatty Liver Disease in a Population Based Study: The Multi-Ethnic Study of Atherosclerosis. *Academic Radiology*. 2012;19(7):811-818. doi:10.1016/j.acra.2012.02.022

27. Liu W, Zhang Y, Yu C-M, et al. Current understanding of coronary artery calcification. *J Geriatr Cardiol*. 2015;12(6):668-675. doi:10.11909/j.issn.1671-5411.2015.06.012

28. Kim D, Choi S-Y, Park EH, et al. Nonalcoholic fatty liver disease is associated with coronary artery calcification. *Hepatology*. 2012;56(2):605-613. doi:10.1002/hep.25593

29. Park HE, Kwak M-S, Kim D, Kim M-K, Cha M-J, Choi S-Y. Nonalcoholic Fatty Liver Disease Is Associated With Coronary Artery Calcification Development: A Longitudinal Study. *The Journal Of Clinical Endocrinology And Metabolism*. 2016;101(8):3134-3143. doi:10.1210/jc.2016-1525

30. Sinn DH, Cho SJ, Gu S, et al. Persistent Nonalcoholic Fatty Liver Disease Increases Risk for Carotid Atherosclerosis. *Gastroenterology*. 2016;151(3):481-488.e1. doi:10.1053/j.gastro.2016.06.001

31. Fernández-Sánchez A, Madrigal-Santillán E, Bautista M, et al. Inflammation, Oxidative Stress, and Obesity. *International Journal of Molecular Sciences*. 2011;12(5):3117-3132. doi:10.3390/ijms12053117

32. Esposito Katherine, Nappo Francesco, Marfella Raffaele, et al. Inflammatory Cytokine Concentrations Are Acutely Increased by Hyperglycemia in Humans. *Circulation*. 2002;106(16):2067-2072. doi:10.1161/01.CIR.0000034509.14906.AE

33. Stefan N, Häring H-U. The role of hepatokines in metabolism. *Nature Reviews Endocrinology*. 2013;9(3):144-152. doi:10.1038/nrendo.2012.258

34. Francque SM, van der Graaff D, Kwanten WJ. Non-alcoholic fatty liver disease and cardiovascular risk: Pathophysiological mechanisms and implications. *Journal of Hepatology*. 2016;65(2):425-443. doi:10.1016/j.jhep.2016.04.005

35. Harrison D, Griendling KK, Landmesser U, Hornig B, Drexler H. Role of oxidative stress in atherosclerosis. *The American Journal of Cardiology*. 2003;91(3, Supplement):7-11. doi:10.1016/S0002-9149(02)03144-2

36. Madamanchi Nageswara R., Vendrov Aleksandr, Runge Marschall S. Oxidative Stress and Vascular Disease. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 2005;25(1):29-38. doi:10.1161/01.ATV.0000150649.39934.13

37. Zivkovic AM, German JB, Sanyal AJ. Comparative review of diets for the metabolic syndrome: Implications for nonalcoholic fatty liver disease. *Am J Clin Nutr*. 2007;86(2):285-300. doi:10.1093/ajcn/86.2.285

38. Hashemi kani A, Alavian SM, Esmaillzadeh A, Adibi P, Azadbakht L. Dietary Quality Indices and Biochemical Parameters Among Patients With Non Alcoholic Fatty Liver Disease (NAFLD). *Hepat Mon*. 2013;13(7). doi:10.5812/hepatmon.10943

39. Wehmeyer MH, Zyriax B-C, Jagemann B, et al. Nonalcoholic fatty liver disease is associated with excessive calorie intake rather than a distinctive dietary pattern. *Medicine (Baltimore)*. 2016;95(23). doi:10.1097/MD.0000000000003887

40. Zelber-Sagi S, Nitzan-Kaluski D, Goldsmith R, et al. Long term nutritional intake and the risk for non-alcoholic fatty liver disease (NAFLD): A population based study. *Journal of Hepatology*. 2007;47(5):711-717. doi:10.1016/j.jhep.2007.06.020

41. Tilg H, Moschen AR. Evolution of inflammation in nonalcoholic fatty liver disease: The multiple parallel hits hypothesis. *Hepatology*. 2010;52(5):1836-1846. doi:10.1002/hep.24001

42. McCarthy EM, Rinella ME. The Role of Diet and Nutrient Composition in Nonalcoholic Fatty Liver Disease. *Journal of the Academy of Nutrition and Dietetics*. 2012;112(3):401-409. doi:10.1016/j.jada.2011.10.007

43. Cortez-Pinto H, Jesus L, Barros H, Lopes C, Moura MC, Camilo ME. How different is the dietary pattern in non-alcoholic steatohepatitis patients? *Clinical Nutrition*. 2006;25(5):816-823. doi:10.1016/j.clnu.2006.01.027

44. Musso G, Gambino R, De Michieli F, et al. Dietary habits and their relations to insulin resistance and postprandial lipemia in nonalcoholic steatohepatitis. *Hepatology*. 2003;37(4):909-916. doi:10.1053/jhep.2003.50132

45. Sacks FMM, Lichtenstein AHDs, Wu JHY, et al. Dietary Fats and Cardiovascular Disease: A Presidential Advisory From the American Heart Association. [Miscellaneous Article]. *Circulation*. 2017;136(3). doi:10.1161/CIR.0000000000000510

46. Scorletti E, Byrne CD. Omega-3 Fatty Acids, Hepatic Lipid Metabolism, and Nonalcoholic Fatty Liver Disease. *Annual Review of Nutrition*. 2013;33(1):231-248. doi:10.1146/annurev-nutr-071812-161230

47. Kannel WB, McGee D, Gordon T. A general cardiovascular risk profile: The Framingham study. *The American Journal of Cardiology*. 1976;38(1):46-51. doi:10.1016/0002-9149(76)90061-8

48. Johnsen SH, Mathiesen EB. Carotid plaque compared with intima-media thickness as a predictor of coronary and cerebrovascular disease. *Curr Cardiol Rep*. 2009;11(1):21-27. doi:10.1007/s11886-009-0004-1

49. La Grutta L, Marasà M, Toia P, et al. Integrated non-invasive approach to atherosclerosis with cardiac CT and carotid ultrasound in patients with suspected coronary artery disease. *Radiol med*. 2017;122(1):16-21. doi:10.1007/s11547-016-0692-8

50. Horgan RP, Kenny LC. 'Omic' technologies: Genomics, transcriptomics, proteomics and metabolomics: The Obstetrician & Gynaecologist. *The Obstetrician & Gynaecologist*. 2011;13(3):189-195. doi:10.1576/toag.13.3.189.27672

51. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews*. 2007;26(1):51-78. doi:10.1002/mas.20108

52. Fiehn O. Metabolomics: The link between genotypes and phenotypes. *Plant Molecular Biology*.:155–171.

53. Xiao JF, Zhou B, Ressom HW. Metabolite identification and quantitation in LC-MS/MS-based metabolomics. *TrAC Trends in Analytical Chemistry*. 2012;32:1-14. doi:10.1016/j.trac.2011.08.009

54. Putri SP, Yamamoto S, Tsugawa H, Fukusaki E. Current metabolomics: Technological advances. *Journal of Bioscience and Bioengineering*. 2013;116(1):9-16. doi:10.1016/j.jbiosc.2013.01.004

55. Ellis DI, Dunn WB, Griffin JL, Allwood JW, Goodacre R. Metabolic fingerprinting as a diagnostic tool. *Pharmacogenomics*. 2007;8(9):1243-1266. doi:10.2217/14622416.8.9.1243

56. O'Sullivan A, Gibney MJ, Brennan L. Dietary intake patterns are reflected in metabolomic profiles: Potential role in dietary assessment studies. *Am J Clin Nutr*. 2011;93(2):314-321. doi:10.3945/ajcn.110.000950

57. Jin Q, Black A, Kales SN, Vattem D, Ruiz-Canela M, Sotos-Prieto M. Metabolomics and Microbiomes as Potential Tools to Evaluate the Effects of the Mediterranean Diet. *Nutrients*. 2019;11(1). doi:10.3390/nu11010207

58. Bondia-Pons I, Martinez JA, Iglesia R de la, et al. Effects of short- and long-term Mediterranean-based dietary treatment on plasma LC-QTOF/MS metabolic profiling of subjects with metabolic syndrome features: The Metabolic Syndrome Reduction in Navarra (RESMENA) randomized controlled trial. *Molecular Nutrition & Food Research*. 2015;59(4):711-728. doi:10.1002/mnfr.201400309

59. Klimczak I, Gliszczyńska-Świgło A. Comparison of UPLC and HPLC methods for determination of vitamin C. *Food Chemistry*. 2015;175:100-105. doi:10.1016/j.foodchem.2014.11.104

60. Nováková L, Matysová L, Solich P. Advantages of application of UPLC in pharmaceutical analysis. *Talanta*. 2006;68(3):908-918. doi:10.1016/j.talanta.2005.06.035

61. Plumb RS, Johnson KA, Rainville P, et al. UPLC/MSE; a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Communications in Mass Spectrometry*. 2006;20(13):1989-1994. doi:10.1002/rcm.2550

62. Price P. Standard Definitions of Terms Relating to Mass Spectrometry: A report from the Committee on Measurements and Standards of the American Society for Mass Spectrometry. *Journal of the American Society for Mass Spectrometry*. 1991;2(4):336-348. doi:10.1016/1044-0305(91)80025-3

63. Atmospheric Pressure Chemical Ionizationâ€ " MagLab. https://nationalmaglab.org/user-facilities/icr/techniques/apci. Accessed May 20, 2019.

64. Ho C, Lam C, Chan M, et al. Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *Clin Biochem Rev*. 2003;24(1):3-12.

65. Banerjee S, Mazumdar S. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. International Journal of Analytical Chemistry. doi:10.1155/2012/282574

66. Hoffmann E de. Mass Spectrometry. In: *Kirk-Othmer Encyclopedia of Chemical Technology*. American Cancer Society; 2005. doi:10.1002/0471238961.1301191913151518.a01.pub2

67. Gross JH. *Mass Spectrometry: A Textbook*. Springer Science & Business Media; 2006.

68. Morris HR, Paxton T, Dell A, et al. High Sensitivity Collisionally-activated Decomposition Tandem Mass Spectrometry on a Novel Quadrupole/Orthogonal-acceleration Time-of-flight Mass Spectrometer. *Rapid Communications in Mass Spectrometry*. 1996;10(8):889-896. doi:10.1002/(SICI)1097-0231(19960610)10:8<889::AID-RCM615>3.0.CO;2-F

69. Dawson PH. *Quadrupole Mass Spectrometry and Its Applications*. Elsevier; 2013.

70. March RE. Ion Trap Mass Spectrometers. In: Lindon JC, ed. *Encyclopedia of Spectroscopy and Spectrometry*. Oxford: Elsevier; 1999:1000-1009. doi:10.1006/rwsp.2000.0143

71. Sleno L, Volmer DA. Ion activation methods for tandem mass spectrometry. *Journal of Mass Spectrometry*. 2004;39(10):1091-1112. doi:10.1002/jms.703

72. Hoffmann E de, Stroobant V. *Mass Spectrometry: Principles and Applications*. John Wiley & Sons; 2013.

73. Uphoff A, Grotemeyer J. The Secrets of Time-of Flight Mass Spectrometry Revealed. *Eur J Mass Spectrom (Chichester)*. 2003;9(3):151-164. doi:10.1255/ejms.545

74. Fjeldsted J. Time-of-Flight Mass Spectrometry. :12.

75. Simon TG, Trejo MEP, McClelland R, et al. Circulating Interleukin-6 is a biomarker for coronary atherosclerosis in nonalcoholic fatty liver disease: Results from the Multi-Ethnic Study of Atherosclerosis. *International Journal of Cardiology*. 2018;259:198-204. doi:10.1016/j.ijcard.2018.01.046

76. Elsheikh E, Younoszai Z, Otgonsuren M, Hunt S, Raybuck B, Younossi ZM. Markers of endothelial dysfunction in patients with non-alcoholic fatty liver

disease and coronary artery disease. *Journal of Gastroenterology and Hepatology*. 2014;29(7):1528-1534. doi:10.1111/jgh.12549

77. E. Yannell K, R. Ferreira C, E. Tichy S, Graham Cooks R. Multiple reaction monitoring (MRM)-profiling with biomarker identification by LC-QTOF to characterize coronary artery disease. *Analyst*. 2018;143(20):5014-5022. doi:10.1039/C8AN01017J

78. Li R, Li F, Feng Q, et al. An LC-MS based untargeted metabolomics study identified novel biomarkers for coronary heart disease. *Molecular BioSystems*. 2016;12(11):3425-3434. doi:10.1039/C6MB00339G

79. Shah Svati H., Bain James R., Muehlbauer Michael J., et al. Association of a Peripheral Blood Metabolic Profile With Coronary Artery Disease and Risk of Subsequent Cardiovascular Events. *Circulation: Cardiovascular Genetics*. 2010;3(2):207-214. doi:10.1161/CIRCGENETICS.109.852814

80. Wang Zhe, Zhu Cong, Nambi Vijay, et al. Metabolomic Pattern Predicts Incident Coronary Heart Disease. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 0(0):ATVBAHA.118.312236. doi:10.1161/ATVBAHA.118.312236

81. Abdallah IA, Hammell DC, Stinchcomb AL, Hassan HE. A fully validated LC–MS/MS method for simultaneous determination of nicotine and its metabolite cotinine in human serum and its application to a pharmacokinetic study after using nicotine transdermal delivery systems with standard heat application in adult smokers. *Journal of Chromatography B*. 2016;1020:67-77. doi:10.1016/j.jchromb.2016.03.020

82. Deininger S-O, Cornett DS, Paape R, et al. Normalization in MALDI-TOF imaging datasets of proteins: Practical considerations. *Anal Bioanal Chem*. 2011;401(1):167-181. doi:10.1007/s00216-011-4929-z

83. Leys C, Ley C, Klein O, Bernard P, Licata L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*. 2013;49(4):764-766. doi:10.1016/j.jesp.2013.03.013

84. Rousseeuw PJ, Croux C. Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association*. 1993;88(424):1273-1283. doi:10.2307/2291267

85. Westerhuis JA, Hoefsloot HCJ, Smit S, et al. Assessment of PLSDA cross validation. *Metabolomics*. 2008;4(1):81-89. doi:10.1007/s11306-007-0099-6

86. Yu L, Aa J, Xu J, et al. Metabolomic phenotype of gastric cancer and precancerous stages based on gas chromatography time-of-flight mass spectrometry. *Journal of Gastroenterology and Hepatology*. 2011;26(8):1290-1297. doi:10.1111/j.1440-1746.2011.06724.x

87. Worley B, Powers R. Multivariate Analysis in Metabolomics. *Curr Metabolomics*. 2013;1(1):92-107. doi:10.2174/2213235X11301010092

88. Zhang S, Cao J. A close examination of double filtering with fold change and t test in microarray analysis. *BMC Bioinformatics*. 2009;10:402. doi:10.1186/1471-2105-10-402

89. Sauer M, Jakob A, Nordheim A, Hochholdinger F. Proteomic analysis of shoot-borne root initiation in maize (Zea mays L.). *PROTEOMICS*. 2006;6(8):2530-2541. doi:10.1002/pmic.200500564

90. Long FH. Chapter 19â€ ” Multivariate Analysis for Metabolomics and Proteomics Data. In: Issaq HJ, Veenstra TD, eds. *Proteomic and Metabolomic Approaches to Biomarker Discovery*. Boston: Academic Press; 2013:299-311. doi:10.1016/B978-0-12-394446-7.00019-4

91. Ramadan Z, Jacobs D, Grigorov M, Kochhar S. Metabolic profiling using principal component analysis, discriminant partial least squares, and genetic algorithms. *Talanta*. 2006;68(5):1683-1691. doi:10.1016/j.talanta.2005.08.042

92. Uarrota VG, Moresco R, Coelho B, et al. Metabolomics combined with chemometric tools (PCA, HCA, PLS-DA and SVM) for screening cassava (Manihot esculenta Crantz) roots during postharvest physiological deterioration. *Food Chemistry*. 2014;161:67-78. doi:10.1016/j.foodchem.2014.03.110

93. Lewis GD, Asnani A, Gerszten RE. Application of Metabolomics to Cardiovascular Biomarker and Pathway Discovery. *Journal of the American College of Cardiology*. 2008;52(2):117-123. doi:10.1016/j.jacc.2008.03.043

94. Raamsdonk LM, Teusink B, Broadhurst D, et al. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnology*. 2001;19(1):45. doi:10.1038/83496

95. Gao X, Ke C, Liu H, et al. Large-scale Metabolomic Analysis Reveals Potential Biomarkers for Early Stage Coronary Atherosclerosis. *Sci Rep*. 2017;7. doi:10.1038/s41598-017-12254-1

96. Feng Q, Liu Z, Zhong S, et al. Integrated metabolomics and metagenomics analysis of plasma and urine identified microbial metabolites associated with coronary heart disease. *Scientific Reports*. 2016;6:22525. doi:10.1038/srep22525

97. Zhang W, Zhao PX. Quality evaluation of extracted ion chromatograms and chromatographic peaks in liquid chromatography/mass spectrometry-based metabolomics data. *BMC Bioinformatics*. 2014;15(Suppl 11):S5. doi:10.1186/1471-2105-15-S11-S5

98. Morales A, Gao W, Lu J, Xing J, Li J. Muscle cyclo-oxygenase-2 pathway contributes to the exaggerated muscle mechanoreflex in rats with congestive heart failure. *Exp Physiol*. 2012;97(8):943-954. doi:10.1113/expphysiol.2012.065425

99. Ochi T, Motoyama Y, Goto T. The analgesic effect profile of FR122047, a selective cyclooxygenase-1 inhibitor, in chemical nociceptive models. *European Journal of Pharmacology*. 2000;391(1):49-54. doi:10.1016/S0014-2999(00)00051-0

100. Lennernäs H. Clinical Pharmacokinetics of Atorvastatin. *Clin Pharmacokinet*. 2003;42(13):1141-1160. doi:10.2165/00003088-200342130-00005

101. Davidson M, McKenney J, Stein E, et al. Comparison of One-Year Efficacy and Safety of Atorvastatin Versus Lovastatin in Primary Hypercholesterolemia. *The American Journal of Cardiology*. 1997;79(11):1475-1481. doi:10.1016/S0002-9149(97)00174-4

102. Jacobsen WD, Kuhn B, Soldner A, et al. Lactonization is the critical first step in the disposition of the 3-hydroxy-3-methylglutaryl-CoA reductase inhibitor atorvastatin. *Drug metabolism and disposition: The biological fate of chemicals*. 2000;28(11):1369-1378.

103. Kantola T, Kivistö KT, Neuvonen PJ. Effect of itraconazole on the pharmacokinetics of atorvastatin. *Clinical Pharmacology & Therapeutics*. 1998;64(1):58-65. doi:10.1016/S0009-9236(98)90023-6

104. Xu Y-F, Lu W, Rabinowitz JD. Avoiding Misannotation of In-Source Fragmentation Products as Cellular Metabolites in Liquid Chromatography–Mass Spectrometry-Based Metabolomics. *Anal Chem*. 2015;87(4):2273-2281. doi:10.1021/ac504118y

105. Abrankó L, García-Reyes JF, Molina-Díaz A. In-source fragmentation and accurate mass analysis of multiclass flavonoid conjugates by electrospray ionization time-of-flight mass spectrometry. *Journal of Mass Spectrometry*. 2011;46(5):478-488. doi:10.1002/jms.1914

BIOGRAPHY

Fang Chiu graduated from Wu-Ling Senior High School, Taoyuan, Taiwan, in 2013. She received her Bachelor of Science from Taipei Medical University in 2017. She is a registered dietitian in Taiwan and completed the practice program in Mackay Memorial Hospital.